Coursera Capstone
IBM Applied Data Science Capstone
Selecting Area to Open a New Shopping Center in Jakarta, Indonesia
By: Eldo Umboto Gultom
February 2020

Introduction

For many people, visiting shopping center is a great way to relax and enjoy themselves, mainly during weekend and holidays. They can do grocery shopping, dine, fashion shopping, watch movies etc. Shopping centers are like a one-stop destination for all types of people.
For retailers, the usually central location and the crowd at the shopping centers provides a great opportunity to market their products and services. Property developers are also taking advantage of this trend. Opening shopping centers allows possibility for property developers to earn consistent rental income.
As a result, there are many shopping centers in the city of Jakarta, even though they are currently chose more selectively which location to be built for they are requires serious consideration. Particularly, the location of the shopping center is - in many aspects - will determine whether the center will be a success or a failure.

Business Problem

The objective of this project is to analyze and select the best locations in the city of Jakarta, Indonesia to open a new shopping center. Using data science methodology, this project aims to provide solutions to answer the business question: In Jakarta, Indonesia, if you are looking to open a new shopping center, where should you open it?

Target Audience of this project

This project is particularly useful to property developers or investors looking to open or invest in new shopping centers in Jakarta, Indonesia, currently capital city of Indonesia and soon to be ex capital city but still a business center city of Indonesia in many years to come.

Data

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

- Population per neighborhoods.
- Income per neighborhoods. The first two data (population and income) will be used to assume whether a neighborhood is opt out or not for candidacy.
- List of neighborhoods (Kelurahan) in Jakarta.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping centers. We will use this data to perform clustering on the neighborhoods.

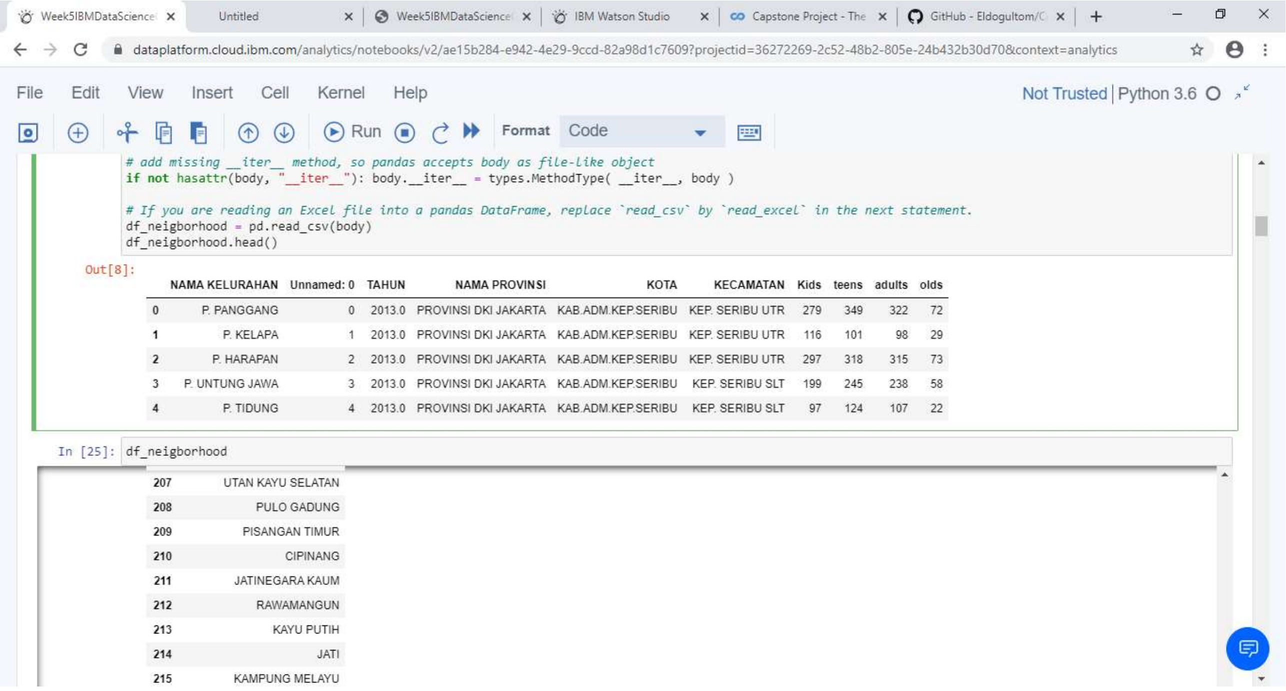
Sources of data and methods to extract them

This City-provided web-page
(http://data.jakarta.go.id/eu/dataset/jumlahpendudukberdasarkanusiaperkelurahandkijakarta/resource/7a6be211-4e8b-487c-a67a-bc796e793eb0)
For visualization, I will use the provided neighborhood list to get geographical coordinates of the neighborhoods (kelurahan) using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods (kelurahan).

