Finding Coronavirus Hotspots

Shikhar Gupta

Introduction

It's no secret that we are currently living in a time of crisis, where the COVID-19 virus has become a global pandemic, forcing billions of people into quarantine for the last 3 months. Even with social distancing and isolation measures, the rates of infection continue to rise, and more and more people are getting the disease. Texas is one of the states in the U.S. that has implemented "reopening" measures, in which many nonessential businesses and recreational centers are allowed to open again. However, this process has led to a spike in infection rates as more people have started to go out again.

Thus, this project seeks to answer these essential questions:

1. *What counties in Texas have the highest infection rates?*
2. *What are the most popular types of venues for each county in Texas?*
3. *Is there a correlation between specific types of venues and infection rates?*
4. *Can we create a ranked list of venues that correlate to the highest infection rates based on this data?*

The audience for this project would be the general public, who would want to reduce the risk of coronavirus infection by avoiding particular types of popular venues when going out. Though there are many other factors that also come in to play in deciding infection rates for a county (such as testing rates, income, population, etc), this could give a good indication of places to avoid for concerned citizens who want to determine best practices for staying safe during the pandemic.

Data

Foursquare API *(from developer.foursquare.com)* - The Foursquare API provides detailed and comprehensive location data that can be used to determine popular venues and details about those venues in a given area. This data will be particularly helpful for obtaining information

about popular venues for each county in Texas and comparing it with the coronavirus case data in order to determine how coronavirus rates relate to different types of venues. The data will be extracted through calls to the API within the Jupyter Notebook and received as a JSON file, which we will convert into pandas data frame format for further processing.

Texas County Coronavirus Cases Data *(from dshs.texas.gov/coronavirus/additionaldata/)* - This comprehensive dataset is updated daily, and gives the total population as well as the total number of confirmed cases for each county. This can be used to determine the infection rate for each county, and it can be processed to aid in creating heatmaps for each county as well as correlating different types of venues to infection rates. This dataset is read in as an excel file, and this will be converted to pandas data frame format for further processing.

Texas County Centroid Map

(from https://data.texas.gov/dataset/Texas-Counties-Centroid-Map/ups3-9e8m) - This dataset provides the central latitude and longitude data for each county, helpful for inputting data for search by the Foursquare API. This is given in CSV format and will be converted into pandas data frame format for further processing.

Texas County Boundaries Map

(from http://gis-txdot.opendata.arcgis.com/datasets/texas-county-boundaries) - Finally, this data set is a GeoJson file that will help with visualizing the data through the Folium library, as well as creating heatmaps when combined with coronavirus data.