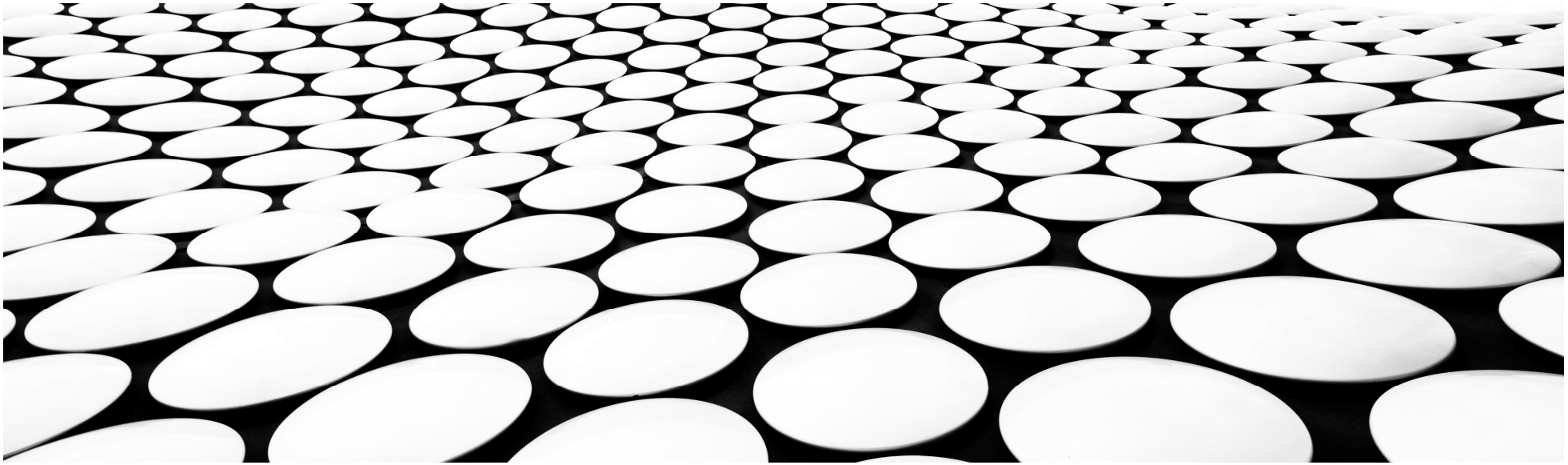
ESTABLISHING ADDITIONAL
GROCERY STORES IN THE DISTRICTS OF MANILA

COURSERA CAPSTONE

APPLIED DATA SCIENCE CAPSTONE (IBM)

ALLEN PATRICK F. CABIGAN



OBJECTIVE OF THE RESEARCH

- This paper aims to identify Districts or Clustered Districts in Manila were interested stakeholders can establish grocery stores with optimal profits;

PYTHON LIBRARIES UTILIZED

- PANDAS
- NUMPY
- JSON
- GEOPY
- BEAUTIFUL SOUP
- SKLEARN
- FOLIUM
- FOURSQUARE API

DATA AND METHODOLOGY

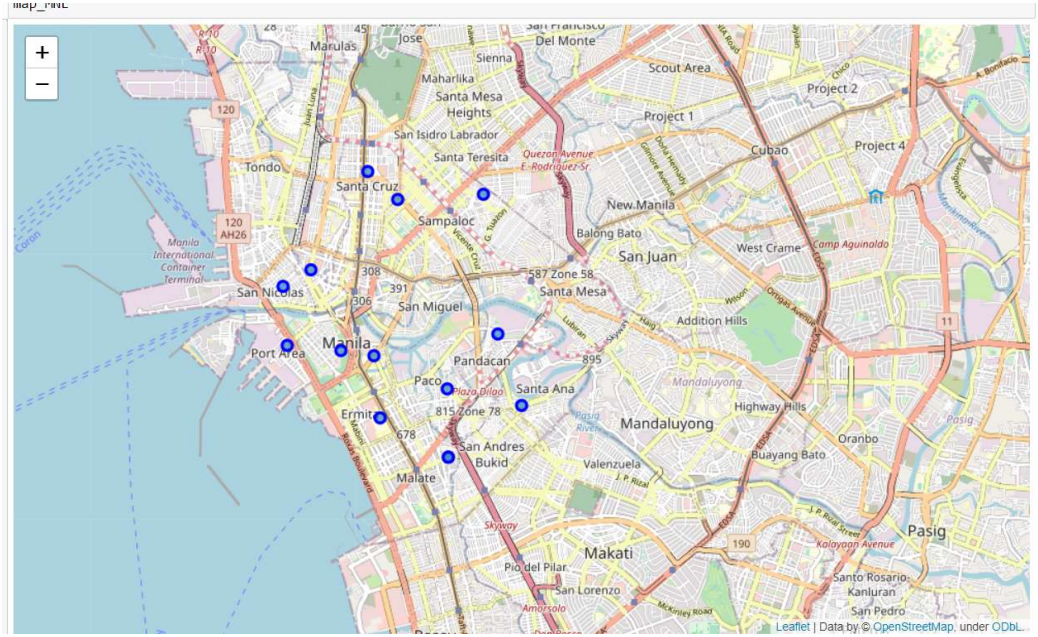
- Data Source: https://en.wikipedia.org/wiki/Manila – used Beautiful Soup to parse the Districts
- Data Acquired:
  - Districts of Manila
  - Coordinates via Geocoder
  - Map of Manila with plotted Districts via Folium
  - Clustered Districts via SKLearn
  - Venues related to grocery stores via Four Square API

