Sustainable Cities and Society 83 (2022) 103858

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

OpenComm: Open community platform for data integration and privacy preserving for 311 calls

Duy H. Ho [a], Yugyung Lee [a,*], Srichakradhar Nagireddy [a], Charan Thota [a], Brent Never [b], Ye Wang [c]

[a] Department of Computer Science Electrical Engineering, School of Computing and Engineering, University of Missouri, Kansas City, MO, USA
[b] Department of Public Affairs, the Henry W. Bloch School of Management, University of Missouri, Kansas City, MO, USA
[c] Department of Communication Studies, College of Arts and Sciences, University of Missouri, Kansas City, MO, USA

ARTICLE INFO

Keywords:
311 calls
Data curation
Data privacy
Data integration
Categorization
Data visualization
Machine learning

ABSTRACT

Local governments are increasingly leveraging administrative data to drive performance. Likewise, cities are interested in improving responsiveness to citizens' demands and cost savings through data analytics. However, city managers face many challenges when utilizing secondary data, such as 311 call records and the US Census. The challenge of interest to the current study is boundary issues as a result of data being collected at divergent geographic levels over different time horizons. Accordingly, an inductive analytical methodology was developed to create units of analysis that were both pragmatically and analytically appropriate for city managers and local policymakers. We created an open data analytics framework called OpenComm to harmonize administrative and secondary data using administrative data derived from Kansas City, Missouri. This framework produced robust inferences regarding the spatial and temporal aspects for the communities. Privacy-preserving technology, in particular, has been applied to public data to protect community privacy. The findings illustrate the power of inductive data aggregation, leading to empirical insights into hidden patterns of city service disparity over a decade-long time horizon. An application for the Open Data Platform is available at http://kc311.herokuapp.com/.

1. Introduction

Building a highly usable open data platform for urban sustainability requires a deep understanding of a society's socially embedded activities, mediated by technology (Bibri & Krogstie, 2017). Any of these platforms will need to achieve a balance between ecological and technical advancements. As pointed out by Yigitcanlar et al. (2019), a city cannot be smart without investment in the growth of human, societal, and environmental capitals that generate sustainable development. The most critical obstacles to building smart, sustainable cities are the low quality of public data, the challenges of data integration across heterogeneous sources, and the challenge of interoperability with diverse stakeholders (Duvier, Anand, & Oltean-Dumbrava, 2018).

Data is complex and continuously changing in terms of attributes or themes, spatial information, and temporal information (Calka, Nowak Da Costa, & Bielecka, 2017; Gregory & Ell, 2005; Monteiro, Martins, Murrieta-Flores, & Pires, 2019; Pavía & Cantarino, 2017; Pittaluga, 2020). Studies on spatial analytics point to the complexities inherent in citizen-generated secondary data: there is no simple universal solution due to the changes in jurisdictional boundaries or mismatches between spatial boundaries. The key motivation of integrating data through connecting physical boundaries is to leverage the ever-expanding amount of data published by multiple data providers and integrate such data by relaxing the physical constraints set by their geographic coding and standards.

Traditionally, these sources are analyzed separately, often leading to isolated inference and limited scope. For example, a 311 call system assigns a location by neighborhoods, counties, council districts, and police districts. At the same time, the Census Bureau aggregates its population data to the country, state, tract, and block group levels. Moreover, the Google Maps API (Google Inc, 2021) provides consistent results for neighborhoods, addresses, counties, and states. However, it falls short at lower levels of localities, such as block groups or school districts. These sources have been analyzed separately and generated isolated inferences with limited scope.

* Corresponding author.
E-mail addresses: dhh3hb@mail.umkc.edu (D.H. Ho), leeyu@umkc.edu (Y. Lee), snp8b@mail.umkc.edu (S. Nagireddy), ctpcc@mail.umkc.edu (C. Thota), brent.never@umkc.edu (B. Never), WanYe@umkc.edu (Y. Wang).
URL: https://info.umkc.edu/leeyu/ (Y. Lee).

https://doi.org/10.1016/j.scs.2022.103858
Received 28 April 2021; Received in revised form 21 March 2022; Accepted 23 March 2022
Available online 5 May 2022

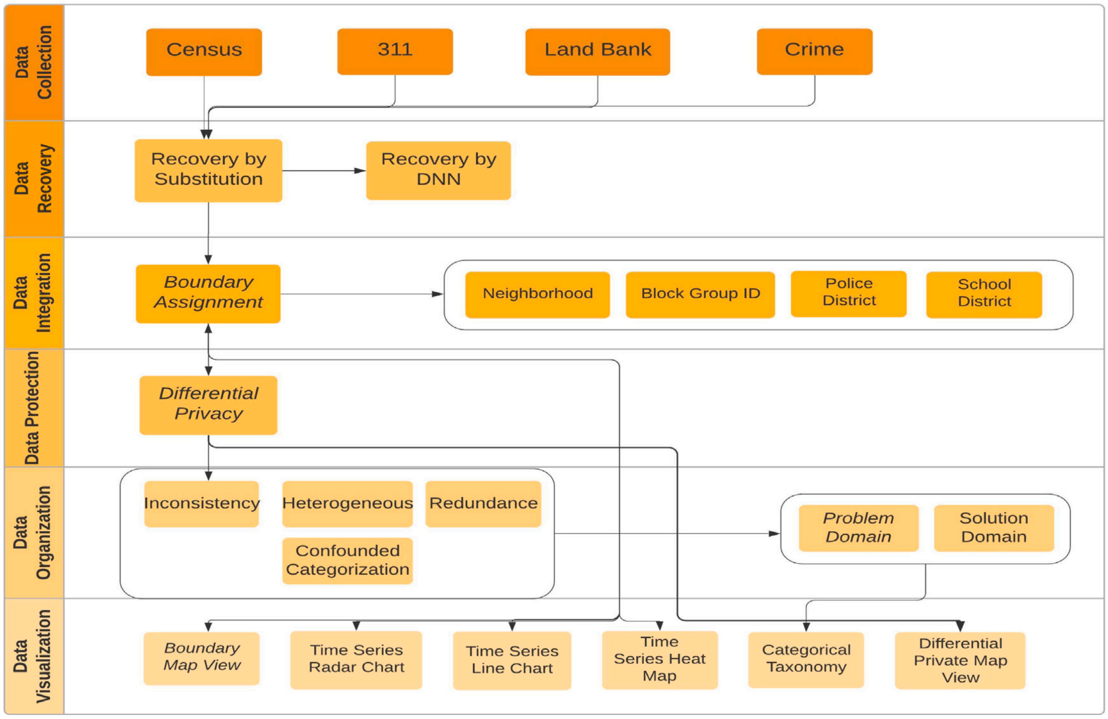

Fig. 1. Workflow Design of OpenComm: data collection, data recovery, data integration, data protection, data categorization, and data visualization.

Highly diverse data sources and existing models, i.e., heterogeneous data sources in dynamic contexts (social, economic, crime and housing characteristics), are not semantically integrated. Therefore, the development of data-driven models or applications becomes challenging. Recently, there has been a high demand for building or sharing such data-driven solutions across multiple agencies, including federated AI (Barclay, Preece, Taylor, & Verma, 2019), policy-based assemble approaches (Verma, Bertino, Russo, Calo, & Singla, 2020), and services and edge computing for federated AI (Lim et al., 2020; Verma, White, & de Mel, 2019).

We hypothesize that data integration will be a solution to the previously mentioned challenges. This paper presents critical research questions for data integration to advance research in the solution area. In order to construct analytic-centric solutions to intractable urban problems, we must address the critical issues associated with publicly-available data: data quality, privacy, segregated data, classification. As a result, we pose the following questions:

- How could we integrate public data from diverse sources and build an open data platform to curate and share them?
- How will the data platform be useful to data analysts, who process relevant, diverse, and complex infrastructure- and community-related data?
- Ultimately, how could we use automation to generate new data types from existing ones given contexts and ultimate goals?

Our proposed solution is to build an open community platform *OpenComm* to share and integrate datasets from different sources and build data-driven applications that would be useful for an innovative, sustainable community. In particular, the platform could be leveraged to share curated and integrated data efficiently and securely with researchers, citizens, public officials, and local stakeholders in order to facilitate smart connected communities. We thoroughly automate the *OpenComm*'s workflow of data recovery, integration, privacy, categorization, and visualization (see Fig. 1). Our case study uses Kansas City, Missouri's 311 calls and the Census Bureau data to demonstrate the challenges and the proposed solutions to tackle the challenges.

Our innovations include (1) an efficient data cleaning model that can quickly identify human data entry errors and neutralize them; (2) a multi-level categorization model capable of organizing structured data from multiple jurisdictional boundaries, such as 311 calls' neighborhoods and Census Bureau's block groups; (3) the novel privacy-preserving technique that permits the interchange of publicly available data while maintaining data privacy; (4) a model of process intelligence capable of mapping and connecting structured data between domains, such as the 311 neighborhoods and police districts (see Fig. 2 and Table 1). *OpenComm* presently provides solutions for those five distinct datasets (described in Table 2). The findings presented in this research are mostly based on the 311 calls and Census Bureau data.

2. Related work

This section reviews the primary challenges associated with developing an open data platform that can be used to support innovative, sustainable cities. Fundamental obstacles to data sharing and interoperability are addressed, such as insufficient data quality and a lack of capabilities for integrating data from multiple sources. Ultimately, these discussions point to our proposed solutions for an open data platform for sharing data with diverse stakeholders (Duvier et al., 2018).

2.1. Open data platforms

Danneels, Viaene, and Van den Bergh (2017) defines open data platforms in terms of three different platform types: (1) cognitivist as "standardized management of information", (2) connectionist as "management of standardized information through communities", and, (3) autopoietic as "management of data through individual people". Neves, de Castro Neto, and Aparicio (2020) analyzed open data impact factors in smart city dimensions and found a big gap between the impacts and sustainable development of smart cities.

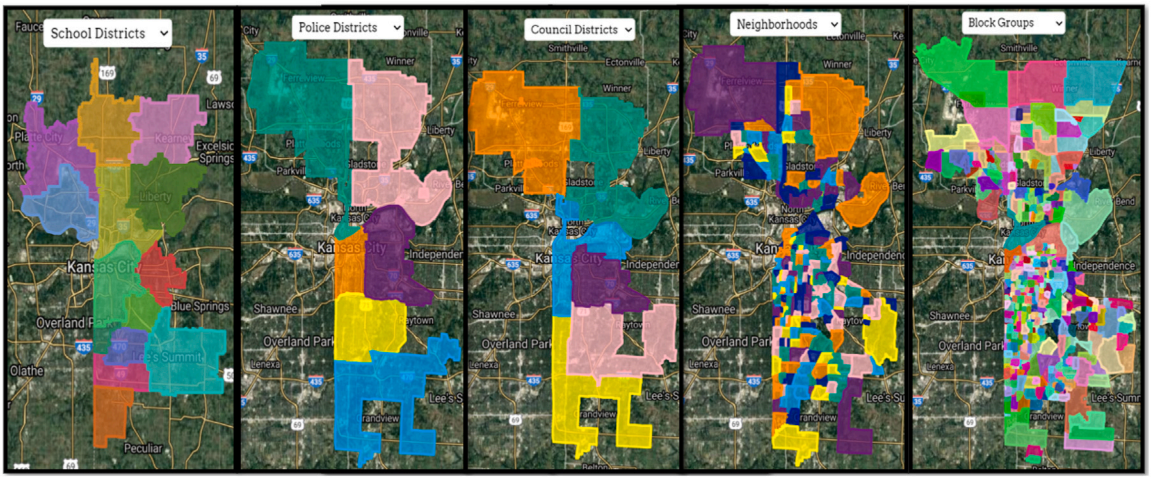

Fig. 2. 5 Types of Geographic Boundaries, ordered by average single-unit area (left to right): school districts, police divisions, council districts, neighborhoods, and block groups.

Table 1
KCMO's five types of geographic boundaries. Fig. 2 shows the example of boundaries; each color is representing a boundary for each boundary type.

Boundary type	Number	Description
School district	14	There are 14 school districts that comprise public schools in Kansas City, Missouri (KCMO). The original data contained many districts in both Kansas and Missouri, which was then filtered down to only the KCMO area.
Police district	6	There are six police divisions throughout Kansas City.
Council district	6	There are six council districts in Kansas City. The data also contains the names of the councilman/councilwoman and the at-large councilman/councilwoman of each district.
Neighborhood	246	The neighborhood geographic data was obtained via the OpenData KC portal as a GeoJSON file, allowing for the most precise polygonal representation of the neighborhoods on a digital map. According to the data, there are 246 neighborhoods with a corresponding id, name, and multi-polygonal annotations in terms of latitude and longitude coordinates. Five out of 246 neighborhoods have no name.
Block group	465	Block groups are an important unit for this project because it enables data integration with other the Census Bureau data products. Block group is the lowest geographic unit that the Census has published in the detailed tables of its annual 5-year American Community Survey (ACS) since 2013, which contains over 20,000 variables on various aspects of the population. For Kansas City, there are 465 block groups, the majority of which overlap heavily with the 246 neighborhoods mentioned in the previous subsection.

Table 2
KCMO data used for open communication platform.

Data	Size	Time	#Feature	Description
311 call record (KCMO, 2021a)	1.5M	2007–2020	32	case id, source, department, work group, request type, category, type, detail, creation date, creation time, creation month, creation year, status, exceeded est timeframe, closed date, closed month, closed year, days to close, street address, address with geocode, zip code, neighborhood, county, council district, police district, parcel id no, latitude, longitude, case URL, 30-60-90 days open window
311 call description (Data, 2021a)	299,582	2015–2020	38	source, category, type, cattype, group311, vis_street, detail, status, county, latitude, longitude, case id, department, work group, request type, creation date, creation time, creation month, creation year, exceeded est timeframe, closed date, closed month, closed year, days to close, street address, address with geocode, zip code, neighborhood, council district, police district, parcel id no, case url, 30-60-90 days open window, nbh_id, nbh_name, description, parent_dept, parent_category
Census data (Python Package Index (PyPI), 2021)	4,186	2013–2020	14	Total Population, Median Income, Median Home Value, Total population age 25+ years with a bachelor's degree or higher, White alone, Black or African American alone, Asian alone, Hispanic or Latino, Total civilian labor force, Total unemployed, Number of households, Total Vacant, Total Renter Occupied, Households with income below poverty
Land Bank data (Data, 2021b)	6,256	2012–2020	16	Parcel Number, Property Class, Property Status, Address, City, State, Postal Code, County, Neighborhood, Council District, Sold Date, School District, Market Value Year, Market Value, Date Evaluated
Crime data (KCPD, 2021)	927,879	2013–2020	11	Address, City, Zipcode, Description, Offense, Reported Time, Age, Sex, Race, District.

The term "open data" refers to data that is freely available online in a machine-readable format for anybody to use, re-use, and redistribute (Davies & Perini, 2016). For example, in recent research, user actions were provided as citizen-generated data via a 311 system (Wu, 2020). They studied the causal link between technological adoption, resident satisfaction, and frequency of public service use to address

Table 3
New York city 311 (Minkoff, 2016).

Government goods	Street condition; Street light condition; Water system; Sewer; Dirty conditions; Sanitation condition; Damaged tree; Rodent; Missed garbage/recycling collection; Derelict vehicle; Overgrown tree/branches; Broken muni meter; Root/sewer/sidewalk condition; Dead tree; Sidewalk condition; Snow; Broken parking meter; Street sign—Damaged; Street sign—Missing; Water conservation; Maintenance or facility; Litter basket/request; Industrial waste; Highway condition; Water quality; Street sweeping; Street sign—Dangling; Recycling enforcement; Overflowing litter baskets; Standing water; Wear and tear; Indoor sewage; Beach/pool/sauna complaint; Bridge condition; Bike rack condition; Highway sign—Damaged; Curb condition; Highway sign—Missing
Noise	General; Street/sidewalk; Vehicle; Park; House of worship
Graffiti	Graffiti

data quality concerns in the 311 system and close the gap between consumer- and user-oriented service in the 311 system.

The 311 service is a non-emergency service for residents to report problems and ask city-related questions in many cities, such as Baltimore, Buffalo, Chicago, New York City, San Francisco, Seattle, and Washington, D.C. (Minkoff, 2016). Some studies indicate that an improvement in a city government's performance was achieved using 311 call services (Clark & Guzman, 2017; Clark et al., 2014; Hartmann, Mainka, & Stock, 2017). The 311 calls were also used to enable citizen engagement with their local government (Gao, 2018; O'Brien, Offenhuber, Baldwin-Philippi, Sands, & Gordon, 2017), fostering public services targeted to serve citizen needs (Clark et al., 2014). Several studies (Minkoff, 2016; O'Brien, 2015; T. O'Brien, 2016; White & Trump, 2018) combined 311 customer data with the census tract-level demographic data in order to get insight into community members' 311 usage habits.

Chatfield and Reddick (2018) highlight the practical application of big data analytics via a case study of Houston's 311 call system for a *data-driven government*. Minkoff examines the distribution of physical conditions associated with government services using 311 data from New York City (Minkoff, 2016). The tract-level analysis of the census and 311 data identified issues in specific geographic areas, such as older housing stock, increased vehicle traffic, and rapid population growth. O'Brien's study using Boston's 311 data (O'Brien, 2016) discovers that neighborhood members are more proactive in reporting difficulties within two blocks of their homes; people were three times more likely to report issues within two blocks of their houses. Dawes, Vidiasova, and Parkhimovich (2016) emphasize the critical nature of a standard open platform for community studies. The focus of this study, inspired by the previous works, is to support communities in building open data platforms to share community data.

2.2. Privacy preserving

There are ongoing studies on location privacy in the literature, as well as advanced techniques for location privacy protection, such as location perturbation by adding noise to locations (Shokri, Theodorakopoulos, Le Boudec, & Hubaux, 2011), spatial cloaking by concealing areas within a shrouded region (Damiani, Bertino, Silvestri, et al., 2010; Xue, Kalnis, & Pung, 2009), and dummy location generation to hide the user's actual location (Niu, Li, Zhu, Cao, & Li, 2014, 2015). However, these safeguards have introduced additional difficulties. For instance, determining the users' real locations is not straightforward. Additionally, appropriate security and privacy protection approaches, such as anonymization, homomorphic encryption, masking, differential privacy, and hashing, were recommended to address potential privacy leakage threats.

Security and privacy protection has been strengthened in a variety of domains: preserving personalized location privacy in ride-hailing services (Khazbak, Fan, Zhu, & Cao, 2018, 2020), quantifying uncertainty in decision-making problems using fuzzy soft sets (Bhardwaj & Sharma, 2021), managing dataflow in the Internet of Things for sensing, control, and security (Wei et al., 2021), and probabilistic graphical models for inference attacks on genomic data (He & Zhou, 2020).

Our research is focused on developing privacy-preserving technologies that protect the privacy and security of public data that is associated with communities rather than individuals. Due to the benefits of community-level differential privacy based on deep learning, our method is more relevant and robust than other existing security and privacy preservation solutions.

2.3. Data curation, integration, and visualization

Data curation (Stonebraker et al., 2013) was defined as the process of data discovery (Deng et al., 2017; Fernandez et al., 2018; Miller et al., 2018; Nargesian, Zhu, Pu, & Miller, 2018), data cleaning (Dallachiesa et al., 2013; Fan & Geerts, 2012; Rekatsinas, Chu, Ilyas, & Ré, 2017), and data integration (Halevy & Doan, 2012; Konda et al., 2016; Stonebraker & Ilyas, 2018) that is required for data analytics. Recent advances in deep learning have been made in data curation (Thirumuruganathan, Tang, Ouzzani, & Doan, 2020). Integrating data is critical for gaining a greater understanding of urban systems to promote economic growth, social fairness, and environmental sustainability (Lai, 2020). The transportation data analytics framework was proposed to maximize efficiency and flexibility while integrating data from disparate sources (Cui, Henrickson, Biancardo, Pu, & Wang, 2020).

The relationship between space, housing, and economic development was examined in a study of New York City's 311 system (Minkoff, 2016). The variables from the census and NYC 311 call categories utilized in their analysis are listed in Table 3. The NYC 311 categories were determined by analyzing service-request calls from residential tracts from 2007 to 2012. Three regression models were used to assess 311 calls, adjusting for geographical and serial dependencies, such as government-provided goods, graffiti, and noise.

311 service requests from 2008 to 2017 in Columbus, OH, were analyzed to reveal hotspots for community distress and opioid-use disorder (Li et al., 2020). Their prediction model was trained on ten different 311 query types to discover parameters associated with the detection of opiate overdoses. The prediction model was developed using the dominating factors, including code violation, public health, and street lighting, and demonstrated an accuracy of 0.92, 0.89, and 0.83, respectively.

Bloch (2020) used Google Street View and 311 calls to map visual and municipal spatial big data to measure neighborhood characteristics and evaluated the limitations of data from existing methods. Zuo et al. (2020) recently used an interactive data visualization and analytics tool to examine mobility and sociability patterns during COVID-19. Schuman et al. (2020) reported on the efficacy of group content learning and scientific techniques for application with touchable visualizations of ocean data. Lee, Choe, Isenberg, Marriott, and Stasko (2020) emphasized the critical role of data visualization in reaching a broader audience.

The current study builds upon these existing studies applied machine learning, data analytics, and visualization approaches to address data challenges. Unfortunately, the four urban data platforms outlined in Barns (2018) (shown in Table 4) are insufficient to provide a solution for the open community platform. In comparison to these works, OpenComm's open community platform incorporates critical areas of data curation, integration, classification, data privacy, and data visualization.

Table 4
Urban data platforms (Barns, 2018) and our OpenComm.

Category	Data repositories	Data showcase	CityScores	Data marketplaces	Open Community (ours)
Type	Open Data Portals	City Dashboards	Score Cards	Datastores	Data Communication
Objectives	Data service innovation, transparency	Data visibility, transparency	Performance monitoring	Data service innovation	Data curation, Data integration, Data privacy preservation, Data categorization, Data visibility
Examples	New York Citizen Dashboard, Socrata Dashboards, CKAN Dashboards	Dublin Dashboard, London Dashboard, Sydney Dashboard	Boston CityScore, GSC Dashboard	London Dashboard, City Data Exchange	KCMO 311 calls, Census data, Crime data, Land bank
Accessibility	Machine readable	Machine readable & not available	not available	Machine readable	Machine readable & downloadable
Creator	City governments	City governments & Educational institutes	City governments	City governments & Private sectors	City governments & Educational institutes
Performance monitoring	✗	✗	Set targets	Data usage	Data usage
Privacy preservation	✗	✗	✗	✗	Differential Privacy
Data curation	✗	✗	✗	✗	Error, missing value, normalization
Data categorization	✗	✗	✗	✗	Semiautomatic
Data integration	✗	✗	✗	✗	Geocode based
Data visualization	✗	Urban policy priorities	✗	✗	Chart, spatial, temporal

3. Research questions

Open data platforms that facilitate data exchange and integration will serve as critical infrastructure for the development of data-driven techniques that can address a range of smart city concerns and provide decision-makers, academics, and residents with timely and relevant information (Barns, 2018).

We are examining several intriguing problems, including (i) the integration of heterogeneous datasets by *breaking boundaries*, (ii) the taxonomy of public data to increase the transparency and efficiency of 311, recovering or correcting any invalid data or missing values, (iii) identifying any potential issues or risks associated with the integration of data collected in different contexts, and (iv) the visualization of data from complex and diverse infrastructures. Finally, we outline several critical issues that must be addressed while constructing an open data platform.

Q1. How can we establish an open platform that enables efficient and secure interchange of segregated public data for sustainable cities and communities?

If the analysis is to be extended beyond the bounds of a single region (boundary) appropriately, we would be able to automatically exploit relevant data or technical solutions based on massive volumes of diverse and complex infrastructure and community-related data. Furthermore, based on the context and ultimate goals, new data types may be automatically generated from existing ones.

Q2. Are there any reservations about the privacy of publicly accessible data? How is the privacy of publicly available data protected?

We must evaluate the privacy implications of publicly available data. Publicly accessible data is not immune from data privacy concerns. Maintaining privacy in this context is more challenging since it is prone to diminish public confidence. We need to understand the purpose of public domain data, the constraints that govern its use, and the best practices to benefit the citizens. A proper solution should address possible threats to the privacy of public data to foster public trust (Razaque & Rizvi, 2017).

Q3. Is there a taxonomy that is generally applicable to various data sources? Why is a standardized taxonomy necessary? How does one create a unified taxonomy?

We hypothesize that a standardized or generic taxonomy is unnecessary because data contexts vary across time and location. For example, 311 call data is more likely to be associated with specific concerns and responses from a city office in a particular area. However, the process of categorizing could be standardized and automated. The great majority of taxonomies in use today were created manually by domain experts or operational staff. As a result, the manual categorization is highly likely to be incorrect or imbalanced. Thus, an inductive data-driven method is more effective in developing a usable and consistent taxonomy that can be applied to real-world analysis, explanation, prediction, and explanation (Gregor, 2006).

There is no taxonomy or process for consistently classifying 311 call services in practice. As a result, the architecture and methods employed by cities' 311 phone systems vary significantly. These divisions inhibit learning from other cities' experiences and may be prone to or skewed towards representing just a subset of real-world problems. As a result, there may be significant differences between genuine resident complaints and agency responses in a 311 call taxonomy. For example, classifying 1.5 million unique inputs into 84 categories regularly and accurately by operational personnel is difficult for Kansas City's 311 system, let alone organizing the calls in various ways.

Q4. Why is visualization beneficial for interpreting data context? To what degree does it accurately portray the circumstances surrounding the collection of records? Which data visualization strategies are most effective in capturing the contexts of the data? Is there a mechanism for displaying spatial and temporal data in an integrated manner that is useful?

Graphs and maps of time series data are frequently used to depict data from a temporal viewpoint. However, because observing from geographical and temporal perspectives simultaneously offers considerable obstacles, it is critical to identify the most effective methods for displaying essential contextual information. Therefore, we assert that temporal and special data should be logically linked to exhibit or explain the context of the data and data sources.

4. Methodology

We propose novel algorithms and modeling techniques for the proposed open community framework in resolving specific challenges associated with comprehending and utilizing large volumes of diverse and complex community data.

