

Coursera Capstone

IBM Applied Data Science Capstone

A New Shopping Mall in Cebu City, Philippines

Johnry B. Morales

July 2020

I. Introduction

Philippines is country with a population of approximately 108 million, with much of the working population concentrated on major metropolitan cities and territories (the National Capital Region, metro Cebu and metro Davao). Filipinos are known for its sociable culture and demeanor with much of the people spending most of their free time on parks, open spaces and shopping malls. This is supported by the fact that four of the largest shopping malls in the world are located in the country; the SM Megamall, SM Mall of Asia, SM Seaside Cebu and SM City North. Visiting shopping malls is a great way to relax and enjoy during weekends and holidays. People can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities or even host events. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. In addition, the government aims to decentralize the already crowded National Capital Region (NCR) and urge the mall owners, developers and other investors to refocus development to other places. With this, property developers are also taking advantage of this trend to build more shopping malls to cater to the demand in the provinces. As a result, building of shopping malls has shifted from the NCR to other major and minor cities to provinces. With the current situation; COVID-19 pandemic, it is a challenge to pursue such project considering the weight of the effect on the economy the pandemic has brought both on the domestic and the international market. However, to be able to restart the domestic economy, the public and the private sector must encourage spending especially on the domestic level in order for the local economy to start-up fresh. With this, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. The location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Definition of Terms

Shopping Mall - A shopping mall (or simply mall) is a North American term for a large indoor shopping center, usually anchored by department stores. The term originally meaning a pedestrian promenade with shops along it, but in the late 1960s began to be used as a generic term for the large enclosed shopping centers becoming common at that time.

Local Government Unit (LGU) – there are two levels of governance in the Philippines; national governance and local governance. Local government unit oversee local governance in 81 provinces, 144 cities, 1,490 municipalities, and 42,028

barangays across the country. LGU officials are responsible for providing direct and basic services to the people and ensuring peace and order within communities.

II. Presentation of the Research Problems and Data Preparation

The objective of this capstone project is to examine, select and calculate the best locations in Cebu City, Philippines to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: considering the existence of large shopping malls like the Ayala Malls and the SM Malls, is there any location in Cebu City that would be suitable in opening a new shopping mall that could attract many people.

Target Population

This project is primarily useful to property developers and investors looking to open or invest in new shopping malls in the Philippines. With possible choice of cities in the country, developers and investors has a range of choice on which location, aside from the NCR, to establish their property. This project is timely as the government aims to decentralize both the public and private infrastructure investment to sprawl development in the provinces. Moreover, this study could also benefit the local constituents of the province of Cebu as this could add to employment generation and additional source of income thereby could possibly help in their development. The local government unit both of the Province of Cebu and the city would also benefit from this study considering that this is an additional source of income for the local government. In addition, this will help lessen the number of individuals without employment and this could add to the province and the city's already booming tourism. Data from the Philippine Statistics Authority pegged the NCR's density at approximately 27,000 persons/km², making the NCR as one of the mostly densely populated cluster of cities in the world. Although the concentration of these financial and business establishments in the NCR, relatively, this kind of business is not affected that much considering the behavior of the populace making the prices of real estate not compromised. With the Philippine economy expected to further expand, consumerism in the country will still sprawl therefore the demand for further development such building of establishments.

Data

To solve the problem, we will need the following data:

- List of municipalities and cities in the province of Cebu. This defines the scope and limitations of this project which is confined to the city of Cebu and its nearby or adjacent municipalities and cities.
- Latitude and longitude coordinates of these municipalities and cities. This is

required in order to plot the map and also to get the venue data.

- The population of each cities and municipalities of the province was utilized in order to present the status of each local government units
- The Annual Revenue Income of each local government units were also retrieved from the data of the Bureau of Treasury, specifically from the Bureau of Local Government Finance (BLGF)
- Area of each cities and municipalities were also obtained to understand the possibility and feasibility of the construction of this establishment
- Venue data, particularly data related to shopping malls. This data was used to perform clustering on the municipalities/cities.

