ANALYSIS OF COVID-19 CASES BY ECONOMIC, HEALTH, EDUCATION AND RACE INDICATORS OF COMMUNITIES

Azene Zenebe, Bowie State University
Nega Lakew, Bowie State University
LaTanya Brown-Robertson, Bowie State University

ABSTRACT

COVID-19 has been sweeping our globe. To keep track of all tests, positive cases, hospitalizations, recoveries, and casualties, researchers use data repository to collect, manage, and store data sets for data analytics. This research attempted to find out how Covid-19 is affecting communities by their healthcare coverage, and economic status as measured by household income, poverty status, economic vulnerability level, race, and other economic. The main research question was, have communities with majority people of color, a high poverty rate, and a high number of people without health insurance been highly impacted by COVID-19? The research used the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology to gather, clean, and integrate data from several sources, and then assess, design, and develop interactive data analytic results in dashboard on Tableau public cloud. The analytics results reveal that higher COVID-19 positive cases are associated with communities with majority people of color, a high poverty rate, a high population density, and a high number of people without health insurance.

INTRODUCTION

The Covid-19 pandemic has altered almost all aspects of life as we knew it. The pandemic has also magnified many of the socioeconomic disparities in our society for centuries (Beyer, 2020). While the virus is affecting everyone, no one is being infected, more impacted than the black community, especially in the United States (Vasquez, 2020 & Plater, 2020). While the State of Maryland has the wealthiest per capita of black communities in America, like the national trend the community of colors is disproportionately impacted by COVID -19. In this study, we will take a deep dive into the problem using the data science and analytics method using the State of Maryland as a case.

COVID-19 affects many communities with different race, residency, economy, revenue, etc. Most research works are at Country, State and County level. This research focuses on communities at Zip code level of granularity. Community groups from lowest to highest as Block, Census Tract, Neighborhood, Zip Code, County, State, Country, and Continent.

The purpose of the research is to determine extent of Covid-19 cases in communities, study relationships between extent of Covid-19 cases with race, education, health care and economic indicators. The research questions are:

Do communities with more people of color have high positive cases?

Do communities/people without health insurance have high positive cases?

Do communities with high poverty rate have more cases?

Do communities with low education level have more cases?

Do communities with high population density have more cases?

We expected that communities without health insurance will be more likely to test positive for the virus. Second, for the State of Maryland, we expect that there will be a higher percentage of positive COVID cases among communities with Black majority. Third, high poverty areas/communities in Maryland will more likely have higher COVID case numbers than areas with low poverty. Fourth, we predict that lower-income Maryland households will have a higher rate of catching the virus than those above the median income. Lastly, areas in Maryland with high living density rates will be more likely to acquire the virus than low-density areas.

The paper is organized into six sections. Section 2 provides review of literature. Section 3 presents the methodology followed by results in Section 4. Section 5 presents conclusion and recommendation followed by limitations and future direction of the research.

LITERATURE REVIEW

Covid-19 is a worldwide pandemic that started in Wuhan, China, in 2019. Not only does Covid-19 itself cause death but, it is also putting considerable stress on health systems with large case numbers and the economy in general. Over the past year, vast health laws and vaccination campaigns have been set in place to mitigate the spread of the virus. While mitigation factors are in place, the National Institute of Health has highlighted that minorities and disadvantaged communities are less likely to adopt measures; due to historical issues related to trust in the health care system, economic circumstances, and the lack of quality healthcare accommodations.

Covid-19 has impacted the Black community significantly due to their association with underlying health conditions such as diabetes, heart, and lung disease (Samuels et al., 2020)—making Americans living in counties with above-average black populations three times as likely to die from the coronavirus as those in above-average white counties (Samuels et al., 2020)

While studying the effects of Covid-19, ICIC estimated that about 78 percent of high-poverty neighborhoods in the nation are highly vulnerable, but 15 percent of low-poverty communities are highly susceptible to the economic impact of the crisis (Eberhardt et al., 2020). Racial and economic segregation exists across the nation. ICIC uses its data to explain further how race plays a role in high poverty and economic vulnerability at national level. In the District of Columbia (DC), for example, the people of color accompany most high-poverty communities, which are highly vulnerable. Their analysis shows that 90 percent of high-poverty neighborhoods whose residents are primarily people of color are highly vulnerable (Finch & Hernández Finch, 2020). In comparison, only 56 percent of high-poverty tracts, at least 50 percent white, are highly vulnerable. Among low-poverty neighborhoods, 39 percent of those who are primarily people of color are highly vulnerable, compared to only 8 percent of those at least 50 percent white. Even though white people are in high-poverty neighborhoods, most of the space is for people of color (Finch & Hernández Finch, 2020). Almost all the residents of color are highly vulnerable compared to the whites, which only have a little over half of the neighborhood. Also, low-poverty, people of color reside in low-poverty areas, but most residents are white, and less than a tenth is highly vulnerable (Eberhardt et al., 2020).

