Contoso

IBM DATA SCIENCE CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS –

ANALYZED ON AMSTERDAM

By Analyst Tse Kit Yee 23/05/2021

BUSINESS PROBLEM

1. Is the place suitable for opening a restaurant?
2. Is it a good place for a constracter to start his/her own business, i.e. to setup his/her office?

OUR GOAL

To analyze: Is Amsterdam being a well multicultural place to satisfy the above two main business problem?

Contoso

INTRODUCTION ON AMSTERDAM

- The largest city in the Netherlands
- Have a population of 1.16 million
- This place with approximately 178 different cultural backgrounds
- It hosts many cultural events throughout the year, which emphasizes their worldly view

DESCRIPTION OF DATASET

- Data scraping from https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam to get **Neighbourhoods of Amsterdam**
- Use geopy library to get the **latitude and longitude values of Amsterdam**
- Use Foursquare API to get data of venue

METHODOLOGY

- Explore and load dataset from website link
- Get latitude and longitude values of Amsterdam and plot the map
- Data transformation by pandas
- Use one-hot encoding to do features extraction
- Apply FourSquare API in all the neighbourhoods on Amsterdam
- Use k-means clustering to visualize the clusters

PROCESS

Contoso

EXPLORE AND LOAD DATASET FROM WEBSITE LINK

In [67]:
```
source = requests.get('https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam').text
source
soup = BeautifulSoup(source, 'lxml')
soup
```


```
ods_of_Amsterdam","wgTitle":"Neighbourhoods of Amsterdam","wgCurRevisionId":645887402,"wgRevisionId":645887402,"wgArticleId":
3311930,"wgIsArticle":!0,"wgIsRedirect":!1,"wgAction":"view","wgUserName":null,"wgUserGroups":["*"],"wgCategories":["Neighbou
rhoods in the Netherlands by city","Geography of Amsterdam","Boroughs of Amsterdam"],"wgPageContentLanguage":"en","wgPageCont
entModel":"wikitext","wgRelevantPageName":"Category:Neighbourhoods_of_Amsterdam","wgRelevantArticleId":3311930,
"wgIsProbablyEditable":!0,"wgRelevantPageIsProbablyEditable":!0,"wgRestrictionEdit":[],"wgRestrictionMove":[],"wgFlaggedRevsP
arams":{"tags":{"status":{"levels":-1}}},"wgMediaViewerOnClick":!0,"wgMediaViewerEnabledByDefault":!0,"wgPopupsFlags":10,"wgV
isualEditor":{"pageLanguageCode":"en","pageLanguageDir":"ltr","pageVariantFallbacks":"en"},"wgMFDisplayWikibaseDescriptions":
{"search":!0,"nearby":!0,"watchlist":!0,"tagline":!1},"wgWMESchemaEditAttemptStepOversample":!1,"wgULSCurrentAutonym":"Englis
h","wgNoticeProject":"wikipedia","wgCentralAuthMobileDomain":!1,"wgEditSubmitButtonLabelPublish":!0,"wgULSPosition":"interlan
guage","wgGENewcomerTasksGuidanceEnabled":!0,"wgGEAskQuestionEnabled":!1,"wgGELinkRecommendationsFrontendEnabled":!1,"wgWikib
aseItemId":"Q8671092"};RLSTATE={"ext.globalCssJs.user.styles":"ready","site.styles":"ready","noscript":"ready","user.style
s":"ready","ext.globalCssJs.user":"ready","user":"ready","user.options":"loading",
"mediawiki.action.view.categoryPage.styles":"ready","mediawiki.helplink":"ready","skins.vector.styles.legacy":"ready","ext.tm
h.thumbnail.styles":"ready","ext.visualEditor.desktopArticleTarget.noscript":"ready","ext.uls.interlanguage":"ready","ext.wik
imediaBadges":"ready","wikibase.client.init":"ready"};RLPAGEMODULES=["site","mediawiki.page.ready","skins.vector.legacy.j
s","ext.gadget.ReferenceTooltips","ext.gadget.charinsert","ext.gadget.extra-toolbar-buttons","ext.gadget.refToolbar","ext.gad
get.switcher","mw.MediaWikiPlayer.loader","mw.PopUpMediaTransform","ext.centralauth.centralautologin","ext.popups","ext.visua
lEditor.desktopArticleTarget.init","ext.visualEditor.targetLoader","ext.eventLogging","ext.wikimediaEvents","ext.navigationTi
ming","ext.uls.compactlinks","ext.uls.interface","ext.cx.eventlogging.campaigns","ext.centralNotice.geoIP","ext.centralNotic
e.startUp"];</script>
```