4.1. Data recovery

4.1.1. Fixing missing and invalid values

Public data are likely to contain various semantic and numerical errors and faults introduced by humans and systems. Inadequate evaluation of these issues will result in naive assumptions and erroneous data analysis, particularly troublesome when performing key prediction and decision-making activities. Individuals make several errors in 311 call data, which results in mismatches, erroneous, and duplicate data. These inaccuracies may include typographical errors in addresses, neighborhood names, neighborhood identification numbers, neighborhood mapping, department resolution, category assignment, request type assignment, type/detail description, and geocoding. Additionally, manual errors are frequently reflected in the numerous occurrences that resulted in the data categorization difficulties detailed in Section 4.4, mainly owing to the overwhelming call volume. The challenges get magnified when multiple sources are used.

To address this issue, we designed the boundary assignment algorithm (Algorithm 1), breaking the 311 call data neighborhood borders and the census data block group boundaries and integrating data from disparate institutional sources that have previously determined their own boundaries. There are connections between a location and the boundaries of various districts. If a point is on a boundary, we consider it similar to a border in the actual world. Real-world boundaries shape this perspective. There is a considerable motivation for and benefit from this approach for data integration via boundary mapping.

Algorithm 1 Boundary Assignment

```
Input: set of boundaries B & an input coordinate p
Output: boundary b

procedure AssignBoundary(p, B)                 ▷ boundary map function
    while B ≠ empty do                         ▷ list of polygonal boundaries
        for b in B do                          ▷ iterate through boundary
            for lat, lon in b do               ▷ loop through each coordinate in polygon
                if p[lat] > lat ∧ p[lon] < lon then
                    tl = true                  ▷ top-left check
                if p[lat] < lat ∧ p[lon] < lon then
                    tr = true                  ▷ top-right check
                if p[lat] > lat ∧ p[lon] > lon then
                    bl = true                  ▷ bottom-left check
                if p[lat] < lat ∧ p[lon] > lon then
                    br = true                  ▷ bottom-right check
                if tl ∧ tr ∧ bl ∧ br then
                    return b                   ▷ returns boundary if point aligns
                else return null
    return null
```

Despite its robustness and breadth of coverage, the American Community Survey data is not perfect, as many demographic and economic variables include missing or erroneous values. For example, during the data collection process for socioeconomic variables for Missouri block groups, 322 block groups out of 4506 (about 7%) had incomplete information. These numbers can be negative (−666,666,666), zero, or "N/A". If only one or two records have an invalid value, the record is considered recoverable. However, a record will be deemed invalid and destroyed if the population and other key factors (such as median income, poverty, education, and demographics) return invalid values due to the neighborhood's scarcity of data.

Around 7% of the socioeconomic data in the American Community Survey (5-Year) has erroneous or missing values; eliminating them may result in the loss of useful information. Given that most of these errors originate in the median home value field, the most logical course of action is to recover it, given the rest of the provided information. The quickest and most straightforward method is to perform multivariate linear regression on this dataset. The median house value is the dependent variable, and the remaining variables are independent. However, this strategy has the disadvantage of misfitting non-linear relationships between variables, prompting us to design another basic yet effective alternative. Given the data space for diverse and complex infrastructure- and community-related data, we developed deep learning-based algorithms based on state-of-the-art research for recovering missing values or correcting invalid data (Khayati, Lerner, Tymchenko, & Cudré-Mauroux, 2020; Tang et al., 2020) that are more effective than conventional approaches such as regression or mean averaging.

A deep neural network (DNN) model was developed to recover missing values or handle erroneous data. The DNN is composed of nine fully connected layers (each of which comprises 32, 64, 128, 256, 256, 128, 64, 32, 16 neurons) and nine activation functions (e.g., ReLU and Leaky ReLU), totaling 153,473 trainable parameters. The model is meant to perform multivariate regression on missing median home value data to impute and recover it. 70% of Missouri's 4452 complete block group data are used for training, 10% for validation, and 20% for testing. All block group values are normalized to their corresponding population data and rescaled appropriately between 0 and 1 (see Section 4.1.2). The model was trained using 100 epochs, a 64-batch size, a 0.0001 learning rate, and an Adam optimizer. The loss function was defined as the Mean Squared Error (MSE). Fig. 3 illustrates the architecture overview and training session.

4.1.2. Normalization and sampling algorithm

We encountered significant data heterogeneity or imbalance concerns with this complex infrastructure data. To address these issues, we developed robust sampling techniques. In addition, we extensively used data enhancement and normalization techniques established by the Machine Learning and statistics communities. To enable the network to train and converge more quickly and with fewer epochs, all data must undergo a two-step preparation process: normalization and scaling to ensure that all input values are between 0 and 1. First, economic, ethnic, population, and education counts are divided by the total population statistics to obtain a frequency proportional to the total population. Following that, a conventional min–max scaler is applied to all numeric data to guarantee that each variable's min and max values remain securely inside the range of 0 and 1. There are several well-known normalization approaches, including Min–Max (Eq. (1)) and Z-score (Eq. (2)), where a value O_v, Min, Max, Mean, Standard deviation are given.

$$N_{mm} = \frac{O_v - Min}{Max - Min} \tag{1}$$

$$X_{norm} = \frac{X - \mu}{\sigma} \tag{2}$$

We applied the practical data sampling approaches that we previously developed using region-based clustering and sampling methods to the data distribution in multidimensional space (Gaikwad, 2020). In addition, it optimized the data utilization during training to increase the model's accuracy. Algorithm 2 describes our technique for normalizing and scaling.

4.2. Data integration with boundary assignment

To avoid geographic mismatches and ensure that numerical data is consistent across locations, we assess the degree of composition at each level and identify all potential values. For instance, if a neighborhood has three-block groups, it is required to consider the conditions associated with each of the three-block groups. As a result, the neighborhood's status will be approximated using the mean of the metrics (i.e., socioeconomic).

To address this issue and provide an intuitive way to interpret the data, our project analyzed both governmental and non-governmental sources simultaneously by extracting boundaries, automatically marking critical regions, and developing optimization algorithms to accurately map coordinates and merge datasets consistent with their corresponding physical constraints. As a result, substantial interoperability

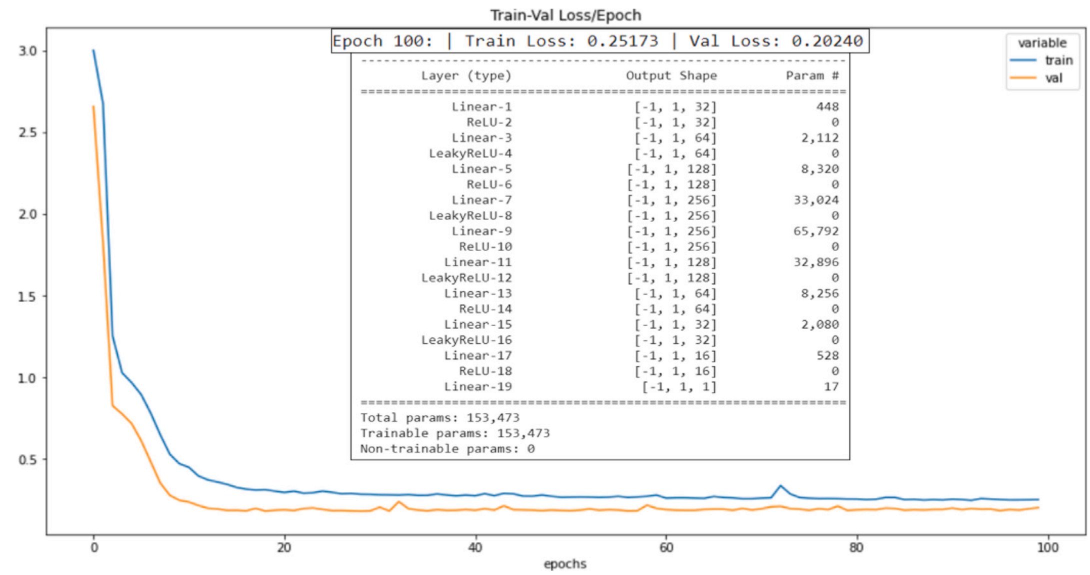

Fig. 3. Model architecture summary and training/validation accuracy. This model is used to impute missing and invalid socioeconomic data from the Census API.

Algorithm 2 Data Normalization and Scaling

Input: X^{nxm} ▷ matrix of n rows x m features
Output: $\hat{X}^{nxm}$ ▷ normalized matrix
Imports: $MeanScaler$ ▷ *Scikit Learn* library

procedure Normalize-Scale(X^{nxm})
  $\hat{X} \leftarrow Arr[n, m]$
  for i in n **do**
    $p \leftarrow$ X[*population*] ▷ total population of neighborhood
    $r_{norm} \leftarrow$ Arr[m]
    for j in m **do**
      $v_{norm} \leftarrow X[i,j]/p$ ▷ normalize values based on neighborhood's population
      r_{norm}.push(v_{norm})
    $\hat{X}$.append(r_{norm})
  $MeanScaler$.fit($\hat{X}$) ▷ re-scale values w/t μ & σ
  $\hat{X} \leftarrow MeanScaler$.transform($\hat{X}$))
  return $\hat{X}$

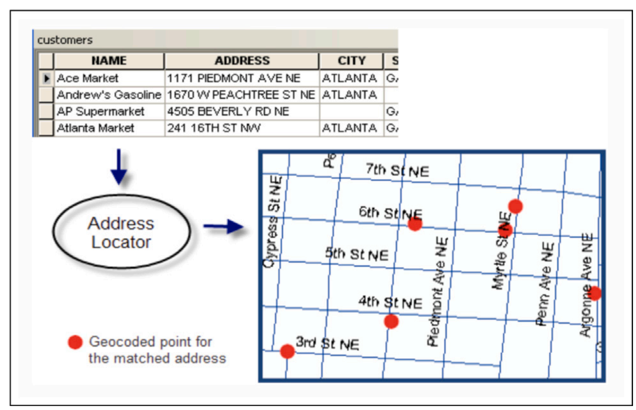

Fig. 4. ArcMap geocoding example.

can be achieved by utilizing five geographic boundaries. Neighborhoods, block groups, police divisions, council districts, and school districts are all examples of sub-localities. More precisely, in Kansas City, police divisions, council districts, and school districts all share the same outer boundaries: the city's overall form, despite their distinct internal boundary configurations (see Fig. 2).

When working with these boundaries, the primary challenge is to handle numerical data with care and caution, particularly when determining socioeconomic characteristics (such as median income, median home value, education, ethnicity, and housing occupancy) for the various locations. In addition, it can be challenging to infer these values from their original bounds to their new ones, as many of these borders do not have a clear one-to-one relationship. For example, a block group or a portion of a block group may be part of many neighborhoods. In addition, an area may encompass multiple school districts, council districts, and police divisions. Thus, if not taken into account, these coincidences add complexity to the task and result in data inconsistency. This issue should be resolved to understand neighborhoods and do predictive analysis on individuals or groups of boundaries. Our boundary mapping technique contributes to the solution by considering the type of boundary data; for example, it might be school, police, or council districts.

The geocoding (as shown in Fig. 4) is generated for a given address, a set of latitude and longitude (x, y) coordinates of a physical address, which can then be plotted or displayed on a map. A boundary is represented as a series of geocodes to specify a home region against neighboring regions. Table 1 shows some boundary examples in KCMO. Since a single region faces multiple neighboring regions, each face (line) of a boundary is represented as a pair of latitude and longitude (x_1, y_1), (x_2, y_2). Each line divides each face of a boundary into two regions. In reality, the shape of a region is not a rectangle or square, and thus, each region has many interfaces with other regions. Therefore, the boundary of each region is very complex. In the boundary assignment algorithm, let B be a master geocode set including all geo-boundaries for a given data space like block group of the census data. A block group boundary $b \in B$ is represented as a set of geocodes $b = \{(x_1, y_1), (x_2, y_2) \ldots (x_k, y_k)\}$, where k faces in the census block group b. Let p be a geocode (x_p, y_p) of an address from a 311 call record. The boundary mapping function h could be represented as follows:

$$h(p, b) = tl(p, b) \wedge bl(p, b) \wedge tr(p, b) \wedge br(p, b) \quad (3)$$

According to Eq. (4), a geocode (x_p, y_p) must be mapped at least as a point in the given boundaries which sits with lower latitude and greater

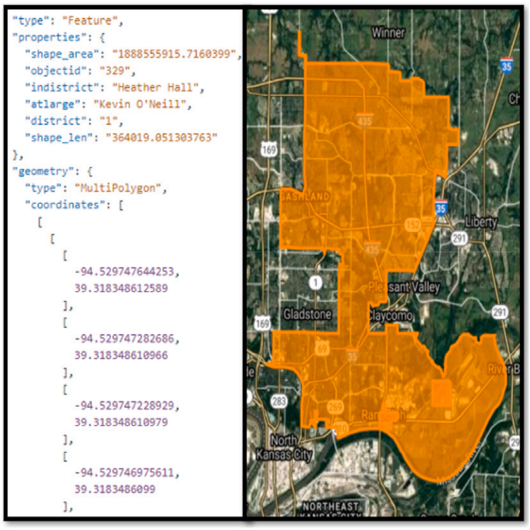

Fig. 5. Example of an geospatial boundary in map and GeoJSON.

longitude.

$$tl(p,b)=\begin{cases}1 & \exists i: 1\le i\le k \ \mid \ [x_i,y_i]\in b\\ & [x_p,y_p]\in p\\ & (x_p>x_i \ \& \ y_p<y_i)\\ 0 & Otherwise\end{cases} \tag{4}$$

According to Eq. (5), a geocode (x_p, y_p) must be mapped at least as a point in the given boundaries which sits with lower latitude and lower longitude.

$$bl(p,b)=\begin{cases}1 & \exists i: 1\le i\le k \ \mid \ [x_i,y_i]\in b\\ & [x_p,y_p]\in p\\ & (x_p>x_i \ \& \ y_p>y_i)\\ 0 & Otherwise\end{cases} \tag{5}$$

According to Eq. (6), a geocode (x_p, y_p) must be mapped as at least a point in the given boundaries that sits with greater latitude and greater longitude.

$$tr(p,b)=\begin{cases}1 & \exists i: 1\le i\le k \ \mid \ [x_i,y_i]\in b\\ & [x_p,y_p]\in p\\ & (x_p<x_i \ \& \ y_p<y_i)\\ 0 & Otherwise\end{cases} \tag{6}$$

According to Eq. (7), a geocode (x_p, y_p) must be mapped at least as a point in the given boundaries that sits with greater latitude and lower longitude.

$$br(p,b)=\begin{cases}1 & \exists i: 1\le i\le k \ \mid \ [x_i,y_i]\in b\\ & [x_p,y_p]\in p\\ & (x_p<x_i \ \& \ y_p>y_i)\\ 0 & Otherwise\end{cases} \tag{7}$$

This algorithm was created to associate each occurrence of a data point with all appropriate boundaries (e.g., neighborhood, block group, police district) depending on its address. The Google Geocoding API (Google Inc, 2021) is used to convert addresses to geographic coordinates (latitude and longitude), which can then be used to position the map for forward geocoding, reverse geocoding, and searching data. This API is sufficiently strong to correct for address typos or to create based on incomplete information. However, there are some duplicate records in the database due to using the same geocodes across multiple forms.

However, some duplicate records exist due to directing the same geocodes to multiple forms. Duplicated geo points do not resolve the geolocation of the area of our interest when such records are considered. We used the street address and zip code to locate and determine the geo points (latitude and longitude) using the Google geocoding API to overcome this issue. In addition, we used a boundary mapping approach to resolve the neighborhood of our interest because we have geolocation. As a result, we have recovered nearly all of the documents associated with that particular issue.

The boundaries of five distinct organizations in KCMO are depicted in Fig. 2, including 14 school districts, 6 police districts, 6 council districts, 246 neighborhoods, and 465 block groups. These borders are highly overlapping but incompatible. For example, a block group or a portion of it may be part of many neighborhoods, and a neighborhood may encompass multiple school districts, council districts, and police divisions. Fig. 5 shows an example of a geographical boundary in map and GeoJSON views. By taking this challenge into account, the boundary data associated with the boundary type, our boundary mapping technique in Algorithm 1 assists in resolving the problem.

4.3. Privacy preserving for public data

In this paper, we propose differential privacy (DP) to protect the sharing of joint datasets and deep learning (DL) models, as well as differential access to DL models for public trust. First, differential privacy is effective for mitigating risk when statistical databases are joined (Dwork, 2008). We studied advanced differential privacy preservation strategies in DL models (Abadi et al., 2016; Phan, Wu, Hu, & Dou, 2017) that have demonstrated remarkable results across a broad range of applications.

Second, while data is already in the public domain, differential access to the data or models should be protected by registration. Authorized researchers will have complete access to models, allowing them to replicate the models on other datasets. Only relevant change agents, such as community leaders, city management, and other stakeholders, should access data or models.

Algorithm 3 Data Transformation with Differential Privacy

Input: X^{nxm}, ϵ
▷ matrix of n rows x m features and a privacy budget
Output: DP-normalized matrix
Imports: $DPScaler$, $MeanScaler$
▷ IBM's *diffprivlib* and *Scikit Learn* library

procedure DP-Transformation(X^{nxm}, ϵ)
$MeanScaler$.fit(X^{nxm}) ▷ find true μ and σ
$Z \leftarrow DPScaler$.fit_transform(X^{nxm}, ϵ)
▷ scores relative to μ_{DP} and σ_{DP}
$X_{DP} \leftarrow MeanScaler$.inverse_transform($Z$)
▷ generate new values from true μ and σ
return X_{DP}

Privacy is integral in every study and analysis regarding publicly available data collection and publication. Even though the Census Bureau only provides the aggregated demographic data within its geographic hierarchy, a large portion of the data can still be recovered and definitively traced back to individual identities by using just the mean and median values, typically by powerful computers and algorithms (i.e., Database Reconstruction Theorem and Attacks) (Dinur & Nissim, 2003; Garfinkel, Abowd, & Martindale, 2019). Over the past three decades, the Census Bureau has attempted to reduce the risk of full exposure by adding random noise to their publicly-available datasets such as the American Community Survey, tables, but failed to either effectively quantify how much confidentiality has been guaranteed or how much has been violated for an individual.

Table 5
Example of 311 calls' categorization issues.

Issue	Description
Confounded categorization	There has been duplication in the categories, leading to confusion and a uniform path for creating the 311 service requests and analyzing. Some examples are Storm Water, Storm Water/Sewer, Sewer. A review of these categories tells us they are the same but with multiple entries, and by looking at the user's request, they are all addressed by the same department. These category types are also changed with time, lacking a uniformity in the data to present with automated solutions.
Heterogeneous categorization	On a high level, the data does not indicate the semantics of each field other than storage. From a logical point of view, the data seemingly interweaves the problem description and the solution together for each record in the form of the six columns above. In Fig. 7, the Request Type column is either a concatenated sequence of or an origin of the next three columns, Category, Type, and Detail. The taxonomy further shows that most links are a 1-to-1 relationship (highlighted in yellow). The problematic links indicate these request types only exist when the same category, type, and detail are present, suggesting a strong recurring pattern across the dataset that can lead to space inefficiency and low column significance, except for certain special cases (such as Property Violations). Therefore, the Request Type column is primarily deducible.
Inconsistent categorization	Inconsistency occurs when instances of the same category are assigned to multiple departments. In the taxonomy of Fig. 9(a), each category is known to have established a connection to more than one associated department. This problem can contribute to a lack of organization and centralization across the 311 solution domain because no primary department takes full responsibility for a certain problem.
Overlapping categorization	The Request Type and Category column contain the same information in exceptional cases. In Fig. 9(b), Property Violations exists on two levels (request type and category), which shows that it can appear twice potentially in one 311 record, contributing to further redundancy and ambiguity.

The 2020 census data represents a paradigm shift in terms of protecting individuals' privacy, as it incorporates a new disclosure avoidance system (DAS) that leverages *Differential Privacy* (DP) to objectively limit the amount of precise information given to the public. However, because this work also makes use of prior-year census data, specifically from 2013 to 2019, these demographic statistics must undergo a well-defined, privacy-protected pre-processing procedure. As a result, we chose to apply DP directly on our training data, shielding the Deep Neural Network (DNN) from the actual statistics. Additionally, we show the findings and evaluations of multiple experiments using a range of different values for the privacy budget ϵ.

The DP transformation method is described in detail in Algorithm 3. Due to the fact that DP is performed during the pre-processing step, all training data is normalized and rescaled accordingly. As with Section 4.1.2, we normalize the socioeconomic metrics using a DP standard scaler, which requires DP calculation. The *fit* function computes the differentially private mean and standard deviation for a DP standard scaler. Following that, the standard score of sample x is generated using the DP mean and standard deviation in the scaler's *transform* function using the formula indicated in Eq. (8).

$$Z_{DP} = \frac{x - \mu_{DP}}{\sigma_{DP}} \tag{8}$$

where μ_{DP} represents the differentially private mean and σ_{DP} represents the differentially private standard deviation. The value of ϵ, defined as the privacy budget allocated to learning about the mean and variance of the training dataset, is a significant aspect in determining the level of privacy in this scaler. The privacy budget is then divided evenly between mean and variance learning. In other words, it denotes a threshold for privacy loss that an individual is willing to accept. A low value for ϵ indicates a trade-off between privacy and accuracy. Additionally, it can be loosely referred to as the scaler's overall sensitivity to individual differential data changes (such as removing and adding a record to the data samples). For instance, a ϵ of 0.01 constrains the scaler's budget for learning about the mean and variance of the entire training data. Still, a value of 1 makes subsequent modifications to the data unimportant. The *fit_transform* is a one-step transformation that computes the mean and standard deviation, as well as centers and scales the data around these parameters using ϵ. Apart from being utilized for training a DNN regressor in Section 5.3, the Z_{DP} scores are used to depict the data in Fig. 6 to effectively demonstrate the differences in the distributions of the original and modified data. As the product of the *inverse_transform method* (which assists in transforming the Z scores of other standard scalers into a differentially private output), the returned matrix of Algorithm 3, X_{DP}, is used to visualize the map.

With a range of ϵ differing in the magnitude of 10, Fig. 6 presents the kernel density estimate (KDE) distribution for data observations from nine Census variables (*Black or African American alone, White alone, Hispanic or Latino, Asian alone, Bachelor's degree or higher, Households with income below poverty, Total Renter Occupied, Total Vacant, and Total unemployed*). These observations are taken from 4,185 valid block groups across all counties in Missouri. As a result, this visual example exhibits a generalizable pattern: the lower the privacy loss value, the greater the disparity between the DP and normalized data. Additionally, ϵ determines the maximum distance between all of the dataset's records. A small ϵ value causes the general distribution of data to become more clustered around the mean and hence more comparable to one another (green curve with $\epsilon = 0.01$). In comparison, a large value produces a more similar distribution to the original normalized distribution (purple curve with $\epsilon = 100$). In Section 5.3, we will present maps comparing the original and transformed Census datasets for a range of ϵ values.

4.4. Data categorization

Due to the nature of public data, address mapping has been hampered. As an illustration of this problem, 42K records contain an identical geo point. To discover the actual address associated with these entries, a reverse google mapping was performed, yielding a range of addresses identified from the given Geolocation and then corrected using data collected from the Google Map API (Google Inc, 2021).

The current 311 data structure has six columns about case details. These are *Category*, *Type*, *Request Type*, *Department*, *Work Group*, *Detail*. However, the meanings of these fields were not fully elaborated, creating uncertainty for researchers and developers. Moreover, the fields contain heavily mutual information for many records, and inconsistency possibly occurred from manual input. To clarify the ambiguity and increase the interpretability, we propose developing a unified categorization to make the data more distinguished while delivering the exact significance of the case. For example, there are the six different categorizations in KCMO 311 requests. Specifically, we can simplify these into two different categorizations such as the problem domain (*Category*, *Type*, *Detail*, *Request Type*) and solution domain (*Department*, *Work Group*). The meanings of these fields were not clearly defined, creating ambiguity. The challenges of categorization with complexity are well discussed in Fass and Feldman (2003). Moreover, the data contain errors due to manual input. Table 5 shows some categorization errors, such as heterogeneity, redundancy, and inconsistency; for example, stormwater, stormwater/sewer, sewer look all addressed by the same department but classified with different requests types raised by the user. The proposed algorithm will validate if there is any inconsistency

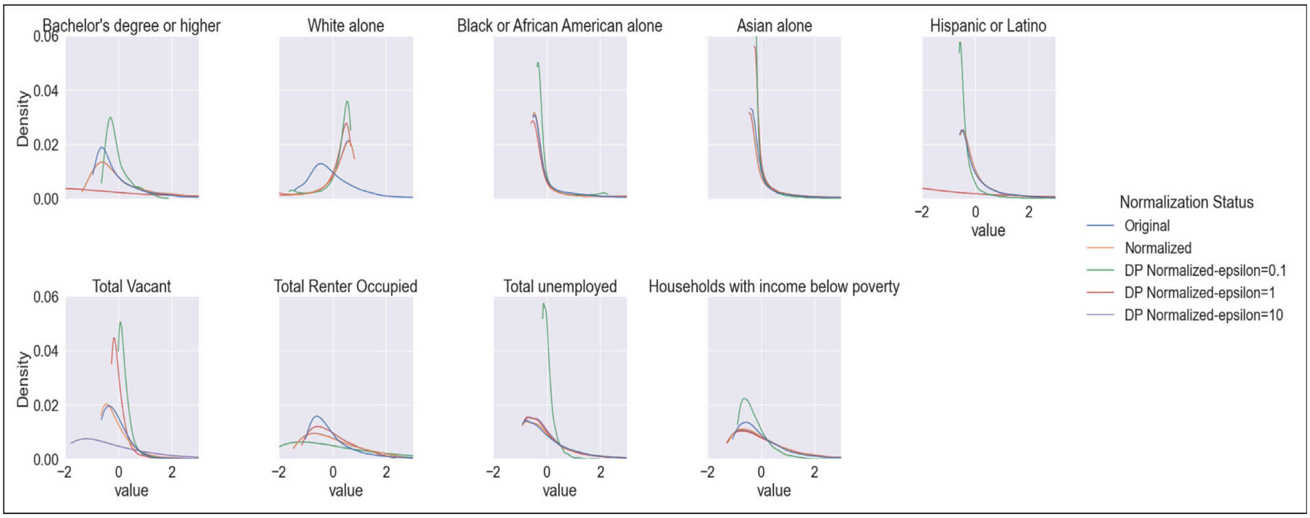

Fig. 6. Data Distribution of Census Variables. Variables are normalized with a mean and standard deviation using a standard scaler. The X-axis depicts mean converted data values with unit variance. The Y-axis depicts the distribution of a variable's z-scores. The different colors denote the various stages of normalization. The differentially private mean and standard deviation are incorporated into DP-Normalization. A lower ϵ value indicates increased privacy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

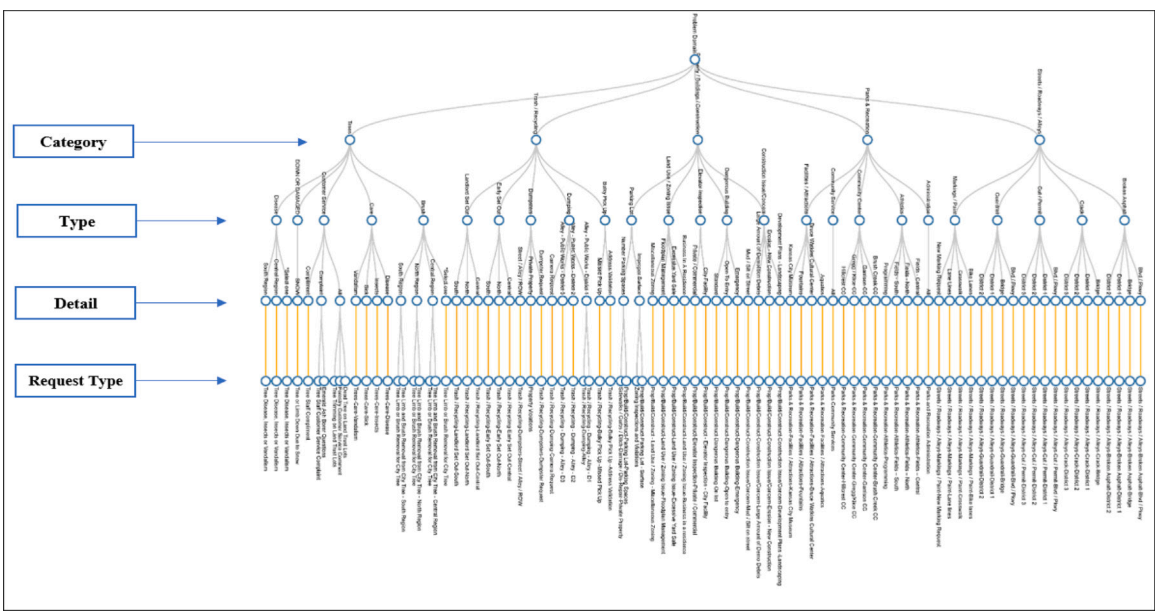

Fig. 7. 311 Call Problem Taxonomy (Tree View). The yellow links indicate a potential issue with diverse categorization in the existing 311 structure, where request types aremostly composed of the concatenation of category, type, and detail, resulting in mostly direct one-to-one linkages. The Problem Taxonomy is detailed in Fig. 25, as is the highresolution figure and its link. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

or duplication in the categorizations. Figs. 7 and 8 show the problem taxonomy and solution taxonomy of the 311 calls, respectively.

