Michael Gonzalez
DSC 680 Fall 2020
Professor Williams

Comparing Cities in Los Angeles County and their COVID-19 rates

Abstract

This project will look at the COVID-19 cases in Los Angeles County. With the evidence that a second wave of COVID-19 is coming in the next few months. I wanted to focus on the number of cases within this county, since it has many well know cities that reside in it. I also wanted to work on a project that relates to the COVID-19 pandemic and my beloved Los Angeles County. In this second project, I am hoping to use COVID-19 dataset to compare cities within Los Angeles County and their number of cases.

Background

The thought process of selecting a topic for the second project was based on a comment that was made from my professor. The professor's comment was about the lack of a project related to COVID-19. The process was also related to the fact that our lives are still being affected this current state of mind. We are at a point where the number of deaths is as not as it was in the beginning. I want to focus on the number of COVID-19 cases from March to September.

The selected dataset comes from the Autonomous Network Research Group and University of Southern California's GitHub repository that used Los Angeles Department of Public Health's data of COVID-19 information. The dataset has five attributes that are based on useful information on the cities' location and cases of COVID-19. The selected dataset has data starting from March 16, 2020 to September 26, 2020. This dataset has data points that include time stamp, region, latitude, longitude, and number of cases. These data points will help in giving insights on the factors that can provide me the progress of COVID-19 cases over the past months. Below is a link to the dataset:

https://github.com/ANRGUSC/lacounty_covid19_data/blob/master/data/Covid-19-aggregated.csv

I want to find out which cities have high number of cases and if these cities are low income? Is your Neighborhood in LA County on the List of COVID-19 Hot Beds? Does it matter where you live in Los Angeles county and the likelihood you contract the COVID-19 virus? With the help of the dataset's attributes, I will compare the number of COVID-19 cases from different cities located in Los Angeles County. The benefit of this project could help local city officials to develop a plan to provide resources for their citizens. This will relate to how cities in Los Angeles County will be able to deal with the upcoming second wave of COVID-19. In addition to providing cities with a history of high case numbers to be better prepared in the following months.

The Walk Through

Starting with the dataset from the Autonomous Network Research Group and University of Southern California's GitHub repository, which was in a CSV format. I decided to use the Pandas package to read the dataset. I have also used the Seaborn, Matplotlib, Power BI, and Plotly packages to create data visualizations that will help in answering my research questions. Going back to the dataset, I want to talk about it and give you the reader a better understanding of it. This dataset has five variables and 45,002 rows of data that includes all the cities in Los Angeles County and the number of cases from March 16, 2020 to September 26, 2020.

Data Wrangling Process:

When I first looked at the dataset, I had noticed that the dates of cases after the month of March were out of order. I initially thought that it was due to backlog record keeping that only affected a few months. I was wrong about that. Every month after March until September had a week or a few days out of place in the dataset. I had to manually place the out of order dates in the correct manner, which took some time out of my exploration of the dataset. This manual replacement was done in Microsoft's Excel program. This part of the process was not a problem, just a small inconvenience. Once I confirmed that all the dates were in the correct order. It was time to perform some checklists functions on the correctly ordered dataset.

Performing Dataset Checks:

I used the Pandas package to read the dataset. Then I checked the shape of the dataset, which gave me the needed information about the number of variables and rows. I moved onto to using the Pandas info function to learn about the dataset's data types. This bit of information was typical for the setting of this dataset. I also did a num sum check to see if the dataset had any missing data. I was lucky to find out that this dataset did not have any missing data. I continued to use the Pandas head function to see the first rows of the dataset to check the data.

Data Visualization Prologue:

When I was doing research for this project, I found many reference materials that focused on the national rate of cases or on a global scale. I wanted to know the COVID-19 cases on a local level, since I did not see it. In the time frame for this project, I wanted to focus my attention on the number of COVID-19 cases in Los Angeles County. I am aware that people may only know of a few cities that are in Los Angeles County. A few of the data visualizations will use a chart or charts, when making a side by side comparison of COVID-19 cases. I will include a brief description of the cities used for comparison. These brief descriptions will include the demographics and the income level of the citizens of the selected city. I will also include a Choropleth map of COVID-19 cases in all cities of Los Angeles County. This will also have a brief description on the county, and it will be in the appendix section of this paper.

The Walk Through (Data Visualizations)

In this section, I will display a few visualizations that were used to help me understand the dataset. As I mentioned earlier in this paper, I used the Seaborn, Matplotlib, Power BI, and Plotly packages to create these visualizations. For the first visualization, it is a heatmap that displays the amount of COVID-19 cases in the cities of Los Angeles County. Seen in Figure 1. Looking at the heatmap's legend, the deeper blue color represents the higher number of cases of COVID-19 in those cities. The mid to light blue color range represents the middle number of COVID-19 cases. The light color range represents the little to low cases if COVID-19 in those cities.