METHODOLOGY

The research uses the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology. This method is a well-proven, structured approach to planning and executing a data analytics and mining project. The six phases, presented in Figure 1, are Domain/Business Understanding, Data Understanding, Data Preparation/Feature Extraction, Modeling, Evaluation, and Deployment.

See figure below:

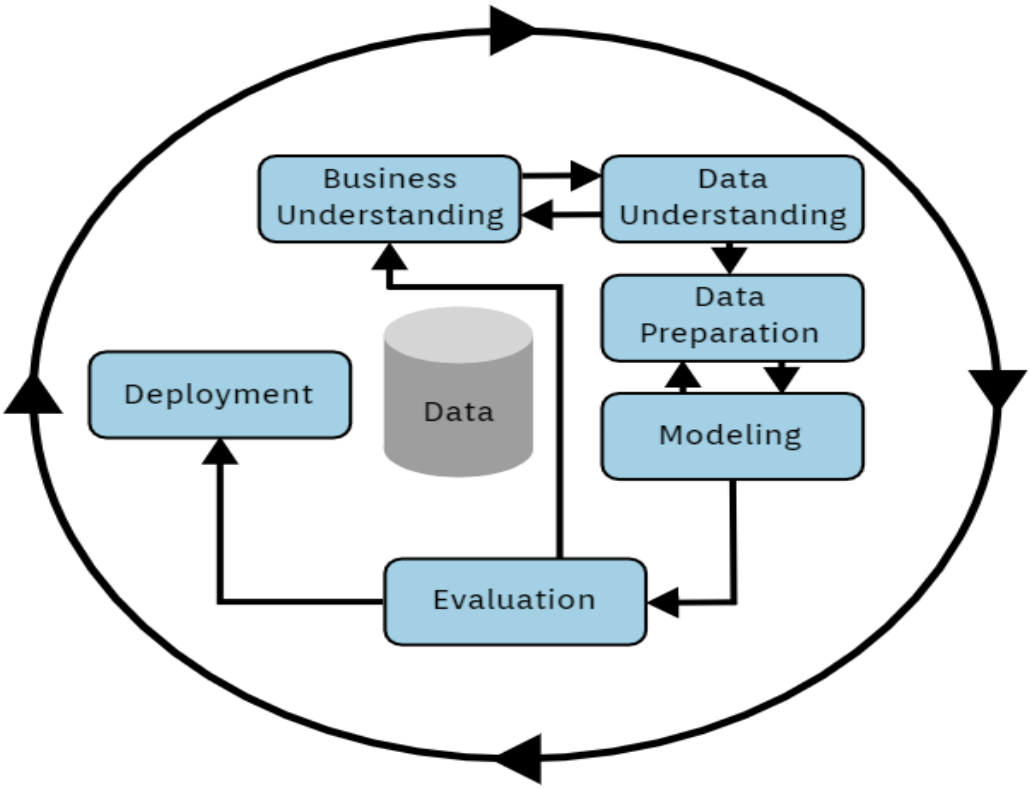

Figure 1. CRISP-DM (Cross Industry Standard Process for Data Mining)
(Source: https://www.oreilly.com/library/view/hands-on-artificial-intelligence/9781788836067/2a71e5a1-3efb-4da7-b5c8-39ab410f9cbf.xhtml)

DATA SOURCES AND FEATURES

Covid-19 data exists in various formats and with different features. Those features included total death by care facilities, total cases by race, total cases by ward, total deaths by age, total deaths by race, total deaths by sex, total deaths by ward, etc. Figure 2 presents the data sources: the U.S. Department of Housing and Urban Development's ZIP Code Crosswalk Files (https://www.huduser.gov/portal/datasets/usps_crosswalk.html), Maryland COID-19 cases by Zip code (https://coronavirus.maryland.gov/datasets/mdcovid19-master-zip-code-cases/explore) with three attributes: Object ID, Zip code, Total Positive Cases as of 7-17-2020, and American Community Survey (ACS) for the Economic, Health and Demographic Data (https://www.census.gov/programs-surveys/acs).

To access ACS data, we use https://www.arcgis.com/home/item.html?id=03b0f660a7a74719930e753428b5cfe2&view=list&sortOrder=true&sortField=defaultFSOrder#data. Figure 3 presents the attributes from these sources.

COVID-19 cases

HUD.GOV

Census Track to Zip Code – Crossover Table

American Community Survey 2018-
Economic Vulnerability Indicators

Figure 2. Data Sources

DATA CLEANING AND INTEGRATION PROCESSES

To integrate the economic and social indicators with the Covid-19 data, Tract-Zip code crosswalk is used. The data model for data integration is presented in Figure 3.