Sources of data and methods to extract them

This website page (<https://www.citypopulation.de/php/philippines-visayas-admin.php?adm1id=0722>) contains a list of municipalities and cities in the Province of Cebu, which is comprised of 1 metropolitan city (Cebu), 8 independent component cities and 44 municipalities with a population of 4,632,359. We will use web scraping techniques to extract the data from the web page, with the help of Python requests and BeautifulSoup packages or if not possible, we will build the data frame manually with the help of MS Excel. Then we will retrieve the geographical coordinates of these municipalities/cities using Python Geocoder package which will give us the latitude and longitude coordinates. After that, we will use Foursquare API to get the venue data for these malls. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, in which in this case, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (web pages), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering), excel formatting and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

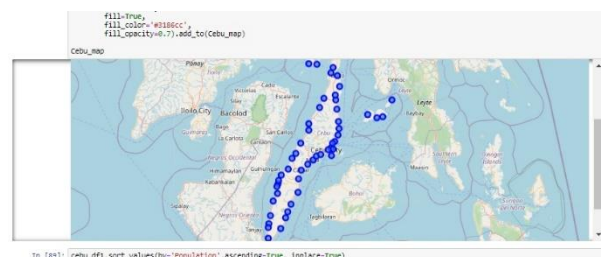
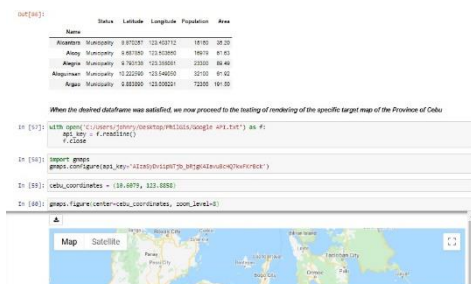
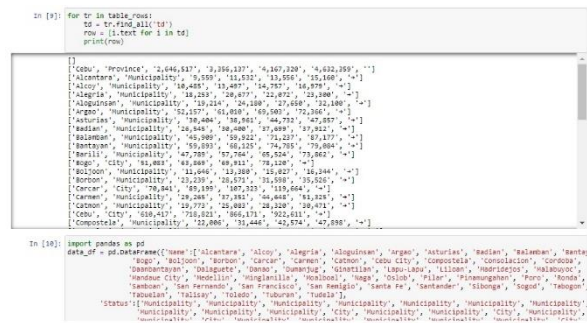
III. Methodology

Creating this kind of study with limited time could have been easily done using the easily accessible data. However, the researcher would want to apply such knowledge in the local setting; which is in the Philippines. However, it is of the failure of the researcher to take into consideration the availability of some of the data or tools needed - such as maps, in the feasibility of such study. Therefore, some of the technique used in this research utilizes the use of other platforms such as excel and google maps.

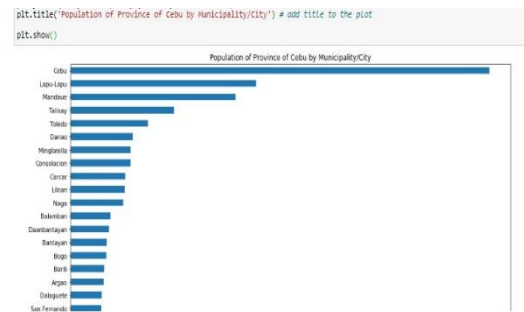
To obtain the data needed, the researcher must have the knowledge about the place where he or she intends to conduct the study. The researcher has lived in Cebu for three (3) years which is enough to have the knowledge and capability to pinpoint the accuracy of the needed data. This research does not need the complicated algorithms to produce the desired

output but is still very useful in the future use. First, the researcher obtained the list of municipalities and cities of the province of Cebu. To be able to scrape the data from the website, the researcher has to study some of the basics in determining the right html structure to be used in Python to be able to extract the data or table successfully. The researcher had a hard time trying to extract the table considering that the structure of the web html with the most accurate data has a different structure than those which can be seen in Wikipedia. Which is why the researcher utilized a different approach using the .find_all() approach and proceeded to extract the data in a text format instead.

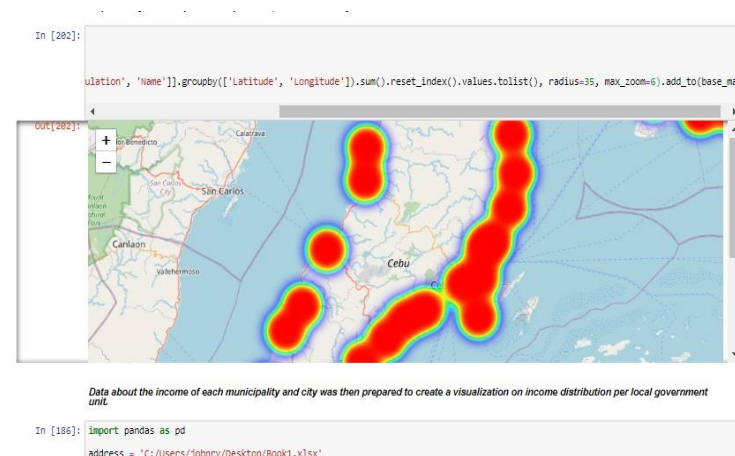
The researcher then built the dataframe manually and saved it in an excel file for it to be easily called whenever needed. For the research to be able to utilize the accuracy of the data, the dataframe has to be cleaned; e.g. dropping those unnecessary columns. For us to be able to present the intended research, mapping is needed and so the researcher utilized the use of gmaps and folium by creating a map with labels to be able to pinpoint the location of