When looking at this heatmap, I see that in the month of March. For most this is the beginning of the pandemic, there were low numbers of COVID-19 cases reported. This could be many people did not know the symptoms of COVID-19. From the months of April to June, most cities had growing number of COVID-19 cases. As the dates move to the month of September, this heatmap shows that some cities in Los Angeles County, have an alarming higher number of COVID-19 cases. This is a good base for me to see the progress of COVID-19 cases. From this heatmap, I saw that the cities with the higher COVID-19 cases tend to be poor to lower middle class cities of Los Angeles County.

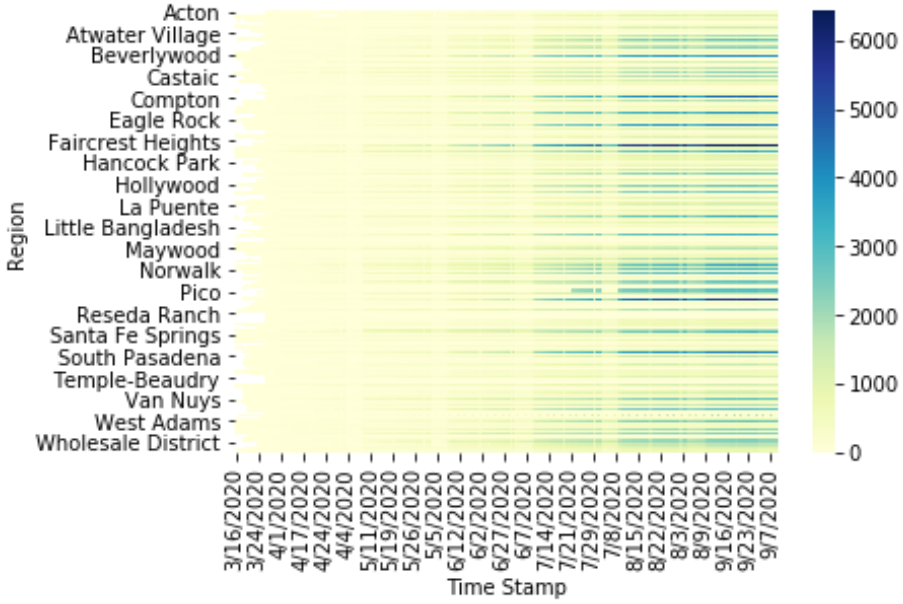


Figure 1: Heatmap of COVID-19 cases in Los Angeles County from March 2020 to September 2020.

The following visualization is a density plot, which looked at the overall cases of COVID-19 from all cities. This density plot displays an interesting plot that could give me more insights. Seen in Figure 2, the higher density plot can lead me to the cities that have exceptionally high number of COVID-19 cases. From this density plot, it displays a high-density metric for the lower number of cases and that metric lowers as the number cases rises. I will look for the cities of Los Angeles County that will have higher case numbers.

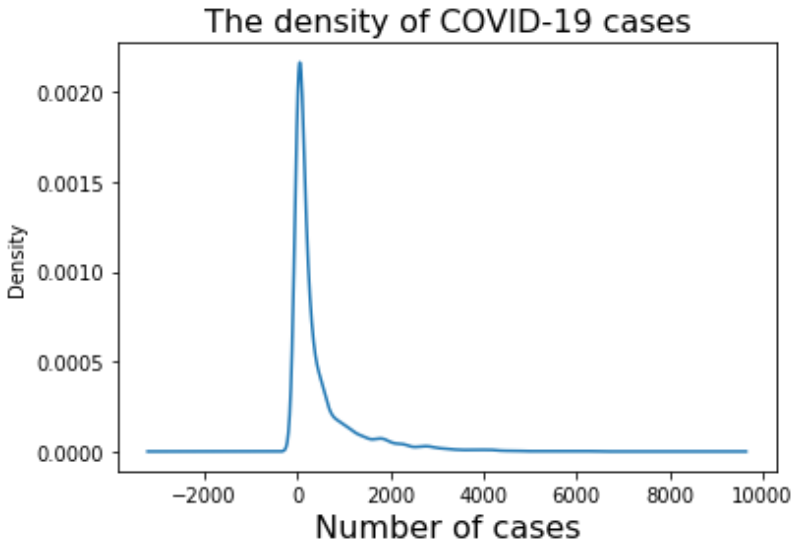


Figure 2: Density plot of overall cases in Los Angeles County.

The next visualization is a clustered bar chart that gave me more insights of the data that displayed a select number of cities that had higher number of COVID-19 cases. Keeping this dataset in mind, this clustered bar chart displays five cities from Los Angeles County that have high number of COVID-19 cases. This clustered bar chart displays the cities of Florence-Firestone, East Los Angeles, Boyle Heights, Downey, and Palmdale. The reason for this comparison was to visually see which cities of Los Angeles County are not doing well with COVID-19. This clustered bar chart displays the areas of Los Angeles County that are not able to deal with this disease, it can be seen in Figure 3.