After that, I will use Foursquare API to get the venue data for those sufficient neighborhoods. Foursquare API will provide many categories of the venue data, I particularly interested in the shopping center category in order to help us to solve the business problem.

This project will make use of several data science skills, like working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next week, I will present the methodology, the steps taken in this project, the data analysis and the machine learning technique that was used.

2. Get data from CSV into a DataFrame



```
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

# If you are reading an Excel file into a pandas DataFrame, replace `read_csv` by `read_excel` in the next statement.
df_neigborhood = pd.read_csv(body)
df_neigborhood.head()
```

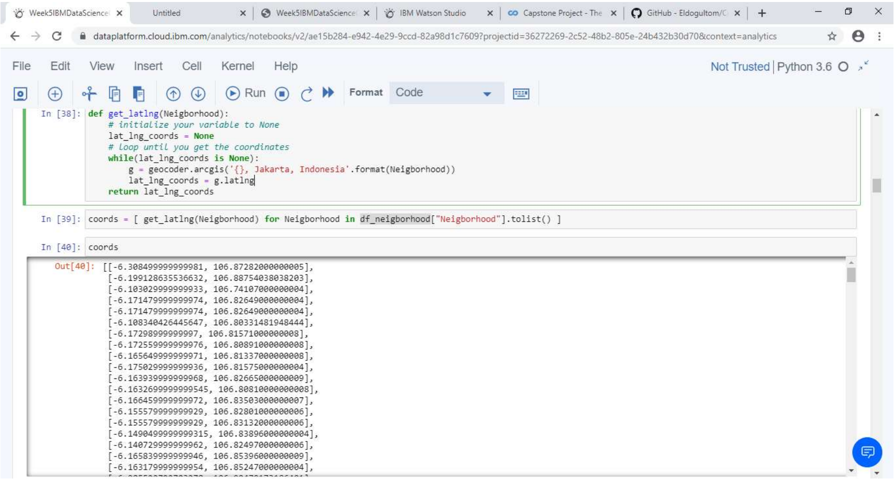
Out[8]:

	NAMA KELURAHAN	Unnamed: 0	TAHUN	NAMA PROVINSI	KOTA	KECAMATAN	Kids	teens	adults	olds
0	P. PANGGANG	0	2013.0	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	279	349	322	72
1	P. KELAPA	1	2013.0	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	116	101	98	29
2	P. HARAPAN	2	2013.0	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	297	318	315	73
3	P. UNTUNG JAWA	3	2013.0	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	199	245	238	58
4	P. TIDUNG	4	2013.0	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	97	124	107	22

```
In [25]: df_neigborhood
```

