ESTABLISHING ADDITIONAL GROCERY STORES IN THE DISTRICTS OF MANILA

COURSERA CAPSTONE

APPLIED DATA SCIENCE CAPSTONE (IBM)

What is the story?

Manila being a melting pot of culture, tradition and trends offers a great level of diversity packaged into one for everyones' access. If we want to find the best and the biggest, we can usually find it in Manila. From parks, restaurants to shopping malls, Manila can offer and deliver it.

This all changed when the pandemic COVID 2019 struck Manila. Diversity was no longer the focal point but consumers shifted from luxury goods to the basic necessities. COVID 2019 higlighted the importance of basic commodities. The essential things needed for families to survive during isolation or quarantine.

These basic commodities are usually provided by grocery stores. Long cues, even after few weeks after the implementation of the Community Quarantine, are initial indications that grocery stores are not yet sufficient to handle the demands in the districts or communities.

Business Problem

This paper aims to recommend the best district cluster in Manila where establishing additional grocery stores would be most profitable for businessmen and property developers. The initial step was to paint a picture of the landscape of grocery stores throughout the districts in Manila. Once the locations of the said grocery stores were identified, cluster them together and label the clusters. When finally grouped, identify the clusters were developers and potential businessmen can establish to optimize profits.

Data and Methodology

To achieve the objectives of this paper, the researcher plans to determine all the districts of Manila. The researcher plans to parse the needed data from https://en.wikipedia.org/wiki/Manila using ***Beautiful Soup***. Once the districts are determined, the coordinates of the said districts shall be determined using ***Geocoder.*** And via ***Pandas***, the researcher will merge the Districts in Manila together with its respective coordinates (latitude and longitude) into a dataframe.

The dataframe shall be passed onto *Folium* so that the Districts respective coordinates will be posted into a visual map. Using ***Four Square API***, the researcher plans to get all the possible matches to the category 'grocery stores' and group the data according to the said category. Using ***SKLearn,*** the grouped category shall be clustered according to their similarities further preparing the data to generate conclusive results.

Results

After implementation of ***Beautiful Soup*** Library to the website address https://en.wikipedia.org/wiki/Manila, the researcher was able to extract the fourteen (14) Districts of Manila and with ***Pandas*** was able to convert it into a dataframe.

Out[9]:

	District
0	Barangay
1	Binondo
2	Ermita, Manila
3	Intramuros
4	Malate, Manila
5	Paco, Manila
6	Pandacan, Manila
7	Port Area, Manila
8	Quiapo, Manila
9	Sampaloc, Manila
10	San Andres, Manila
11	San Miguel, Manila
12	San Nicolas, Manila
13	Santa Ana, Manila
14	Santa Cruz, Manila
15	Santa Mesa
16	Tondo, Manila

The next step was to retrieve the coordinates of the Districts via **Geocoder** and merge it with the dataframe of Districts as seen in the output below.

	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

Once the coordinates of the Districts are plugged in, the researcher used **Folium** to make the map of Manila and plot the Districts of Manila to have an idea of the landscape. To retrieve all the locations of grocery stores, the researcher used **Four Square API** and again via **Pandas** the grocery stores were grouped by Districts. To prep the data further, we used **Kmeans** to cluster the grocery stores per Districts together. Once done, three (3) clusters were formed as seen below.

	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

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.