IBM Data Science Professional Certificate Project – Report

INTRODUCTION

Essential businesses such as grocery stores, markets, pharmacies, hospitals, etcetera, can aid in resource distribution during the fluid environment of the COVID-19 year.

This project will aim to examine clusters of essential business in the state of Michigan. The scope is to determine which cities and counties have better access to these essential business (by proximity to the cluster densities of these essential businesses), and which ones do not (by the lack of close proximity to the clusters of essential businesses).

DATA

Two set of excel worksheets will be loaded. 1 - A set of Zipcode, City, and County data compiled from the following website: https://www.zipcodestogo.com/Michigan/ 2 - A set of Zipcode, Latitude and Longitude data compiled from the following github page: https://gist.github.com/erichurst/7882666

These two sets of data will be the foundation for building the final table which will be linked to the Foursquare API venue data, on which K-means will be built on.

METHODOLOGY

1) Defining the initial DataFrame of Zip Code, City and County.
   This will be the backbone of the data onto which all other elements will be built on to create the cluster mapping and k-means iterations.
2) Importing the Latititude and Longitude of each Zip Code.
   Here a list of latitudes and longitudes for each zipcode from a separate file will be gathered and merged on the table from the first step.
3) Merging the latitudes/longitudes of each zip code to the original table into one final table.
4) Exploring the county regions of Michigan.
   At this point the counties can be examined at the city level with a map visualization of Michigan.
5) Visualizing the Map of Michigan and it's corresponding zipcodes by City, County.
6) Connecting to Foursquare API and pulling categories of essential businesses.
   Essential business will defined as markets, pharmacies and hostpitals, for the purpose of this project to keep things simplied.

ANALYSIS

Using "one hot encoding", Michigan cities will be matched to the Michigan venues (which in this case are the defined essential business).

Analysis includes looking at the following Top 10 most common essential business of these categories above

Lastly, the K-Means clustering and iterations are performed on the data. This is perhaps one of the most interesting parts of the project as it shows where the essential business are clustered respective to the cities.

CONCLUSION

During this exercise it was determined that perhaps the essential business need to be more specific in order to identify clustering better. The data was far too spread in order to create significant clustering of the essential business towards certain cities versus others. This means that essential business in Michigan are well spread throughout the state, except in some obscure more isolated areas.