Capstone Project -
Advises for Dinner Takeout and Grocery Shopping During the Pandemic Period

Mei Ding

May 7th, 2020

1. Introduction

During the Covid 19 pandemic period, many restaurants are closed, and only limited restaurants are open for takeout. Where to order dinner while maintaining low risk of exposure to coronavirus? Many markets are closed. Where to shop for grocery safely? In this capstone project, I will help you find safe places for dinner takeout and grocery shopping.

This project targets to provide advices to people live in Bergen County, NJ. The location could be easily changed to any district/location in the US, so the information can help audience anywhere in the US.

This project is divided into three parts.

- Collect data, including coronavirus cases, geospatial data for 70 boroughs, restaurant and supermarket rating and open status.
- Segment and cluster supermarkets based on # of coronavirus cases, venue rating, and distance to my home.
- Segment and cluster restaurants based on # of coronavirus cases, venue rating, and distance to my home.

2. Data

Here are the data used in this project:

- Coronavirus cases

  This data is stored in a csv file and uploaded to the data asset on IBM Watson.
- Geospatial data for 70 boroughs in Bergen County, NJ

  This data is retrieved using the geocode library.
- Venue information

  Venue information is retrieved through FourSquare API https://foursquare.com/developers/api. I use regular calls to retrieve venue name, latitude, and longitude. And I use premium calls to retrieve rating and open status. Since many places are shut down, open status helps to filter out venues currently open in the pandemic period.
- Distance between venue and my home

  This is calculated using pyproj geod library.

3. Methodologies

Here are methodologies/tools used in this project:

- K-Means clustering

  Used to cluster restaurants and supermarkets based on # of coronavirus cases, venue rating, and distance to my home.
- Z-score normalization

  Used to normalize data before applying K-Means clustering
- Folium map

  Used to visually present the clustering and segmentation. The plot empowers a user to quickly choose a venue.

4. Data Analysis and Data Modeling

Here is the data analysis and data modeling process:

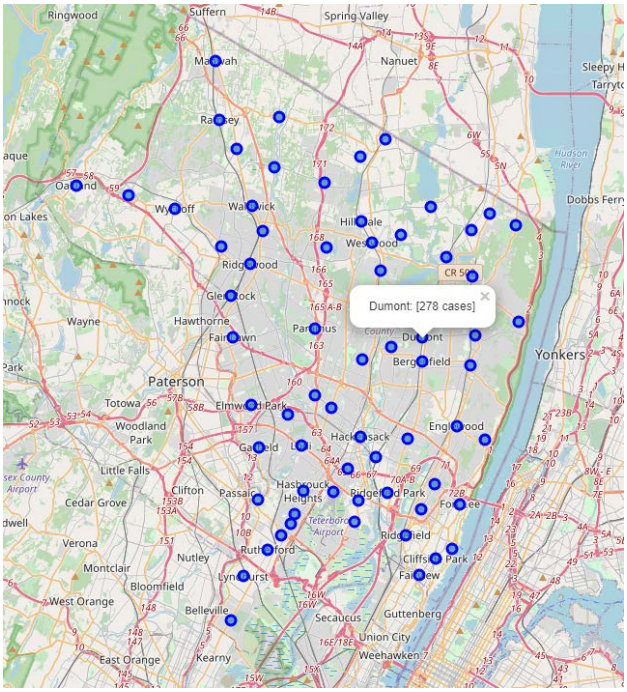
1. Retrieve Coronavirus data

   I uploaded a csv file with coronavirus data to the data asset section in IBM Watson. This csv file contains number of coronavirus cases for 70 boroughs in Bergen County, NJ. And then I read the csv file to Pandas data frame.

2. Retrieve latitude and longitude for every borough in Bergen County, NJ
   I used geocoder library in this step.

	Borough	Coronavirus_Positive_Cases	Latitude	Longitude
0	Allendale	54	41.041486	-74.129032
1	Alpine	21	40.948550	-73.928114
2	Bergenfield	726	40.927599	-73.997361
3	Bogota	162	40.876211	-74.029862
4	Carlstadt	89	40.840378	-74.090697