Intuitively, the solution to the heterogeneity and redundancy is to delete the *Request Type* column entirely from the original structure, thereby increasing data size and coherence by keeping the fields as specific as possible. On the other hand, inconsistent and perplexing categorization necessitates additional analysis and reasoning. As a result, an efficient solution requires extensive restructuring of the current structure, including disintegration and rearrangement. We proposed a *Problem and Solution* substructure for consistency, with each column containing two columns with varying degrees of specificity. The *Problem* domain will have a high-level overview of the problem and a low-level label for it.

Following that, the *Solution* domain will be assigned depending on the problem's specifics, often in the form of a department or workgroup accountable for the reported issue. Additionally, to tackle the 81-category problem, we created two semantically correct taxonomies at the category and department levels (see Fig. 7). We added 21 new values to describe the problem's overall high-level label while retaining the 81 original categories as low-level field possibilities. Similarly, in Fig. 8, there are 7 broad divisions and 26 subordinate divisions. Finally, this structure will aid in the organization of records by making them space-efficient and self-explanatory.

5. Results and evaluation

We demonstrate the feasibility of the proposed OpenComm framework through a case study involving public datasets (described in Table 2), such as data pulled from 311 calls and the American Community Survey. The results demonstrate that the suggested framework

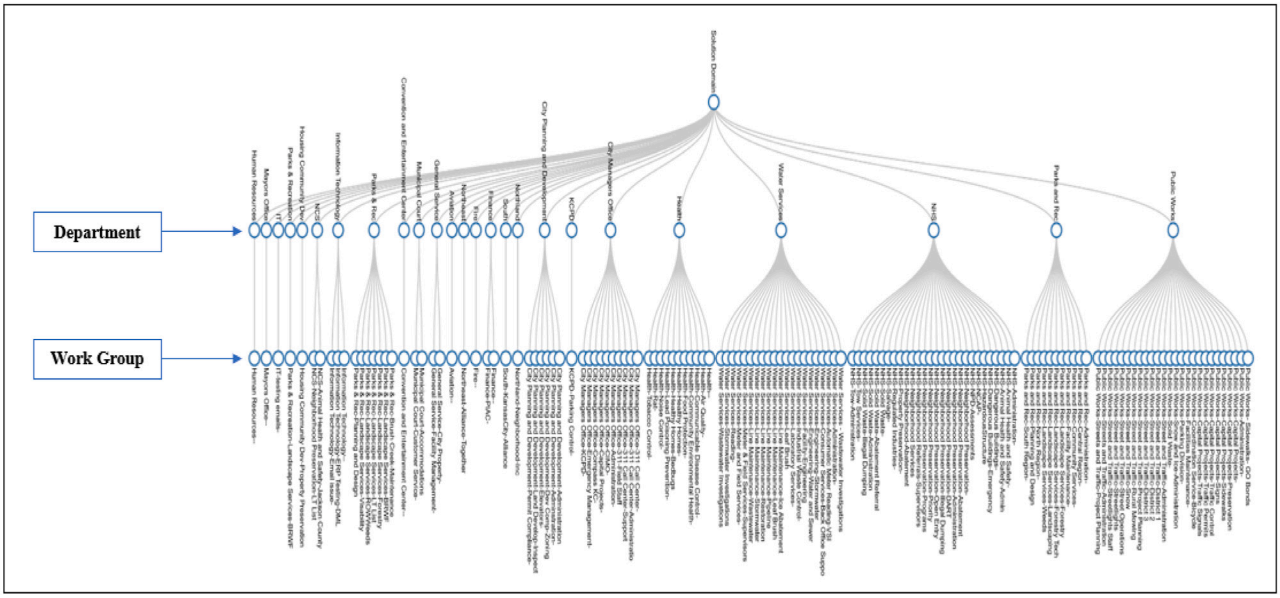

Fig. 8. Call Solution Taxonomy (Tree View): The details of the Solution Taxonomy, the high resolution figure is available in Fig. 26.

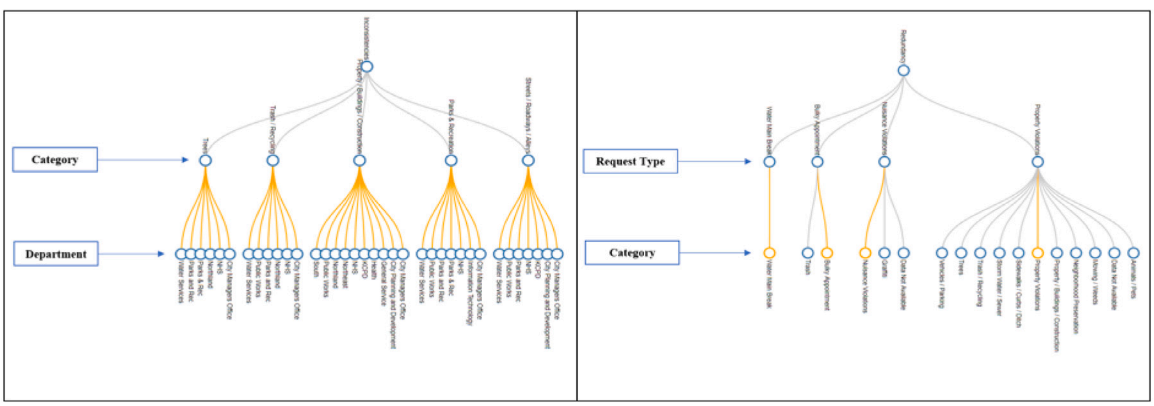

Fig. 9. 311 Data Categorization Issues: (a) Inconsistency — a single issue falls under the purview of many departments; (b) Redundancy — the same terms are used in both the request type and the category descriptions. The yellow links indicate an inconsistency or redundancy issue of the current 311 categorization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 4 Categorization Algorithm

Input: D, c_1, c_2, c_3, c_4 ▷ data and level of specificity
Output: 4-level taxonomy

procedure TAXONOMY(Input)
  $T = \{i: \text{new dict()} \mid i \in \mathrm{D}[c_1]\}$ ▷ root represents field c_1
  for $r \in D$ **do** ▷ record in data
    $k_1 \leftarrow r[c_1]$ ▷ identifier at level 1
    $k_2 \leftarrow r[c_2]$
    $k_3 \leftarrow r[c_3]$
    $k_4 \leftarrow r[c_4]$
    $l_1 \leftarrow T[k_1]$
    if $k_2 \notin l_1$ **then** $l_1[k_2] = \text{new dict()}$
    $l_2 \leftarrow l_1[k_2]$
    if $k_3 \notin l_2$ **then** $l_2[k_3] = \text{new set()}$
    $l_3 \leftarrow l_2[k_3]$
    if $k_4 \notin l_3$ **then** $l_3 \cup k_4$
  return T ▷ taxonomy is modified in-place

effectively cleans, integrates, preserves, organizes, and visualizes data in publicly available datasets.

It is critical to evaluate the model's correctness to determine how well it predicts the models in our study, which include missing value recovery (Section 5.1.1), data normalization (Section 5.1.2), and differential privacy (Section 5.3). Our regression model will analyze errors by studying how predictions are made. In regression analysis, the most often used metrics for model performance are (i) Mean Absolute Error (MAE), which is calculated by averaging the absolute difference between the original and predicted targets; (ii) Mean Squared Error (MSE) is the average difference between the original and anticipated targets; (iii) Coefficient of Determination (R^2) is a number between 0 and 1 that indicates how well the values suit the original values. The greater the coefficient value, the more accurate the model.

5.1. Data recovery: Results and evaluation

5.1.1. Missing value recovery

At the end of the training, the training loss and R^2 were 0.22936 and 0.79303, while validation loss and R^2 were 0.18126 and 0.78, respectively. We analyze the training and testing sets to ensure that our predictive model is not too fitted to the training data. The test

Table 6
ML results for missing median home income (Partially Missing Value) recovery.

Type	Data	MSE loss	R^2
Training	70%	0.2242	0.7952
Validation	10%	0.1781	0.7919
Testing	20%	0.2033	0.7795

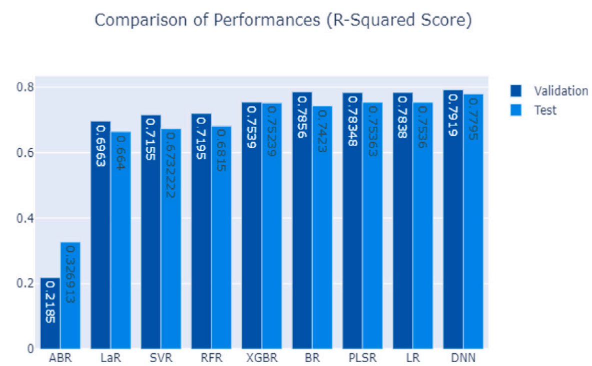

Fig. 10. Comparison of performances between different recovery methods (including traditional regression techniques and DNN) in terms of R^2.

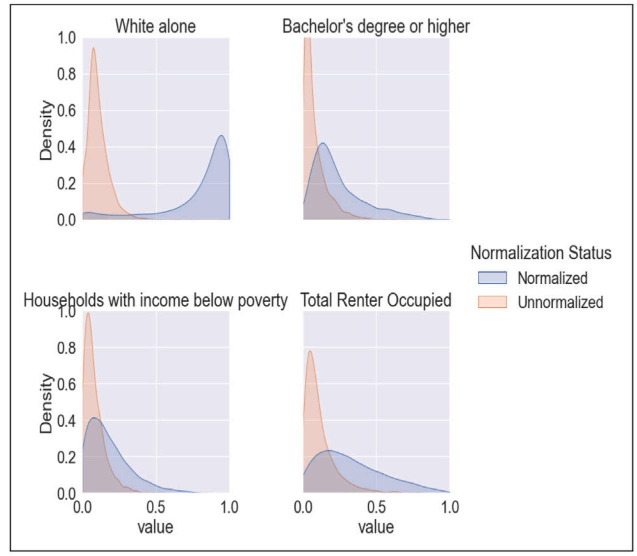

Fig. 11. Comparison of Census Data Distribution by Normalization Status. Kernel Density Estimation is used to represent the distributions (KDE). To emphasize the effect, values are generated using a Min–Max scaler.

loss and correlation coefficient were 0.20909 and 0.77326, respectively (see Table 6 and Fig. 10). The trained model is then utilized to recover the missing median home values for each of the 36 block groups by utilizing the remaining fields (ethnicity, education, vacancy, poverty, and income).

The DNN model outperforms unsupervised techniques such as collaborative filtering and standard statistical methods (such as linear regression or means averaging), which are incapable of efficiently capturing non-linear correlations between variables.

We chose eight alternative regression models for quantitative comparison, including linear regression, bagging, boosting, and optimal variations. Linear regression (LR) is usually considered first when confronted with a regression problem due to its interpretability. Despite its simplicity, the model performed effectively, earning R^2 scores of 0.7834 for validation and 0.7536 for testing. Lasso Regression (LaR) is preferable because it is a variant of linear regression that incorporates data reduction and regularization. We utilized the lowest MSE with 10-fold cross-validation on the training data to calculate the number of components for dimensionality reduction in Partial Least Squares Regression (PLSR). The results are comparable to those obtained using linear regression, the second-best performing approach.

For standardization and consistency, tagging and boosting families use the Decision Tree Regressor as the basis estimator. There are two types of boosting regressions: AdaBoost Regression (ABR) and Extreme Gradient Boosting Regression (xGBR). Boosting makes corrections to previous learners using the weights from the prior regression phase. Because xGBR is employed significantly more frequently in competitions due to optimization, its performance is superior to more fundamental techniques such as AdaBoost, which performs poorly with 0.2185 and 0.3269 on validation and testing, respectively. However, the accuracy of xGBR is still inferior to that of linear regression.

Bagging is a family of regression techniques, including Random Forest Regression (RFR) and Bagging Regression (BR). Unlike boosting, these bagging algorithms operate parallelly by training many weak learners and then merging them during the inference phase.

BR surpassed xBGR in terms of accuracy (0.7856 vs. 0.7539), although this tendency was not evident in testing (0.7423 vs. 0.7523). Despite this, neither of these strategies outperformed LR in any phase. Support Vector Regression obtained 0.7155 and 0.6732 for training and testing accuracy but fell short of producing a compelling outcome compared to the other methods.

As a result, none of the regression strategies stated previously outperformed our DNN (see Table 7). They appear insufficient to interpret the census data's more complex properties. The support vector, boosting, and bagging estimators are too shallow for the provided data. Although LR and PLSR are similar, DNN can take advantage of the non-linearity, increasing validation and test scores.

5.1.2. Data normalization: Results and evaluation

As illustrated in Fig. 11, normalization does have an effect on the data properties during the pre-processing step, which can influence the deep learning process and the overall improvement of the data recovery. Additionally, the variables whose distributions shift significantly include *White alone*, *Total population with Bachelor's degree or above*, *Total Renter Occupied*, and *Households with income below poverty*, with a magnitude of at least 0.1 difference in both proportion to the total population (X-axis) and density (Y-axis). For example, a majority of the people in 4452 complete block groups in Missouri is white, as the peak density of the *White alone* variable is roughly 0.95 on the X-axis. However, the density estimation on the unnormalized population reveals a considerable disparity because the *White alone* population count was only subjected to a naive rescaling pre-processing step in $[0, 1]$. It also was not calculated relative to the total population of its respective block group, resulting in a misleading and inaccurate representation of the entire demographics.

5.2. Data integration: Results and evaluation

The algorithm for assigning boundaries is designed in the following manner: The algorithms incorporate massive amounts of diverse and complex data about infrastructure and communities. A boundary example is shown in Fig. 5. This figure depicts its characteristics, including information about the locality (e.g., neighborhood) and the geometry, a multi-polygonal geographic representation using latitude and longitude coordinate pairs. Each data point (311 call request) is assigned to its limits based on its geocode (latitude and longitude coordinate pairs). We combined approximately 1.5 million 311 call records collected during 2007 and 2020 with various KCMO datasets, including the Census, Crime, Poverty, and Land Bank (Housing).

Geographic boundaries are primarily derived from the publicly accessible webpage OpenData KC. The site provides data for the entire

Table 7
Evaluation of performance of missing data recovery between multiple machine learning algorithms, in terms of MAE, MSE, and R^2.

Algorithm	R^2		MAE		MSE	
	Validation	Test	Validation	Test	Validation	Test
Linear Regression (LR)	0.7834	0.7536	0.3202	0.3705	0.2190	0.2708
Random Forest Regression (RFR)	0.7195	0.6815	0.3820	0.4245	0.2838	0.3500
Extreme Gradient Boosting Regression (xGBR)	0.7320	0.7524	0.3307	0.3486	0.2712	0.2722
Support Vector Regression (SVR)	0.7155	0.6732	0.3309	0.3957	0.2879	0.3592
AdaBoost Regression (ABR)	0.2186	0.3269	0.7616	0.7549	0.7907	0.7398
Bagging Regression (BR)	0.7856	0.7424	0.3147	0.3510	0.2169	0.2832
Lasso Regression (LaR)	0.6963	0.6640	0.3631	0.4218	0.3073	0.3693
Partial Least Squares Regression (PLSR)	0.7835	0.7536	0.3202	0.3705	0.2190	0.2708
Deep Neural Network (DNN)	**0.7919**	**0.7795**	**0.2838**	**0.3143**	**0.1781**	**0.2033**

Table 8
Example of invalid data in 311 calls and Census Bureau.

Error	Description	Approach	Field	Original	Recovered
Typo	Misspelling of street name, city name, zip code, neighborhood name, etc.	Google geocoding API to resolve actual address pertaining to geo location	address tag	777,099	729,432
Mismatch	Mismatched between neighborhood name vs. neighborhood ID	Assignment algorithm for determining neighborhood ID and name.	nbh_id, nbh_name	25,821	15,673
Assignment Error	request-type vs. type, request-type vs. categories, categories vs. department, categories vs. work group	categorization for fixing category and department assignment	department, category, type, work group, retail, request type	263,796	263,796
Out of boundary	The geocode of the call is out of KCMO boundaries.	Boundary Assignment algorithm for neighborhoods and blockgroups (if a 311 call is out of the KCMO neighborhoods or blockgroup boundaries, no 311 service will be provided.)	address with geocode, nbh_id, nbh_name, blockgroup	47,667	0
Duplicates	Same geocodes for different addresses	Google geocoding API for fixing addresses and geocodes	address with geocode, latitude, longitude	40610	40239
Missing Value	Census: Negative number (−666,666,666), 0, or “N/A”.	Partially Missing Values: 267 via substitution, 36 via DNN	Median home value	303	303
		Heavily Missing Values	Many (e.g., Median Income, Median Home Value)	19	1

city in a variety of formats, both common (such as Comma-separated Values (CSV), Javascript Object Notation (JSON), or GeoJSON) and proprietary (such as ArcGIS GeoDatabases (GDB)). Table 2 lists the KCMO datasets that were used in this study.

The boundary mapping algorithm considers the geolocation from the 311 data and a review of the master data from all specific addresses to the location; considering the parameters of the defined algorithm, it generates the name of the boundary that corresponds to the input geo points. The algorithm is implemented using Python (version 3) library that performs reference checks against the conditions to return the geolocations' boundaries. Because of our method, we could correct some data issues, such as typos, mismatches, assignment errors, duplication, and missing values. In addition, out-of-boundary calls were detected when a call originates beyond the KCMO 311 service boundary. Table 8 displays the number of mistakes before and following the correction of the 311 data and the Census Bureau.

5.3. Differential privacy: Results and evaluation

The DNN model was trained and tested 100 times using the same architecture as described Section 4.1, each time with a different ϵ uniformly distributed over the range $[10^{-2}, 10^{2}]$. Accuracy is depicted in Fig. 12 in terms of R^2 score. In this case, the baseline (dotted orange line) is the DNN trained with regular normalization and without DP, which achieves an accuracy of 0.78019. The blue line denotes the correctness of each ϵ value. While the range is subject to fluctuations, an established pattern should not be overlooked: an exceptionally low value of ϵ results in a strongly negative value, indicating that the model learns input that is significantly diverged from the original. If data is not appropriately represented, it could have a negative impact on the results. A positive number greater than two always reduces accuracy to the baseline, jeopardizing individual privacy. While the chart demonstrates a high correlation between ϵ and accuracy, it also shows a trade-off between absolute confidentiality and model accuracy maximization. A ϵ value of 1 with a score of 0.7115 can serve as an acceptable trade-off point that maintains appropriate privacy while providing statistically relevant data for learning and analysis.

Fig. 13 illustrates a depiction of differential privacy maps with five epsilons in the range [0.01-100] for the median home value in Kansas City, Missouri's 462 block groups. The red color scheme denotes incremental positive numerical differences from the initial values, whereas the blue color scheme denotes negative numerical differences. The intensity of the colors implies a significant divergence from the truth value. A low value of ϵ renders the data useless, as all values are at least 50% larger or smaller than the genuine mean deduced from the highly dark red and blue block groups. As we progress to the right, the colors become lighter, indicating that increasing ϵ results in only a minor modification to the representation of Kansas City's Census data, but may risk leaking sensitive information and attracting unwanted exploitation if not well regulated.

More precisely, Fig. 14 illustrates the maps with five different values of ϵ in the range [0.01-100]: Median Home Value (in dollars) for a single block group. This block group's ground truth median home value is $305,800. After performing DP, values changed in a negative

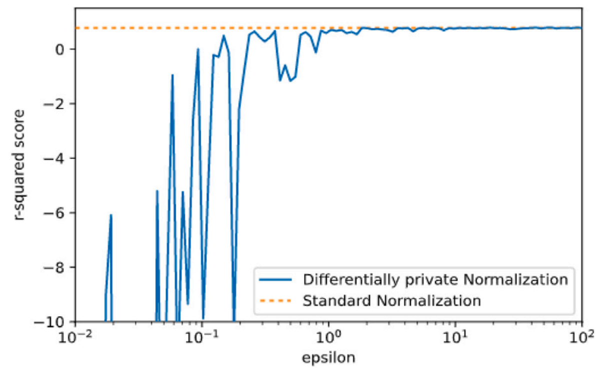

Fig. 12. DNN Test Accuracy with ϵ in the range [0.01 - 100]. A smaller ϵ value means increased privacy but decreased accuracy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

direction with varying magnitudes. This is because all differentially private data are less than the original data but become more accurate as ϵ increases, allowing for the disclosure of important information to the public. As a result, a ϵ of 0.01 results in only $67,545, which is an erroneous representation of the block group by 77.91. On the other hand, a large number of ϵ suggests a generous privacy budget, making the differentially private data much closer to the actual mean (approximately 2.41% difference) but potentially negating the privacy preservation purpose.

5.4. Data categorization: Results and evaluation

The taxonomies shown in Figs. 7 and 8 are meant to help domain experts quickly analyze the data and identify any problems in the 311 call structure. In addition, the obtained statistics and the categorization will serve as a visual overview of a large amount of data, including millions of records and crucial data analytics and system automation.

However, the taxonomy's development is critical because it defines its use and interpretability. Naturally, taxonomy should be top-down (i.e., general topics to specific cases). However, assume we develop a taxonomy based on the existing 311 columns. In that instance, the tree's size, intricacy, and several existing duplicates may result in an incomprehensible structure. This is because ordering by the *Request Type* field (which appears first in the database as a problem assignment, followed by *Category*, *Type*, and *Detail*) results in a significant number of duplicate nodes at each parent level. In addition, *Request Types* are typically unique and have a one-to-one relationship with the other three columns; thus, they are the most particular field. As a result, it should be placed as a leaf node rather than a root.

By designating the *Category* field as the first-level parent, we can ensure that the top-down taxonomy's characteristics are adequately satisfied. For example, in Fig. 15, two taxonomies are generated to describe the identified subset of "Trash/Recycling" category calls, with the left following the original data order and the right following the amended one. Visually, the updated taxonomy more closely resembles a tree where both parents have at least one child. The original taxonomy consists primarily of parents with a single offspring, except the root. Additionally, the latter version is more optimized and comprehensible since it minimizes the number of nodes by increasing their specificity from the root.

The complete categorization findings are presented in Table 9, which includes *Category*, *Type*, *Details*, and *Request Type* for the *Problem* taxonomy, as well as *Departments* and *Workgroups* for the *Solution* taxonomy. By using unique categories as top-level parents, we can reduce the number of nodes by 93.77% at the *Category* level, 73.87% at the *Type* level, and −28.04% at the *Details* level, at the cost of a 20% increase at the *Request Type* level. The overall number of nodes, on the other hand, was drastically reduced by 43.93%.

Table 9
KCMO 311 call taxonomy node count by level.

Type	Field	# of unique values
Problem Taxonomy	Category	84
	Type	293
	Detail	569
	Request Type	1229
Solution Taxonomy	Department	25
	Work Group	143

Similarly, the *Solution* taxonomy is generated using the original data hierarchy's *Department* and *Workgroup*, with the latter representing an improved level of specificity. In total, 143 *Workgroups* are affiliated with 25 distinct *Departments*. Some Departments have only one *Workgroup*, which is frequently a duplicate of the *Department* itself (e.g., *Human Resources*, *Mayors Office*, *Fire*, *Convention and Entertainment Center*, and *Aviation*) while others have well over 15 distinct Workgroups (e.g., *Water Services*, *NHS*, and *Public Works*), possibly due to the diversity of reported cases, complexity of the tasks, and volume of work. By examining the hierarchy and relationships between these categories, we can ascertain which *Department* is currently overburdened and explain why such *Departments* often exhibit a delay pattern due to the high volume of incoming traffic. Following that, an efficient option is to either offer additional resources and support to individuals experiencing stress or to further divide *Departments* into more manageable sub-departments with fewer *Workgroups* to keep workload distribution under control.

