CAPSTONE PROJECT:

BATTLE OF THE NEIGHBORHOODS

I. PURPOSE

This document is the report of the IBM Data Science Professional Certificate program – Coursera Capstone.

II. INTRODUCTION

Singapore is a small country and one of the most visited countries in Asia. There are lots of websites where travelers can recieve recommendations of places to stay or visit. However, most of these websites provides recommendation simply based on usual tourist attractions or key residential areas that are mostly expensive or already known for travelers based on certain keywords like "Hotel", or "Backpackers" etc. The purpose of this project is to collect and provide a data driven recommendation that can supplement the recommendation with statistical data. This will also be utilizing data retrieved from Singapore open data sources and FourSquare API venue recommendations.

The sample recommender in this notebook will provide the following use case scenario:
    * A person planning to visit Singapore as a Tourist or an Expat and looking for a reasonable accommodation.
    * The user wants to receive venue recommendation where he can stay or rent an HDB apartment with close proximity to places of interest or search category option.
    * The recommendation should not only present the most viable option, but also present a comparison table of all possible town venues.

For this demonstration, this notebook will make use of the following data:
    * Singapore Median Rental Prices by town.
    * Popular Food venues in the vicinity. (Sample category selection)

Note: While this demo makes use of Food Venue Category, Other possible categories Can also be used for the same implementation such as checking categories like:
    * Outdoors and Recreation
    * Nightlife
    * Nearby Schools, etc.
I will limit the scope of this search as FourSquare API only allows 50 free venue query limit per day when using a free user access.

III. DATA ACQUISITION

This demonstration will make use of the following data sources:

Singapore Towns and median residential rental prices.

Data will retrieved from Singapore open dataset from [median rent by town and flattype](#) from https://data.gov.sg website.

The original data source contains median rental prices of Singapore HDB units from 2005 up to 2nd quarter of 2018. I will retrieve rental the most recent recorded rental prices from this data source (Q2 2018) being the most relevant price available at this time. For this demonstration, I will simplify the analysis by using the average rental prices of all available flat type.

Singapore Towns location data retrieved using Google maps API.

Data coordinates of Town Venues will be retrieved using google API. I also make use of MRT stations coordinate as a more important center of for all towns included in venue recommendations.

Singapore Top Venue Recommendations from FourSquare API

(FourSquare website: www.foursquare.com)

I will be using the FourSquare API to explore neighborhoods in selected towns in Singapore. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

* Venue ID
* Venue Name
* Coordinates : Latitude and Longitude
* Category Name

Another venue query will be performed to retrieve venue ratings for each location. Note that rating information is a paid service from FourSquare and we are limited to only 50 queries per day. With this constraint, we limit the category analysis with only one type for this demo. I will try to retrieve as many ratings as possible for each retrieved venue ID.

IV. METHODOLOGY

Singapore Towns List with median residential rental prices.

The source data contains median rental prices of Singapore HDB units from 2005 up to 2nd quarter of 2018. I will retrieve the most recent recorded rental prices from this data source (Q2 2018) being the most relevant price available at this time. For this demonstration, I will simplify the analysis by using the average rental prices of all available flat type.

Data Cleanup and re-grouping. The retrieved table contains some un-wanted entries

and needs some cleanup.
The following tasks will be performed:
  * Drop/ignore cells with missing data.
  * Use most current data record.
   * Fix data types. Post Processed Singapore towns list with and median residential rental prices
  * Adding geographical coordinates of each town location.
Retrieve town coordinates.

Google API was be used to retrieve the coordinates (latitude and longitude of each town centers. For this exercise, I just used the MRT stations as the center points of each evaluated towns. The town coordinates will be used in retrieval of Foursquare API location data.

```
singapore_average_rental_prices_by_town['Latitude'] = 0.0
singapore_average_rental_prices_by_town['Longitude'] = 0.0
for idx,town in singapore_average_rental_prices_by_town['Town'].iteritems():
address = town + " MRT station, Singapore" ; # I prefer to use MRT stations as
more important central location of each town
url =
'https://maps.googleapis.com/maps/api/geocode/json?address={}&key={}'.format(address,
google_key)
lat = requests.get(url).json()["results"][0]["geometry"]["location"]['lat']
lng = requests.get(url).json()["results"][0]["geometry"]["location"]['lng']
singapore_average_rental_prices_by_town.loc[idx,'Latitude'] = lat
singapore_average_rental_prices_by_town.loc[idx,'Longitude'] = lng
```

V. Segmenting and Clustering Towns in Singapore

Retrieving FourSquare Places of interest.

Using the Foursquare API, the **explore** API function was be used to get the most common venue categories in each neighborhood, and then used this feature to group the neighborhoods into clusters. The *k*-means clustering algorithm was used for the analysis. Fnally, the Folium library is used to visualize the recommended neighborhoods and their emerging clusters.

