CAPSTONE PROJECT WEEK 2

FINALE

BACKGROUND

- The london housing market is in a rut.
- Facing possible higher taxes
- the bank of england that U.K. Home values could fall as much as 30 percent in the event of a disorderly exit from the european union.
- hidden price falls, record-low sales, homebuilder exodus and tax hikes addressing overseas buyers of homes in england and wales.

BUSINESS PROBLEM

- The need for homebuyers clientele in London to make wise and effective decisions.
- how could we provide support to homebuyers clientele in to purchase a suitable real estate in London in this uncertain economic and financial scenario?
- To solve this business problem,
  - we are going to cluster London neighbourhoods in order to recommend venues
  - the current average price of real estate where homebuyers can make a real estate investment.

  Recommendation: We will recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

DATA SECTION

- Data on London properties and the relative price paid data were extracted from the HM Land Registry (http://landregistry.data.gov.uk/).
- To explore and target recommended locations across different venues according to the presence of amenities and essential facilities,
- we will access data through FourSquare API interface and arrange them as a dataframe for visualization.
- By merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable real estate investments.

METHODOLOGY SECTION

FOUR STAGES:

- 1. COLLECT INSPECTION DATA
- 2. EXPLORE AND UNDERSTAND DATA
- 3. DATA PREPARATION AND PREPROCESSING
- 4. MODELLING

METHODOLOGY SECTION I

- 1. Collect Inspection Data
- After importing the necessary libraries, we download the data from the HM Land Registry website as follows:

```
import os # Operating System
import numpy as np
import pandas as pd
import datetime as dt # Datetime
import json # library to handle JSON files

!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

!conda install -c conda-forge folium=0.5.0 --yes
import folium #import folium # map rendering library

print('Libraries imported.')
```

METHODOLOGY SECTION II

- EXPLORE AND UNDERSTAND DATA
- WE READ THE DATASET THAT WE COLLECTED FROM THE HM LAND REGISTRY WEBSITE INTO A PANDAS' DATA FRAME AND DISPLAY THE FIRST FIVE ROWS OF IT AS FOLLOWS:

In [3]:
```
df_ppd.head(5)
```

Out[3]:

	{6DA0844A-2DB9-30F2-E053-6B04A8C05F3B}	597000	2018-05-04 00:00	W2 6BN	F	N	L	58B	Unnamed: 8	GLOUCESTER GARDENS	Unnamed: 10	LONDON	CITY OF WESTMINSTER	GREATER LONDON	A	A.1
0	{6DA0844A-2DBA-30F2-E053-6B04A8C05F3B}	3400000	2018-05-23 00:00	NW6 1HS	D	N	F	37	NaN	CREDITON HILL	NaN	LONDON	CAMDEN	GREATER LONDON	A	A
1	{6DA0844A-2DBB-30F2-E053-6B04A8C05F3B}	431000	2018-04-17 00:00	SW1P 4HN	F	N	L	DUKES HOUSE	FLAT 17	VINCENT STREET	NaN	LONDON	CITY OF WESTMINSTER	GREATER LONDON	A	A
2	{6DA0844A-2DBC-30F2-E053-6B04A8C05F3B}	430000	2018-05-11 00:00	N5 2UA	F	N	L	50	NaN	HIGHBURY QUADRANT	NaN	LONDON	ISLINGTON	GREATER LONDON	A	A
3	{6DA0844A-2DBD-30F2-E053-6B04A8C05F3B}	462000	2018-05-09 00:00	N19 4JR	F	N	L	73C	NaN	LANDSEER ROAD	NaN	LONDON	ISLINGTON	GREATER LONDON	A	A
4	{6DA0844A-2DBE-30F2-E053-6B04A8C05F3B}	585000	2018-05-02 00:00	W12 0AD	T	N	F	204	NaN	WULFSTAN STREET	NaN	LONDON	HAMMERSMITH AND FULHAM	GREATER LONDON	A	A

METHODOLOGY SECTION III

- DATA PREPARATION
  - At this stage, we prepare our dataset for the modeling process, opting for the most suitable machine learning algorithm for our scope. Accordingly, we perform the following steps:

1. Rename the column names
2. Format the date column
3. Sort data by date of sale
4. Select data only for the city of London
5. Make a list of street names in London
6. Calculate the street-wise average price of the property
7. Read the street-wise coordinates into a data frame, eliminating recurring word London from individual names
8. Join the data to find the coordinates of locations which fit into client's budget
9. Plot recommended locations on London map along with current market prices

METHODOLOGY SECTION III

- DATA PREPARATION

```
In [10]: # Display the dataframe
df_affordable
```

Out[10]:

	Street	Avg_Price
20	ABBOTSBURY CLOSE	2.367093e+06
178	ALBION SQUARE	2.450000e+06
355	ANHALT ROAD	2.435000e+06
368	ANSDELL TERRACE	2.250000e+06
381	APPLEGARTH ROAD	2.400000e+06
617	AYLESTONE AVENUE	2.286667e+06
753	BARONSMEAD ROAD	2.375000e+06
867	BEAUCLERC ROAD	2.480000e+06
1079	BICKENHALL STREET	2.351667e+06
1094	BILLING ROAD	2.200000e+06
1108	BIRCHLANDS AVENUE	2.217000e+06
1310	BOWERDEAN STREET	2.300000e+06
1371	BRAMPTON GROVE	2.475833e+06
1439	BRIARDALE GARDENS	2.397132e+06
1605	BROWNING CLOSE	2.320000e+06
1820	CALLCOTT STREET	2.375000e+06

```
In [17]: df_affordable
```

Out[17]:

	Street	Avg_Price	city_coord	Latitude	Longitude
20	ABBOTSBURY CLOSE	2.367093e+06	(51.5322588, -0.0061531)	51.532259	-0.006153
178	ALBION SQUARE	2.450000e+06	(-41.27375755, 173.289393239104)	-41.273758	173.289393
355	ANHALT ROAD	2.435000e+06	(51.4803265, -0.1667607)	51.480326	-0.166761
368	ANSDELL TERRACE	2.250000e+06	(51.4998899, -0.1891027)	51.499890	-0.189103
381	APPLEGARTH ROAD	2.400000e+06	(53.7486539, -0.3266704)	53.748654	-0.326670
617	AYLESTONE AVENUE	2.286667e+06	(51.5409157, -0.2178742)	51.540916	-0.217874
753	BARONSMEAD ROAD	2.375000e+06	(51.4773147, -0.239457)	51.477315	-0.239457
867	BEAUCLERC ROAD	2.480000e+06	(51.4995771, -0.2290331)	51.499577	-0.229033
1079	BICKENHALL STREET	2.351667e+06	(51.5211969, -0.1589341)	51.521197	-0.158934
1094	BILLING ROAD	2.200000e+06	(51.4818833, -0.1878624)	51.481883	-0.187862
1108	BIRCHLANDS AVENUE	2.217000e+06	(51.4483941, -0.1604676)	51.448394	-0.160468
1310	BOWERDEAN STREET	2.300000e+06	(51.4727099, -0.1924853)	51.472710	-0.192485
1371	BRAMPTON GROVE	2.475833e+06	(51.5703648, -0.2833944)	51.570365	-0.283394
1439	BRIARDALE GARDENS	2.397132e+06	(51.5601748, -0.1954305)	51.560175	-0.195431
1605	BROWNING CLOSE	2.320000e+06	(51.8858497, 0.8560813)	51.885850	0.856081
1820	CALLCOTT STREET	2.375000e+06	(51.5083499, -0.1983276)	51.508350	-0.198328
1871	CAMPDEN HILL ROAD	2.352889e+06	(51.5064605, -0.1988955)	51.506461	-0.198896
1889	CANFIELD GARDENS	2.278000e+06	(51.5467987, -0.1797091)	51.546799	-0.179709
1894	CANNING PLACE	2.425000e+06	(51.4995696, -0.1842477)	51.499570	-0.184248

```
In [22]: #Define Foursquare Credentials and Version

CLIENT_ID = 'KI3TR0QO4JOKMFELOMF3WSOOI3HFNBF5YLW354MYWBKDHEX3' # Foursquare ID
CLIENT_SECRET = 'QF4ZBLJRBV4BQX52DVWUPEHJ14A2UJABPCZARZQZYTKIISUD' # Foursquare Secret
VERSION = '20181206' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: KI3TR0QO4JOKMFELOMF3WSOOI3HFNBF5YLW354MYWBKDHEX3
CLIENT_SECRET:QF4ZBLJRBV4BQX52DVWUPEHJ14A2UJABPCZARZQZYTKIISUD
```

METHODOLOGY SECTION IV

- MODELLING
  - AFTER EXPLORING THE DATASET AND GAINING INSIGHTS INTO IT, WE ARE READY TO USE THE CLUSTERING METHODOLOGY TO ANALYZE REAL ESTATES. WE WILL USE THE K-MEANS CLUSTERING TECHNIQUE AS IT IS FAST AND EFFICIENT IN TERMS OF COMPUTATIONAL COST, IS HIGHLY FLEXIBLE TO ACCOUNT FOR MUTATIONS IN REAL ESTATE MARKET IN LONDON AND IS ACCURATE.