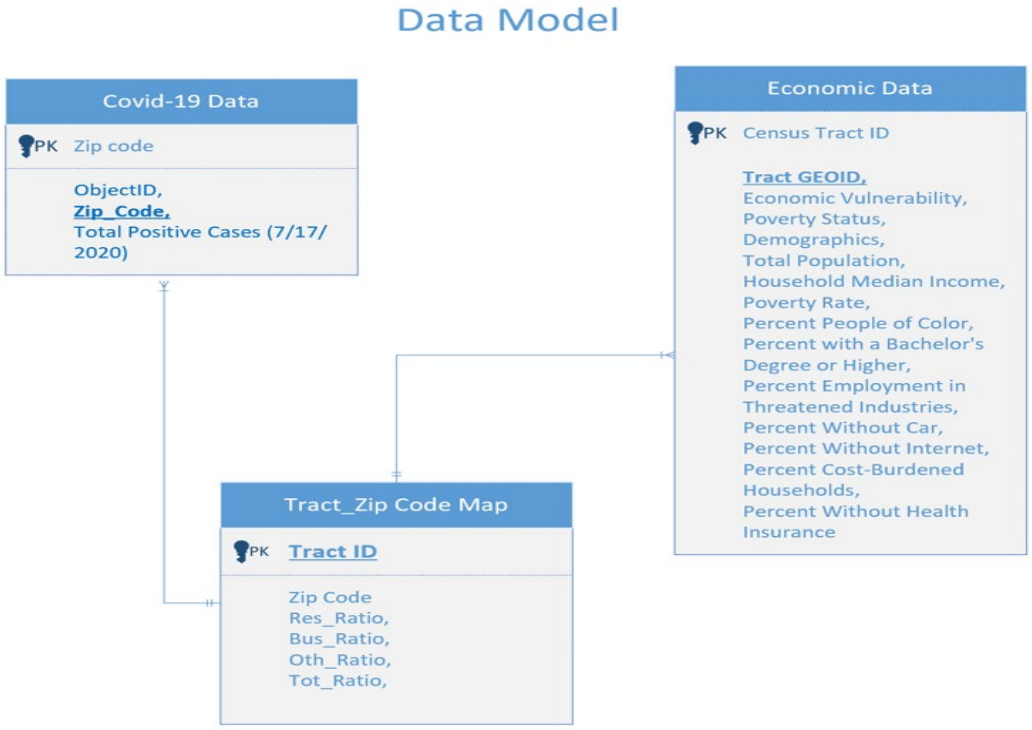

Figure 3. Covid-19 and Economic and Social Indicator Data Integration at Zip Code Level

Figure 4. The cleaning and joining process using Tableau Prep

Figure 4 shows the cleaning and integration of the data sources using Tableau Prep. The process started with inputting different data sets, each data set was individually cleaned by changing data types, removing unnecessary fields, and making the data easier to look at and be prepared to join in the next step. That next step included joining those two data sources using the ZIP Code Crosswalk File. We joined those two different data sources by a common field. Then, we added another clean step to get rid of the duplicated field, like two zip code fields and other fields that were not necessary for the analysis. Table 1 shows a sample of the final output data produced.

Table 1. Integrated COVID-19 and Economic, Education, Health and Demographic Data at Zip Code Level

