IBM Data Science Capstone Project

Done by: Angelina

Description of Problem

Background

Singapore is known for its gastronomical delights. If a businessman wants to open a restaurant in Singapore, how does he start his research?

Problem Statement

How can we leverage Foursquare location data and machine learning to help the businessman decide he can open his restaurant? What type of restaurants are popular in Singapore?

Jewel Changi Airport

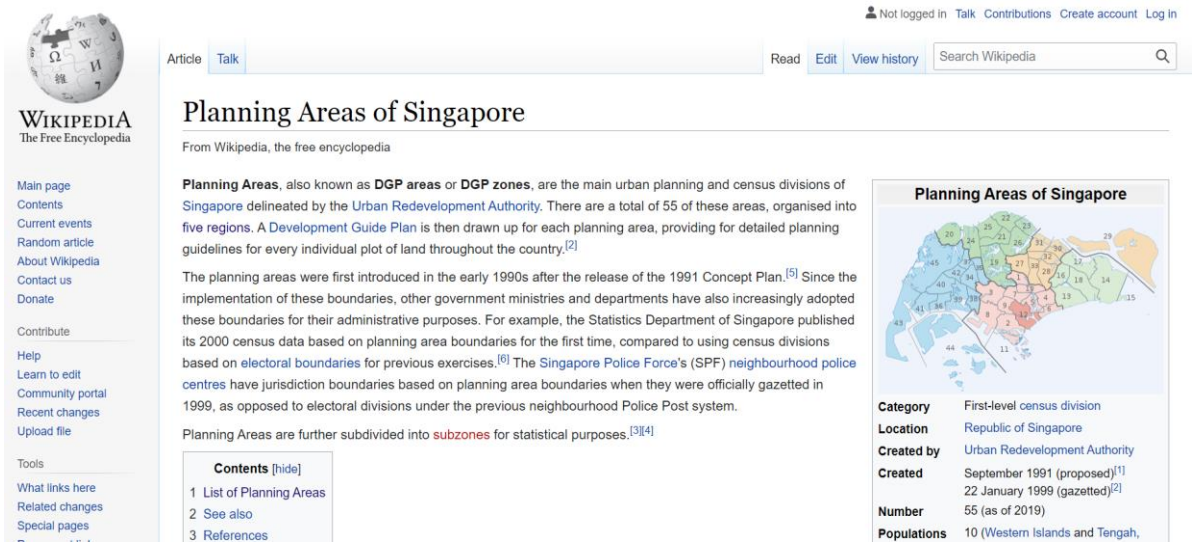
How Data Will Be Used

Foursquare

Foursquare data about restaurants in Singapore will be used to extract information such as its location, name and category of food

Planning Areas of Singapore

The overview of planning areas in Singapore will be extracted from Wikipedia: https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages

Planning Areas of Singapore

From Wikipedia, the free encyclopedia

Planning Areas, also known as **DGP areas** or **DGP zones**, are the main urban planning and census divisions of Singapore delineated by the Urban Redevelopment Authority. There are a total of 55 of these areas, organised into five regions. A Development Guide Plan is then drawn up for each planning area, providing for detailed planning guidelines for every individual plot of land throughout the country.[2]

The planning areas were first introduced in the early 1990s after the release of the 1991 Concept Plan.[5] Since the implementation of these boundaries, other government ministries and departments have also increasingly adopted these boundaries for their administrative purposes. For example, the Statistics Department of Singapore published its 2000 census data based on planning area boundaries for the first time, compared to using census divisions based on electoral boundaries for previous exercises.[6] The Singapore Police Force's (SPF) neighbourhood police centres have jurisdiction boundaries based on planning area boundaries when they were officially gazetted in 1999, as opposed to electoral divisions under the previous neighbourhood Police Post system.