3. Draw a Folium map for Bergen County, NJ



4. Use FourSquare API to get all venues for every borough

```
https://api.foursquare.com/v2/venues/explore?client_id=DTI3F3LYBQK3MGGWQFEWIZGSQWKUC0ON4ZO
OZXQPFWPOJPXC&client_secret=CDLNKMXNCTLPUKSXHCBWUE3CRLQ1RAYQFPCY0W53YPJI0DUI&ll=40.8871438
,-74.0410865&v=20180605&radius=500&limit=100
```

   There are total 1,395 venues. The sample data looks like following:

	Neighborhood	Coronavirus Positive Cases	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue id	Venue Latitude	Venue Longitude	Venue Category
0	Bergenfield	726	40.927599	-73.997361	Chapala Grill	4baff2e6f964a520b02d3ce3	40.926046	-73.996719	Mexican Restaurant
1	Bergenfield	726	40.927599	-73.997361	Christine's Bake Shop	4b49f4d3f964a520b07626e3	40.928778	-73.997970	Bakery
2	Bergenfield	726	40.927599	-73.997361	Tommy Fox's	4b3617f3f964a520163025e3	40.926707	-73.996453	Pub
3	Bergenfield	726	40.927599	-73.997361	U.M.A. United Martial Arts Taekwondo	535acd65498e70b6d315b6fc	40.926948	-73.996311	Martial Arts Dojo
4	Bergenfield	726	40.927599	-73.997361	Empanada Mania	4ed92ffb6c2547bc612ccc96	40.925710	-73.996562	Empanada Restaurant

5. Filter supermarkets out of all the venues. There are 33 supermarkets.

6. User FourSquare API to retrieve information for every supermarket, including rating and open status. Since many markets are closed during pandemic shutdown, open status is a very important information for shoppers.

```
https://api.foursquare.com/v2/venues/4b5762e8f964a520243528e3?client_id=DTI3F3LYBQK3MGGWQF
EWIZGSQWKUC0ON4ZOOZXQPFWPOJPXC&client_secret=CDLNKMXNCTLPUKSXHCBWUE3CRLQ1RAYQFPCY0W53YPJI0
DUI&v=20180605&oauth_token=X10MOF5IU5WTDA3QC0ORNA4GULI2N3EF4E1UYRU0DS0SNOKZ
```

After this step, there are 19 open restaurants.

7. Calculate the distance between restaurant and my home.
I used pyproj geod library to calculate the distance based on latitude and longitude.

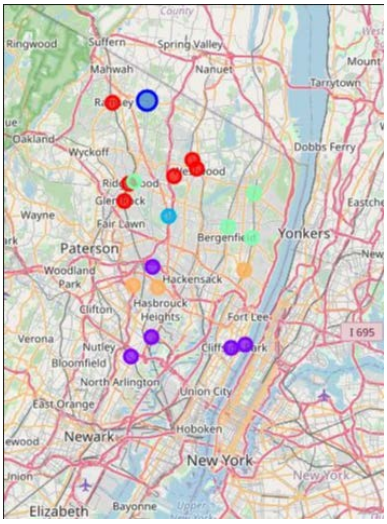
8. Sort all the restaurants based on # of coronavirus cases, venue rating, and distance to my home.