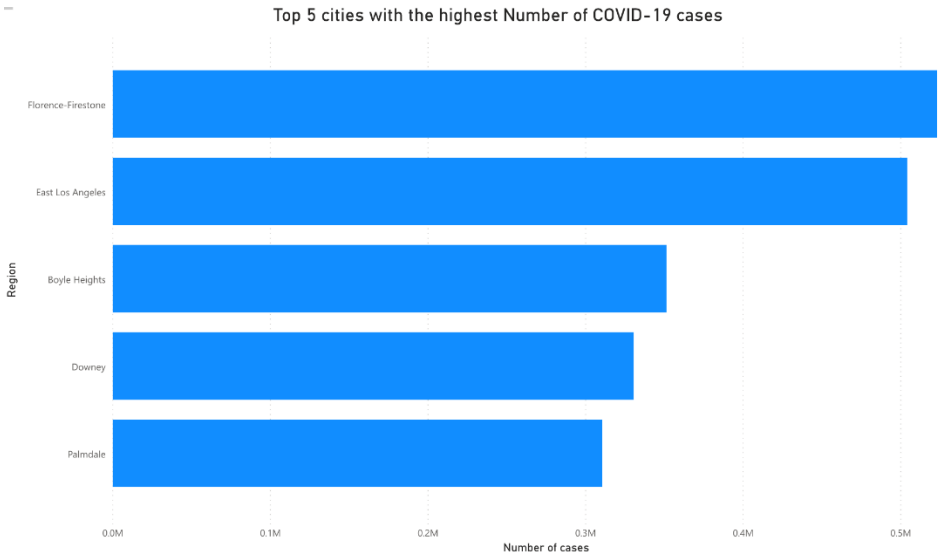


Figure 3: Cluster bar chart that displays the five cities of Los Angeles County that are have high

COVID-19 cases.

It seems that these five cities are poor to lower middle class areas that have a mixture of different income levels. This is giving an idea that could help in answering my research questions.

Conclusion

The focus of this project was to look at the number of COVID-19 cases in the cities that are in Los Angeles County. The reason is based on, we are going into flu season and a second phase of COVID-19 in Los Angeles County. With that in mind, it is more helpful to look at the number of COVID-19 cases, since the most common way of transmission is through asymptomatic people. In Los Angeles County, the number of cases of COVID-19 have lowered a bit in the summer. In the dataset, it displays that a steady growth of COVID-19 cases from the months of June to September. I have discovered the cities with low income levels and limited resources are being affected by this disease the most. I will be comparing well-known cities such as West Los Angeles, Hollywood, Beverly Hills, and Santa Monica. I will also compare other lesser known cities such as Compton, Downey, Florence-Firestone, and South Gate. These mentioned cities are all located in Los Angeles county.

From these comparisons, I can propose a strategy with the public health department for these cities to improve their cases of COVID-19. The best course of action is to wear a mask and Physically distance yourself from people who do not live in the same household now. There must be a better way to combat this disease, since in Los Angeles County there are a few cities that need improve their COVID-19 cases. Finding a way to get people to help lowering the number of COVID-19 will allow them to a somewhat kind of normal life with less restrictions. This is could be another reason to make these insights more appealing to everyone since many people are at their breaking point with COVID-19 and the current safety measures.

References

[1] Ramachandran, G.S. LA County COVID-19 Data Set and Tools for Data Scientists. N. D. https://github.com/ANRGUSC/lacounty_covid19_data (Information on COVID-19 data in Los Angeles County)

[2] Kiamari, M., Nguyen, N. Ramachandran, G. & Krishnamachari, B. LA County COVID-19 Data Set and Tools for Data Scientists. September 27, 2020. https://anrg.usc.edu/www/covid19.html (Information status of COVID-19 data for data scientists)

[3] Smarthardik. COVID-19 Data Visualization using matplotlib in Python. August 06, 2020. https://www.geeksforgeeks.org/covid-19-data-visualization-using-matplotlib-in-python/ (Article on visualizing COVID-19)

[4] Wikipedia. Downey, California. September 12, 2020. https://en.wikipedia.org/wiki/Downey,_California (Article on the history of Downey, California)

[5] Wikipedia. Los Angeles County, California. October 09, 2020. https://en.wikipedia.org/wiki/Los_Angeles_County,_California (Article on the history of Los Angeles County, California)

[6] Wikipedia. Florence-Firestone, California. August 20, 2020. https://en.wikipedia.org/wiki/Florence-Graham,_California (Article on the history of Florence-Firestone, California)