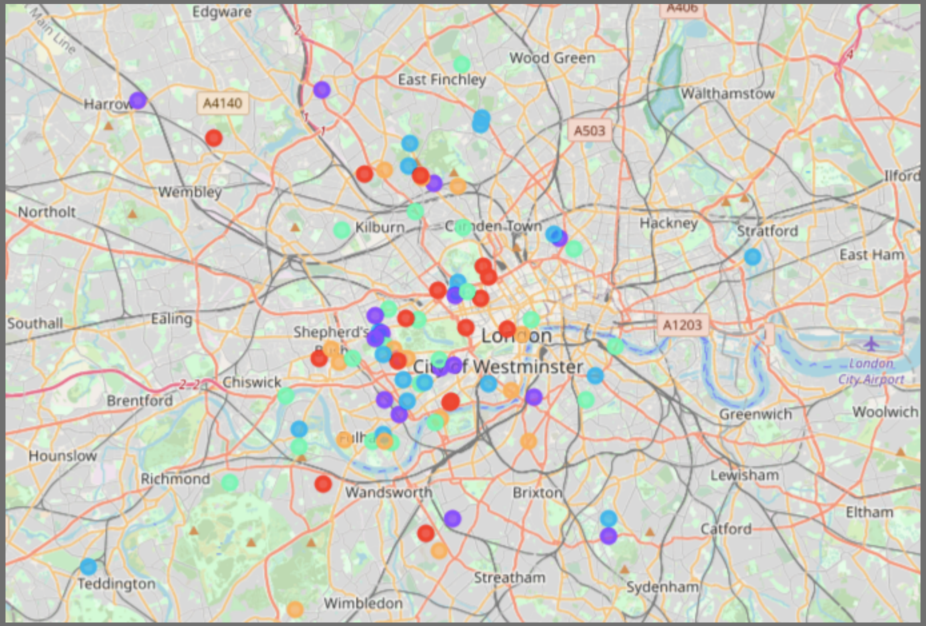
In [46]:
```
# add clustering labels
london_grouped_clustering['Cluster Labels'] = kmeans.labels_

# merge london_grouped with london_data to add latitude/longitude for each neighborhood
london_grouped_clustering = london_grouped_clustering.join(venues_sorted.set_index('Street'), on='Street')

london_grouped_clustering.head(30) # check the last columns!
```

Out[46]:

	Street	Avg_Price	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
20	ABBOTSBURY CLOSE	2.367093e+06	51.532259	-0.006153	2	Grocery Store	Park	Waterfront	Hotel	Thai Restaurant	Farm	Eastern European Restaurant
178	ALBION SQUARE	2.450000e+06	-41.273758	173.289393	0	Café	Restaurant	Indian Restaurant	Bar	Coffee Shop	Pub	New American Restaurant
355	ANHALT ROAD	2.435000e+06	51.480326	-0.166761	4	Pub	Plaza	Pizza Place	Grocery Store	Japanese Restaurant	French Restaurant	English Restaurant
368	ANSDELL TERRACE	2.250000e+06	51.499890	-0.189103	1	Clothing Store	Italian Restaurant	Café	English Restaurant	Pub	Juice Bar	Hotel
381	APPLEGARTH ROAD	2.400000e+06	53.748654	-0.326670	4	Bar	Pub	Casino	Nightclub	Fast Food Restaurant	English Restaurant	Event Space
617	AYLESTONE AVENUE	2.286667e+06	51.540916	-0.217874	3	Park	Movie Theater	Bus Stop	Café	Fish & Chips Shop	Event Space	Exhibit
753	BARONSMEAD ROAD	2.375000e+06	51.477315	-0.239457	2	Food & Drink Shop	Restaurant	Breakfast Spot	Nature Preserve	Bookstore	Farmers Market	Thai Restaurant
867	BEAUCLERC ROAD	2.480000e+06	51.499577	-0.229033	0	Coffee Shop	Pub	Hotel	Thai Restaurant	Grocery Store	Chinese Restaurant	Cocktail Bar
1079	BICKENHALL STREET	2.351667e+06	51.521197	-0.158934	2	Hotel	Chinese Restaurant	Gastropub	Restaurant	Italian Restaurant	Pizza Place	Pub



RESULTS & DISCUSSION

1. It is interesting to note that, although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores,

    - South-West London (Wandsworth, Balham) and North-West London (Isliington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair.

2. Second, we may analyze our results according to the five clusters we have produced.

- Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns.
    - The first pattern we are referring to, i.e. Clusters 0, 2 and 4, may target home buyers prone to live in 'green' areas with parks, waterfronts.
    - Instead, the second pattern we are referring to, i.e. Clusters 1 and 3, may target individuals who love pubs, theatres and soccer.

CONCLUSION

- It is still an "ever-green" for business.
- We discussed our results under two main perspectives.
  - First, we examined them according to neighborhoods/London areas. although West London and North-West London might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues, South-West London and North-West London are arising as next future elite venues with a wide range of amenities and facilities.
  - Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair.
  - Second, we analyzed our results according to the five clusters we produced. While Clusters 0, 2 and 4 may target home buyers prone to live in 'green' areas with parks, waterfronts, Clusters 1 and 3 may target individuals who love pubs, theatres and soccer.