these cities/municipalities. Along the way, the researcher found out an interesting way to obtain easily the latitude and longitude of a location without the need for geocoder. Excel can provide latitude and longitude data by converting the names of an area into a geographic data which gives an option for a column filled with the said data. Moreover, such approach can also provide the population data in excel. Next, the plotting of these data locations on a map took time as the challenge regarding this research has been apparent: the limited geodata of the Philippines. Luckily, by using folium, it was able to render a quality image of the Province of Cebu. By plotting the longitude and latitude of these cities and municipalities, it is the intention



of the researcher to give an idea to the readers the location of each local government units, especially the cities which are a key to the location of the construction of this possible new shopping mall. Population is also a factor in choosing where to construct the said establishment considering the return of investment (RoI) in the mind of the investors and developers. Moreover, this would also affect the planning considering in mind the accessibility to water and



electric supply, public transportation, sewerage system and others.

To give emphasis further on the distribution of the population, a heatmap was generated by the researcher utilizing the folium plugins. This would further highlight the concentration of the

population of Cebu with focus on the cluster of the cities of Cebu, Lapu-Lapu and Mandaue. In the business environment, one of the most important factors for a business to succeed is the manpower, in short, population would matter. Besides, you will not be able to run your business without manpower and at the same time the business will need customers. Just like in the situation of the NCR, Cebu city's population doubles or triples on business days and hours, which means, that a 900,000+ population could reach 2.7 million in a day between 8 am to 5 pm not taking into account the tourists and other visitors in the city; which is very attractive for business opportunity. However, of course, we are not limiting the location to Cebu city, the neighboring cities of Mandaue and Lapu-Lapu city is also attractive considering the short

distance between these cities and is highly accessible in terms of public transportation.

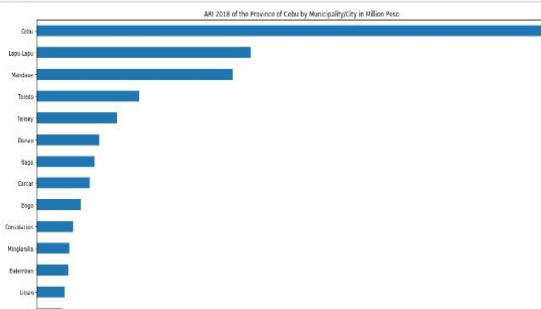
The next data to be utilized is the income classification and annual regular income of these local government units. This is an important data because it will give us an awareness about the capability of a local government unit in terms of infrastructure development and financing. It also means the presence of investors and establishments in an area which is related to how developed a certain

```
In [186]: import pandas as pd
address = 'C:/Users/Johnny/Desktop/Book1.xlsx'
cebu_income = pd.read_excel(address)
cebu_income

Out[186]:
```

	Name	Status	ARI_2018 in Million Peso
0	Alcantara	Municipality	69 02000
1	Alcoy	Municipality	77 73000
2	Alegria	Municipality	99 12000
3	Aligunsa	Municipality	92 34000
4	Argao	Municipality	192 80000
5	Asuncion	Municipality	157 30000
6	Badian	Municipality	142 54000
7	Balamban	Municipality	337 54000
8	Bantayan	Municipality	208 29000
9	Barili	Municipality	191 40000
10	Bogo	City	487 82000
11	Boljoan	Municipality	81 40000
12	Borbon	Municipality	102 80000
13	Carcar	City	889 14000
14	Carmen	Municipality	142 42000
15	Casson	Municipality	102 00000

```
plt.figure(figsize=(10, 10))
plt.title('ARI 2018 of the Province of Cebu by Municipality/City in Million Peso') # add title to the plot
plt.show()
```