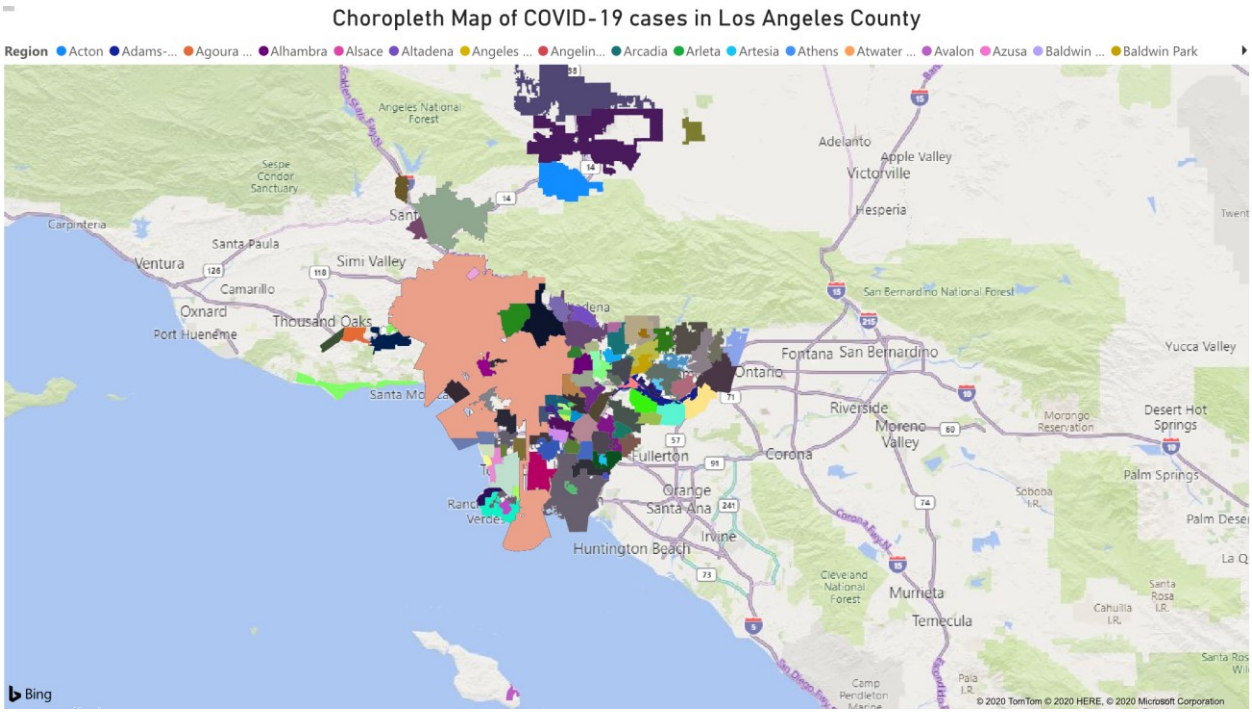
[7] Wikipedia. South Gate, California. September 10, 2020. https://en.wikipedia.org/wiki/South_Gate,_California (Article on the history of South Gate, California)

[8] Wikipedia. Compton, California. September 30, 2020. https://en.wikipedia.org/wiki/Compton,_California (Article on the history of Compton, California)

[9] Fard, B. Mapping the Spread of Coronavirus COVID-19 with python and Plotly. March 08, 2020. https://medium.com/analytics-vidhya/mapping-the-spread-of-coronavirus-covid-19-d7830c4282e (Article on mapping COVID-19 data)

[10] Gupta, A. Using Python to visualize COVID-19 projections. April 21, 2020. https://opensource.com/article/20/4/python-data-covid-19 (Article on visualizing COVID-19 data)

Appendix



Los Angeles County covers a good portion of Southern California. This county is broken down to eight areas that are listed as Antelope valley, Central Los Angeles, Gateway cities, San Fernando valley, San Gabriel valley, Santa Clarita valley, South bay, and Westside cities. Within these eight areas, there are 88 cities that make up this county and boost it to being the second largest county in the United States. Los Angeles County had a population of 9,818,605 based on the 2010 Census. The median personal earnings for all workers 16 and older in Los Angeles County are $30,654, slightly below the US median. This can change from your location within the county. This county has a mixture of different cultures co-existing with each other.

10 Questions

1. Why did you focus on the COVID-19 cases for Los Angeles County?
2. What were the reason for selecting this topic?
3. Why didn't you look at the hospitalization rates?
4. Why was the reason for using the selected dataset?
5. Did you have to make any adjustments to your dataset for the insights?
6. Why didn't you look at the ethnicity of the COVID-19 cases?
7. Did you find any correlations with COVID-19 and income levels?
8. Besides the cities you listed, what other cities had high numbers of COVID-19 cases?
9. What are you going to do with the insights you found?
10. Are there any areas of your project that you would like to improve upon?