Contoso

GET NEIGHBORHOODS OF AMSTERDAM

Out[69]:

	Neighborhood
0	Admiralenbuurt
1	Amsteldorp
2	Amsterdam Oud-West
3	Amsterdam Oud-Zuid
4	Amsterdam Science Park
5	Apollobuurt
6	Betondorp
7	Bijlmermeer
8	Binnenstad (Amsterdam)
9	Bos en Lommer
10	Buiksloot

GET LATITUDE AND LONGITUDE VALUES OF AMSTERDAM AND PLOT THE MAP

In [72]:
```
adf = df
adf['Latitude'] = lat
adf['Longitude'] = lng
adf.head(10)
```

Out[72]:

	Neighborhood	Latitude	Longitude
0	Admiralenbuurt	52.372752	4.856359
1	Amsteldorp	52.339680	4.918740
2	Amsterdam Oud-West	52.365390	4.870220
3	Amsterdam Oud-Zuid	52.352350	4.877880
4	Amsterdam Science Park	52.354300	4.958010
5	Apollobuurt	52.350294	4.867990
6	Betondorp	52.423405	4.833395
7	Bijlmermeer	52.307031	4.969744
8	Binnenstad (Amsterdam)	52.369930	4.907880
9	Bos en Lommer	52.379190	4.851740

AFTER SET UP FOURSQUARE CREDENTIALS, PLOT MAP

Then, get:

Latitude and longitude values of Admiralenbuurt are 52.3727517353822, 4.85635890037192, to ensure map-plotting is right to continue.

Keep taking Admiralenbuurt as example:

Create and display URL returned by Foursquare API, next we could get the nearby venues of Admiralenbuurt

```
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

```
<ipython-input-129-561c05f0fdd1>:3: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize inste
ad
  nearby_venues = json_normalize(venues) # flatten JSON
```

Out[129]:

	name	categories	lat	lng
0	Sapporo Ramen Sora	Ramen Restaurant	52.371294	4.855144
1	Deli-caat	Deli / Bodega	52.371221	4.856056
2	Rein Cityspa	Spa	52.371217	4.855969
3	Café Cook	Pub	52.371208	4.852792
4	Maz Mez	Lebanese Restaurant	52.371231	4.857968

USE ONE-HOT ENCODING TO DO FEATURES EXTRACTION

- Analyze Each Neighborhood:

In [138]:
```
# one hot encoding
a_onehot = pd.get_dummies(a_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
a_onehot['Neighborhood'] = a_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [a_onehot.columns[-1]] + list(a_onehot.columns[:-1])
a_onehot = a_onehot[fixed_columns]