To practically eliminate human error from the daily burden, we developed a user-friendly chatbot that automates case reports and data entry by connecting with several state-of-the-art systems that assist users in reporting cases fast and accurately. Additionally, it can accept calls in both text and audio formats, as it is equipped with text-to-speech and speech-to-text APIs for a bidirectional discussion. Further, by utilizing GPS and Geocoding APIs, the chatbot can establish a user's present position and automatically fill in any missing geographic parameters from the user's input, considerably lowering the likelihood of incorrect addresses and typographical errors.

6. Data visualization

To better understand communities' traits and situations, it is prudent to closely examine their distinctive behaviors and patterns. For 311 data, the designs may be reviewed in both the problem and solution domains, which indicate statistically what problems a neighborhood generally confronts and whose agency is responsible and how well it handled the issue.

The problem domain's call category and call frequency are assessed. The call category is normalized based on the total volume of calls (total number of calls). Call frequency is expressed as a percentage of the population of the corresponding neighborhood as a direct indication of its call intensity. The department assignment and response time fields define the solution domain. Department assignment was normalized similarly to the call category. It is expressed as a percentage of total call volume, whereas the response time required to take an additional step is expressed in milliseconds. Because response time values span a large number of days (in days) and contain several errors (e.g., negative numbers) and outliers (e.g., 1000+ to 4000+ days), the field must be regrouped and standardized. As a result, we have reorganized them into periods of 1, 2, 7, 14, 30, 60, 180, 365, and 365+ days to facilitate intuitive reading.

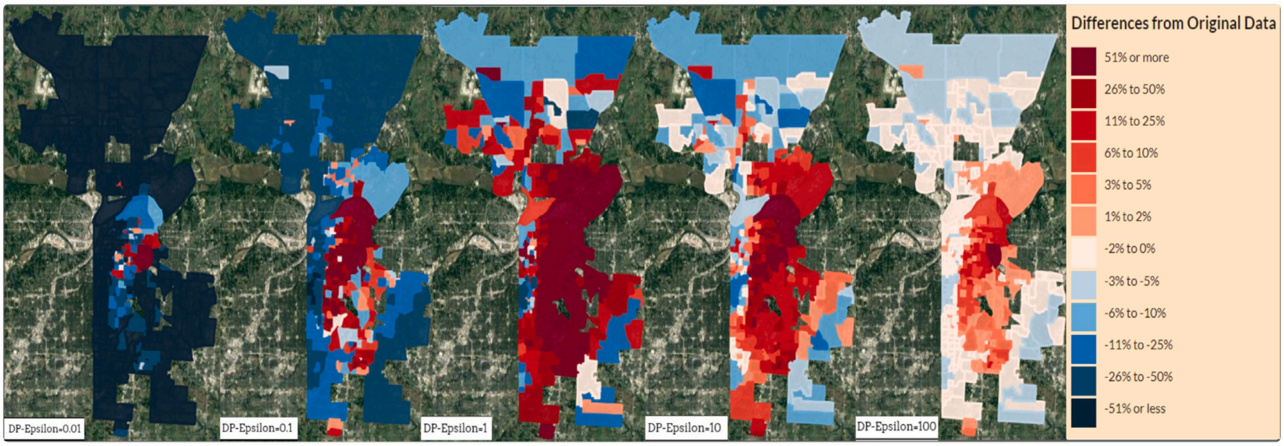

Fig. 13. Visualization of the Median Home Value of 462 Block Groups in Kansas City, Missouri using Differential Privacy with five ϵ in the range [0.01-100]. The red color scheme denotes incremental positive numerical differences from the initial values, whereas the blue color scheme denotes negative numerical differences. In general, a lower ϵ value indicates a greater departure from the original data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

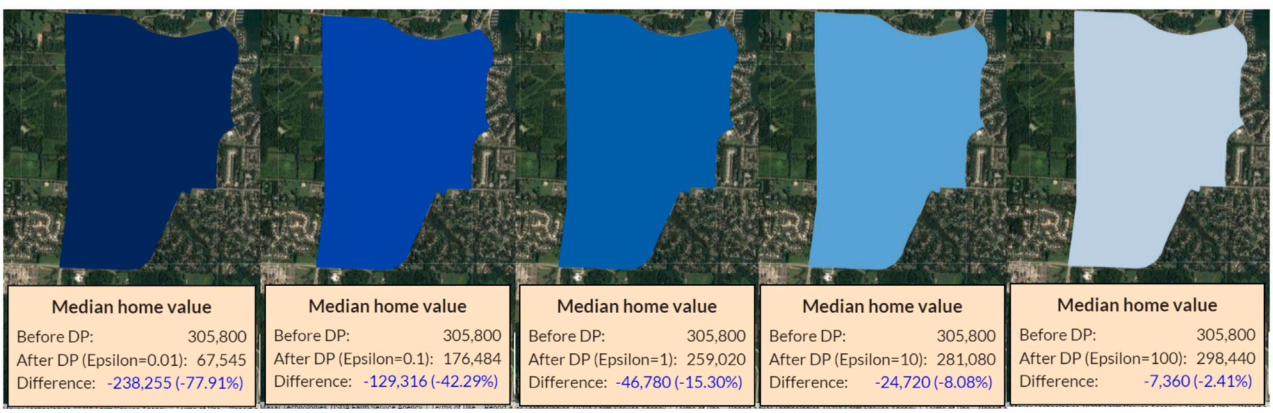

Fig. 14. Differential Privacy Map with 5 ϵ in [0.01-100]: Median Home Value (in dollars) of a single block group. All changes are negative relative to the original data, but their magnitudes differ. A smaller ϵ value suggests a greater divergence.

6.1. Predictive models for public data

To obtain a comprehensive perspective that encompasses numerous potentially interconnected spatial relationships between neighborhoods and block groups throughout Kansas City, we used K-Means clustering to accurately identify which neighborhoods have developed similar behaviors for utilizing and handling non-emergency 311 service calls. After data collection, standardization, normalization, and recovery as described in Section 4.1, clustering is performed based on six distinct conditions: *311 Call Category, 311 Assigned Departments, 311 Response Time, 311 Call Frequency, Census Socioeconomic Metrics, and KCPD Crime Data*, each of which is critical to not only civic leaders responsible for neighborhoods, but also authorities, investors, police, analysts, experts, and social scientists (see Fig. 16). We implemented k from 3 to 9 in our application for comprehension; however, for interpretability, we included just 3 to 6 clusters (see Fig. 17). Each cluster includes a variety of neighborhoods with similar characteristics and behaviors in terms of 311 services, both within and externally. Internal conditions can be classified based on call frequency, call category, socioeconomic factors, and criminality. External departments and response times may be considered allocated because they are not within the jurisdiction of any one neighborhood.

We believe that clustering neighborhoods based on both conditions will enable us to take a more objective look at both the *Problem* and *Solution* domains. A community cannot be evaluated solely based on its issues but on the city's ability to address them on time. As a result, for clarity, we additionally compare these clusters statistically for a given condition. Fig. 18 shows how each of the four groups created using the K-Means clustering approach with *311 Response Time* differs. Based on the date and time, this clearly shows how quickly their difficulties were resolved. While there is no clear-cut ranking system for individual clusters, various variables can be considered since the priority should be based on the urgency of the challenges and the city's resources and assistance. Analytically, we may state that Clusters 1 and 2 are highly similar because their top three time windows are 7 days, 1 day, and on time, even though their top ten time windows are 180 days and 365+ days, respectively. On the other hand, Clusters 3 and 4 exhibit a more obvious differential, with the third most frequent time window for resolving a 311 problem being 14 days (for Cluster 3) rather than 30 days (for Cluster 4). (for Cluster 4). As a result, experts and leaders can use this data to alter their plans and reallocate resources to better meet the demands of their residences.

6.2. Map visualization

Due to the nature of call data, we were able to plot the requests on a map and evaluate them using spatial hierarchical clustering (as shown in Fig. 19). It identifies the cluster's center using a greedy hierarchical clustering technique and collects the volume of calls at a specific zoom level. The locations of the calls become more distinct when the zoom level is increased (more granular).

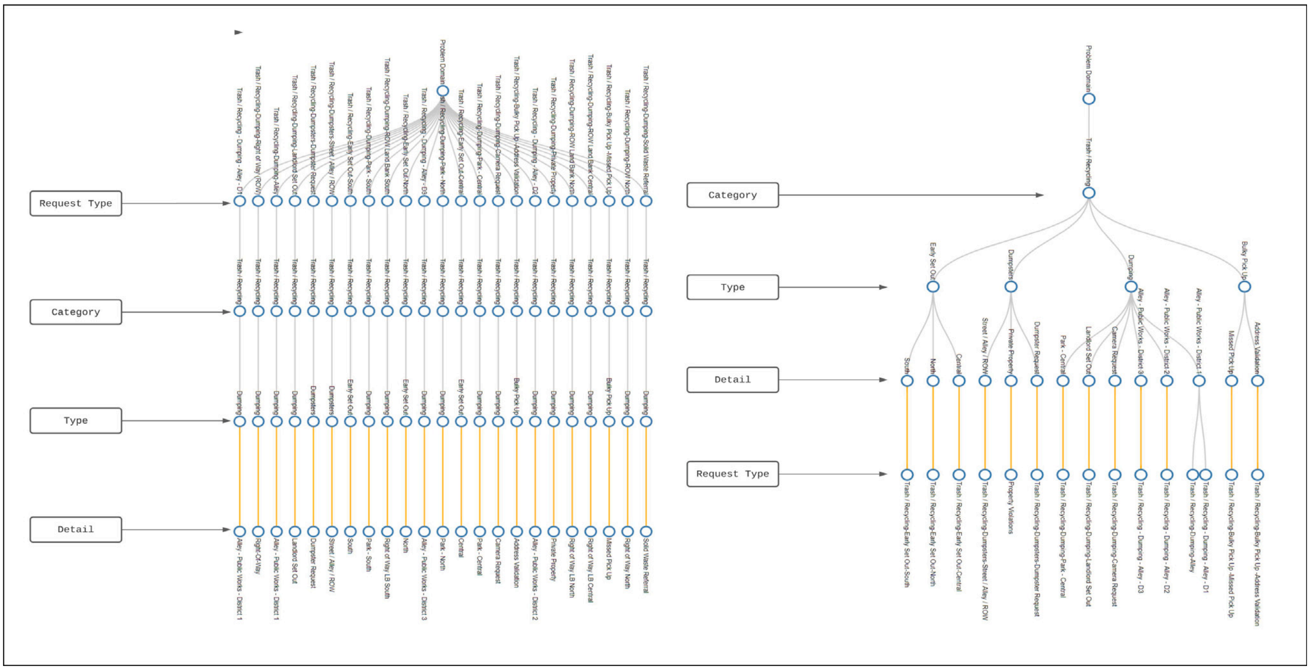

Fig. 15. Two versions of taxonomy: left is ordered by request type first (original order from 311 call structure), and right is ordered by category first (our revised structure). Both graphs represent a taxonomy for the "Trash/Recycling" category. The details of the Trash/Recycling are available in Appendix.

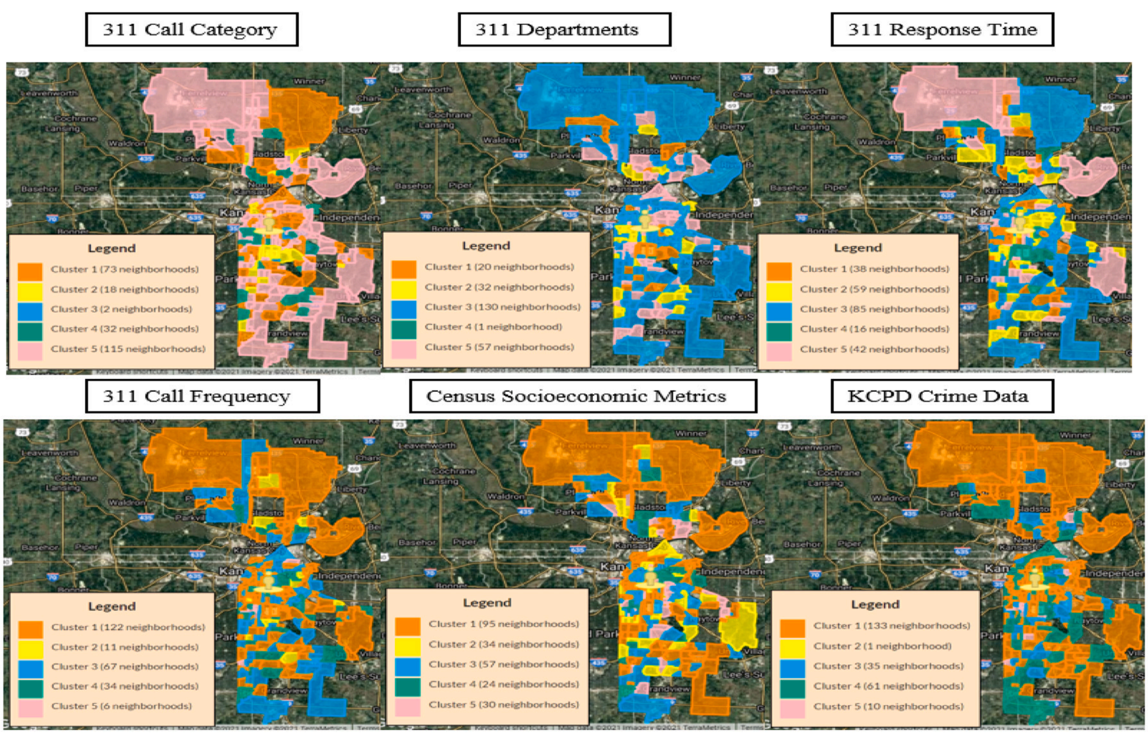

Fig. 16. K-Means Clustering of 246 Kansas City Neighborhoods according to *311 Call Category, 311 Assigned Departments, 311 Response Time, 311 Call Frequency, Census Socioeconomic Metrics, and KCPD Crime Data*. For all circumstances, the number of clusters is five.

Due to the regional and temporal characteristics of 311 call data, visualization of 311 calls physically and temporally would be used to uncover significant trends. When combined with the power of charts, it becomes a highly effective tool for gaining insights on both analytical and semantic levels. Combining structured and unstructured data and spatial and temporal data proved to be an efficient method for constructing a comprehensive picture of the state of the geographic entity under study (state/neighborhood/block-group).

Each block's color represents the call volume for that geographic unit (i.e., neighborhood), ranging from brighter to darker shades. It serves as a first-level indicator of the neighborhood's performance. The neighborhood's color naturally draws the user's attention to the neighborhood with the highest call volume, corresponding to a more significant number of difficulties. To conduct extensive research, one can use the graphs to determine patterns at the departmental, categorized, and total volume levels, using bar charts and line charts.

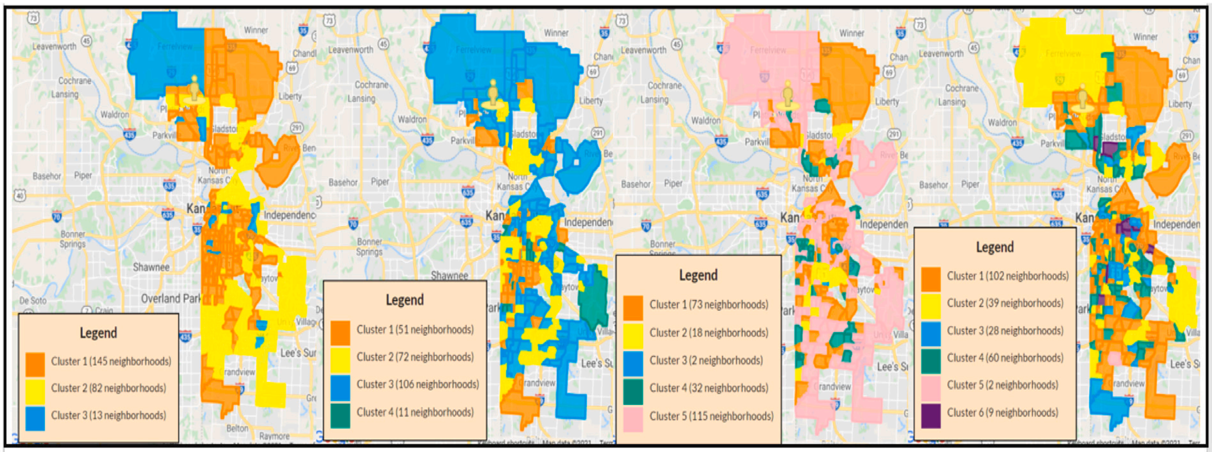

Fig. 17. K-Means Clustering of 246 Neighborhoods in Kansas City. k ∈ [3, 6].Clustering category is *311 Call Category*.

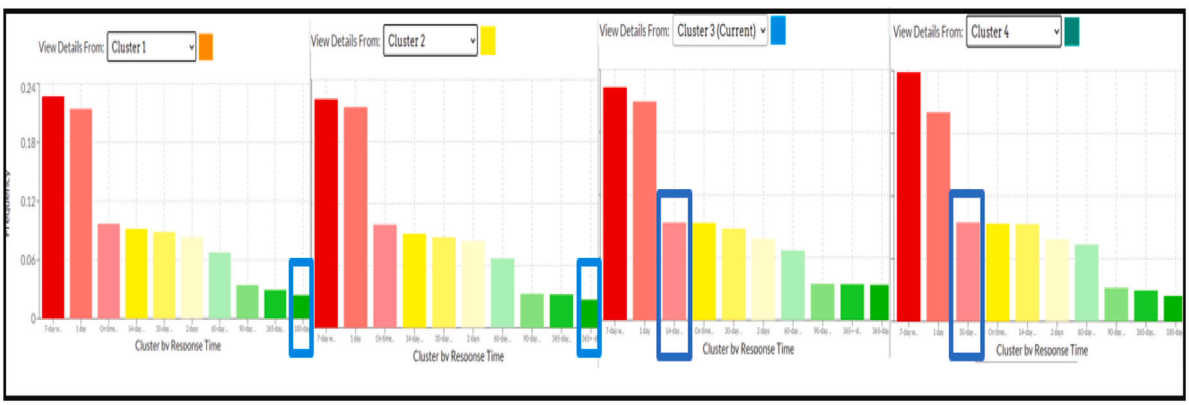

Fig. 18. *311 Response Time* comparison of four clusters of Kansas City's 246 neighborhoods, ordered by the top ten time frames. While each of the four clusters varies slightly in terms of how quickly cases are resolved, they all share the common trait of cases being treated more frequently within seven days of the call being recorded.

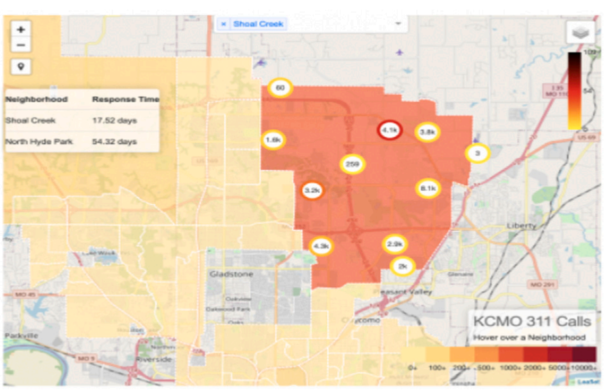

Fig. 19. 311 Call Map Visualization for 311 Call with Spatial Partition.

Algorithm 5 Spatial Partition Algorithm

```
procedure PARTITIONS(n, level, parts)
    medians ← set()
    while n > level do
        for part ← iterator(parts) do
            medians ← medians + MEDIAN(parts)
        PARTITIONS(n − 1, medians)
    return PARTITIONS(n − 1, level, medians)
```

As seen in Algorithm 5, the location of 311 calls is used to cluster the calls according to the zoom level. The user can access 19 levels, beginning at 0 (for the entire planet) and ending at 18. (for the most granular tile). The partitioning technique provides a different number of clusters depending on the region of the geographic entity under consideration. Clustering of partitions occurs faster when the area is small than when the area is large. Clusters are constructed bottom-up by examining the Euclidean distance of positional data at a lower level and selecting central locations from the median points of the clusters in the layer below.

A temporal filter connects all charts and maps, allowing for time-based evaluation of 311 data at both the annual and monthly granularity levels. After selecting a year and month, the filter is applied to trends in call volume, departmental, and category. Additionally, the filtered data can be used for further analysis. Call volume analysis begins with a logical assessment of departments and category-based research. For example, the volume of calls received by each department shows the departments' success in resolving neighborhood issues. Likewise, the call volume by category provides insight into the effectiveness of departments in resolving local issues.

6.2.1. Results and evaluation

The overall application serves as a single-point-of-contact (dashboard) for analyzing the progress of 311 calls in a city on a neighborhood, block-group, and county level. First, at a high level, it provides a quantitative analysis of neighborhood-level requests and complaints based on updated publicly available data and a snapshot of governance

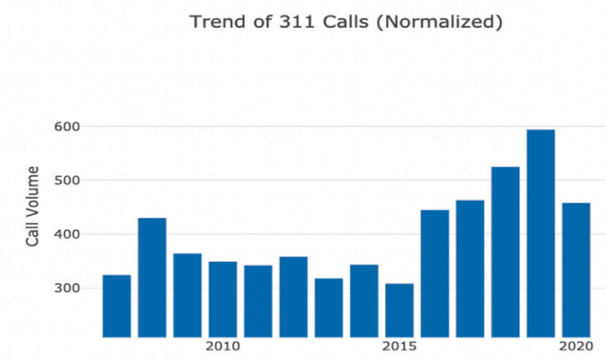

Fig. 20. 311 Call volume trends for years from 2007 to 2020.

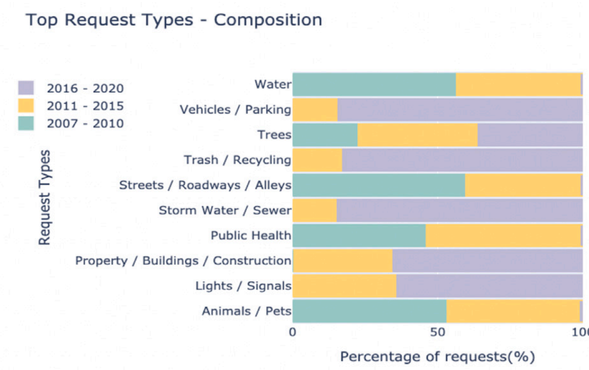

Fig. 21. Yearly Composition of 311 Call types for three bins — (1) 2007–2010 (Green), (2) 2011–2015 (Orange) and (3) 2016–2020 (Purple). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

performance across multiple metrics (call volume, categories, departments, etc.). Secondly, on a more granular level, it works as a tool for drilling down and determining the explanation for the visible picture by examining the 311 calls on an individual basis.

The spatial partitioning approach automatically extracts fine-grained characteristics and enables a more comprehensive understanding of the geographical distribution of 311 calls. The considered 311 call data comprises information about multiple categories, cleansed and mapped to 246 in the final version. At the base level, a collection of 16 to 64 nodes is clustered together, compounding the number of nodes at each level: 16 -> 16^2 -> 16^4 -> 16^8 and so on. Handling individual calls would necessitate a significant investment in processing resources at a higher level. This approach normalizes the number of data points at each level, which simplifies the computation of the required computing power.

6.3. Trend visualization

Call Volume Trend: The call volume trend is depicted in Fig. 20, which depicts the number of calls received over the specified period in the filters. The number of calls can serve as a direct indicator of neighborhood activity.

Yearly Composition of 311 Call Types: Fig. 21 illustrates the composition of 311 call types from 2007 to 2020, grouped into three bins: (1) 2007–2010 (Green), (2) 2011–2015 (Orange), and (3) 2016–2020 (Orange) (Purple). Each bin represents the percentage of 311 calls of a particular type as a percentage of total calls received. It sheds light on how the composition has changed throughout time. If a call had more incidents in the previous bin than in the current bin, it might indicate that the issue has been resolved or that a solution has been implemented to mitigate the issue's effect. Suppose the percentage of call volume is greater in the recent bin than in the previous years' bin. In that case, this indicates a relatively recent development and the need to address the issue to prevent it from expanding further.

Distribution of 311 Call Timings:

It is particularly effective for scaling existing public services, as it assists in finding and raising maximum utilization to provide a higher level of service to the public. To illustrate this concept, we used a radar chart to display call activity on a clock in Fig. 22. This chart is live and updates as the time and neighborhood criteria are changed. The peaks in this graph show times when the 311 calls service is hectic. Increasing the resources available during peak hours would allow more people to be served while raising the satisfaction score.

Departmental Call Volume Trend: When a 311 call is received, it is directed to the appropriate department as part of the resolution process. The department then takes over and solves the issue, providing a solution and closing it down. The amount of calls handled by each department provides insight into the workload, resource demand,

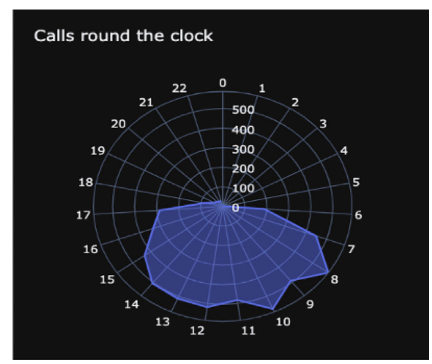

Fig. 22. Round clock of 311 calls on typical work day running from 7 AM to 5 PM.

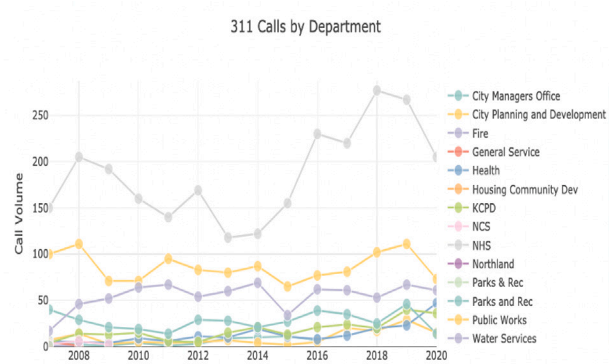

Fig. 23. Departmental Call Volume Trend for KCMO Core Departments.

and the need for collaboration, all of which are essential criteria for effective governance. As a result, call volume trends for various departments can be viewed as indicators of the corresponding department's performance. Fig. 23 depicts a visual comparison of call volume per department.

Categorical Call Volume Trend: Each call is assigned a category, similar to how it is assigned to a department. Departments define these categories to rapidly identify, categorize, and arrange the different sorts of 311 calls, allowing for a more efficient and timely resolution of concerns. These categories correspond to the type of case. Analyzing call volume based on diversity reveals information about the type and severity of problems in a community — the greater the number of difficulties, the greater the necessity to address and focus on the

Table 10
KCMO's 311 call problem taxonomy: categories and types.