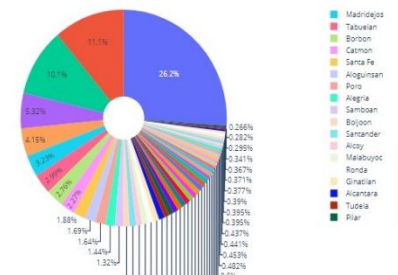


area is. The cities of Cebu, Lapu-Lapu and Mandaue has a combined Annual Regular Income (ARI) of approximately 8 billion pesos. However, for the sake of technicality, however, will not be discussed here, ARI and IRA are different. IRA stands for Internal Revenue Allotment; this refers to the

share of the local government given from the national government. ARI on the other hand, refers to the net income of the local government unit in a given year. This data was obtained

from the national government's archive, specifically from the Bureau of Treasury (BoT) which is under the Department of Finance (DOF). The data was then converted and rounded of the create a per million chart presentation to avoid crowding because of the large figures. The latter was presented in a horizontal bar chart to give a visual

```
# use 'hole' to create a donut-like pie chart
fig = go.Figure(data=[go.Pie(labels=labels, values=values, hole=.2)])
fig.show()
```



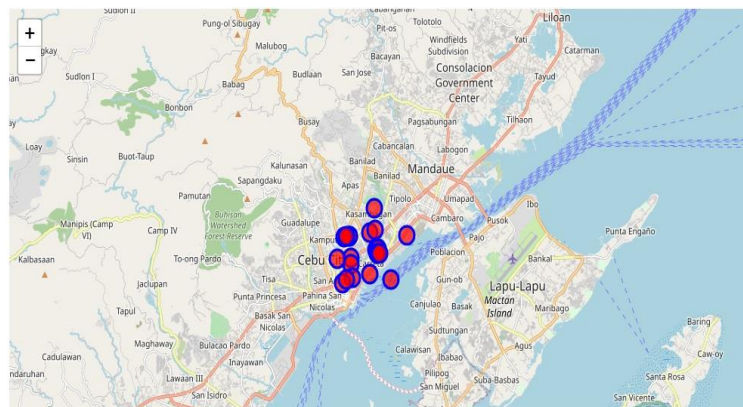
presentation and a pie chart to give emphasis on the percentage of contribution to the total ARI of the province of Cebu. Then of course, with the Foursquare API, we were able to filter through search queries the shopping malls around the cities of the Province of Cebu and were able to render a map with plots to present the locations of these existing malls. This is also important to determine the saturation of an area and to answer our problem question.

IV. Data Analysis and Presentation

With the initial structuring of our dataframe the researcher has already some conclusions about the data. However, such will not be of use if not rendered on a map or a visual presentation. Besides the main purpose of the study of Data Science is for the researchers to be able to gather, test, interpret and more importantly present the data.

Here are some observations gathered from this research:

- 1) Cebu, Lapu-Lapu and Mandaue city enjoys the biggest revenue income among the cities in the Province of Cebu, this is due to Cebu city being the capital of the province being tagged as the queen city of the south. Adjacent to the city is the city of Lapu-Lapu which is home to Cebu's international airport (Mactan-Cebu International Airport) and Mandaue city in which some of the major factories are located,
- 2) These three cities (Cebu, Lapu-Lapu and Mandaue) also has the top three largest population,
- 3) Considering that these three has the highest annual revenue income, it is also conclusive that in terms of human development and purchasing power of the population, these three also has the highest,
- 4) Using the Foursquare API, Cebu city is a bit saturated in terms of the presence of shopping malls,
- 5) Data in some of the venues given by the Foursquare API is not accurate and has to be verified and cleaned further by the administrators and the researcher has to manually clean the obtained data from Foursquare API through manual verification and cleaning using excel,
- 6) Lapu-Lapu and Mandaue only ranks 39th and 48th respectively in terms of land area and only Cebu city sits on the top 10



V. Conclusion and Recommendation

As we can see, if we will look back at the map above which highlights the locations of these cities and municipalities of the Province of Cebu, There are top three locations to choose from in which an investor or developer may put up a new establishment like a shopping mall; these are the cities of Cebu, Mandaue and Lapu-Lapu. These cities also are highly developed as shown in the figures of their Annual Revenue Income and this is further supported by the data taken from Foursquare that most of the large establishments in the province is located on these clusters. Population wise also, these three cities are largely populated, in fact, they occupy the top 3 populated areas in the province. However, the problem is that Lapu-Lapu and Mandaue only ranks 39th and 48th respectively in terms of land area and only Cebu city sits on the top 10. We can make a conclusion, that Cebu city is the most suited to have another mall on their neighborhood. The researcher would like to suggest however, to explore other nearby cities or municipalities to put up a mall, like the city of Talisay or in Consolacion. Moreover, the researcher observed the inaccuracy of some of the data obtained from Foursquare, it is highly suggested by the researcher for purposes of future studies to explore other the same API service and to engage in verification of data obtained from these sources and to update geographical data from the Philippines.