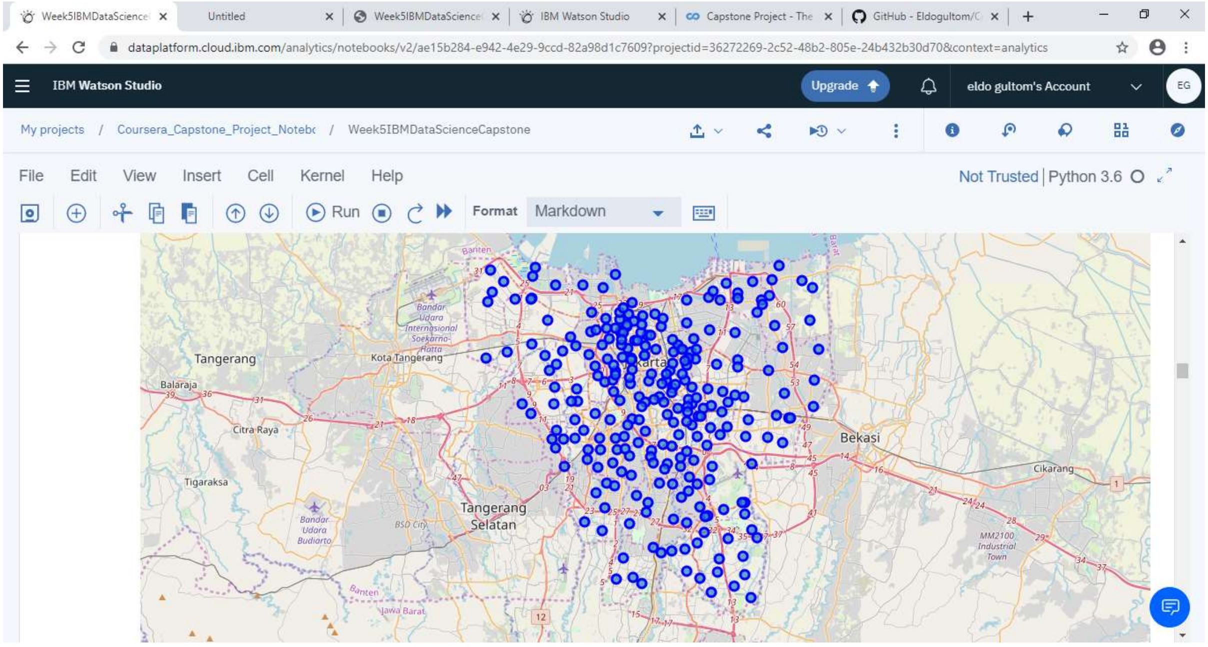
207	UTAN KAYU SELATAN
208	PULO GADUNG
209	PISANGAN TIMUR
210	CIPINANG
211	JATINEGARA KAUM
212	RAWAMANGUN
213	KAYU PUTIH
214	JATI
215	KAMPUNG MELAYU

3. Get the geographical coordinates

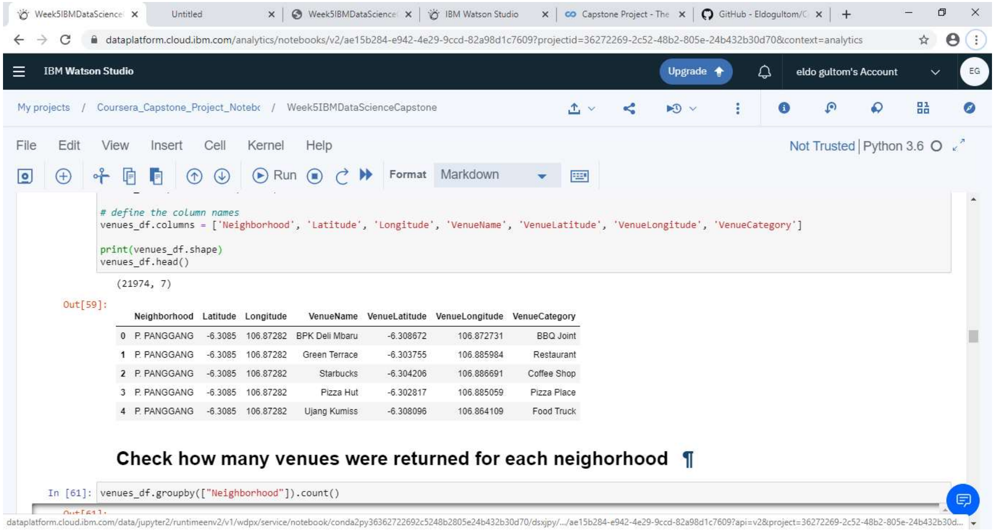


```
In [38]: def get_latlng(Neigborhood):
             # initialize your variable to None
             lat_lng_coords = None
             # loop until you get the coordinates
             while(lat_lng_coords is None):
                 g = geocoder.arcgis('{}, Jakarta, Indonesia'.format(Neigborhood))
                 lat_lng_coords = g.latlng
             return lat_lng_coords

In [39]: coords = [ get_latlng(Neigborhood) for Neigborhood in df_neigborhood["Neigborhood"].tolist() ]

In [40]: coords

Out[40]: [[-6.308499999999981, 106.87282000000005],
          [-6.199128635536632, 106.88754038038203],
          [-6.103029999999933, 106.74107000000004],
          [-6.171479999999974, 106.82649000000004],
          [-6.171479999999974, 106.82649000000004],
          [-6.108340426445647, 106.80331481948444],
          [-6.17298999999997, 106.81571000000008],
          [-6.172559999999976, 106.80891000000008],
          [-6.165649999999971, 106.81337000000008],
          [-6.175029999999936, 106.81575000000004],
          [-6.163939999999968, 106.82665000000009],
          [-6.1632699999999545, 106.80810000000008],
          [-6.166459999999972, 106.83503000000007],
          [-6.155579999999929, 106.82801000000006],
          [-6.155579999999929, 106.83132000000006],
          [-6.1490499999999315, 106.83896000000004],
          [-6.140729999999962, 106.82497000000006],
          [-6.165839999999946, 106.85396000000009],
          [-6.163179999999954, 106.85247000000004],
```