CATEGORY (84)	Air Quality, Animal, Animals/Pets, Aviation, Boulevard Parks and Rec, Bridge, Bulky Appointment, Capital Projects, City Clerks Office, City Facilities, City Managers Office, City Planning & Development, Cleaning, Consumer Services, Convention & Entertainment Center, Data Not Available, Ditch, Downtown Parking, Facilities, Finance, Fire, Food Protection, Government, Graffiti, Guard Rail, Health, Housing, Information Request, Information Technology, Law, Lights/Signals, Maintenance, Markings, Mayors Office, Mowing/Weeds, Municipal Court, Neighborhood & Community Services, Neighborhood Preservation, Noise Control, Nuisance Violations, Park, Park Maintenance, Parking Meter, Parks & Recreation, Permit, Pothole, Property/Buildings/Construction, Property & Nuisance Violations, Property Preservation, Property Violations, Public Health, Public Safety, Public Works, Right of Way, Sealing, Sewer, Sidewalk, Sidewalks/Curbs/Ditch, Signal, Signs, Snow & Ice, Steel Plate, Storm Water, Storm Water/Sewer, Street, Street Light, Street Sign, Streets/Roadways/Alleys, Sweeping, Traffic Sign, Traffic Study, Trash, Trash/Recycling, Trees, Vehicle, Vehicles/Parking, Wastewater, Water, Water Leak, Water Main Break, Water Quality, Water Services, Weeds, Zoning
TYPE (293)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 311 Administration, Abandoned, Abandoned On Street, Account, Action Center, ADA, ADA Accessibility, Administration, Airports, All, Alley, Ambulance, Animal, Assessment Dispute, Athletics, Automatic Meter Reading, Banners, Barking Dog, Barrier Removal Program, Bicycle Pedestrian, Billing, Bite, Blvd/Pkwy, Board Up, Boulevards and Parkways, Broken Asphalt, Bruce R. Watkins Freeway, Brush, Building/Property, Bulky, Bulky Pick Up, Capital Improvement Management Office, Capital Projects Office, Care, Catch Basin, City, City Communications, City Manager Äòs Office, City Property, Cleaning, Commercial Parking, Commercial Signs/Ads, Commercial Vehicle, Communicable Disease, Communications, Community Center, Community Service, Construction Issue/Concern, Construction Repair, Contractor, Contractor Restoration, Crack, Cruelty or Neglect, Culvert, Curb Box, Customer Service, Cut/Permit, Damage, Damage/Dis-Repair, Dangerous Building, Data Not Available, Dead Animal, Disabled/Unlicensed on Private Property, Disease, Disease Control, Disrepair, Ditch, DOWN OR DAMAGED, Dumping, Dumpsters, Early Set Out, Elevator Inspection, Emerald Ash, Emerald Ash Borer, Emergency, Emergency Management, Encroachment, Engineering, Facilities/Attractions, Finance, Fire, Flood Barricade, Food Establishment, Google Fiber Project, Graffiti, Guardrail, Hazardous, Hotel/Motel, Human Resources (HR), Hydrant Repair, Icing, Illegal Dumping, Illegally Parked, Illegally Parked Vehicle, Improved Channel, Industrial, Injury or Cruelty involving an Animal, Insect Problem, Inspection for NPD, Intersection, Investigation, Investigations, IT/Websites, Jackson County, KC Police, Laboratory Service, Land Bank, Land Development, Land Trust, Land Use/Zoning Issue, Landlord Set Out, Landscape, Landscaping, Law, Lead Poisoning Prevention, Leaf/Brush, Leaf & Brush, Leak, Maintenance, Malfunction, Manhole, Markings/Paint, Mayor Äòs Office, Meter, MFS Referral (Meter Field Services), Minor Home Repair, Missing, Municipal Court, Neighborhood Abatement Program, Neighborhood Preservation, New, New Request, NHS (Neighborhood Housing Services), No Dumping Sign, No Water/Pressure, Noise, Not Provided By KCMO, Nuisance, Obstructed/Closure, Other, Other-Maintenance, Outage, Outdoor Air Quality, Owned or Stray at Large, Paint Program, Paper Street, Park Maintenance, Park Property, Parked on Unapproved Surface, Parking, Parking Lot, Parking Meter, Parking Ticket Issue, Parks, Parks Facilities, Pavement Restoration, Permit, Permit/License, PIAC, Pipeline Barricade, Pipeline Referral, Pipeline Repair, Pipeline Restoration Concerns, Planning, Planting, Plate, Police, Pollutants, Pool/Hot Tub, Pothole, Priority, Private/Commercial, Private Property, Privately Owned Property, Property Maintenance, Public Facilities, Public Improvement Advisory Committee, Public Property, Public Restroom, Public Works, Quality, Questionable Activity, Railroad Property, Rat Control and Treatment, Rat Treatment, Recycle, Recycling, Referral, Regulated Industries, Removal, Repair, Replacement, Research, Restoration, Resurfacing, Return Call to Citizen, Revenue Protection, Right of Way, Right of Way (ROW), Rodent, Roll-out Cart, Salvage Yard Investigation, Scooter, Security/Safety, Services, Services (CPD), Services (NPD), Sewer, Sewer Backup/Leak, Sewer Odor, Sewer Referral, Sinkhole, Sinkhole Referral, Smoking/Tobacco, Snow/Ice, Solid Waste Operations, Speed Bump, Storm Clean Up, Storm Damage, Storm Drain/Catch Basin, Storm Water, Storm Water Referral, Stormwater Billing, Stray, Stray Confined, Street, Street and Traffic, Street Clean Up, Street Light, Street Name Signs, Street Preservation, Street Services, Street Sweeping, Streetlight Pole, Stump Removal, Supervisors, Ticket, Tobacco, Tow Services, Traffic Permit, Traffic Sign, Traffic Signal, Traffic Signs, Traffic Study, Trails, Transportation, Trash, Trash Cart, Trash Collection, Treatment, Trimming, Turf Restoration, Unapproved Objects, Valve Repair, Visibility, Volunteer Inspector Program, Walk In Center City Hall, Water, Water Billing, Water Main Repair, WaterServices-ConsumerServices-BackOfficeSupport, Weatherization Program, Wildlife

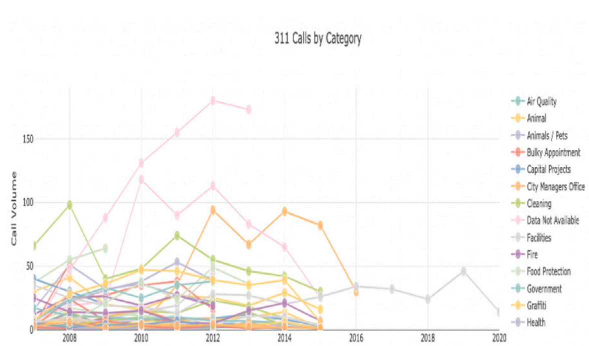

Fig. 24. Categorical Call Volume Trend for KCMO's Major 311 Call categories during 2007 to 2020.

issue. Fig. 24 shows a sample of the definite call volume trend for all neighborhoods.

6.4. Implementation and discussion

The OpenComm framework was implemented using the Dash (https://dash.plotly.com/). Because Plotly's Dash platform provides a collection of online data analytics and visualization capabilities, it was suitable for OpenComm's data visualization and full-stack product development while considering greater control over the data and providing a high-quality interactive depiction of insights. Unlike the most popular data visualization tool, Tableau, which is less flexible, Dash is a highly integrated data visualization platform with scalable front-end and back-end functionality. Dash uses React JS, a component-based development framework for the front-end, and Flask, a Python-based framework for the back-end that is easy to set up and extend, with close interaction. As a result, allowing for shorter development cycles speeds up application development.

Map Component: The leaflet is an open-source map provider that works with React.JS and Dash. Its versatility and active community support have led to widespread adoption. The Leaflet Map Component receives the neighborhood data as GeoJson, an industry standard for encoding geographical mapping and boundary information. Before mapping, the insights generated from the raw data are translated into a compatible format. The user is then given the option of selecting World Imagery, Gray Canvas, Topology, or Terrain base maps. Finally, events triggered by selecting (by clicking) and lingering over an active item (polygon, boundary, layers, etc.) are utilized to update data across all application charts.

Chart Components: Dash consists of elements for Bar Chart (used for Volume Trend), Radar Chart (used for Round the Clock Trend), Line Chart (used for Department and Category Trend), and Horizontal Bar Chart (used for Composition Trend) to deal with various types

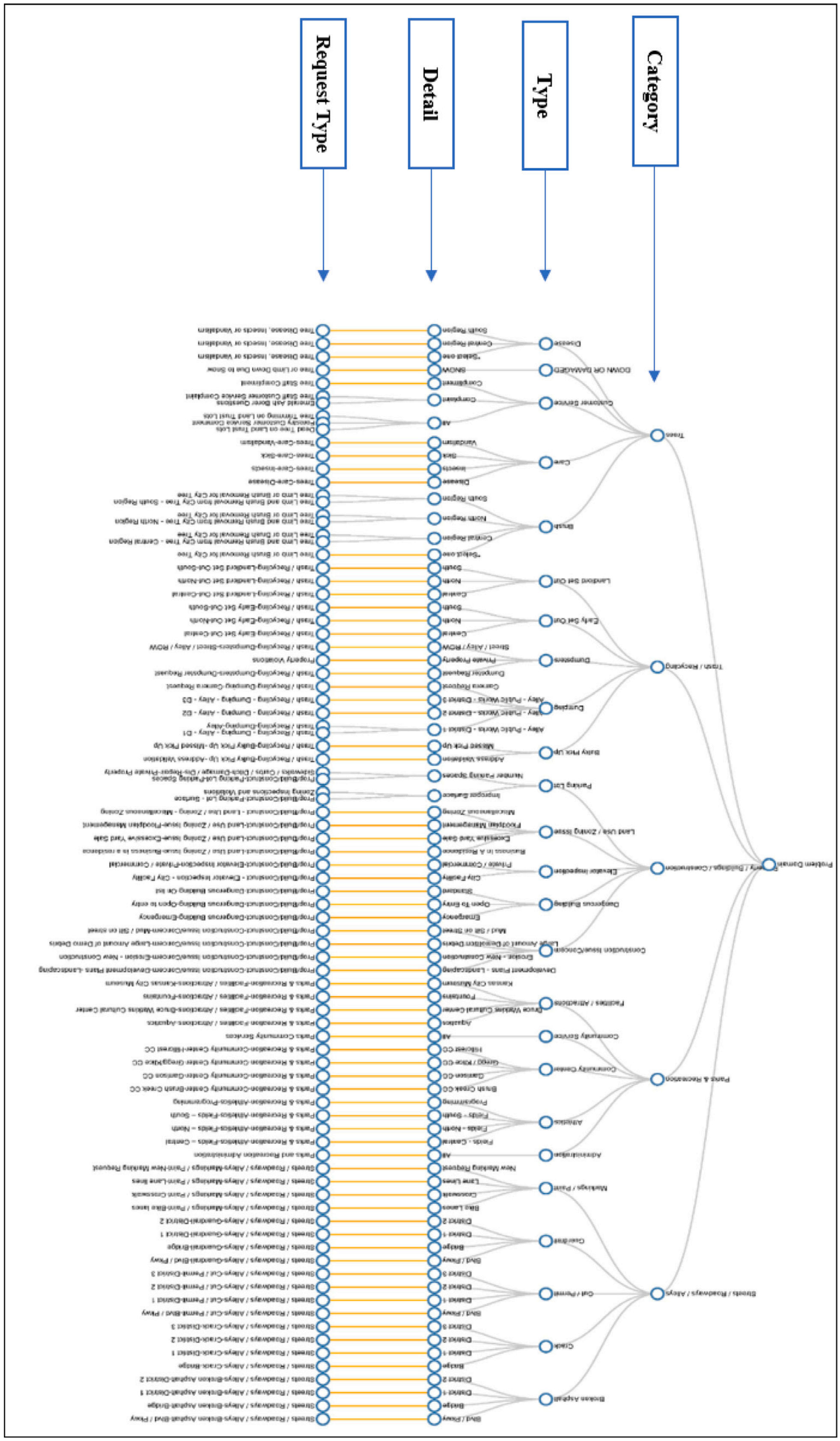

Fig. 25. 311 Call Problem Taxonomy (Tree View). The yellow links indicate a potential heterogeneous categorization issue of the current 311 structure where the request types are primarily the concatenation of category, type, and detail, resulting in mostly direct 1-to −1 relationships. The details of the Problem Taxonomy are available in Appendix. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

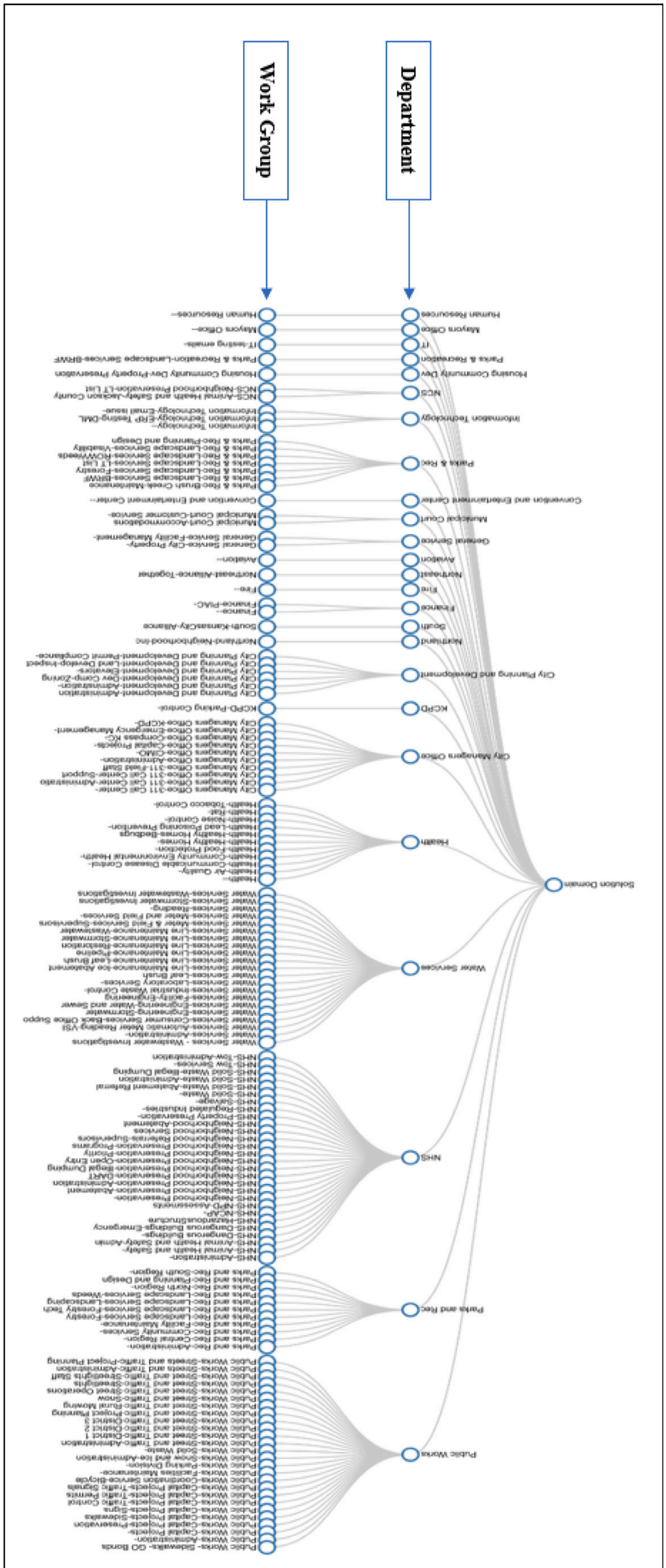

Fig. 26. 311 Call Solution Taxonomy (Tree View): The details of Solution Taxonomy are available in Appendix.

Table 11
KCMO's 311 call problem taxonomy: detail (1).

DETAIL (566)	211, *Select One, 1 Light Out, 2 to 4 Lights Out, 5+ Lights Out, Abandonment, Abuse/Injury/Death, Abused, Accelerated Abatement, Accommodations, Account Services, ADA, Address Change, Address Validation, Address Will Not Validate, Adjustment, Admin Tasks, Administration, Adult Entertainment, Aggressive, Airport ADA, Alcohol Business, ALL, Alley - Public Works - District 1, Alley - Public Works - District 2, Alley - Public Works - District 3, Amusement Business, Animal on Animal, Appearance/Speaker, Appointment, April, Aquatics, Ash Treatment, Assessment Dispute, Assessment for Abatement, At Curb/In Yard, At Large - Aggressive, At Large - Highway, At Large - Non-Aggressive, At Large - Owned, At Large - Pack, Audio, Aviation, Badge Identification, Barking Dog, Barricades, Basement, Basement Backup Prevent Proj, Basketball Goal, Basketball Hoop/Goal, Bat visible in house, Bedbugs, Bike Lanes, Bike Route, Billboard, Billing/Administration, Bin Replacement or Damage, Blinking, Blinking Off and On, Block Pruning, Blocked, Blue Bag Program, Blvd/Pkwy, Blvd/Pkwy - Central, Blvd/Pkwy - North, Blvd/Pkwy - South, Blvd/Pkwy/Park, Boat, Boat Private Property, Bridge, Broken, Bruce Watkins Cultural Center, Brush, Brush Creek, Brush Creek CC, Brush Pickup, Buckner, Building, Bulky, Burning During The Day, Business In A Residence, Business Question on Order, Call Back Request, Camera Request, Car, Car Private Property, Carbon Monoxide, Cart Size Change, Catering, Central, Central (Dept. Use), Central Region, Chicken/Rooster, Citizen Satisfaction Survey, City, City Block, City Construction Site, City Facility, City Hall, City Hall Garage, City Owned, City Owned Trash Container, City Property, City Street, City Vacant Lot, Clogged/Debris, Clogged/Over-Flow, Cock Fighting, Code 1, Collection Notice, Color/Cloudy, Comment or Question, Commercial, Compass KC Assistance, Complaint, Compliance Letter, Compliment, Confined, Construction, Continuance, Contractor, Contractor Restoration, Convenience/Grocery Store, Convention Center, Cooling Center Transportation, County, Creek/Green Solution, Crosswalk, Curb, Curb Box Leak, Curb or Sidewalk, Curb-box/Service Line, Damaged, Damaged - Emergency, Damaged - Non-Emergency, Damaged/Broken, DART, Data Not Available, Day Burn, Day Labor Business, Dead, December, Declining, Demolition and Debris, Department Referral, Design, Detour, Development Plans - Landscaping, Disability Pick Up, Disconnect Service, Disease, District 1, District 2, District 3, District One, District Three, District Two, Document Request, Dog Fighting, Down/Damaged, Dropped Personal Item, Drug Abatement Response Team, Dumping, Dumping in Street, Dumpster Request, EAB, Early Set Out, East, Education/Health, Elevator, Emergency, Emergency Off, Emergency Off/Flooding, Employee Behavior, Engineering, Erosion - New Construction, Erosion Control, Escalation, Event/Closure Request, Excessive Yard Sale, FEB, February, Feedback, Feedback/Concern, Fence and Wall, Fence Issue, Field Investigation, Fields - Central, Fields - North, Fields - South, Financial Loss, Fire Investigation, Fire Station Maintenance, Flood, Floodplan Management, Food Truck, Fountains, Friday - Central, Friday - North, Friday - South, From Manhole, Garrison CC, General, General Sanitation, Good Neighbor Award, Graffiti, Greenwood, Gregg/Klice CC, Grounds, Hazard, Hazardous, Hazardous Materials, Hazardous Structure, Healthy Homes Information, High Bill, High Pressure, Hillcrest CC, Home Repair, Homeless Camp, Honorary/Renaming, How's My Driving, Hydrant, Hydrant Leak, Ice, Ice Due to Leak, Improper Surface, In A House, In Right-Of-Way, In Street, In the Sewer, Inclement Weather, Indoor, Initial, Injured, Injured Cat or Dog, Insects, Inside, Inside Home, Inspection, Integrated Pest Management Class, Intersection, Investigation, Jack, January, June, Kansas City Museum, Kansas City North CC, KCATA/Metro, KCI, KCMO.gov, KCMore, KCPD, Lack of Information, Lacking Info Needed, Lakeside Nature Center, Land Bank, Land Bank (Services), Landlord Set Out, Landscaping, Lane Closed, Lane Lines, Large Amount of Demolition Debris, Lead Poisoning Prevention Info, Leaf/Brush Appointment Missed, Levacy, License, Line Creek CC, Livestock, Lodging, Lone Jack, Loose

Table 12
KCMO's 311 call problem taxonomy: detail (2).

DETAIL (566)	Lot - Structure, Lot - Vacant, Low Pressure, Main Leak, Main Structure, Marathon/Run/Race, March, Marlborough CC, Materials Left At Site, MAY, Medical Tags, Merchandise, Meter, Meter Leak, Miscellaneous Zoning, Missed, Missed by City, Missed by Contractor North, Missed by Contractor South, Missed City Trash and Recycling, Missed North Trash - Recycling, Missed Pick Up, Missed South Trash - Recycling, Missing, Missing - Emergency, Missing - Non-Emergency, Missing/Broken, Missing/Displaced, MODOT, Monday - Central, Monday - North, Monday - South, Monetary Inquiries, Mowing List, Mud/Silt on Street, Multi-Unit, Multiple, Municipal Court, Music, NCS Referral, Negative Comments, Neighborhood Training, New Marking Request, New Sign Installation, New Sign Request, New Signal Request, New Smoke Alarm Request, New Water Service, New/Remove Light, News and Media, No Bill, No Known Property Owner, No Structure on the Property, No Utilities, No Water, Non Emergency, North, North (Dept. Use), North Contractor, North Region, Northeast Alliance, Northland Neighborhoods INC, Not Delivered, Number of Dwelling Units, Number of Handicap Spaces, Number of Residents, Number Parking Spaces, Off Leash - Investigation, Officer Assist, On Demand (PW), On Private Property, Open Storage, Open To Entry, Other, Other City Buildings, Other Garage, Other Light Issue, Other Litigations, Other Priority Issue, Other Private Property, Other Property Issue, Other Websites - Tech, Outdoor, Outside, Over Limit/Hoarding, Owned, Owner Äòs Responsibility, Paint Program, Park - Central, Park - North, Park - South, Park/Blvd/Pkwy, Parking, Parking Surface - Pay Lot, Parks Barricade, Parks Maintained, Parks Website, Pavement, Payment Issues, Pedestrian/Walkway, Permit (Fireworks), Permit Request, Pest Management Class, Pipeline Restoration, Pipeline Restoration Concerns, Plant, Plant City Tree Request, Pole, Pole Down or Leaning, Pool, Positive Comments, Post Demo Issues, Priority Review, Private/Commercial, Private Line Repair, Private Property, Private Property (ROW), Private Property Nuisance, Private Property Pickup, Private Property Violation, Probation, Programming, Project Planning, Property Damage, Public, Public Building, Public Works Barricade, Question, Question/Feedback, Questions, Ramp, Rat, Rat Control Treatment and Baiting, Referral, Referral by Parks and Rec, Referred from a department, Region - Central, Region - North, Region - South, Removal, Remove, Remove Sign, Repair, Repeat Missed, Repeated Report, Replace - City Project, Request, Request New, Rescue Permit, Resident Request, Restaurant, Restoration, Restoration Issue, Review, Right of Way Central, Right of Way LB Central, Right of Way LB North, Right of Way LB South, Right of Way North, Right of Way South, Right-Of-Way, Road Closed, Route and Trail, Sales to Minors, Salvage and Scrap, School, Seasonal, Secondary Structure, Service Issue/Problem, Sewer Projects, Sewer Restoration, Shelter - Central, Shelter - North, Shelter - South, Shelter/Food/Water, Short Term Rental, Sibley, Sick, Sidewalk, Sidewalks/Curbs, Sign, Signal, Signal Flashing, Signal Malfunction, Signal Out, Signal Pole Damage, Signal Timing, Single, Six-plex or less, Smell/Odor, Smell or See Smoke, SNOW, Snow Blocking Drive or Sidewalk, Snow on Sidewalk, Snow Route Verification, Solid Waste Referral, South, South (Dept. Use), South Contractor, South Kansas City Alliance, South Region, Southeast CC, Standard, Starlight Theater, State, Storm Trash, Stormwater, Stormwater Projects, Stray, Street, Street/Alley/ROW, Street and Traffic, Street Leak, Street Light, Street Permit Research, Street Resurfacing, Street/Sidewalk, Streetlight, StreetLight, Stump, Sugar Creek, Supervisors, Sweeping Schedule, Taste, Tax, Theft, Thursday - Central, Thursday - North, Thursday - South, Tobacco Business, Tony Aguirre CC, Tours, Tow Company, Tow Services, Traffic, Traffic Study Request, Traffic Video Request, Trailer, Trailer Private Property, Trash, Trash on Private Property, Trash Truck, Treatment, Treatment Plant, Tree Dead, Tree Debris, Tree Down, Tree Limb/Brush, Tree Limbs, Trim, Tuesday - Central, Tuesday - North, Tuesday - South, Turf, Turn Lanes, Turn Off for Repairs, Turn On for Paid Bill, Turn On Repairs Made, Two to Four, Unapproved Land Use, Unincorporated Areas of Jackson County, Unity Village, Unsanitary Conditions, Utility Cut, Vacant Home, Vacant Lot, Vaccination/Pet License, Valve, Valve Leak, Vandalism, Vehicle, Vehicle Damage, Vehicle For Hire, Vehicles, Violation, Visibility, Water Main, Water Main Projects, Water Service Barricade, Water Services Department, Water Website, WaterServices-ConsumerServices-BackOfficeSupport, Weather Siren, Weatherization Program, Wednesday - Central, Wednesday - North, Wednesday - South, Westport Roanoke CC, Wheeler (Downtown), Wildlife, Wire Down/Hanging Arm, Wire or Arm, Wolf Garage, Work Without Permit, Yard

Table 13
KCMO's 311 call problem taxonomy: request type (1).