ZIP Code	Tract GEOID	City	Econom	latitude	longitud	Total Po	Populati	Household	Percent	Percent	Percent	Percent	Percent	Percent Cost-Burdened House	Percent Without Health Insu
21202	24510110100	Baltimo	5	39.3	-76.61	377	22832	45607.73	0.685	0.403	0.238	0.438	0.299	0.411266667	0.061933333
21663	24041960800	Saint Mi	6	38.78	-76.22	18	3308	76088	0.135	0.452	0.262	0.026	0.17	0.272666667	0.035
20646	24017850904	La Plata	1	38.53	-76.97	189	18890	101217.1	0.42	0.265	0.187	0.027	0.178	0.311181818	0.041909091
21154	24025305100	Street	4	39.65	-76.34	15	6464	87452.2	0.037	0.249	0.222	0.035	0.191	0.2794	0.0382
21082	24005411201	Hydes	1	39.48	-76.47		654	116246.2	0.0604	0.4402	0.1794	0.0096	0.1062	0.2124	0.0226
21757	24021767600	Keymar	4	39.59	-77.26	38	3036	79238.25	0.07	0.208	0.2165	0.025	0.261	0.251	0.0685
20748	24033801801	Temple Hi	9	38.8	-76.94	632	38792	73408.3182	0.953864	0.208273	0.2715	0.117636	0.164364	0.379363636	0.083045455
21631	24019970200	East New	4	38.59	-75.92	12	2731	59176	0.2565	0.1675	0.224	0.0445	0.234	0.3175	0.0555
21771	24031700103	Mount Air	2	39.37	-77.15	270	29563	122327.4615	0.137231	0.394846	0.181615	0.022692	0.103769	0.235615385	0.034
21704	24021772200	Frederick	6	39.42	-77.41	126	13321	111590.25	0.340625	0.4515	0.194875	0.048875	0.103375	0.31325	0.047625
21223	24510180100	Baltimore	10	39.3	-76.61	360	26366	31698.2381	0.843143	0.118952	0.307381	0.459333	0.356619	0.450857143	0.079761905
21911	24015031400	Rising Sur	3	39.69	-76.06	143	11004	79786	0.0638	0.1884	0.2152	0.026	0.1858	0.2522	0.0414
20886	24031700812	Montgom	7	39.18	-77.19	672	33282	89256	0.7039	0.4095	0.2284	0.0733	0.1007	0.3568	0.1109
21286	24005490900	Towson	6	39.39	-76.62	180	19206	82659.4	0.338267	0.462267	0.2264	0.085733	0.123933	0.313	0.050133333
21220	24005451802	Middle Ri	5	39.33	-76.43	503	39199	66013.5833	0.28775	0.193167	0.247083	0.056667	0.169417	0.326916667	0.06325
21236	24005440400	Nottingha	2	39.38	-76.48	320	38474	79946.85	0.35315	0.3469	0.24115	0.04975	0.14055	0.2806	0.0407
20619	24037875600	California	1	38.29	-76.49	43	10503	100948.6	0.227	0.3742	0.1978	0.033	0.151	0.2474	0.0546
21230	24510240400	Baltimore	2	39.3	-76.61	420	33568	72744.3125	0.462938	0.443063	0.217438	0.227125	0.214063	0.3363125	0.058875
20879	24031700101	Gaithersb	1	39.14	-77.21	475	24360	98907.3529	0.662471	0.425176	0.218647	0.054765	0.099471	0.343882353	0.097411765
20877	24031700713	Gaithersb	10	39.14	-77.21	1037	34321	89842.0714	0.672143	0.402357	0.226857	0.101929	0.1595	0.387785714	0.1415
21722	24043010500	Clear Spri	7	39.65	-77.93	16	5545	60804.3333	0.136667	0.132667	0.278667	0.031667	0.272667	0.244666667	0.083
21144	24003740303	Severn	3	39.13	-76.69	283	31884	102470.5455	0.490545	0.359545	0.218909	0.041364	0.106455	0.307	0.044454545
21084	24025304101	Jarrettsvil	2	39.6	-76.47	33	7652	106774.2	0.0618	0.3552	0.1972	0.0318	0.138	0.2338	0.0336
21050	24025303300	Forest Hill	1	39.58	-76.39	182	18202	105900.8	0.0714	0.3711	0.2072	0.0258	0.1035	0.2333	0.026
21093	24005408800	Luthervill	3	39.43	-76.64	259	36465	110630.8	0.22655	0.60235	0.17485	0.05235	0.1107	0.2689	0.04185
21804	24045010400	Salisbury	7	38.37	-75.58	566	38491	55613	0.333857	0.2395	0.292	0.076429	0.242143	0.342071429	0.071428571
21133	24005402604	Randallstc	2	39.37	-76.8	551	29998	81592.25	0.849083	0.347083	0.22975	0.076	0.116	0.3245	0.05125

RESULTS AND DISCUSSION

There are 516 rows/observations for MD – zip code level COVID-19 data. There are 1,385 rows/observations for MD at Tract level economic, health and educational data. There are 172,122 rows/observations for USA – Tract level in the Tract-Zip code crossover file. The integrated data has 1,511 rows detailed at tract level, and 484 aggregated by zip code for MD. The results from applications of Visualizations using Tableau are presented online for interactive access on Tableau public at https://public.tableau.com/app/profile/azene.zenebe as well as presented and discussed as follow.

Figure 5 shows total positive Cases by Zip code with Household Median Income. Low median income (light blue color) communities have high positive cases (larger circle). Figure 6 shows Poverty Rate Compared to Positive Cases, where there is an increase in positive cases as the poverty rate increases. Figure 7 depicts the relationship between household median income and total positive cases. The analysis of data shows that in some instances those under the median income level had high cases but, there were other situations where those above the median income had a higher number of cases.