4. Create a map of Jakarta with neighborhoods superimposed on top



5. Use the Foursquare API to explore the neighborhoods


IBM Watson Studio
My projects / Coursera_Capstone_Project_Notebc / Week5IBMDataScienceCapstone
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.6
Run
Format Markdown

```
# define the column names
venues_df.columns = ['Neighborhood', 'Latitude', 'Longitude', 'VenueName', 'VenueLatitude', 'VenueLongitude', 'VenueCategory']

print(venues_df.shape)
venues_df.head()
```

```
(21974, 7)
```

Out[59]:

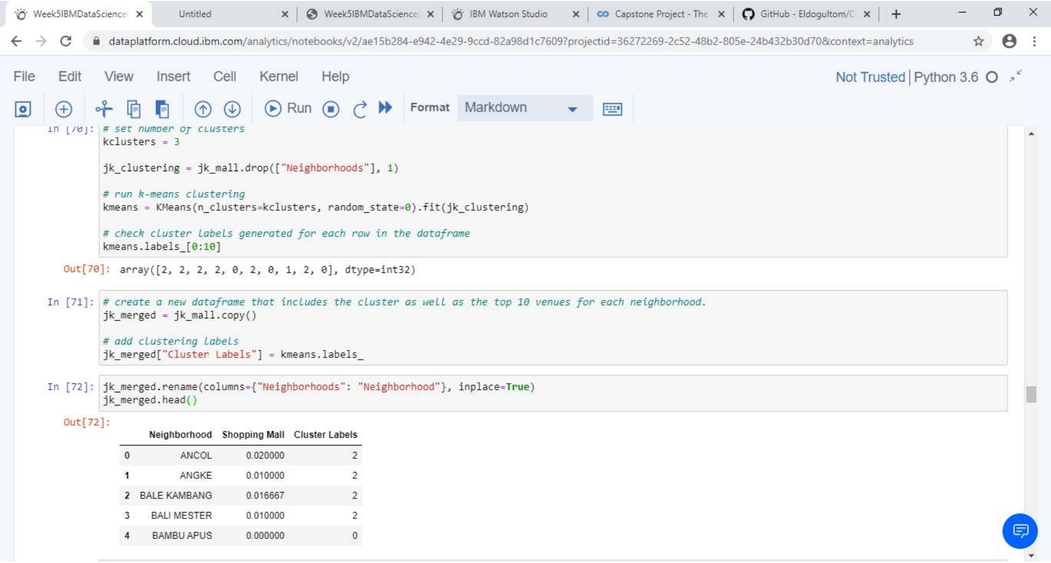
	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	P. PANGGANG	-6.3085	106.87282	BPK Deli Mbaru	-6.308672	106.872731	BBQ Joint
1	P. PANGGANG	-6.3085	106.87282	Green Terrace	-6.303755	106.885984	Restaurant
2	P. PANGGANG	-6.3085	106.87282	Starbucks	-6.304206	106.886691	Coffee Shop
3	P. PANGGANG	-6.3085	106.87282	Pizza Hut	-6.302817	106.885059	Pizza Place
4	P. PANGGANG	-6.3085	106.87282	Ujang Kumiss	-6.308096	106.864109	Food Truck

Check how many venues were returned for each neighorhood ¶

```
In [61]: venues_df.groupby(["Neighborhood"]).count()
```