REQUEST TYPE (1229)	311 Action Center, Abandoned Vehicle, Accelerated Abatement Program, Action Center Escalation, ADA Accessibility, Additional Information Required, Alley Cleaning (North of the River), Alley Cleaning (South of 47th Street and West of BLue PKWY), Alley Cleaning (South of 47th Street), Alley Cleaning (South of River to 47th Street and East of Blue PKWY), Alley Cleaning (South of River to 47th Street), Animal Admin Call Back, Animal At Large, Animal Bite, Animal Concerns Buckner Jackson County, Animal Concerns Greenwood Jackson County, Animal Concerns Levacy Jackson County, Animal Concerns Lone Jack Jackson County, Animal Concerns Sibley Jackson County, Animal Concerns Sugar Creek Jackson County, Animal Concerns Unincorporated Jackson County, Animal Concerns Unity Village Jackson County, Animal Concerns Unicorporated Jackson County, Animal Control, Animal Control Administration, Animal Emergency, Animal Emergency Buckner Jackson County, Animal Emergency Greenwood Jackson County, Animal Emergency Lone Jack Jackson County, Animal Emergency Sibley Jackson County, Animal Emergency Sugar Creek Jackson County, Animal Emergency Unincorporated Jackson County, Animal Emergency Unity Village Jackson County, Animal Investigation, Animals/Pets - Wildlife - Outside, Animals/Pets-Bite-Animal on Animal, Animals/Pets-Bite-Owned, Animals/Pets-Bite-Stray, Animals/Pets-Bite-Wildlife, Animals/Pets-Cruelty or Neglect-Abandonment, Animals/Pets-Cruelty or Neglect-Abuse/Injury/Death, Animals/Pets-Cruelty or Neglect-Inclement Weather, Animals/Pets-Cruelty or Neglect-Shelter/Food/Water, Animals/Pets-Dead Animal -Curb, Animals/Pets-Dead Animal -Private Property Pickup, Animals/Pets-Dead Animal -Street, Animals/Pets-Permit/License-Chicken/Rooster, Animals/Pets-Permit/License-Livestock, Animals/Pets-Permit/License-Other, Animals/Pets-Permit/License-Permit Request, Animals/Pets-Permit/License-Rescue Permit, Animals/Pets-Permit/License-Vaccination/Pet License, Animals/Pets-Questionable Activity-Cock fighting, Animals/Pets-Questionable Activity-Dog fighting, Animals/Pets-Questionable Activity-Other, Animals/Pets-Questionable Activity-Over Limit/Hoarding, Animals/Pets-Rat Treatment-City Block, Animals/Pets-Rat Treatment-City Const Site, Animals/Pets-Rat Treatment-In the Sewer, Animals/Pets-Rat Treatment-Inside Residence, Animals/Pets-Rat Treatment-Outside, Animals/Pets-Rat Treatment-Vacant Home, Animals/Pets-Rat Treatment-Vacant Property, Animals/Pets-Services-Barking Dog Letter, Animals/Pets-Services-Call Back Request, Animals/Pets-Services-Feedback, Animals/Pets-Services-Off Leash, Animals/Pets-Services-Off Leash - Investigation, Animals/Pets-Services-Officer Assist, Animals/Pets-Services-Question, Animals/Pets-Services-Repeated Report, Animals/Pets-Services-Service Issue/Problem, Animals/Pets-Stray-At Large - Aggressive, Animals/Pets-Stray-At Large - Highway, Animals/Pets-Stray-At Large - Non-Aggressive, Animals/Pets-Stray-At Large - Owned, Animals/Pets-Stray-At Large - Pack, Animals/Pets-Stray-Confined, Animals/Pets-Stray-Injured, Animals/Pets-Wildlife-Injured, Animals/Pets-Wildlife-Inside, Automatic Meter Reading, Automatic Meter Reading - VSI, Aviation - All, Back Office Support, Barking Dog, Barrier Removal Program, Bedbugs Problem, Bicycle or Pedestrian Trail, Bridge Repair (All Districts), Bruce R Watkins Freeway, Brush Creek Maintenance, Building Code and Permit Violations, Bulky Address Validation, Bulky Appointment, Bulky Appointment DEC 10, Bulky Appointment - April 27, Bulky Appointment - April 28, Bulky Appointment - August 11, Bulky Appointment - August 12, Bulky Appointment - August 13, Bulky Appointment - DEC 26, Bulky Appointment - FEB 17, Bulky Appointment - FEB 18, Bulky Appointment - FEB 19, Bulky Appointment - FEB 20, Bulky Appointment - FEB 21, Bulky Appointment - July 14, Bulky Appointment - May 1, Bulky Appointment - May 11, Bulky Appointment - May 12, Bulky Appointment - May 13, Bulky Appointment - May 14, Bulky Appointment - May 15, Bulky Appointment - May 18, Bulky Appointment - May 19, Bulky Appointment - May 20, Bulky Appointment - May 21, Bulky Appointment - May 22, Bulky Appointment - May 26, Bulky Appointment - May 27, Bulky Appointment - May 28, Bulky Appointment - May 29, Bulky Appointment - May 30, Bulky Appointment - May 4, Bulky Appointment - May 5, Bulky Appointment - May 6, Bulky Appointment - May 7, Bulky Appointment - May 8, Bulky Appointment DEC 02, Bulky Appointment DEC 03, Bulky Appointment DEC 04, Bulky Appointment DEC 05, Bulky Appointment DEC 08, Bulky Appointment DEC 09, Bulky Appointment DEC 11, Bulky Appointment DEC 12, Bulky Appointment DEC 13, Bulky Appointment DEC 15, Bulky Appointment DEC 16, Bulky Appointment DEC 17, Bulky Appointment DEC 18, Bulky Appointment DEC 19, Bulky Appointment DEC 20, Bulky Appointment DEC 21, Bulky Appointment DEC 22, Bulky Appointment DEC 23, Bulky Appointment DEC 24, Bulky Appointment DEC 27, Bulky Appointment December 01, Bulky Appointment FEB 02, Bulky Appointment FEB 03, Bulky Appointment FEB 04, Bulky Appointment FEB 05, Bulky Appointment FEB 06, Bulky Appointment FEB 09, Bulky Appointment FEB 10, Bulky Appointment FEB 11, Bulky Appointment FEB 12, Bulky Appointment FEB 13, Bulky Appointment JAN 02, Bulky Appointment JAN 03, Bulky Appointment JAN 05, Bulky Appointment JAN 06, Bulky Appointment JAN 07, Bulky Appointment JAN 08, Bulky Appointment JAN 09, Bulky Appointment JAN 12, Bulky Appointment JAN 13, Bulky Appointment JAN 14, Bulky Appointment JAN 15, Bulky Appointment JAN 16, Bulky Appointment JAN 20, Bulky Appointment JAN 21, Bulky Appointment JAN 22, Bulky Appointment JAN 23, Bulky Appointment JAN 26, Bulky Appointment JAN 27, Bulky Appointment JAN 28, Bulky Appointment JAN24, Bulky Appointment- August 10, Bulky Appointment- August 17, Bulky Appointment- August 21, Bulky Appointment- August 24, Bulky Appointment- August 5, Bulky Appointment- Apr 1, Bulky Appointment- Apr 10, Bulky Appointment- Apr 13, Bulky Appointment- Apr 14, Bulky Appointment- Apr 15, Bulky Appointment- Apr 16, Bulky Appointment- Apr 17, Bulky Appointment- Apr 2, Bulky Appointment- Apr 20, Bulky Appointment- Apr 21, Bulky Appointment- Apr 22, Bulky Appointment- Apr 23, Bulky Appointment- Apr 24, Bulky Appointment- Apr 3

of data and fit a variety of visual insights. When combined with the power of the AWS cloud, Flask's lightweight and fast server allows the application to come to life in a flash.

Deployment: AWS EC2 is a strong platform for web application deployment. A t2.medium tier instance is excellent for handling the amount and velocity of data to insights in our OpenComm. Our full version of the OpenComm platform will be deployed due to its dependability, availability, and low cost. Meanwhile, the lightweight version of the OpenComm Platform application can be available at http://kc311.herokuapp.com/.

7. Discussion

The OpenComm framework focuses on overcoming challenges caused by data fragmentation and public data security and privacy leakage. We investigated machine learning solutions for data integration, classification, privacy protection, and data visualization. The framework offers significant contributions in terms of public data, such as the 311 call domain and the Census Bureau. Our contributions are summarized below:

First, several types of missing value or data mistakes in the public domain have been identified. We developed recovery methods based on machine learning approaches to address the challenges. Furthermore, we compared several machine learning techniques for missing data recovery, measuring their performance in MAE, MSE, and R^2. We discovered that the Convolutional Neural Network model outperformed other machine learning models.

Second, our contribution is to break down the multi-domain boundary barriers through the geocode-based assignment. We applied the method to five distinct domains: neighborhood, school districts, police districts, council districts, and census block groupings. This technique looked impressive in integrating multiple data sources relevant to the 311 call domain while resolving boundary challenges.

Third, we created deep learning-based privacy-preserving with Differential Privacy (DP), where distinct privacy budgets ϵ are uniformly spaced across the range of $[10^{-2}, 10^{2}]$ with the best accuracy of 0.78019. This study found a substantial link between ϵ and accuracy. In addition, we proposed a $epsilon$ of 1 with a score of 0.7115 as an appropriate trade-off point for privacy-preserving and meaning machine

Table 14
TKCMO's 311 call problem taxonomy: request type (2).

REQUEST TYPE (1229)	Bulky Appointment- Apr 6, Bulky Appointment- Apr 7, Bulky Appointment- Apr 8, Bulky Appointment- Apr 9, Bulky Appointment- August 14, Bulky Appointment- FEB 23, Bulky Appointment- FEB 24, Bulky Appointment- FEB 25, Bulky Appointment- FEB 26, Bulky Appointment- FEB 27, Bulky Appointment- July 1, Bulky Appointment- July 10, Bulky Appointment- July 13, Bulky Appointment- July 15, Bulky Appointment- July 16, Bulky Appointment- July 17, Bulky Appointment- July 20, Bulky Appointment- July 21, Bulky Appointment- July 24, Bulky Appointment- July 27, Bulky Appointment- July 3, Bulky Appointment- July 6, Bulky Appointment- July 7, Bulky Appointment- July 8, Bulky Appointment- July 9, Bulky Appointment- June 1, Bulky Appointment- June 10, Bulky Appointment- June 11, Bulky Appointment- June 12, Bulky Appointment- June 15, Bulky Appointment- June 16, Bulky Appointment- June 17, Bulky Appointment- June 18, Bulky Appointment- June 19, Bulky Appointment- June 2, Bulky Appointment- June 22, Bulky Appointment- June 23, Bulky Appointment- June 24, Bulky Appointment- June 25, Bulky Appointment- June 26, Bulky Appointment- June 3, Bulky Appointment- June 4, Bulky Appointment- June 5, Bulky Appointment- June 8, Bulky Appointment- June 9, Bulky Appointment- MAR 10, Bulky Appointment- MAR 11, Bulky Appointment- MAR 12, Bulky Appointment- MAR 13, Bulky Appointment- MAR 16, Bulky Appointment- MAR 17, Bulky Appointment- MAR 18, Bulky Appointment- MAR 19, Bulky Appointment- MAR 2, Bulky Appointment- MAR 20, Bulky Appointment- MAR 23, Bulky Appointment- MAR 24, Bulky Appointment- MAR 25, Bulky Appointment- MAR 26, Bulky Appointment- MAR 27, Bulky Appointment- MAR 3, Bulky Appointment- MAR 4, Bulky Appointment- MAR 5, Bulky Appointment- MAR 6, Bulky Appointment- MAR 9, Bulky Pick Up Due to Flood Damage, Bulky Pick-up Missed, Capital Projects Office, Capital Projects-Parks-Question/Feedback, Capital Projects-Parks-Service Issue/Problem, Capital Projects-PIAC-Department Referral, Capital Projects-PIAC-Question/Feedback, Capital Projects-PIAC-Resident Request, Capital Projects-Public Works-Question/Feedback, Capital Projects-Public Works-Service Issue/Problem, Capital Projects-Sewer-Engineering, Capital Projects-Sewer-Question/Feedback, Capital Projects-Sewer-Service Issue/Problem, Capital Projects-Storm Water-Engineering, Capital Projects-Storm Water-Question/Feedback, Capital Projects-Storm Water-Service Issue/Problem, Capital Projects-Water-Basement Backup Prev Prog, Capital Projects-Water-Engineering, Capital Projects-Water-Question/Feedback, Capital Projects-Water-Service Issue/Problem, Catch Basins and Storm Water Concerns, Childhood Lead Poisoning Prevention, City Clerks Office-All, City Facilities-Airports-KCI, City Facilities-Airports-Wheeler (Downtown), City Facilities-Building/Property-Aviation, City Facilities-Building/Property-City Hall, City Facilities-Building/Property-Convention Center, City Facilities-Building/Property-Fire Station Maintenance, City Facilities-Building/Property-Municipal Court, City Facilities-Building/Property-Other City Buildings, City Facilities-Building/Property-Water Services Department, City Facilities-Parking-City Hall Garage, City Facilities-Parking-Other Garage, City Facilities-Parking-Wolf Garage, City Maintained Property Addition, City Maintained Property List, City Manager's Office-Administration, City Manager's Office-Capital Improvement Management Office, City Manager's Office-Emergency Management, City Manager's Office-Public Improvement Advisory Committee, City Managers Office-Action Center, City Managers Office-Administration, City Managers Office-Capital Improvement Management Office, City Managers Office-City Communications, City Managers Office-Emergency Management, City Managers Office-Public Improvement Advisory Committee, City Planning & Development-Administration, Cleaning City Property, Cleaning vacant Land Trust Lots, Communicable Disease Control, Completed Contractor Restoration Concerns, Convention and Entertainment Center-All, Cooling Center Transportation, Crack Sealing -(South of 47th Street and West of Blue PKWY), Crack Sealing -(South of River to 47th Street and East of Blue PKWY), Crack Sealing (North of River), Crack Sealing (South of 47th Street), Crack Sealing (South of River to 47th Street and East of Blue PKWY), Crack Sealing (South of River to 47th Street), Dangerous Building, Dangerous Building and Emergency Board Up, Dangerous Building Emergency, Dead Animal Pick-up, Dead Tree on Land Trust Lots, Demolition Debris and Landfill, Dispute of Assessment for NCS Property Abatement, Ditch Cleaning (North from 27th Street), Ditch Cleaning (North of River), Ditch Cleaning (South of 27th Street), Ditch Cleaning (South of 47th Street), Ditch Cleaning (South of River to 47th Street), Drug Abatement Response Team, Elevator Inspection, Emerald Ash Borer Questions, Emerald Ash Borer Tree Removal, Emergency Animal Levacy Jackson County, Emergency Home Repair Program, Emergency Water Turn Off, Erosion or Sediment, Facilities - City Owned, Fence or Wall, Field Services Escalation, Finance - All, Fire - All, Food Establishment Complaints, Forestry Customer Service Comment, Good Neighbor Award Nomination, Google's FTTH Project Contractor Verification, Government - ADA - Airport, Government - Not Provided by KCMO - KCATA/Metro, Government-311 Administration-Appearance/Speaker, Government-311 Administration-Document Request, Government-311 Administration-Feedback, Government-311 Administration-Field Investigation, Government-311 Administration-Lack of information, Government-311 Administration-Review, Government-ADA-Sidewalks/Curbs, Government-City Manager Äôs Office-All, Government-City Managers Office-Compass KC, Government-Communications-Citizen Satisfaction Survey, Government-Communications-KCMO.gov, Government-Communications-KCMore, Government-Communications-News and Media, Government-Finance-Collection Notice, Government-Finance-Feedback, Government-Finance-Tax, Government-Human Resources (HR)-All, Government-IT/Websites-Other Websites - Tech, Government-IT/Websites-Parks Website, Government-IT/Websites-Water Website, Government-Law-Financial Loss, Government-Law-Other Litigations, Government-Law-Property Damage, Government-Law-Vehicle Damage, Government-Mayor Äôs Office-Feedback, Government-NHS (Neighborhood Housing Services)-Assessment Dispute, Government-NHS (Neighborhood Housing Services)-Feedback, Government-NHS (Neighborhood Housing Services)-Good Neighbor Award, Government-NHS (Neighborhood Housing Services)-Home Repair, Government-NHS (Neighborhood Housing Services)-No Known Property Owner, Government-NHS (Neighborhood Housing Services)-Paint Program, Government-NHS (Neighborhood Housing Services)-Weatherization Program, Government-Not Provided By KCMO-211, Government-Not Provided By KCMO-County, Government-Not Provided By KCMO-KCPD, Government-Not Provided By KCMO-MODOT, Government-Not Provided By KCMO-State, Government-Public Works-Call Back Request, Government-Public Works-Feedback, Government-Public Works-Question, Government-Public Works-Service Issue/Problem, Government-Security/Safety-Employee Behavior, Government-Security/Safety-How's My Driving, Government-Security/Safety-Tours

learning. The visualization validated this for the DP case study, which had 462 block groups in Kansas City, Missouri, and a median home value (in dollars) of 1 block group. Applying DP to publicly available data and showing the degree of privacy and prediction model accuracy is an intriguing and significant contribution for interpretation.

Fourth, we contributed to categorize the problem domain and solution (department) domain in 311 calls. The categorization's primary accomplishment is mapping and grouping four distinct categories: Type, Category, Request Type, and Details. Many of the 311 domain's features would benefit from this categorization, as would essential elements of 311 data interoperability for sharing varied community concerns and solutions.

Finally, we used a variety of data visualization approaches to display tabular data in temporal and spatial domains in the 311 call domain. The data visualization tool is search-based for looking at specific datasets or visualizations. The time series analysis is mainly performed dynamically in response to a user query, and the interactive map chart of the 311 calls is constructed using hierarchical closest neighbor partitioning techniques. We found that the visualization is quite natural to comprehend the underlying context of the residences due to 311 calls for multiple neighborhoods.

Because of the complexity and scope of our work, the proposed framework is limited: First, we used specified data from KCMO 311 data for a given period, namely 2007 to 2020. As a result, we were unable to conduct a comparison with other frameworks or use other datasets. As

Table 15
KCMO's 311 call problem taxonomy: request type (3).

REQUEST TYPE (1229)	Graffiti - Park Property Central Region, Graffiti - Park Property North Region, Graffiti - Park Property South Region, Graffiti Abatement Training, Graffiti on City Property, Graffiti on Private Property, Graffiti on Signs, Graffiti on Streetlight Pole, Graffiti Removed by Citizen, Guard Rail Repair (North of River), Guard Rail Repair (South of 47th Street and West of Blue PKWY), Guard Rail Repair (South of 47th Street), Guard Rail Repair (South of the River to 47th Street and East of Blue PKWY), Guard Rail Repair (South of the River to 47th Street), Hazardous Structure, Health-All, Healthy Homes Program, High Bill Adjustment, High Water Bill Inspection, Homeless Camp, Hotel and Motel Inspections, Hydrant Repair Referral, Icing Caused by Water Leak, Illegal Dumping on Property, Illegal Dumping on Right of Way, Info Center Escalation, Information Request - Assessment for Private Property, Information Request-Neighborhood, Information Request-Railroad Property, Information Technology-All, Injury or Cruelty Involving an Animal, Integrated Pest Management Class, Issues on City Maintained Property, Land Trust Mowing List, Landscaping and Blvd/Pkwy Mowing, Law-All, Leaf and Brush Pick-up Missed, Lights/Signals-Services-Call Back Request, Lights/Signals-Services-Feedback, Lights/Signals-Services-Question, Lights/Signals-Services-Service Issue/Problem, Lights/Signals-Street Light-1 light out, Lights/Signals-Street Light-2 to 4 lights out, Lights/Signals-Street Light-5+ lights out, Lights/Signals-Street Light-Blinking, Lights/Signals-Street Light-Burning during the day, Lights/Signals-Street Light-Graffiti, Lights/Signals-Street Light-New Light Request, Lights/Signals-Street Light-New/Remove Light, Lights/Signals-Street Light-Other Light Issue, Lights/Signals-Street Light-Parks Maintained, Lights/Signals-Street Light-Pole down or leaning, Lights/Signals-Street Light-Wire Down/Hanging Arm, Lights/Signals-Traffic Signal-Audio, Lights/Signals-Traffic Signal-Bike Route, Lights/Signals-Traffic Signal-Graffiti, Lights/Signals-Traffic Signal-New Signal Request, Lights/Signals-Traffic Signal-Pedestrian/Walkway, Lights/Signals-Traffic Signal-Signal flashing, Lights/Signals-Traffic Signal-Signal malfunction, Lights/Signals-Traffic Signal-Signal out, Lights/Signals-Traffic Signal-Signal pole damage, Lights/Signals-Traffic Signal-Signal timing, Mayors Office-All, Metal Plate in Street (North of River), Metal Plate in Street (South of 47th Street), Metal Plate in Street (South of River to 47th Street), Meter Leak or Problem, Minor Home Repair Program, Mowing/Weeds - Visibility - Intersection, Mowing/Weeds-Alley-District 1, Mowing/Weeds-Alley-District 2, Mowing/Weeds-Alley-District 3, Mowing/Weeds-Blvd/Pkwy-Central, Mowing/Weeds-Blvd/Pkwy-North, Mowing/Weeds-Blvd/Pkwy-South, Mowing/Weeds-Paper Street-District 1, Mowing/Weeds-Paper Street-District 2, Mowing/Weeds-Paper Street-District 3, Mowing/Weeds-Parks-Central, Mowing/Weeds-Parks-Feedback, Mowing/Weeds-Parks-North, Mowing/Weeds-Parks-South, Mowing/Weeds-Public Property-City Facility, Mowing/Weeds-Public Property-City Vacant Lot, Mowing/Weeds-Public Property-Land Bank, Mowing/Weeds-Right of Way (ROW)-On Demand (PW), Mowing/Weeds-Right of Way (ROW)-Region - Central, Mowing/Weeds-Right of Way (ROW)-Region - North, Mowing/Weeds-Right of Way (ROW)-Region - South, Mowing City Property, Mowing of Private Property with No Structure, Mowing on Land Trust Lots, Mowing on the Right of Way, Mud and silt, Neighborhood and Community Services Administration, Neighborhood Preservation Administration, Neighborhood Preservation Priority Review, New Water Service, No Known Property Owner, No Water Bill Received, Noise Complaints, Nuisance Violations, Nuisance Violations on Private Property Vacant Structure, Outdoor Air Quality, Paint Program, Park Maintenance Central Region, Park Maintenance North Region, Park Maintenance South Region, Parking Meter Repair, Parking Meter Repair or Damage, Parking Ticket Complaint, Parking Ticket Issue, Parks & Recreation-Athletics-Fields ÄíCentral, Parks & Recreation-Athletics-Fields ÄíNorth, Parks & Recreation-Athletics-Fields ÄíSouth, Parks & Recreation-Athletics-Programming, Parks & Recreation-Community Center-Brush Creek CC, Parks & Recreation-Community Center-Garrison CC, Parks & Recreation-Community Center-Gregg/Klice CC, Parks & Recreation-Community Center-Hillcrest CC, Parks & Recreation-Community Center-Kansas City North CC, Parks & Recreation-Community Center-Line Creek CC, Parks & Recreation-Community Center-Marlborough CC, Parks & Recreation-Community Center-Southeast CC, Parks & Recreation-Community Center-Tony Aguirre CC, Parks & Recreation-Community Center-Westport Roanoke CC, Parks & Recreation-Facilities/Attractions-Aquatics, Parks & Recreation-Facilities/Attractions-Bruce Watkins Cultural Center, Parks & Recreation-Facilities/Attractions-Fountains, Parks & Recreation-Facilities/Attractions-Kansas City Museum, Parks & Recreation-Facilities/Attractions-Lakeside Nature Center, Parks & Recreation-Facilities/Attractions-Other, Parks & Recreation-Facilities/Attractions-Shelter ÄíCentral, Parks & Recreation-Facilities/Attractions-Shelter ÄíNorth, Parks & Recreation-Facilities/Attractions-Shelter ÄíSouth, Parks & Recreation-Facilities/Attractions-Starlight Theater, Parks & Recreation-Graffiti-Central, Parks & Recreation-Graffiti-North, Parks & Recreation-Graffiti-South, Parks & Recreation-Landscaping-All, Parks & Recreation-Landscaping-Central (Dept. Use), Parks & Recreation-Landscaping-North (Dept. Use), Parks & Recreation-Landscaping-South (Dept. Use), Parks & Recreation-Park Maintenance-Central, Parks & Recreation-Park Maintenance-North, Parks & Recreation-Park Maintenance-South, Parks & Recreation-Services-Call Back Request, Parks & Recreation-Services-Feedback, Parks & Recreation-Services-Question, Parks & Recreation-Services-Service Issue/Problem, Parks & Recreation-Trails-Central, Parks & Recreation-Trails-North, Parks & Recreation-Trails-South, Parks and Recreation Administration, Parks Capital Project Referral, Parks Community Services, Parks Drinking Fountain Central Referral, Parks Drinking Fountain North Referral, Parks Drinking Fountain South Referral, Parks Facility Indoor Plumbing North Referral, Parks Facility Maintenance C Referral, Parks Facility Maintenance N Referral, Parks Facility Maintenance Referral, Parks Facility Maintenance S Refer, Parks Facility Outdoor Plumbing Central Referral, Parks Lighting Central Referral, Parks Lighting North Referral, Parks Lighting South Referral, Pending Restoration Referral, Permit for Private Removal of a City Tree, Permit to prune, remove, or plant a City Tree, PIAC Referral From Department, Pipeline Barricade Referral, Pipeline Code 1 Repair - Referral, Pipeline Code 2 Repair - Referral, Pipeline Code 3 Repair - Referral, Pipeline Contractor Restoration, Pipeline Pavement Restoration Referral, Pipeline Repair or Restoration Concerns, Pipeline Repair Referral, Pipeline Turf Referral, Plate Hazard, Plate in Street, Pothole (North of River), Pothole (River south to 47th Street and East of Blue PKWY), Pothole (River south to 47th Street), Pothole (South of 47th Street and West Of Blue PKWY), Pothole (South of 47th Street), Pothole (South of River to 47th Street), Prob/Build/Construct-Property Maintenance-Northeast, Project Planning for Streets and Traffic, Prop/Build/Construct - Elevator Inspection - City Facility, Prop/Build/Construct - Land Use/Zoning - Miscellaneous Zoning, Prop/Build/Construct - Land Use/Zoning - Unapproved Land Use, Prop/Build/Construct-Construction Issue/Concern-Development Plans -Landscaping, Prop/Build/Construct-Construction Issue/Concern-Erosion - New Construction, Prop/Build/Construct-Construction Issue/Concern-Large Amount of Demo Debris, Prop/Build/Construct-Construction Issue/Concern-Mud/Silt on street, Prop/Build/Construct-Construction Issue/Concern-Other

a result, the framework may not be adequate for generating additional data from diverse sources, and further evaluation is required to confirm the significance. Second, some proposed algorithms or methodologies are implemented or extended from existing ones. We could not contribute much to the design of fresh approaches due to the suggested framework's wide variety of viewpoints. Third, the framework is primarily designed for tabular data. A multi-modal framework capable of dealing with text data and visuals might be a more exciting avenue for this research.