	Venue	Venue Category	Borough	Coronavirus Positive Cases	Rating	Distance	Venue Latitude	Venue Longitude	Open Status
1	Stop & Shop	Supermarket	Closter	72	5.9	15097	40.969031	-73.962986	Closed until 7:30 AM tomorrow
17	Super Foodtown	Supermarket	Washington Twp.	89	7.3	15097	40.986631	-74.064956	Closed until 7:00 AM tomorrow
7	Kilroy's Wonder Market	Grocery Store	Glen Rock	99	7.6	15093	40.962113	-74.127409	Closed until 8:00 AM tomorrow
8	Kings Food Markets	Grocery Store	Hillsdale	101	7.6	15099	41.001234	-74.041039	Closed until 7:00 AM tomorrow
12	Ramsey Farmer's Market	Farmers Market	Ramsey	109	8.8	15103	41.056289	-74.142212	Closed until 9:00 AM Sunday
4	Whole Foods Market	Grocery Store	Edgewater	119	8.5	15082	40.824706	-73.973828	Closed until 9:00 AM tomorrow
3	Aldi's	Supermarket	East Rutherford	123	7.2	15080	40.831393	-74.092718	Closed until 9:00 AM tomorrow
16	Stop & Shop	Supermarket	Tenafly	141	5.7	15093	40.926974	-73.965214	Closed until 7:30 AM tomorrow
18	Trader Joe's	Grocery Store	Westwood	177	8.3	15098	40.992950	-74.035835	Closed until 10:00 AM tomorrow
13	Whole Foods Market	Grocery Store	Ridgewood	262	7.9	15095	40.978418	-74.122593	Closed until 9:00 AM tomorrow
14	Super Stop & Shop	Supermarket	Ridgewood	262	5.8	15095	40.980982	-74.115379	Closed until 7:30 AM tomorrow
15	Acme	Supermarket	Saddle Brook	267	7.1	15087	40.898949	-74.092395	Closed until 7:00 AM tomorrow
2	Super Stop & Shop	Supermarket	Dumont	278	6.8	15093	40.937412	-73.996043	Closed until 7:30 AM tomorrow
10	ShopRite	Supermarket	Lyndhurst	331	6.9	15078	40.812360	-74.118955	Open until Midnight
0	Food Bazaar Supermarket	Supermarket	Cliffside Park	415	8.0	15081	40.821489	-73.991550	Open until 10:30 PM
9	ShopRite of Lodi	Grocery Store	Lodi	574	7.3	15085	40.879279	-74.084812	Open until 11:59 PM
6	Price Rite Of Garfield	Grocery Store	Garfield	669	6.9	15085	40.881520	-74.116071	Closed until 8:00 AM tomorrow
5	ShopRite of Englewood	Grocery Store	Englewood	690	7.2	15090	40.896316	-73.975366	Open
11	Trader Joe's	Grocery Store	Paramus	831	9.2	15093	40.948220	-74.070867	Closed until 9:00 AM tomorrow

9. Normalize data
I used Z-score to normalize # of coronavirus cases, venue rating, and distance to my home.
After data normalization, data is ready for clustering and segmentation.

10. Data clustering and segmentation
I used K-Means to segment data (k=5).

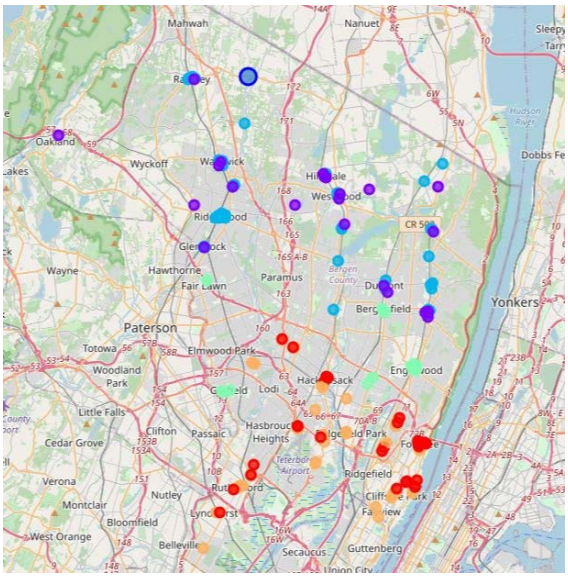
11. Draw a folium map to visually present data clustering.



12. It turns cluster 0 has the best restaurants for grocery shopping. They have low coronavirus cases, good rating, and short driving distance.

	Venue	Venue Category	Borough	Coronavirus Positive Cases	Rating	Distance
7	Kilroy's Wonder Market	Grocery Store	Glen Rock	99	7.6	15093
8	Kings Food Markets	Grocery Store	Hillsdale	101	7.6	15099
12	Ramsey Farmer's Market	Farmers Market	Ramsey	109	8.8	15103
13	Whole Foods Market	Grocery Store	Ridgewood	262	7.9	15095
17	Super Foodtown	Supermarket	Washington Twp.	89	7.3	15097
18	Trader Joe's	Grocery Store	Westwood	177	8.3	15098

13. Repeat step 5-12 to analyze restaurants. There are 480 restaurants in total, and 205 are open. Here is the folium map after restaurant clustering.



5. Conclusion

In this study, I analyzed restaurants and supermarkets data at Bergen County, NJ. I built data clustering and segmentation models to find the venues with good ratings, low coronavirus cases, and short driving distance. These models can be very useful in advising people as to where to go for dinner takeout and grocery shopping while keeping less exposure to coronavirus. And these models could be easily adapted to any location/district in the US.