Cluster Neighborhoods

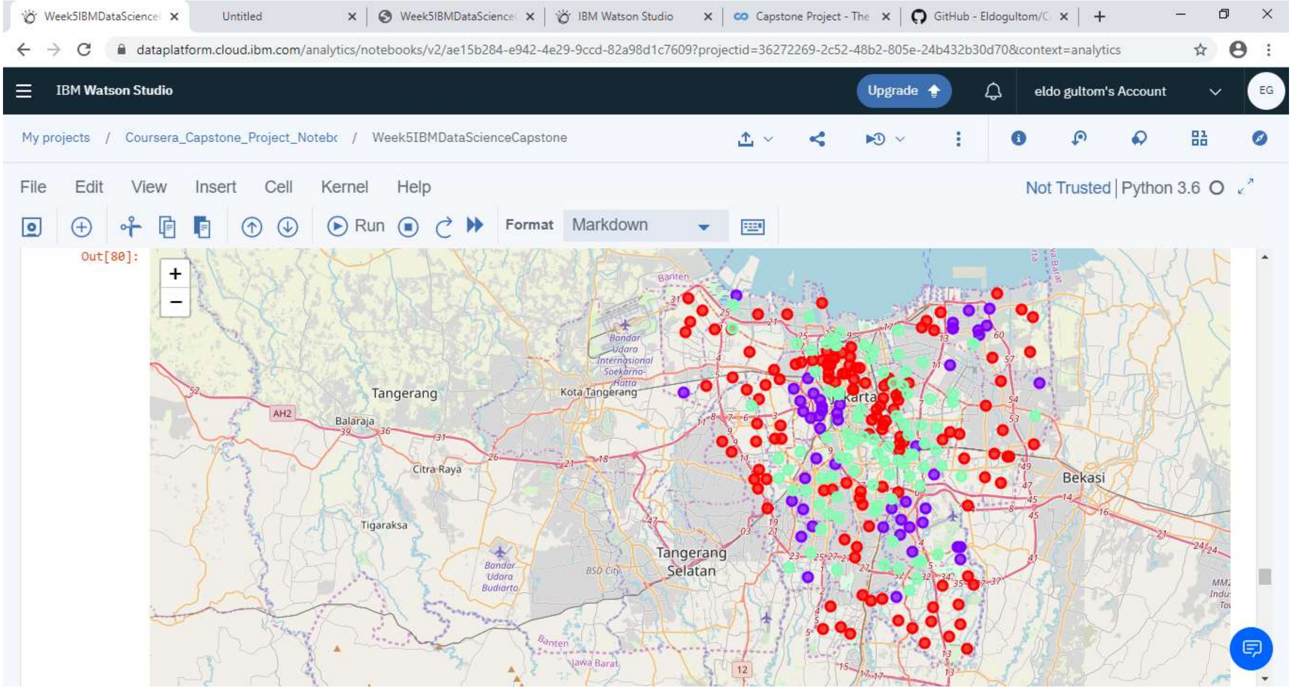
Run k-means to cluster the neighborhoods in Jakarta into 3 clusters



```
In [70]: # set number of clusters
         kclusters = 3

         jk_clustering = jk_mall.drop(["Neighborhoods"], 1)

         # run k-means clustering
         kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(jk_clustering)

         # check cluster labels generated for each row in the dataframe
         kmeans.labels_[0:10]

Out[70]: array([2, 2, 2, 2, 0, 2, 0, 1, 2, 0], dtype=int32)

In [71]: # create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
         jk_merged = jk_mall.copy()

         # add clustering labels
         jk_merged["Cluster Labels"] = kmeans.labels_

In [72]: jk_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
         jk_merged.head()

Out[72]:
```

	Neighborhood	Shopping Mall	Cluster Labels
0	ANCOL	0.020000	2
1	ANGKE	0.010000	2
2	BALE KAMBANG	0.016667	2
3	BALI MESTER	0.010000	2
4	BAMBU APUS	0.000000	0

Visualize the resulting clusters



Discussion and Conclusion

Most of the shopping malls are dispersed but mainly in the central area of Jakarta, with the highest number in cluster 1 and slighty moderate number in cluster 2. On the other hand, cluster 0 has very low number to totally no shopping mall in the neighborhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. From another perspective, this also shows that the dispersion of shopping malls in Jakarta mostly happened because in the central usually act as business center of the city thus have not so much shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 2 with moderate competition. Lastly, property developers are still be able to develop neighborhoods in cluster 1 if applicable, since concentration of shopping malls are far from from intense