Planning Areas are further subdivided into subzones for statistical purposes.[3][4]

Contents [hide]
1 List of Planning Areas
2 See also
3 References

Planning Areas of Singapore

Category	First-level census division
Location	Republic of Singapore
Created by	Urban Redevelopment Authority
Created	September 1991 (proposed)[1] 22 January 1999 (gazetted)[2]
Number	55 (as of 2019)
Populations	10 (Western Islands and Tengah,

Methodology

Scrapping of Data into a Dataframe

Irrelevant columns will be dropped to get a nice dataframe with relevant columns

```
In [5]: df = df.drop(columns=['Malay', 'Chinese', 'Pinyin','Tamil'])

In [6]: df.shape
Out[6]: (55, 5)

In [7]: df.head()
Out[7]:
```

	Name (English)	Region	Area (km2)	Population[7]	Density (/km2)
0	Ang Mo Kio	North-East	13.94	163950	13400
1	Bedok	East	21.69	279380	13000
2	Bishan	Central	7.62	88010	12000
3	Boon Lay	West	8.23	30	3.6
4	Bukit Batok	West	11.13	153740	14000

Methodology

Obtain Coordinates of Planning Areas: Geopy Client

Used the nominatim function to add geospatial data to the data frame

Out[13]:

	Name (English)	Region	Area (km2)	Population[7]	Density (/km2)	Latitude	Longitude
0	Ang Mo Kio	North-East	13.94	163950	13400	1.370080	103.849523
1	Bedok	East	21.69	279380	13000	1.323976	103.930216
2	Bishan	Central	7.62	88010	12000	1.350986	103.848255
3	Boon Lay	West	8.23	30	3.6	1.338550	103.705812
4	Bukit Batok	West	11.13	153740	14000	1.349057	103.749591
5	Bukit Merah	Central	14.34	151980	11000	4.561694	101.024037
6	Bukit Panjang	West	8.99	139280	15000	1.379149	103.761413
7	Bukit Timah	Central	17.53	77430	4400	1.354690	103.776372
8	Central Water Catchment	North	37.15	*	*	1.375708	103.801743
9	Changi	East	40.61	1830	80.62	43.880078	126.564903
10	Changi Bay	East	1.70	*	*	1.316850	104.020649
11	Choa Chu Kang	West	6.11	190890	30000	1.385317	103.744325
12	Clementi	West	9.49	92420	9800	1.315100	103.765231
13	Downtown Core	Central	4.34	2720	680	1.287475	103.856033
14	Geylang	Central	9.64	110200	11400	1.318186	103.887056
15	Hougang	North-East	13.93	226240	16000	1.370801	103.892544
16	Jurong East	West	17.83	79240	4400	1.333108	103.742294
17	Jurong West	West	14.69	264860	18000	1.339636	103.707339
18	Kallang	Central	9.17	101520	11000	1.310759	103.866262