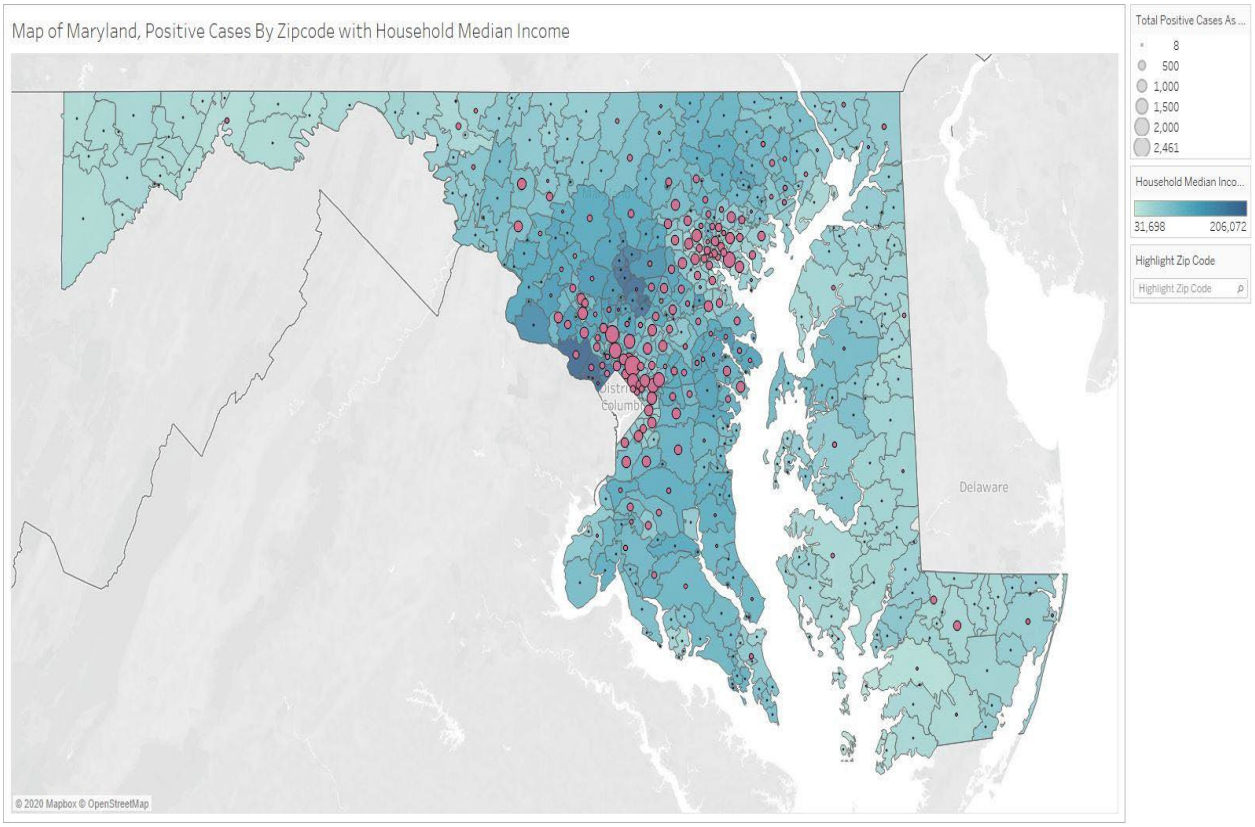

Figure 5. Positive Cases with Household Median Income by Zip code

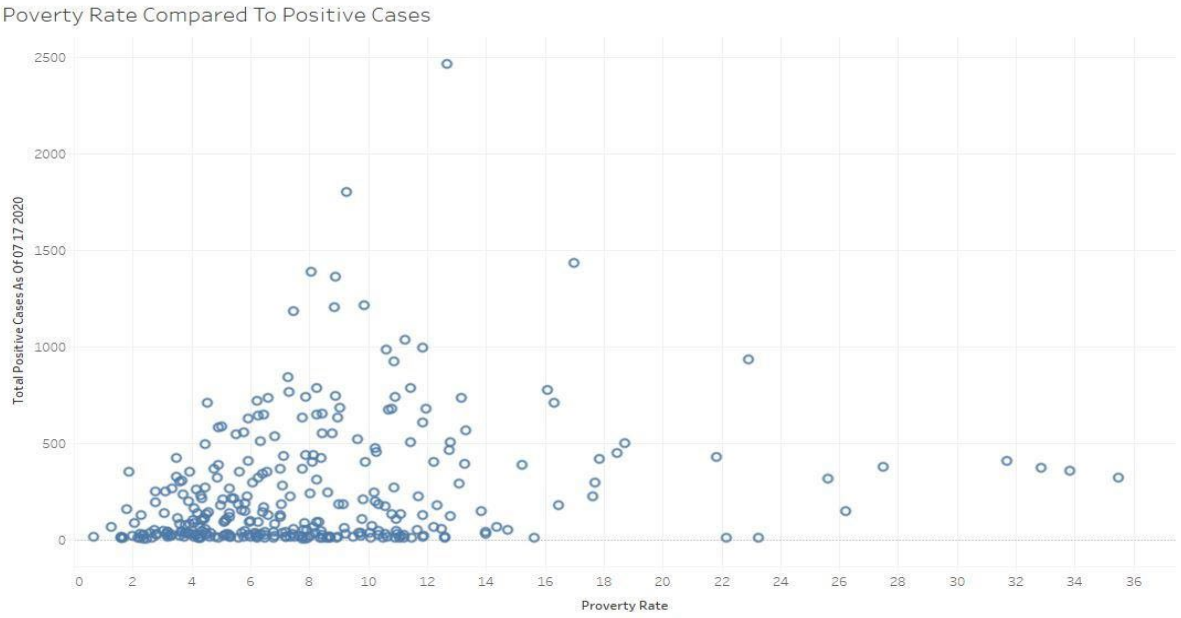

Figure 6. Poverty Rate Compared to Positive Cases

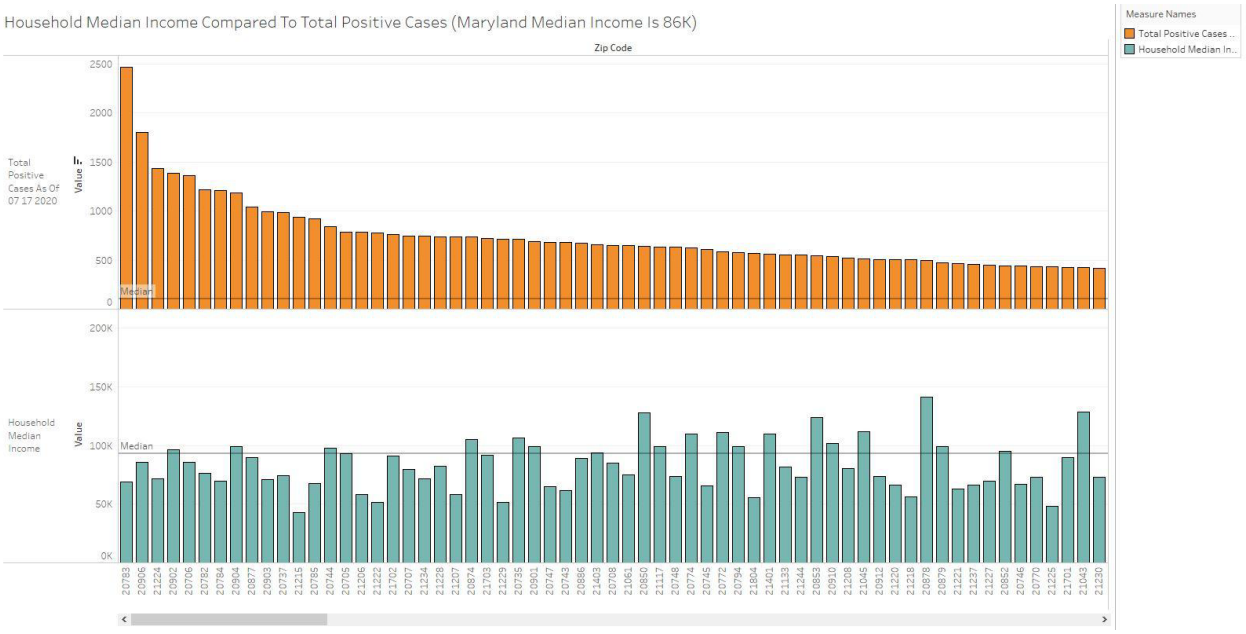

Figure 7. Household Median Income Compared to Total Positive Cases (MD Median Income = 86K)

Figure 8 shows the Percentage without Health Insurance Compared to Total Positive Cases, where there is an increase in positive cases as the uninsured rate increases. Figure 9 presents the Population size By Total Positive Cases for different communities at zip code level. Highly dense communities have higher positive cases; those with a larger population tended to average at 700 or more total positive cases.