THE DISTRICT AND ITS COORINATES

	District	Latitude	Longitude
0	Binondo	14.602370	120.974070
1	Ermita, Manila	14.578630	120.985550
2	Intramuros	14.589480	120.979030
3	Malate, Manila	14.572210	120.996820
4	Paco, Manila	14.583170	120.996640
5	Pandacan, Manila	14.592120	121.005090
6	Port Area, Manila	14.590220	120.970060
7	Quiapo, Manila	14.588640	120.984540
8	Sampaloc, Manila	14.613700	120.988540
9	San Andres, Manila	14.572219	120.996821
10	San Miguel, Manila	14.614556	121.002770
11	San Nicolas, Manila	14.599640	120.969370
12	Santa Ana, Manila	14.580580	121.009130
13	Santa Cruz, Manila	14.618100	120.983570

After parsing the Districts from Wikipedia (via Beautiful Soup) and acquiring the coordinates via Geocoder, the researcher used Pandas to construct this dataframe.

MAP OF MANILA GENERATED BY FOLIUM



The constructed dataframe with Districts and coordinates were used to plot the said Districts in the Map of Manila via Folium.

CLUSTERED DISTRICTS AS A RESULT OF FOURSQUARE API AND SKLEARN

	District	Grocery Store	Cluster Labels	Latitude	Longitude
12	Santa Ana, Manila	0.02	0	14.58058	121.00913

	District	Grocery Store	Cluster Labels	Latitude	Longitude
1	Ermita, Manila	0.0	1	14.578630	120.985550
3	Malate, Manila	0.0	1	14.572210	120.996820
8	Sampaloc, Manila	0.0	1	14.613700	120.988540
9	San Andres, Manila	0.0	1	14.572219	120.996821
10	San Miguel, Manila	0.0	1	14.614556	121.002770
13	Santa Cruz, Manila	0.0	1	14.618100	120.983570

	District	Grocery Store	Cluster Labels	Latitude	Longitude
0	Binondo	0.010000	2	14.60237	120.97407
2	Intramuros	0.010000	2	14.58948	120.97903
4	Paco, Manila	0.010000	2	14.58317	120.99664
5	Pandacan, Manila	0.012821	2	14.59212	121.00509
6	Port Area, Manila	0.013514	2	14.59022	120.97006
7	Quiapo, Manila	0.010000	2	14.58864	120.98454
11	San Nicolas, Manila	0.010000	2	14.59964	120.96937

After using the FourSquare API, the researcher was able to acquire Nearby venues in the Districts. Pandas was then applied to to limit The categories to grocery store and then group them by Districts.

SKLearn was then applied to generate the cluster displayed above.

DISTRICTS PER CLUSTER

A diagram of the said cluster.

CONCLUSION

- After prepping and analyzing the data, this paper is confident to recommend that any businessman or property developer can profit from additional grocery stores in any of the Districts of Manila. But for optimal returns, we suggest that the interested stakeholders prioritize development in **Cluster 1** since, there is no available grocery store in the area. Though there is a limited number present in **Cluster 2,** establishing additional grocery stores will still result in gains.