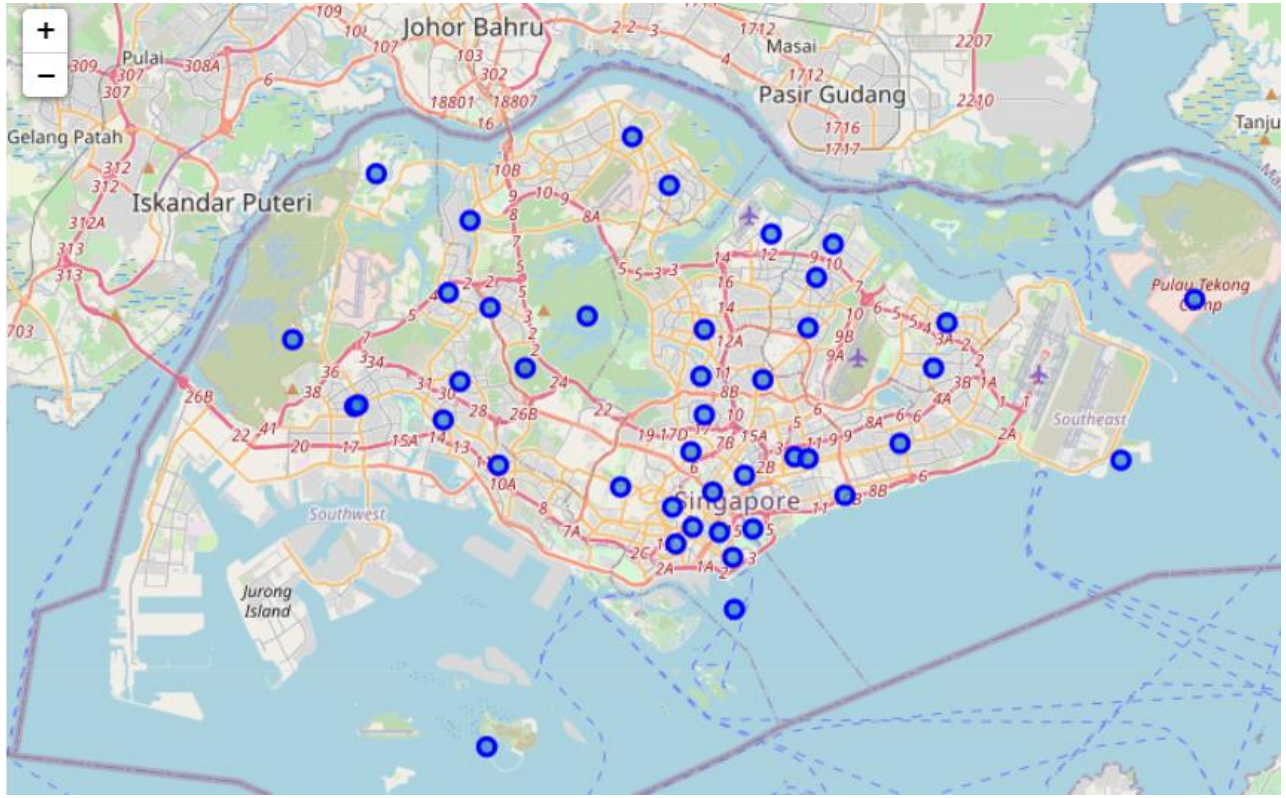
Methodology

Create Map with Folium

In [14]:
```
# create map of Singapore  using latitude and longitude values
map_singapore = folium.Map(location=[latitude, longitude], zoom_start=11)

 # add markers to map
for lat, lng, label in zip(df['Latitude'], df['Longitude'], df['Name (English)']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_singapore)

map_singapore
```



Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA)

