

***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 everyone's 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 highlighted 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 where 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.