a_onehot.head()
```

Out[138]:

	Zoo Exhibit	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	A R
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

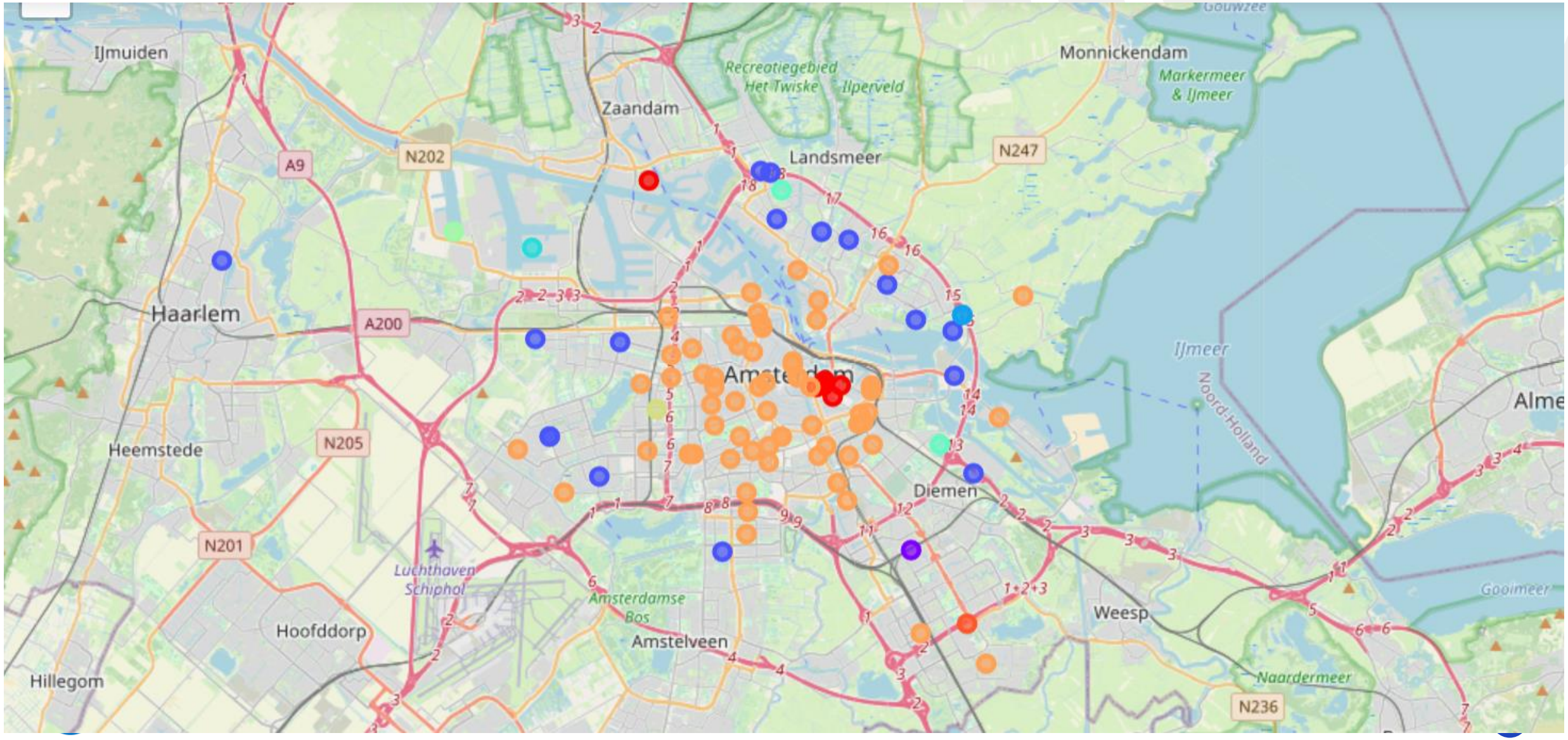
APPLY FOURSQUARE API IN ALL THE NEIGHBOURHOODS ON AMSTERDAM

```
# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = a_grouped['Neighborhood']

for ind in np.arange(a_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(a_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[144]:

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Admiralenbuurt	Restaurant	Snack Place	Café	Bar	Coffee Shop	Deli / Bodega	Supermarket	Plaza	Pet Café	Scandinavian Restaurant
1	Amsteldorp	Furniture / Home Store	Sports Club	Indoor Play Area	Dutch Restaurant	Café	French Restaurant	Gastropub	Park	Italian Restaurant	Brasserie
2	Amsterdam Oud-West	Café	Coffee Shop	Restaurant	Italian Restaurant	Bar	Pizza Place	Indonesian Restaurant	Movie Theater	Drugstore	Mediterranean Restaurant
3	Amsterdam Oud-Zuid	Restaurant	Hotel	Bar	Bakery	Plaza	Juice Bar	French Restaurant	Tram Station	Concert Hall	Spa
4	Amsterdam Science Park	Bus Stop	Coffee Shop	Convenience Store	Restaurant	Zoo	Empanada Restaurant	Dog Run	Doner Restaurant	Drugstore	Dutch Restaurant

USE K-MEANS CLUSTERING TO DO CLUSTER NEIGHBORHOODS



DISCUSSIONS

- Both red cluster and orange cluster are shown high concentration of restaurant in that neighborhoods, so we would not consider to start a food business there.
- Deep-Blue cluster is more discrete, which shown the neighborhoods in Amsterdam has less or no existence of restaurant.
- For the remaining cluster, they are also a choice to start Chinese food business, because those areas are without overlapping clustering with others.

IMPROVEMENTS

- The latest mapping are included all types of restaurant or shops provided food
- Since I want to analyze on is there possible to start a Chinese restaurant that attract multicultural people to visit, so the map may better to focus on Chinese food shop/ restaurant in those selected neighborhoods. So, I may map a map, only consider on the shops provided Chinese food, not all types of food.
- However, consider on all types of food shop has one advantage because people are more eager to have normal meal rather than trying new types of Chinese food in weekdays, so selected locations among tourism area may better attract multicultural people in weekends.

Contoso

THANK YOU!

End of week 5’s task

Contoso