Capstone Project – The Battle of Neighborhoods

*by*

[www.coursera.org](http://www.coursera.org)

**“COVID-19: Numbers & Rates in TX State”**

**by**

**Ing. Luis Javier García Murillo**

*Applied Data Science Capstone*

*by*

*Alex Aklson*

May 12th, 2020

Monterrey, NL, Mexico

INDEX

1. Introduction / Business problem ----------------------------------------- 3
2. Data sources / Pre-processing -------------------------------------------- 4
3. Methodology ----------------------------------------------------------------- 6
4. Results & discussion -------------------------------------------------------- 8
5. Conclusion -------------------------------------------------------------------- 9
6. Bibliography ------------------------------------------------------------------ 10

*Keywords:*

*COVID-19, China, United States, Texas, Infected rate, Death rate, Recovery rate, EDA, Foursquare, County, Population, Cinema.*

I. Introduction

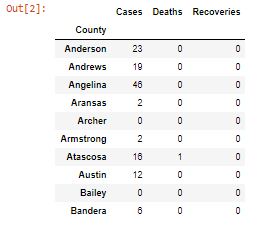
During this year, the world has witnessed a new virus called Coronavirus-19 (COVID-19). The outbreak began in late December last year in a city from China called Wuhan. This virus spreads in three possible ways; direct contact, droplet spray in short range or aerosol in long-range. Thanks to its high propagation capacity, during the first quarter of this year, and the high level of economic activity in China. The virus was able to spread in other countries such as Europe, Italy and the United States, respectively.

In this report we will talk about how this virus has infected the North American population, specifically in the state of Texas, as the state closest to the border with Mexico and therefore the main source of infection for our country. The ten most infected counties in the state of Texas will be revealed, important data will be shown such as; the infected rate, the death rate, and the recovery rate by county.

In addition, an Exploratory Data Analysis (EDA) will be conducted with information from Foursquare to seek to find a relationship or cause for these 10 counties to become the most infected states in Texas. To conclude, we will discuss the results of the analyzes and seek to reach a coherent conclusion for all readers.

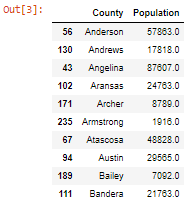
II. Data sources / Pre-processing

For the preparation of this report, we set out to use the skills that this course taught us. We started by looking for the COVID-19 cases as up to date as possible and that it was separated by county in the State of Texas, USA. We found a table inside the Wikipedia web page [1] and we proceeded to extract and clean it using the Pandas library as we were taught in this course. The result of the data frame can be seen in Figure 1.



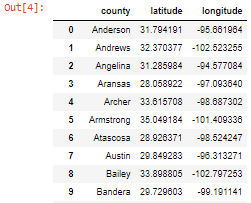
**Figure 1. COVID-19 numbers per county in Texas.** (source: Wikipedia [1])

Subsequently, we had to find the total population by county to obtain the indices mentioned in the introduction to this report. A table was found within the official Texas Demographics page [2] and we proceeded to extract it, clean it and order it to be able to join it with our first table (fig. 1). The extraction result can be seen in Figure 2 below.



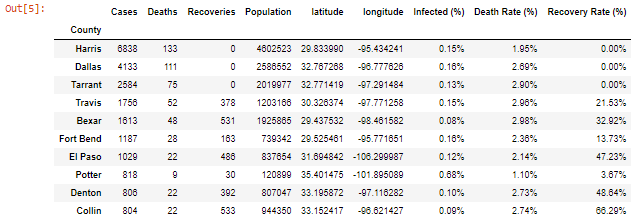
**Figure 2. Population per county in Texas.** (source: Texas Demographics [2])

Then we proceeded to search the latitude and longitude data for each of the counties in Texas. The research was not easy, but we found in Gaslamp Media [3] our information. First, a list was downloaded and then loaded on the Jupyter platform to be able to upload it to our notebook. The result is shown in Figure 3. The latter to be able to relate our data to some Foursquare data that we seek.



**Figure 3. Latitude & Longitude per county in Texas.** (source: Gaslamp Media [3])

In order to begin carrying out our analyzes, we proceeded to clean up our final data frame and for this, all counties where there are no records of cases infected with the virus were removed. Then, they were ordered to obtain the first 10 counties with the highest number of COVID-19 cases registered. The result of our final data frame with some columns calculated from the data in Figs. 1 and 2 are shown below in Figure 4.



**Figure 4. Top-10 Counties in Texas with the greatest number of COVID-19 cases registered.**