In the ipynb notebook, the function **getNearbyVenues** extracts the following information for the dataframe it generates:
  * Venue ID
  * Venue Name
  * Coordinates : Latitude and Longitude
  * Category Name

The function **getVenuesByCategory** performs the following:

1. **Category** based venue search to simulate user venue searches based on certain places
of interest. This search extracts the following information:
    * Venue ID
    * Venue Name
    * Coordinates : Latitude and Longitude
    * Category Name
2. For each retrieved **venueID**, retrieve the venues category rating.
The generated data frame in the second function contains the following column:

Search Venues with recommendations on : Food Venues (Restaurants,Fastfoods, etc.)

To demonstrate user selection of places of interest, We will use this Food Venues category in our further analysis.
    * This Foursquare search is expected to collect venues in the following category:
    * category
    * Food Courts
    * Coffee Shops
    * Restaurants
    * Cafés
    * Other food venues
I used the FourSquare API to retrieve venue scores of locations. Note that there is mamax query limit of 50 in FourSquare API for free subscription. So use or query carefully.

Data cleanup un-needed entries
    * Eliminate possible venue duplicates.
   * Improve the quality of our venue selection by removing venues with no ratings or 0.0

Column Name	Description
Town	Town Name
Town Latitude	Towns MRT station Latitude
Town Longitude	Town MRT station Latitude
VenueID	FourSquare Venue ID
VenueName	Venue Name
score	FourSquare Venue user rating
category	Category group name
catID	Category ID
latitude	Venue Location - latitude
longitude	Venue Location - longitude

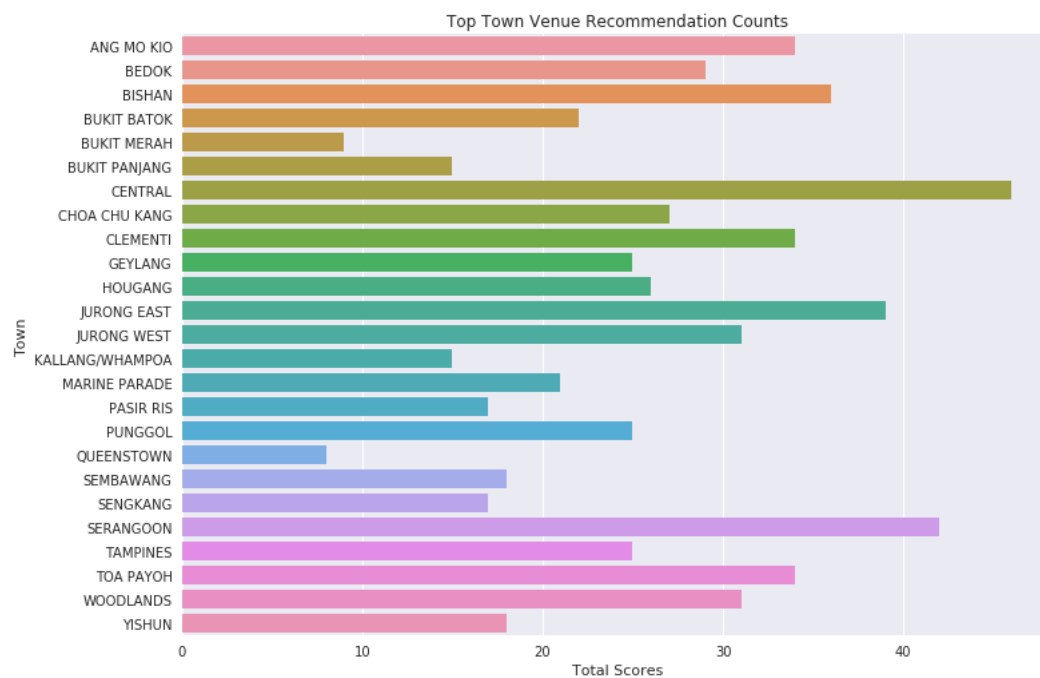### Results: Town Venue Recommendation Counts:

Town	Venue Categoy Counts
ANG MO KIO	34
BEDOK	29
BISHAN	36
BUKIT BATOK	22
BUKIT MERAH	9
BUKIT PANJANG	15
CENTRAL	46
CHOA CHU KANG	27
CLEMENTI	34
GEYLANG	25
HOUGANG	26
JURONG EAST	39
JURONG WEST	31
KALLANG/WHAMPOA	15
MARINE PARADE	21
PASIR RIS	17
PUNGGOL	25
QUEENSTOWN	8
SEMBAWANG	18
SENGKANG	17
SERANGOON	42
TAMPINES	25
TOA PAYOH	34
WOODLANDS	31
YISHUN	18

There are 67 uniques categories *can be curated from all the returned venues.*
What are the top 20 most common venue types?