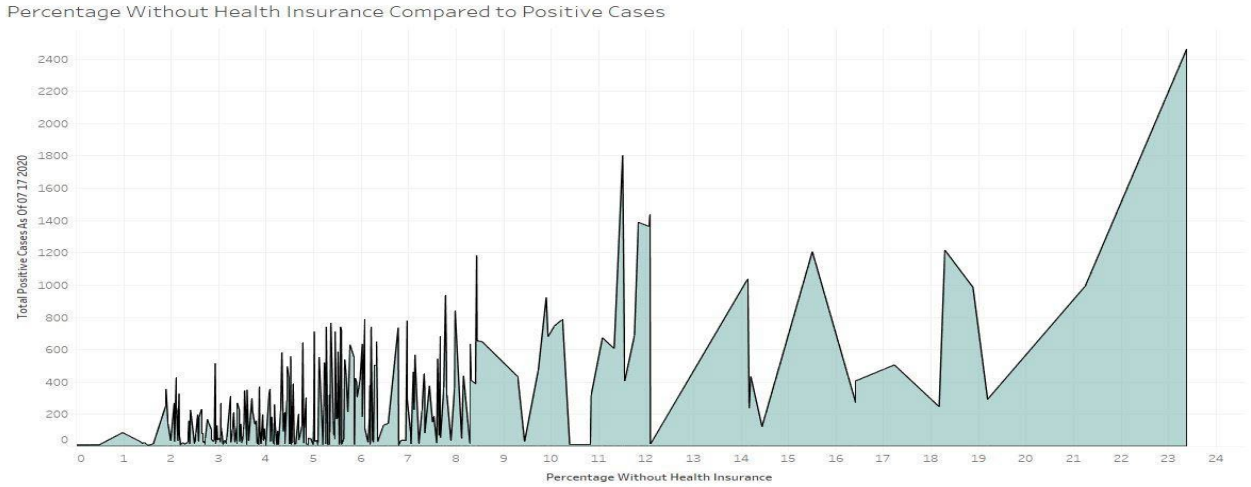

Figure 8. Percentage without Health Insurance Compared to Positive Cases

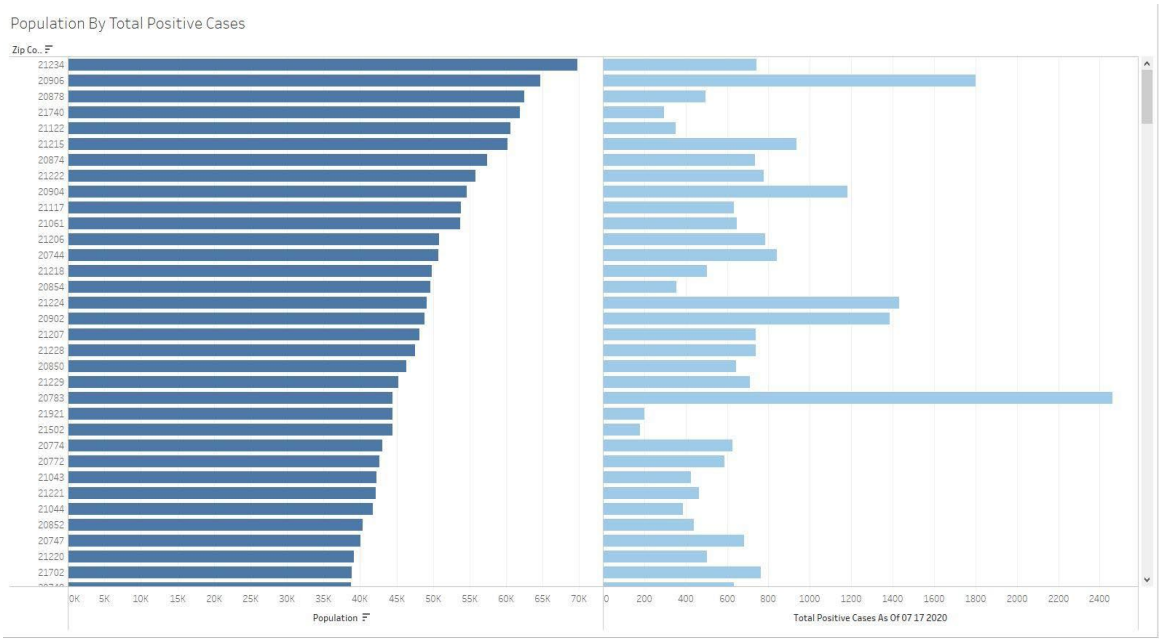

Figure 9. Population Size by Total Positive Cases

Figure 10 shows the relationship between areas with high percentages of people of color compared to total positive cases. The areas with a higher percentage of people of color have higher total positive cases compared to areas with a lower percentage.

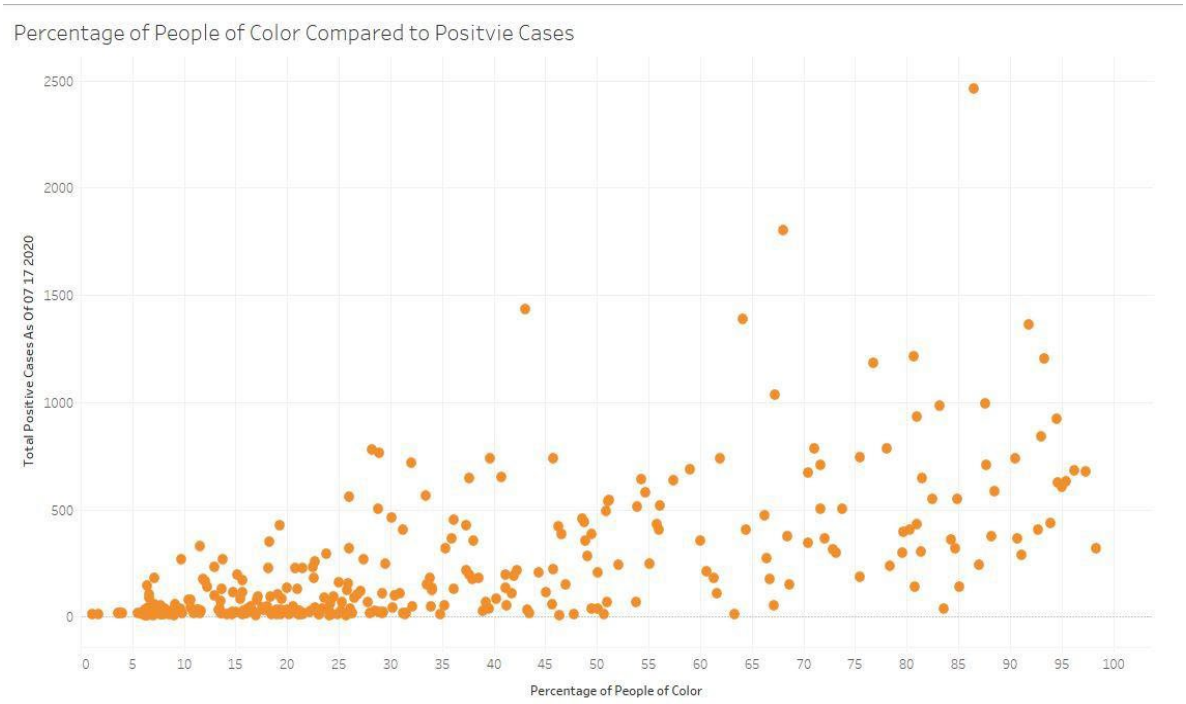

Figure 10. Percentage of People of Color by Positive Cases

CONCLUSION AND RECOMMENDATION

The data analysis results reveal, communities that have had more positive cases of Covid-19 in the State of Maryland are people that do have low access to health insurance, households that have an income that is lower than Maryland's median, large amount of people living in the area (high density of people), and communities with high percentage of people of color. In some areas such as zip codes with a higher percentage rate of 1.4 that do not have health insurance has around 2,500 confirmed cases compared to a lower percentage rate of 0.2 has a little over 500 confirmed cases.