8. Conclusions and future work

Governments in general, particularly urban governments, are awash with data yet have difficulty fully developing insights from divergent data sources. As we have shown, the case of Kansas City, Missouri, is emblematic of a government that has had an aggressive orientation towards leveraging data collection to drive more efficient and effective public service delivery; following an extensive partnership with CISCO in 2015, Kansas City has trumpeted that it is "one of the

Table 16
KCMO's 311 call problem taxonomy: request type (4).

REQUEST TYPE (1229)	Prop/Build/Construct-Construction Issue/Concern-Work without permit, Prop/Build/Construct-Dangerous Building-Emergency, Prop/Build/Construct-Dangerous Building-On list, Prop/Build/Construct-Dangerous Building-Open to entry, Prop/Build/Construct-Elevator Inspection-Private/Commercial, Prop/Build/Construct-Land Use/Zoning Issue-Business in a residence, Prop/Build/Construct-Land Use/Zoning Issue-Excessive Yard Sale, Prop/Build/Construct-Land Use/Zoning Issue-Floodplan Management, Prop/Build/Construct-Land Use/Zoning Issue-Number of Dwelling Units, Prop/Build/Construct-Land Use/Zoning Issue-Number of Residents, Prop/Build/Construct-Land Use/Zoning-Short Term Rental, Prop/Build/Construct-Parking Lot - Surface, Prop/Build/Construct-Parking Lot-Parking Spaces, Prop/Build/Construct-Property Maintenance-DART, Prop/Build/Construct-Property Maintenance-Land Bank, Prop/Build/Construct-Property Maintenance-Land Bank (Services), Prop/Build/Construct-Property Maintenance-NNI, Prop/Build/Construct-Property Maintenance-No Utilities, Prop/Build/Construct-Property Maintenance-SKCA, Prop/Build/Construct-Services (CPD)-Call Back Request, Prop/Build/Construct-Services (CPD)-Feedback, Prop/Build/Construct-Services (CPD)-Post Demo Issues, Prop/Build/Construct-Services (CPD)-Question, Prop/Build/Construct-Services (CPD)-Service Issue/Problem, Prop/Build/Construct-Services (NPD)-Call Back Request, Prop/Build/Construct-Services (NPD)-Feedback, Prop/Build/Construct-Services (NPD)-Hazardous Structure, Prop/Build/Construct-Services (NPD)-Question, Prop/Build/Construct-Services (NPD)-Service Issue/Problem, Property and Nuisance Violations, Property Maintenance, Property Violations, Public Health-Disease Control-All, Public Health-Disease Control-Business Question, Public Health-Food Establishment-Catering, Public Health-Food Establishment-Convenience/Grocery Store, Public Health-Food Establishment-Education/Health, Public Health-Food Establishment-Food Truck, Public Health-Food Establishment-Other, Public Health-Food Establishment-Restaurant, Public Health-Hotel/Motel-Bedbugs, Public Health-Hotel/Motel-Other, Public Health-Hotel/Motel-Unsanitary Conditions, Public Health-Noise-All, Public Health-Noise-Construction, Public Health-Noise-Music, Public Health-Noise-Other, Public Health-Noise-Trash Truck, Public Health-Noise-Vehicles, Public Health-Outdoor Air Quality-All, Public Health-Pool/Hot Tub-Public, Public Health-Public Restroom-Other, Public Health-Public Restroom-Restaurant, Public Health-Public Restroom-School, Public Health-Services-Bedbugs, Public Health-Services-Call Back Request, Public Health-Services-Feedback, Public Health-Services-Healthy Homes Information, Public Health-Services-Lead Poisoning Prevention Info, Public Health-Services-Pest Management Class, Public Health-Services-Question, Public Health-Services-Service Issue/Problem, Public Health-Smoking/Tobacco-Public Building, Public Health-Smoking/Tobacco-Restaurant, Public Health-Smoking/Tobacco-Sales to minors, Public Pool Inspections, Public Restroom Sanitation, Public Safety-Ambulance-Billing/Administration, Public Safety-Emergency Management-Feedback, Public Safety-Emergency Management-Weather Siren, Public Safety-Fire-Carbon monoxide, Public Safety-Fire-Fire Investigation, Public Safety-Fire-Hazardous materials, Public Safety-Fire-New smoke alarm request, Public Safety-Fire-Permit (Fireworks), Public Safety-Fire-Question/Feedback, Public Safety-Fire-Smell or see smoke, Public Safety-Municipal Court-Accommodations, Public Safety-Municipal Court-Address Change, Public Safety-Municipal Court-Compliance Letter, Public Safety-Municipal Court-Continuance, Public Safety-Municipal Court-Monetary Inquiries, Public Safety-Municipal Court-Probation, Public Safety-Municipal Court-Question/Feedback, Public Safety-Police-Homeless camp, Public Safety-Police-Other, Public Safety-Regulated Industries-Adult Entertainment, Public Safety-Regulated Industries-Alcohol Business, Public Safety-Regulated Industries-Amusement Business, Public Safety-Regulated Industries-Day Labor Business, Public Safety-Regulated Industries-Question/Feedback, Public Safety-Regulated Industries-Salvage and Scrap, Public Safety-Regulated Industries-Tobacco Business, Public Safety-Regulated Industries-Vehicle for hire, Public Works Administration, Questions or Comments, Rat Control Education, Rat Control Treatment, Recycle Bin Damaged, Recycle Bin New Service, Recycling Missed, Referral for Abatement of Illegal Dumping, Referral for Sewage on Private Property, Regulated Industries, Rural ROW Mowing, Salvage Yard Investigation, Service or Curb Box Repair Referral, Sewer - Services - Sewer Restoration, Sewer - Sewer Backup - Owner Responsibility, Sewer Assessment Questions, Sewer Cleaning or Deodorizing - Referral, Sewer Cleaning or Repair and Waste Water Concern, Sewer Contractor Restoration Referral, Sewer Deodorizing - Referral, Sewer Investigation and Repairs - Referral, Sewer Investigation Referral, Sewer Pavement Restoration Referral, Sewer Projects, Sewer Repair Referral, Sewer Turf Restoration - Referral, Sewer Water in Basement, Sewer Water in Basement - Referral, Sewer-Manhole-Missing/Broken, Sewer-Sewer Backup/Leak-From Manhole, Sewer-Sewer Backup/Leak-In a house, Sewer-Sewer Backup/Leak-Outside, Sewer-Sewer Backup/Leak-Treatment Plant, Sewer-Sewer Odor-Indoor, Sewer-Sewer Odor-Outdoor, Sewer-Sewer Referral-Investigation, Sewer-Sewer Referral-Private Line, Sewer-Sewer Referral-Repair, Sewer-Sewer Referral-Restoration, Sewer-Sewer Referral-Restoration Issue, Sewer-Sinkhole-Sidewalk, Sewer-Sinkhole-Street, Sewer-Sinkhole-Yard, Sidewalk Hazardous, Sidewalk or Curb in Disrepair, Sidewalk projects that are funded by GO Bond funds, Sidewalk to be replaced Due to City Project, Sidewalk/Curb/Ditch-Park Property-Central, Sidewalk/Curb/Ditch-Park Property-North, Sidewalk/Curb/Ditch-Park Property-South, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-Park/Blvd/Pkwy, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-ADA, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-Blvd/Pkwy, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-City, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-Private Property, Sidewalks/Curbs/Ditch-Damage/Dis-Repair-Replace - City Project, Sidewalks/Curbs/Ditch-Ditch-North, Sidewalks/Curbs/Ditch-Ditch-South, Sidewalks/Curbs/Ditch-New Request-Blvd/Pkwy/Park, Sidewalks/Curbs/Ditch-New Request-City Street, Sidewalks/Curbs/Ditch-Obstructed/Closure-ADA, Sidewalks/Curbs/Ditch-Obstructed/Closure-Blocked, Sidewalks/Curbs/Ditch-Obstructed/Closure-Ramp, Sidewalks/Curbs/Ditch-Snow/Ice-Bridge, Sidewalks/Curbs/Ditch-Snow/Ice-City Owned, Sidewalks/Curbs/Ditch-Snow/Ice-Park - Central, Sidewalks/Curbs/Ditch-Snow/Ice-Park - North, Sidewalks/Curbs/Ditch-Snow/Ice-Park - South, Sidewalks/Curbs/Ditch-Unapproved Objects-Basketball Hoop/Goal, Sidewalks/Curbs/Ditch-Unapproved Objects-Landscaping, Sidewalks/Curbs/Ditch-Unapproved Objects-Merchandise, Sidewalks/Curbs/Ditch-Unapproved Objects-Other, Sign Missing, Sign Missing or Damaged - Non-emergency, Sign Violation Zoning, Signs-Commercial Signs/Ads-Banner/Billboard, Signs-Commercial Signs/Ads-In Right-Of-Way, Signs-No Dumping Sign-District 1, Signs-No Dumping Sign-District 2, Signs-No Dumping Sign-District 3, Signs-Street Name Signs-Down/Damaged, Signs-Street Name Signs-Graffiti, Signs-Street Name Signs-Honorary/Renaming, Signs-Street Name Signs-Missing, Signs-Street Name Signs-New Sign Request, Signs-Street Name Signs-Remove Sign, Signs-Traffic Sign-Bike Route, Signs-Traffic Sign-Damaged ÄíEmergency, Signs-Traffic Sign-Damaged ÄíNon-Emergency, Signs-Traffic Sign-Graffiti, Signs-Traffic Sign-Missing ÄíEmergency, Signs-Traffic Sign-Missing ÄíNon-Emergencyfication Referral

world's most connected Smart City'' (KCMO, 2021b). The extensive use of sensors within the downtown area has led to an explosion in data. Yet, the Kansas City example illustrates the difficulty of integrating data analysis into a usable workflow for public management decision-making.

Here we address the omnipresent issue of needing to mix data from divergent sources to aid in real-world decisions about how to allocate limited resources. An inductive method that relies on machine learning techniques allows for the clustering of data across time and geographic space to dictate areas of a city's landscape where needs are emergent, all without the constraints of jurisdictional boundaries that can serve to segregate data, which in turn impedes an analyst's ability to create a holistic explanation for current and future trends of citizen need.

While the work presented here indicates one path forward, there are limitations in the analysis. First, the differential time horizons of data make it challenging to create accurate temporal analysis. For example, sociodemographic data from the American Community Survey

Table 17
KCMO's 311 call problem taxonomy: request type (5).

REQUEST TYPE (1229)	Signs-Traffic Sign-New Sign Installation, Signs-Traffic Sign-New Sign Request, Signs-Traffic Sign-Remove Sign, Signs/Banners/Removal, Sinkhole or Sewer Repair - Referral, SLA0 Streetlight Pole Down or Leaning, SLB0 Streetlight Wire Down or Hanging Arm, SLD0 Streetlight Out - Multiple Outage of Five or More, SLG0 Streetlight Out - Single Light Out, SLG1 Streetlight Blinking Off and On, SLG2 Streetlight Burning During Day, SLG3 Streetlights Out - Two to Four Lights, SLP0 Streetlight Miscellaneous, SLQ0 Streetlight Problem Identi, Smoking Complaints, Snow and Ice Administration, Snow or Ice on Sidewalk, Snow or Ice on Street (North of River), Snow or Ice on Street (South of 47th Street and West of Blue PKWY), Snow or Ice on Street (South of 47th Street), Snow or Ice on Street (South of River to 47th Street and East of Blu, Snow or Ice on Street (South of River to 47th Street), Snow or Ice on Street District 1, Snow or Ice on Street District 2, Snow or Ice on Street District 3, Solid Waste Customer Service, Staff Complaints for Solid Waste, Staff Compliments for Solid Waste, Storm/Sewer-Storm Water Referral-Repair, Storm/Sewer-Stormwater Referral-Investigation, Storm Damage East, Storm Water/Sewer - Construction Repair - Referral, Storm Water/Sewer - Flood - Parks Barricade, Storm Water/Sewer - Flood - Public Works Barricade, Storm Water/Sewer - Flood - Water Services Barricade, Storm Water/Sewer - Services - Sewer Restoration, Storm Water/Sewer-Culvert-North, Storm Water/Sewer-Culvert-South, Storm Water/Sewer-Improved Channel-Clogged/Debris, Storm Water/Sewer-Improved Channel-Damaged, Storm Water/Sewer-Manhole-Missing/Broken, Storm Water/Sewer-Services-Call Back Request, Storm Water/Sewer-Services-Feedback, Storm Water/Sewer-Services-Question, Storm Water/Sewer-Services-Service Issue/Problem, Storm Water/Sewer-Sewer Backup/Leak-From Manhole, Storm Water/Sewer-Sewer Backup/Leak-In a house, Storm Water/Sewer-Sewer Backup/Leak-Outside, Storm Water/Sewer-Sewer Backup/Leak-Treatment Plant, Storm Water/Sewer-Sewer Odor-Indoor, Storm Water/Sewer-Sewer Odor-Outdoor, Storm Water/Sewer-Sewer Referral-Basement, Storm Water/Sewer-Sewer Referral-Investigation, Storm Water/Sewer-Sewer Referral-Private Line, Storm Water/Sewer-Sewer Referral-Repair, Storm Water/Sewer-Sewer Referral-Restoration, Storm Water/Sewer-Sewer Referral-Restoration Issue, Storm Water/Sewer-Sinkhole-Sidewalk, Storm Water/Sewer-Sinkhole-Street, Storm Water/Sewer-Sinkhole-Yard, Storm Water/Sewer-Storm Drain/Catch Basin-Clogged/Over-Flow, Storm Water/Sewer-Storm Drain/Catch Basin-Damaged/Broken, Storm Water/Sewer-Storm Drain/Catch Basin-Dropped Personal Item, Storm Water/Sewer-Storm Referral-Creek/Green Solution, Storm Water/Sewer-Stormwater Referral-Restoration, Storm Water-Storm Drain/Catch Basin-Clogged/Over-Flow, Stormwater - Flood - Parks Barricade, Stormwater - Flood - Public Works Barricade, Stormwater - Flood - Water Services Barricade, Stormwater/Sewer-Services-Call Back Request, Stormwater/Sewer-Services-Feedback, Stormwater/Sewer-Services-Service Issue/Problem, Stormwater Charges, Stormwater Projects, Stormwater-Culvert-North, Stormwater-Culvert-South, Stormwater-Improved Channel-Clogged/Debris, Stormwater-Improved Channel-Damaged, Stormwater-Storm Drain/Catch Basin-Clogged/Over-Flow, Stormwater-Storm Drain/Catch Basin-Damaged/Broken, Stormwater-Storm Drain/Catch Basin-Dropped Personal Item, Stormwater-Stormwater Referral-Investigation, Stormwater-Stormwater Referral-Repair, Stormwater-Stormwater Referral-Restoration, Stray Animal Confined, Street and Traffic Administration, Street and Traffic Permit, Street Light Malfunctioning, Street Light New, Street Light Out, Street Light Pole Damage, Rust, Vandalism, Street Light Restoration Referral, Street Maintenance General, Street Maintenance General (North of the River), Street Maintenance General (South of 47th Street and West of Blue PKWY), Street Maintenance General (South of 47th Street), Street Maintenance General (South of River to 47th Street and East of Blue PKWY), Street Maintenance General (South of River to 47th Street), Street Markings Painted, Street Preservation or Overlay, Street Sign Damage, Street Sign Missing, Street Sweeping (North of River), Street Sweeping (South of 47th Street), Street Sweeping (South of River to 47th Street), Street Sweeping Citywide, Street/Roadway/Alley - Obstructed/Closure - Marathon/Run/Race, Street/Roadway/Alley-Snow/Ice-Snow Blocking Drive, Streets/Roadways/Alleys-Broken Asphalt-Blvd/Pkwy, Streets/Roadways/Alleys-Broken Asphalt-Bridge, Streets/Roadways/Alleys-Broken Asphalt-District 1, Streets/Roadways/Alleys-Broken Asphalt-District 2, Streets/Roadways/Alleys-Broken Asphalt-District 3, Streets/Roadways/Alleys-Crack-Bridge, Streets/Roadways/Alleys-Crack-District 1, Streets/Roadways/Alleys-Crack-District 2, Streets/Roadways/Alleys-Crack-District 3, Streets/Roadways/Alleys-Cut/Permit-Blvd/Pkwy, Streets/Roadways/Alleys-Cut/Permit-District 1, Streets/Roadways/Alleys-Cut/Permit-District 2, Streets/Roadways/Alleys-Cut/Permit-District 3, Streets/Roadways/Alleys-Guardrail-Blvd/Pkwy, Streets/Roadways/Alleys-Guardrail-Bridge, Streets/Roadways/Alleys-Guardrail-District 1, Streets/Roadways/Alleys-Guardrail-District 2, Streets/Roadways/Alleys-Guardrail-District 3, Streets/Roadways/Alleys-Markings/Paint-Bike lanes, Streets/Roadways/Alleys-Markings/Paint-Crosswalk, Streets/Roadways/Alleys-Markings/Paint-Lane lines, Streets/Roadways/Alleys-Markings/Paint-New Marking Request, Streets/Roadways/Alleys-Markings/Paint-Turn lanes, Streets/Roadways/Alleys-Obstructed/Closure-Detour, Streets/Roadways/Alleys-Obstructed/Closure-Event/Closure Request, Streets/Roadways/Alleys-Obstructed/Closure-Lane Closed, Streets/Roadways/Alleys-Obstructed/Closure-Road Closed, Streets/Roadways/Alleys-Plate-Loose, Streets/Roadways/Alleys-Plate-Missing/Displaced, Streets/Roadways/Alleys-Plate-Request, Streets/Roadways/Alleys-Pothole-Blvd/Pkwy, Streets/Roadways/Alleys-Pothole-Bridge, Streets/Roadways/Alleys-Pothole-District 1, Streets/Roadways/Alleys-Pothole-District 2, Streets/Roadways/Alleys-Pothole-District 3, Streets/Roadways/Alleys-Resurfacing-District 1, Streets/Roadways/Alleys-Resurfacing-District 2, Streets/Roadways/Alleys-Resurfacing-District 3, Streets/Roadways/Alleys-Resurfacing-Service Issue/Problem, Streets/Roadways/Alleys-Sinkhole Referral-District 1, Streets/Roadways/Alleys-Sinkhole Referral-District 2, Streets/Roadways/Alleys-Sinkhole Referral-District 3, Streets/Roadways/Alleys-Snow/Ice-District 1, Streets/Roadways/Alleys-Snow/Ice-District 2, Streets/Roadways/Alleys-Snow/Ice-District 3, Streets/Roadways/Alleys-Snow/Ice-Feedback, Streets/Roadways/Alleys-Speed Bump-Removal, Streets/Roadways/Alleys-Speed Bump-Request New, Streets/Roadways/Alleys-Street Clean Up-Bike Route, Streets/Roadways/Alleys-Street Clean Up-Blvd/Pkwy - Central, Streets/Roadways/Alleys-Street Clean Up-Blvd/Pkwy - North, Streets/Roadways/Alleys-Street Clean Up-Blvd/Pkwy - South, Streets/Roadways/Alleys-Street Clean Up-Bridge, Streets/Roadways/Alleys-Street Clean Up-District 1, Streets/Roadways/Alleys-Street Clean Up-District 2, Streets/Roadways/Alleys-Street Clean Up-District 3, Streets/Roadways/Alleys-Street Clean Up-Sweeping Schedule, Streets/Roadways/Alleys-Street Services-Call Back Request, Streets/Roadways/Alleys-Street Services-Feedback, Streets/Roadways/Alleys-Street Services-Question, Streets/Roadways/Alleys-Street Services-Service Issue/Problem, Streets/Roadways/Alleys-Street Services-Street Permit Research, Streets/Roadways/Alleys-Street Services-Traffic Study Request

aggregated to the Census Bureau's block group level only occurs in 5-year estimates meaning that a study of 311 calls that can be attributed down to the minute must necessarily be joined with data collected over sixty months. Second, the inductive methodology developed here to attempt to account for standard error in some types of data, such as 311 calls where the data represents the entire population of calls over a while; this does not hold for census data drawn from the American Community Survey where sometimes significant standard errors can influence inference. Lastly, the nature of machine learning analytical frameworks requires a method to impute non-existent values in the data; all imputation methods inherently have shortcomings and rely on careful consideration as to the introduction of systemic bias into the data. These challenges deserve deeper exploration in future studies of city administrative and secondary data.

The data revolution is exciting and daunting for city administrators. However, the ability to develop a method to aggregate and ultimately create inference from disparate data sources can aid administrators' quest to provide finite city services to communities in need. In addition,

Table 18
KCMO's 311 call problem taxonomy: request type (6).

REQUEST TYPE (1229)	Streets/Roadways/Alleys-Street Services-Traffic Video Request, Streets/Roadways/Alleys-Other Maintenance-District 1, Streets/Roadways/Alleys-Other Maintenance-District 2, Streets/Roadways/Alleys-Other Maintenance-District 3, Streets/Roadways/Alleys-Snow/Ice-Snow Verify List, Tow Services Administration, Traffic Engineering and Traffic Studies, Traffic Sign Damage, Traffic Sign Down or Missing, Traffic Sign Down or Missing - Emergency, Traffic Signal Malfunction, Traffic Signal Out, Traffic Signal Pole Damage, Rust, Vandalism, Traffic Study for Signs, Reduced Speed, Signal, Crosswalk, Trahs-Recycle Bin Damaged, Trahs-Recycle Bin New Service, Trash - Missed by City, Trash - Missed by Contractor, Trash - Multi Unit Investigation, Trash - Recycle Bin Damaged, Trash - Recycle Bin New Service, Trash - Roll-Out Cart Administration, Trash - Roll-out Cart Damaged or Broken, Trash - Roll-out Cart Missing, Trash - Solid Waste Customer Service, Trash - Staff Complaints for Solid Waste, Trash/Recycling - Dumping - Alley - D1, Trash/Recycling - Dumping - Alley - D2, Trash/Recycling - Dumping - Alley - D3, Trash/Recycling - Trash Collection - City Trash - Recycle, Trash/Recycling - Trash Collection - North Trash - Recycle, Trash/Recycling - Trash Collection - South Trash - Recycle, Trash/Recycling-Bulky Pick Up -Address Validation, Trash/Recycling-Bulky Pick Up -Missed Pick Up, Trash/Recycling-Dumping-Alley, Trash/Recycling-Dumping-Camera Request, Trash/Recycling-Dumping-Landlord Set Out, Trash/Recycling-Dumping-Park - Central, Trash/Recycling-Dumping-Park - North, Trash/Recycling-Dumping-Park - South, Trash/Recycling-Dumping-Private Property, Trash/Recycling-Dumping-Right of Way, Trash/Recycling-Dumping-Right of Way (ROW), Trash/Recycling-Dumping-ROW Central, Trash/Recycling-Dumping-ROW Land Bank Central, Trash/Recycling-Dumping-ROW Land Bank North, Trash/Recycling-Dumping-ROW Land Bank South, Trash/Recycling-Dumping-ROW North, Trash/Recycling-Dumping-ROW South, Trash/Recycling-Dumping-Solid Waste Referral, Trash/Recycling-Dumpsters-Dumpster Request, Trash/Recycling-Dumpsters-Street/Alley/ROW, Trash/Recycling-Early Set Out-Central, Trash/Recycling-Early Set Out-North, Trash/Recycling-Early Set Out-South, Trash/Recycling-Landlord Set Out-Central, Trash/Recycling-Landlord Set Out-North, Trash/Recycling-Landlord Set Out-South, Trash/Recycling-Leaf/Brush-Appointment Missed, Trash/Recycling-Leaf/Brush-Dumping in Street, Trash/Recycling-Leaf/Brush-Missed Pick Up, Trash/Recycling-Nuisance-Early Set Out, Trash/Recycling-Nuisance-Land Bank, Trash/Recycling-Recycling-Missed by City, Trash/Recycling-Recycling-Missed by Contractor North, Trash/Recycling-Recycling-Missed by Contractor South, Trash/Recycling-Recycling-Repeat Missed, Trash/Recycling-Services-Blue Bag Program, Trash/Recycling-Services-Call Back Request, Trash/Recycling-Services-Disability Pick Up, Trash/Recycling-Services-Feedback, Trash/Recycling-Services-Medical Tags, Trash/Recycling-Services-Multi-Unit, Trash/Recycling-Services-Question, Trash/Recycling-Services-Service Issue/Problem, Trash/Recycling-Solid Waste Operations-Brush, Trash/Recycling-Solid Waste Operations-Bulky, Trash/Recycling-Solid Waste Operations-Trash, Trash/Recycling-Trash Cart-Cart Size, Trash/Recycling-Trash Cart-Not Delivered, Trash/Recycling-Trash Collection-City Owned Trash Container, Trash/Recycling-Trash Collection-Missed by City, Trash/Recycling-Trash Collection-Missed by Contractor North, Trash/Recycling-Trash Collection-Missed by Contractor South, Trash/Recycling-Trash Collection-Repeat Missed, Trash Cart Damaged, Trash Cart Missing, Trash Missed by Contractor North, Trash Missed by Contractor South, Trash Multi Unit Investigation, Trash Recycle Missed by Contractor North, Trash Recycle Missed by Contractor South, Trash Recycling Missed by City, Trash Recycling Missed by Contractor, Trash Roll Out Cart Damaged or Missing, Trash Roll Out Cart Missing, Trash Storm Cleanup, Trash was missed for Tuesday pickup for 2 bags., Trash-Leaf and Brush Pick-up Missed, Trash-Missed Bulky, Trash-Recycling Missed, Trash-Staff Compliments for Solid Waste, Trash/Recycling - Leaf/Brush - Friday - Central, Trash/Recycling - Leaf/Brush - Friday - North, Trash/Recycling - Leaf/Brush - Friday - South, Trash/Recycling - Leaf/Brush - Monday - Central, Trash/Recycling - Leaf/Brush - Monday - North, Trash/Recycling - Leaf/Brush - Monday - South, Trash/Recycling - Leaf/Brush - Thursday - Central, Trash/Recycling - Leaf/Brush - Thursday - North, Trash/Recycling - Leaf/Brush - Thursday - South, Trash/Recycling - Leaf/Brush - Tuesday - Central, Trash/Recycling - Leaf/Brush - Tuesday - North, Trash/Recycling - Leaf/Brush - Tuesday - South, Trash/Recycling - Leaf/Brush - Wednesday - Central, Trash/Recycling - Leaf/Brush - Wednesday - North, Trash/Recycling - Leaf/Brush - Wednesday - South, Trash/Recycling - Trash Collection - Storm Trash, Tree Damage to City Tree from Storm Event, Tree Disease, Insects or Vandalism, Tree Inspection on Private Property, Tree Limb and Brush Removal from City Tree - Central Region, Tree Limb and Brush Removal from City Tree - North Region, Tree Limb and Brush Removal from City Tree - South Region, Tree Limb or Brush Removal for City Tree, Tree Live Removal, Tree Maintenance Seasonal, Tree or Limb Down Due to Snow, Tree Planting Central Region, Tree Planting for New Tree, Tree Planting North Region, Tree Planting South Region, Tree Pruning for Block, Tree Removal Central Region, Tree Removal North Region, Tree Removal of a City Tree, Tree Removal South Region, Tree Staff Compliment, Tree Staff Customer Service Complaint, Tree Storm Damage Central Region, Tree Storm Damage North Region, Tree Storm Damage South Region, Tree Stump Removal Central Region, Tree Stump Removal for City Tree, Tree Stump Removal North Region, Tree Stump Removal South Region, Tree Trimming at Intersection, Tree Trimming Central Region, Tree Trimming for City Tree, Tree Trimming from Electric Wire or Street Light, Tree Trimming from Street Light, Tree Trimming North Region, Tree Trimming on Land Trust Lots, Tree Trimming South Region, Trees - Permit - Ash Treatment, Trees-Care-Disease, Trees-Care-Insects, Trees-Care-Sick, Trees-Care-Vandalism, Trees-Emerald Ash-Question/Feedback, Trees-Emerald Ash-Removal, Trees-Emerald Ash-Treatment, Trees-Land Bank-Dead, Trees-Land Bank-Hazardous, Trees-Land Bank-Tree Debris, Trees-Permit-Plant, Trees-Permit-Remove, Trees-Permit-Trim, Trees-Private Property-Hazardous, Trees-Removal-Brush pickup, Trees-Removal-Dangerous, Trees-Removal-Declining, Trees-Removal-Dumping, Trees-Removal-Stump, Trees-Removal-Tree dead, Trees-Removal-Tree down, Trees-Removal-Tree Limbs, Trees-Services-EAB, Trees-Services-Feedback, Trees-Services-Plant City Tree Request, Trees-Services-Question, Trees-Services-Review, Trees-Services-Service Issue/Problem, Trees-Storm Damage-Tree Down, Trees-Storm Damage-Tree Limb/Brush, Trees-Storm Damage-Trim, Trees-Trimming-Block Pruning, Trees-Trimming-Seasonal, Trees-Trimming-Streetlight, Trees-Trimming-Tree Limbs, Trees-Trimming-Visibility, Turn Off Water Service For A Repair, Turn On Water Services - Internal Repairs Completed, Turn On Water Services For Paid Bill, Vacant Structure Open to Entry, Valve Repair Referral, Vehicle Illegally Parked, Vehicle Towed From Lot Without Owner's Permission by Tow Company, Vehicle-Scooter-Feedback/Concern, Vehicles/Parking-Abandoned On Street-Boat, Vehicles/Parking-Abandoned On Street-Car, Vehicles/Parking-Abandoned On Street-Other, Vehicles/Parking-Abandoned On Street-Trailer, Vehicles/Parking-Commercial Parking-Number of Handicap Spaces, Vehicles/Parking-Commercial Parking-Parking Surface - Pay Lot, Vehicles/Parking-Commercial Vehicle-On private property

the continued evolution of the smart city requires such methods so that administrators can continue to innovate in service delivery.