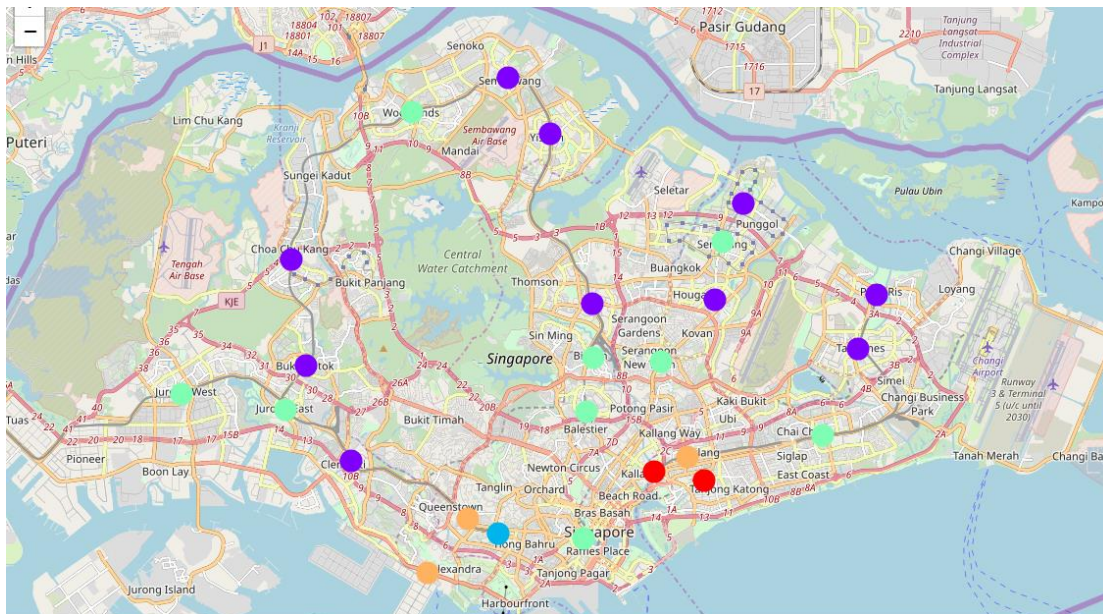Top 10 Recommended Venue in Food Category Type
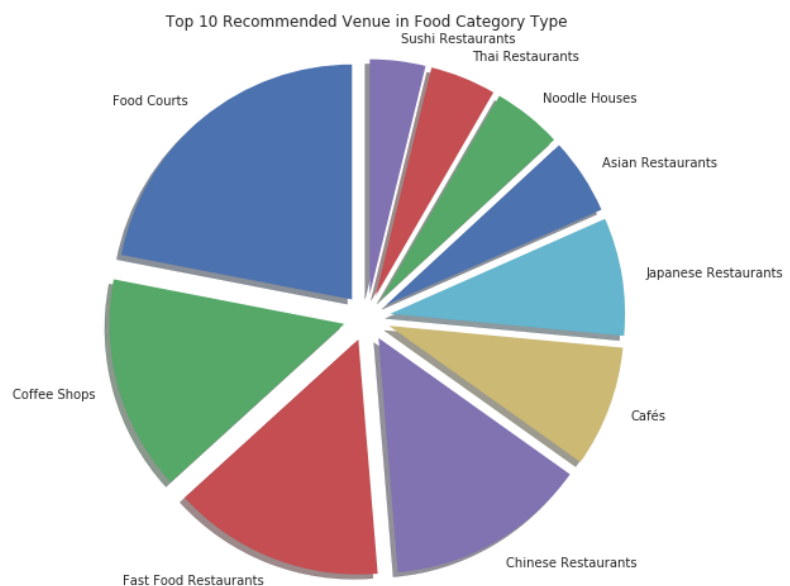
What are the top 10 Venue Recommendation Scores?

Analyze Each Singapore Town nearby recommended venues

Technique : **One Hot Encoding**

After some coding, here is the map:

IV. Conclusion

On this notebook, Analysis of best town venue recommendations based on Food venue category has been presented. Recommendations based on other user searches like available outdoor and recreation areas are also available. As singapore is a small country with a whole host of interesting venues scattered around the town, the information extracted in this notebook present on the town areas, will be a good

supplement to web based recommendations for visitors to find out nearby venues of interest and be a useful aid in deciding a place to stay or where to go during their visits.

Using Foursquare API, we have collected a good amount of venue recommendations in Singapore Towns. Sourcing from the venue recommendations from FourSquare has its limitation; The list of venues is not exhaustive list of all the available venues is the area. Furthermore, not all the venues found in the the area has a stored ratings. For this reason, the number of analyzed venues is only about 50% of all the available venues initially collected. The results therefore may significantly change, when more information are collected on those with missing data.

The generated clusters from our results shows that there are very good and interesting places located in areas where the median rents are cheaper. This kind of results may be very interesting for travelers who are also on budget constraints. Our results also yielded some interesting findings. For instance, The initial assumption among websites providing recommendations is that the Central Area that have the highest median rent also have better food venues. The results however show that while Marine Parade, a cheaper location has better rated food courts. Result shows that most popular food venue among Singaporeans, residents and visitors are **Food Courts, Coffee Shops and Fast Food Restaurants**. The highest rated Food Courts are located in **Marine Parade**, and in **Central Area**.