There is a need for actions and policies that address the factors that are associated with high COVID-19 cases including median household income, percentage of people of Color, population density, and percentage of people with low insurance. Communities with high vulnerability should be provided with more personal protective equipment and resource allocation including testing, personal protective equipment (gloves and face masks), soaps, hand sanitizers, etc.

LIMITATION AND FUTURE RESEARCH

The research covers only Maryland's data, uses only the total number of positive cases, and focuses on economic data. Future work will include other COVID-19 impact indicators such as total number of deaths, the entire east coast and other USA communities, more data on social, housing, and demographic data, more data at lowest granular level – Blocks, Tracts, and hypotheses testing using a larger dataset.

ACKNOWLEDGEMENT

The following undergraduate students Kirk Williams, II, Ramar White, Justan McNair-Sneed provided support as undergraduate research mentees during this research under Summer Undergraduate Research Institute (SURI) program. NSF and BSU also provided financial and infrastructure support.

REFERENCES

Beyer, D. (2020). The Economic State of Black America in 2020. *Https://Www.Jec.Senate.Gov/Public/_cache/Files/Ccf4dbe2-810a-44f8-B3e7-14f7e5143ba6/Economic-State-of-Black-America-2020.Pdf.* Published.

Coronavirus in African Americans and Other People of Color. (2020). Https://Www.Hopkinsmedicine.Org/Health/Conditions-and-Diseases/Coronavirus/Covid19-Racial-Disparities.

Davis, P. (2020, June 22). Maryland reports 297 newly confirmed coronavirus cases, 8 more deaths. *The Baltimore Sun*. https://www.baltimoresun.com/coronavirus/bs-md-coronavirus-update-june22-20200622-hegxk5gbbrb3xe4e6evx3urhuq-story.html

Eberhardt, P., Flanagan, S., Scott, C., Wial, H., & Yee1', D. (2020). *Not the Great Equalizer: Which Neighborhoods Are Most Economically Vulnerable to the Coronavirus Crisis?* initiative for a competitive inner city. https://icic.org/wp-content/uploads/2020/04/ICIC_Coronavirus_EconVul_Brief_web.pdf

Finch, W. H., & Hernández Finch, M. E. (2020). Poverty and Covid-19: Rates of Incidence and Deaths in the United States During the First 10 Weeks of the Pandemic. *Frontiers in Sociology*, *5*. https://doi.org/10.3389/fsoc.2020.00047

Data USA Fort Washington, MD. (2019). DATA USA. https://datausa.io/profile/geo/fort-washington-md#demographics.

Harapan, H., Itoh, N., Yufika, A., Winardi, W., Keam, S., Te, H., Megawati, D., Hayati, Z., Wagner, A. L., & Mudatsir, M. (2020). Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health*, *13*(5), 667–673. https://doi.org/10.1016/j.jiph.2020.03.019

How to Protect Yourself & Others. (2020, April 24). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

Landman, K. (2020, June 1). *What We Can Learn From South Korea's Coronavirus Response*. Medium. https://elemental.medium.com/what-we-can-learn-from-south-koreas-coronavirus-response-97a4db5c9fef

Plater, R. (2020, April 14). *Why the African American Community Is Being Hit Hard by COVID-19*. Healthline. https://www.healthline.com/health-news/covid-19-affecting-people-of-color

Poethig, E. C., Greene, S., Stacy, C. P., Srini, T., & Meixell, B. (2018, April 24). *Inclusive Recovery in US Cities*. Urban Institute. https://www.urban.org/research/publication/inclusive-recovery-us-cities

Samuels, R., Williams, A., Jan, T., & del Real, J. A. (2020, June 8). *How cities lost precious time to protect African American residents from coronavirus*. Lown Institute. https://lowninstitute.org/headlines/how-cities-lost-precious-time-to-protect-african-american-residents-from-coronavirus/

Vasquez Reyes, M. (2020). The Disproportional Impact of COVID-19 on African Americans. *Health and Human Right Journal*, *2*, 299–307. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7762908/

Social Determinants of Health. (2008, August 27). *Closing the gap in a generation: health equity through action on the social determinants of health - Final report of the commission on social determinants of health*. Https://Www.Who.Int/Publications/i/Item/WHO-IER-CSDH-08.1.

Dr. Azene Zenebe, Dr. Nega Lakew, and **Dr. LaTanya Brown-Robertson** are Professors at Bowie State University.