Future studies can conduct a comparative evaluation with other frameworks using other data from various sources as future work. Built upon this framework, researchers can enhance the novelty of the methodology for creating the open community framework and interpret the empirical findings through empirical measurement and assessment. This framework will be extended to improve its handling of free text and images in the long run. These works would make it possible to address the ultimate questions for sustainable communities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 19
KCMO's 311 call problem taxonomy: request type (7).

REQUEST TYPE (1229)	Vehicles/Parking-Illegally Parked-Street/Sidewalk, Vehicles/Parking-Meter-Repair, Vehicles/Parking-Tow Services-Feedback, Vehicles/Parking-Tow Services-Service Issue/Problem, Violations and Encroachments on the Right of Way, Visibility Problems at an Intersection, Volunteer Inspection Program, Wastewater Code 1 - Referral, Wastewater Code 2 - Referral, Wastewater Code 3 - Referral, Wastewater Treatment, Water - MFS Referral - Investigation, Water - Services - Pipeline Restoration, Water Bill Admin Tasks, Water Bill Concerns, Water Facilities and Engineering, Water Facilities and Stormwater Engineering, Water Industrial Waste Control, Water Leak or Pressure Problem, Water Main Break, Water Main Projects, Water Main Repair Referral, Water Payment or Refund, Water Pipeline Barricade, Water Quality - Laboratory Services, Water Repair and Restoration, Water Restoration Contract, Water Services Administration, Water Services Pavement Restoration, Water Services Same Day Service, Water Services Turf Restoration, Water Theft and Fraud, Water-Account-Account Services, Water-Account-Theft, Water-Leak-At Curb/In Yard, Water-Leak-Emergency Off/Flooding, Water-Leak-Hydrant, Water-Leak-Ice Due to Leak, Water-Leak-In Street, Water-Leak-Meter, Water-No Water/Pressure-High Pressure, Water-No Water/Pressure-Low Pressure, Water-No Water/Pressure-No Water, Water-Pipeline Referral-Curb-box/Service Line, Water-Pipeline Referral-Hydrant, Water-Pipeline Referral-Materials left at site, Water-Pipeline Referral-Pipeline/Restoration Issues, Water-Pipeline Referral-Restoration, Water-Pipeline Referral-Restoration Issue, Water-Pipeline Referral-Valve, Water-Pipeline Referral-Water Main, Water-Quality-Color/Cloudy, Water-Quality-Smell/Odor, Water-Quality-Taste, Water-Services-Call Back Request, Water-Services-Feedback, Water-Services-Question, Water-Services-Service Issue/Problem, Weatherization Program, Web SR - Health Code Violation, Web SR - Property Maintenance - Neighborhoods, Web SR - Street Maintenance/Snow, Web SR - Streetlights, Web SR - Trash, Recycle, Bulky, Web SR - Trees (City Trees Only), Wildlife, Zoning Inspections and Violations

Table 20
311 call solution taxonomy: department and work group.

Department (25)	Aviation, City Managers Office, City Planning and Development, Convention and Entertainment Center, Finance, Fire, General Service, Health, Housing Community Dev, Human Resources, Information Technology, IT, KCPD, Mayors Office, Municipal Court, NCS, NHS, Northeast, Northland, Parks & Rec, Parks & Recreation, Parks and Rec, Public Works, South, Water Services
Work Group (143)	Aviation–, City Managers Office-311 Call Center-, City Managers Office-311 Call Center-Administration, City Managers Office-311 Call Center-Support, City Managers Office-311-Field Staff, City Managers Office-Administration-, City Managers Office-Capital Projects-, City Managers Office-CIMO-, City Managers Office-Compass KC-, City Managers Office-Emergency Management-. City Managers Office-KCPD-, City Planning and Development-Administration, City Planning and Development-Administration-, City Planning and Development-Dev Comp-Zoning, City Planning and Development-Elevators-, City Planning and Development-Land Develop-Inspect, City Planning and Development-Permit Compliance-, Convention and Entertainment Center–, Finance–, Finance-PIAC-, Fire–, General Service-City Property-, General Service-Facility Management-, Health–, Health-Air Quality-, Health-Communicable Disease Control-, Health-Community Environmental Health-, Health-Food Protection-, Health-Healthy Homes-, Health-Healthy Homes-Bedbugs, Health-Lead Poisoning Prevention-, Health-Noise Control-, Health-Rat-, Health-Tobacco Control-, Housing Community Dev-Property Preservation, Human Resources–, Information Technology–, Information Technology-Email issue-, Information Technology-ERP Testing-DML, IT-testing emails-, KCPD-Parking Control-, Mayors Office–, Municipal Court-Accommodations, Municipal Court-Customer Service-, NCS-Animal Health and Safety-Jackson County, NCS-Neighborhood Preservation-LT List, NHS-Administration-, NHS-Animal Health and Safety-, NHS-Animal Health and Safety-Admin, NHS-Dangerous Buildings-, NHS-Dangerous Buildings-Emergency, NHS-HazardousStructure, NHS-NCAP-, NHS-Neighborhood Preservation-, NHS-Neighborhood Preservation-Abatement, NHS-Neighborhood Preservation-Administration, NHS-Neighborhood Preservation-DART, NHS-Neighborhood Preservation-Illegal Dumping, NHS-Neighborhood Preservation-Open Entry, NHS-Neighborhood Preservation-Priority, NHS-Neighborhood Preservation-Programs, NHS-Neighborhood Referrals-Supervisors, NHS-Neighborhood Services, NHS-Neighborhood-Abatement, NHS-NPD-Assessments, NHS-Property Preservation-, NHS-Regulated Industries-, NHS-Salvage-, NHS-Solid Waste-, NHS-Solid Waste-Abatement Referral, NHS-Solid Waste-Administration, NHS-Solid Waste-Illegal Dumping, NHS-Tow Services-, NHS-Tow-Administration, Northeast-Alliance-Together, Northland-Neighborhood-Inc, Parks & Rec-Brush Creek-Maintenance, Parks & Rec-Landscape Services-BRWF, Parks & Rec-Landscape Services-Forestry, Parks & Rec-Landscape Services-LT List, Parks & Rec-Landscape Services-ROWWeeds, Parks & Rec-Landscape Services-Visibility, Parks & Rec-Planning and Design, Parks & Recreation-Landscape Services-BRWF, Parks and Rec-Administration-, Parks and Rec-Central Region-, Parks and Rec-Community Services-, Parks and Rec-Facility Maintenance-, Parks and Rec-Landscape Services-Forestry, Parks and Rec-Landscape Services-Forestry Tech, Parks and Rec-Landscape Services-Landscaping, Parks and Rec-Landscape Services-Weeds, Parks and Rec-North Region-, Parks and Rec-Planning and Design, Parks and Rec-South Region-, Public Works- Sidewalks- GO Bonds, Public Works-Administration-, Public Works-Capital Projects-, Public Works-Capital Projects-Preservation, Public Works-Capital Projects-Sidewalks, Public Works-Capital Projects-Signs, Public Works-Capital Projects-Traffic Control, Public Works-Capital Projects-Traffic Permits, Public Works-Capital Projects-Traffic Signals, Public Works-Coordination Service-Bicycle, Public Works-Facilities Maintenance-, Public Works-Parking Division-, Public Works-Snow and Ice-Administration, Public Works-Solid Waste-, Public Works-Street and Traffic-Administration, Public Works-Street and Traffic-District 1, Public Works-Street and Traffic-District 2, Public Works-Street and Traffic-District 3, Public Works-Street and Traffic-Project Planning, Public Works-Street and Traffic-Rural Mowing, Public Works-Street and Traffic-Snow, Public Works-Street and Traffic-Street Operations, Public Works-Street and Traffic-Streetlights, Public Works-Street and Traffic-Streetlights Staff, Public Works-Streets and Traffic-Administration, Public Works-Streets and Traffic-Project Planning, South-KansasCity-Alliance, Water Services - Wastewater Investigations, Water Services-Administration-, Water Services-Automatic Meter Reading-VSI, Water Services-Consumer Services-Back Office Suppo, Water Services-Engineering-Stormwater, Water Services-Engineering-Water and Sewer, Water Services-Facility-Engineering, Water Services-Industrial Waste Control-, Water Services-Laboratory Services-, Water Services-Leaf Brush, Water Services-Line Maintenance-Ice Abatement, Water Services-Line Maintenance-Leaf Brush, Water Services-Line Maintenance-Pipeline, Water Services-Line Maintenance-Restoration, Water Services-Line Maintenance-Stormwater, Water Services-Line Maintenance-Wastewater, Water Services-Meter & Field Services-Supervisors, Water Services-Meter and Field Services-, Water Services-Reading-, Water Services-Stormwater Investigations, Water Services-Wastewater Investigations

Acknowledgment

The coauthor, Yugyung Lee, would like to acknowledge the partial support of the National Science Foundation (NSF) Grant No. 1747751, 1935076, and 1951971.

Appendix. Supplement data for categorization

Table 10 shows 84 Categories and 293 Types and Tables 11 and 12 show 566 Details Tables 13–19 show 1229 Request types in the KCMO 311's Problem Taxonomy (shown in Fig. 7). Table 20 shows 25 Departments and 143 Work groups in the KCMO 311's Solution Taxonomy (shown in Fig. 8).

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., et al. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308–318).

Barclay, I., Preece, A., Taylor, I., & Verma, D. (2019). A conceptual architecture for contractual data sharing in a decentralised environment. In *Artificial intelligence and machine learning for multi-domain operations applications, Vol. 11006* (p. 110060G). International Society for Optics and Photonics.

Barns, S. (2018). Smart cities and urban data platforms: Designing interfaces for smart governance. *City, Culture and Society*, *12*, 5–12.

Bhardwaj, N., & Sharma, P. (2021). An advanced uncertainty measure using fuzzy soft sets: Application to decision-making problems. *Big Data Mining and Analytics*, *4*(2), 94–103.

Bibri, S. E., & Krogstie, J. (2017). On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society. *Sustainable Cities and Society*, *29*, 219–246.

Bloch, S. (2020). An on-the-ground challenge to uses of spatial big data in assessing neighborhood character. *Geographical Review*, *110*(1–2), 210–223.

Calka, B., Nowak Da Costa, J., & Bielecka, E. (2017). Fine scale population density data and its application in risk assessment. *Geomatics, Natural Hazards and Risk*, *8*(2), 1440–1455.

Chatfield, A. T., & Reddick, C. G. (2018). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, *35*(2), 336–347.

Clark, B. Y., & Guzman, T. S. (2017). Does technologically enabled citizen participation lead to budget adjustments? An investigation of Boston, MA, and San Francisco, CA. *The American Review of Public Administration*, *47*(8), 945–961.

Clark, B. Y., et al. (2014). Do 311 systems shape citizen satisfaction with local government? Available At SSRN 2491034.

Cui, Z., Henrickson, K., Biancardo, S. A., Pu, Z., & Wang, Y. (2020). Establishing multisource data-integration framework for transportation data analytics. *Journal of Transportation Engineering, Part A: Systems*, *146*(5), Article 04020024.

Dallachiesa, M., Ebaid, A., Eldawy, A., Elmagarmid, A., Ilyas, I. F., Ouzzani, M., et al. (2013). NADEEF: A commodity data cleaning system. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data* (pp. 541–552).

Damiani, M. L., Bertino, E., Silvestri, C., et al. (2010). The PROBE framework for the personalized cloaking of private locations. *Transactions on Data Privacy*, *3*(2), 123–148.

Danneels, L., Viaene, S., & Van den Bergh, J. (2017). Open data platforms: Discussing alternative knowledge epistemologies. *Government Information Quarterly*, *34*(3), 365–378.

Data, K. (2021a). Find the status of a property or nuisance violation. http://city.kcmo.org/kc/ActionCenterRequest/. [Online; accessed 08-April-2021].

Data, K. (2021b). Land bank data. https://data.kcmo.org/Property/Land-Bank-Data/j7gh-5trq. [Online; accessed 08-April-2021].

Davies, T., & Perini, F. (2016). Researching the emerging impacts of open data: revisiting the ODDC conceptual framework. *The Journal of Community Informatics*, *12*(2).

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27.

Deng, D., Fernandez, R. C., Abedjan, Z., Wang, S., Stonebraker, M., Elmagarmid, A. K., et al. (2017). The data civilizer system. In *Cidr*.

Dinur, I., & Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems* (pp. 202–210). New York, NY, USA: Association for Computing Machinery, ISBN: 1581136706, http://dx.doi.org/10.1145/773153.773173.

Duvier, C., Anand, P. B., & Oltean-Dumbrava, C. (2018). Data quality and governance in a UK social housing initiative: Implications for smart sustainable cities. *Sustainable Cities and Society*, *39*, 358–365.

Dwork, C. (2008). Differential privacy: A survey of results. In *International conference on theory and applications of models of computation* (pp. 1–19). Springer.

Fan, W., & Geerts, F. (2012). Foundations of data quality management. *Synthesis Lectures on Data Management*, *4*(5), 1–217.

Fass, D., & Feldman, J. (2003). Categorization under complexity: A unified MDL account of human learning of regular and irregular categories. *Advances in Neural Information Processing Systems*, 35–42.

Fernandez, R. C., Mansour, E., Qahtan, A. A., Elmagarmid, A., Ilyas, I., Madden, S., et al. (2018). Seeping semantics: Linking datasets using word embeddings for data discovery. In *2018 IEEE 34th international conference on data engineering (ICDE)* (pp. 989–1000). IEEE.

Gaikwad, P. V. (2020). DeepSampling: IMage sampling technique for cost-effective deep learning.

Gao, X. (2018). Networked co-production of 311 services: investigating the use of Twitter in five US cities. *International Journal of Public Administration*, *41*(9), 712–724.

Garfinkel, S., Abowd, J. M., & Martindale, C. (2019). Understanding database reconstruction attacks on public data. *Communications of the ACM*, [ISSN: 0001-0782] *62*(3), 46–53. http://dx.doi.org/10.1145/3287287.

Google Inc (2021). Google maps geocoding API - google maps platform. https://developers.google.com/maps/. [Online; accessed 30-June-2020].

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 611–642.

Gregory, I. N., & Ell, P. S. (2005). Breaking the boundaries: geographical approaches to integrating 200 years of the census. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 419–437.

Halevy, A., & Doan, A. (2012). *ZGI (Autor). Principles of data integration*. Morgan Kaufmann, Waltham.

Hartmann, S., Mainka, A., & Stock, W. G. (2017). Citizen relationship management in local governments: The potential of 311 for public service delivery. In *Beyond bureaucracy* (pp. 337–353). Springer.

He, Z., & Zhou, J. (2020). Inference attacks on genomic data based on probabilistic graphical models. *Big Data Mining and Analytics*, *3*(3), 225–233.

KCMO (2021a). 311 Call Center Service Requests: 2007 - March 2021. https://data.kcmo.org/311/311-Call-Center-Service-Requests-2007-March-2021/7at3-sxhp. [Online; accessed 1-April-2021].

KCMO (2021b). Emerging Technology Initiative (Smart City). https://www.kcmo.gov/programs-initiatives/emerging-technology. [Online; accessed 08-April-2021].

KCPD (2021). KCMO's crime statistics. https://www.kcpd.org/crime/crime-statistics/. [Online; accessed 22-Mar-2021].

Khayati, M., Lerner, A., Tymchenko, Z., & Cudré-Mauroux, P. (2020). Mind the gap: an experimental evaluation of imputation of missing values techniques in time series. *Proceedings of the VLDB Endowment*, *13*(5), 768–782.

Khazbak, Y., Fan, J., Zhu, S., & Cao, G. (2018). Preserving location privacy in ride-hailing service. In *2018 IEEE conference on communications and network security (CNS)* (pp. 1–9). IEEE.

Khazbak, Y., Fan, J., Zhu, S., & Cao, G. (2020). Preserving personalized location privacy in ride-hailing service. *Tsinghua Science and Technology*, *25*(6), 743–757.

Konda, P., Das, S., Suganthan GC, P., Doan, A., Ardalan, A., Ballard, J. R., et al. (2016). Magellan: Toward building entity matching management systems. *Proceedings of the VLDB Endowment*, *9*(12), 1197–1208.

Lai, Y. (2020). Hyper-local urban contextual awareness through open data integration. In *2020 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)* (pp. 1–6). IEEE.

Lee, B., Choe, E. K., Isenberg, P., Marriott, K., & Stasko, J. (2020). Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications*, *40*(2), 82–90.

Li, Y., Hyder, A., Southerland, L. T., Hammond, G., Porr, A., & Miller, H. J. (2020). 311 Service requests as indicators of neighborhood distress and opioid use disorder. *Scientific Reports*, *10*(1), 1–11.

Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y.-C., Yang, Q., et al. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *22*(3), 2031–2063.

Miller, R. J., Nargesian, F., Zhu, E., Christodoulakis, C., Pu, K. Q., & Andritsos, P. (2018). Making open data transparent: Data discovery on open data. *IEEE Data Engineering Bulletin*, *41*(2), 59–70.

Minkoff, S. L. (2016). NYC 311: A tract-level analysis of citizen–government contacting in New York City. *Urban Affairs Review*, *52*(2), 211–246.

Monteiro, J., Martins, B., Murrieta-Flores, P., & Pires, J. M. (2019). Spatial disaggregation of historical census data leveraging multiple sources of ancillary information. *ISPRS International Journal of Geo-Information*, *8*(8), 327.

Nargesian, F., Zhu, E., Pu, K. Q., & Miller, R. J. (2018). Table union search on open data. *Proceedings of the VLDB Endowment*, *11*(7), 813–825.

Neves, F. T., de Castro Neto, M., & Aparicio, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. *Cities*, *106*, Article 102860.

Niu, B., Li, Q., Zhu, X., Cao, G., & Li, H. (2014). Achieving k-anonymity in privacy-aware location-based services. In *IEEE INFOCOM 2014-IEEE conference on computer communications* (pp. 754–762). IEEE.

Niu, B., Li, Q., Zhu, X., Cao, G., & Li, H. (2015). Enhancing privacy through caching in location-based services. In *2015 IEEE conference on computer communications (INFOCOM)* (pp. 1017–1025). IEEE.

O'Brien, D. T. (2015). Custodians and custodianship in urban neighborhoods: A methodology using reports of public issues received by a city's 311 hotline. *Environment and Behavior*, *47*(3), 304–327.

O'Brien, D. T. (2016). 311 Hotlines, territoriality, and the collaborative maintenance of the urban commons: Examining the intersection of a coproduction policy and evolved human behavior. *Evolutionary Behavioral Sciences*, *10*(2), 123.

O'Brien, D. T., Offenhuber, D., Baldwin-Philippi, J., Sands, M., & Gordon, E. (2017). Uncharted territoriality in coproduction: The motivations for 311 reporting. *Journal of Public Administration Research and Theory*, *27*(2), 320–335.

Pavía, J. M., & Cantarino, I. (2017). Can dasymetric mapping significantly improve population data reallocation in a dense urban area? *Geographical Analysis*, *49*(2), 155–174.

Phan, N., Wu, X., Hu, H., & Dou, D. (2017). Adaptive laplace mechanism: Differential privacy preservation in deep learning. In *2017 IEEE international conference on data mining (ICDM)* (pp. 385–394). IEEE.

Pittaluga, P. (2020). Pioneering urban practices in transition spaces. *City, Territory and Architecture*, *7*(1), 1–10.

Python Package Index (PyPI) (2021). Download data from U.S. census API. https://pypi.org/project/CensusData/. [Online; accessed 08-April-2021].

Razaque, A., & Rizvi, S. S. (2017). Privacy preserving model: a new scheme for auditing cloud stakeholders. *Journal of Cloud Computing*, *6*(1), 1–17.

Rekatsinas, T., Chu, X., Ilyas, I. F., & Ré, C. (2017). Holoclean: Holistic data repairs with probabilistic inference. *Proceedings of the VLDB Endowment*, *10*(11), 1190–1201.

Schuman, C., Stofer, K. A., Anthony, L., Neff, H., Chang, P., Soni, N., et al. (2020). Ocean data visualization on a touchtable demonstrates group content learning, science practices use, and potential embodied cognition. *Research in Science Education*, 1–13.

Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., & Hubaux, J.-P. (2011). Quantifying location privacy. In *2011 IEEE symposium on security and privacy* (pp. 247–262). IEEE.

Stonebraker, M., Bruckner, D., Ilyas, I. F., Beskales, G., Cherniack, M., Zdonik, S. B., et al. (2013). Data curation at scale: The data tamer system. In *Cidr, Vol. 2013*. Citeseer.

Stonebraker, M., & Ilyas, I. F. (2018). Data integration: The current status and the way forward. *IEEE Data Engineering Bulletin*, *41*(2), 3–9.

T. O'Brien, D. (2016). Lamp lighters and sidewalk smoothers: How individual residents contribute to the maintenance of the urban commons. *American Journal of Community Psychology*, *58*(3–4), 391–409.

Tang, X., Yao, H., Sun, Y., Aggarwal, C., Mitra, P., & Wang, S. (2020). Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values. In *Proceedings of the AAAI conference on artificial intelligence, Vol. 34, no. 04* (pp. 5956–5963).

Thirumuruganathan, S., Tang, N., Ouzzani, M., & Doan, A. (2020). Data curation with deep learning. In *EDBT* (pp. 277–286).

Verma, D. C., Bertino, E., Russo, A., Calo, S., & Singla, A. (2020). Policy-based ensembles for multi domain operations. In *Artificial intelligence and machine learning for multi-domain operations applications II, Vol. 11413* (p. 114130A). International Society for Optics and Photonics.

Verma, D., White, G., & de Mel, G. (2019). Federated AI for the enterprise: A web services based implementation. In *2019 IEEE international conference on web services (ICWS)* (pp. 20–27). IEEE.

Wei, D., Ning, H., Shi, F., Wan, Y., Xu, J., Yang, S., et al. (2021). Dataflow management in the Internet of Things: Sensing, control, and security. *Tsinghua Science and Technology*, *26*(6), 918–930.

White, A., & Trump, K.-S. (2018). The promises and pitfalls of 311 data. *Urban Affairs Review*, *54*(4), 794–823.

Wu, W.-N. (2020). Determinants of citizen-generated data in a smart city: Analysis of open 311 user behavior. *Sustainable Cities and Society*, Article 102167.

Xue, M., Kalnis, P., & Pung, H. K. (2009). Location diversity: Enhanced privacy protection in location based services. In *International symposium on location-and context-awareness* (pp. 70–87). Springer.

Yigitcanlar, T., Kamruzzaman, M., Foth, M., Sabatini-Marques, J., da Costa, E., & Ioppolo, G. (2019). Can cities become smart without being sustainable? A systematic review of the literature. *Sustainable Cities and Society*, *45*, 348–365.

Zuo, F., Wang, J., Gao, J., Ozbay, K., Ban, X. J., Shen, Y., et al. (2020). An interactive data visualization and analytics tool to evaluate mobility and sociability trends during COVID-19. In *Proceedings of UrbComp 2020: The 9th SIGKDD international workshop on urban computing*.