Create a Dataframe for Restaurants Only

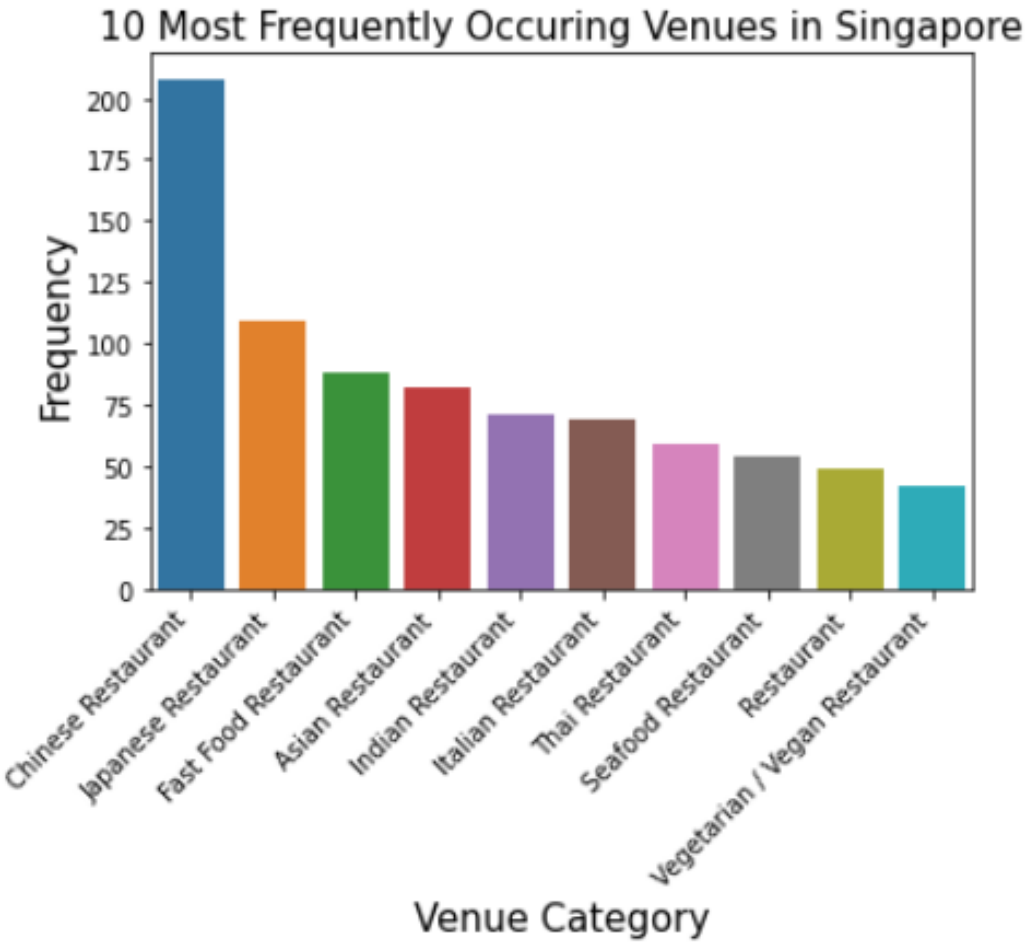
Obtain list of restaurants in Singapore

```
In [37]: print (Singapore_Venues_only_restaurant['Venue Category'].value_counts())
```

```
Chinese Restaurant                208
Japanese Restaurant               109
Fast Food Restaurant               88
Asian Restaurant                   82
Indian Restaurant                  71
Italian Restaurant                 69
Thai Restaurant                    59
Seafood Restaurant                 54
Restaurant                         49
Vegetarian / Vegan Restaurant      42
Sushi Restaurant                   32
French Restaurant                  20
Hotpot Restaurant                  20
American Restaurant                20
Korean Restaurant                  19
Dim Sum Restaurant                 18
Dumpling Restaurant                14
Vietnamese Restaurant              14
Indonesian Restaurant              12
Malay Restaurant                   12
Spanish Restaurant                 11
Hainan Restaurant                  10
Modern European Restaurant         10
Cantonese Restaurant                9
Ramen Restaurant                    8
Halal Restaurant                    7
```

Exploratory Data Analysis (EDA)

Used seaborn to reflect the top 10 most frequently occurring venues in Singapore



Exploratory Data Analysis (EDA)

Obtained the most common venue in each neighbourhood

Out[67]:

	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
0	Ang Mo Kio	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Shaanxi Restaurant	Indian Restaurant
1	Bedok	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Malay Restaurant
2	Bishan	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Szechuan Restaurant	Sushi Restaurant
3	Boon Lay	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant	Restaurant	Thai Restaurant
4	Bukit Batok	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant	Japanese Restaurant	Hainan Restaurant
5	Bukit Merah	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Vietnamese Restaurant	Hainan Restaurant	Kebab Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant
6	Bukit Panjang	Fast Food Restaurant	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant	Japanese Restaurant
7	Bukit Timah	Italian Restaurant	Chinese Restaurant	Korean Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Paella Restaurant

Exploratory Data Analysis (EDA)

K Means Clustering

To put the planning areas into 5 clusters, and add the longitude and latitude and merge the tables

In [100]:
```
Singapore_merged.rename(columns={'Name (English)':'Neighborhood'}, inplace=True)
Singapore_merged = Singapore_merged.merge(neighborhoods_venues_sorted_w_clusters.set_index('Neighborhood'), on='Neighborhood')
Singapore_merged.head()
```

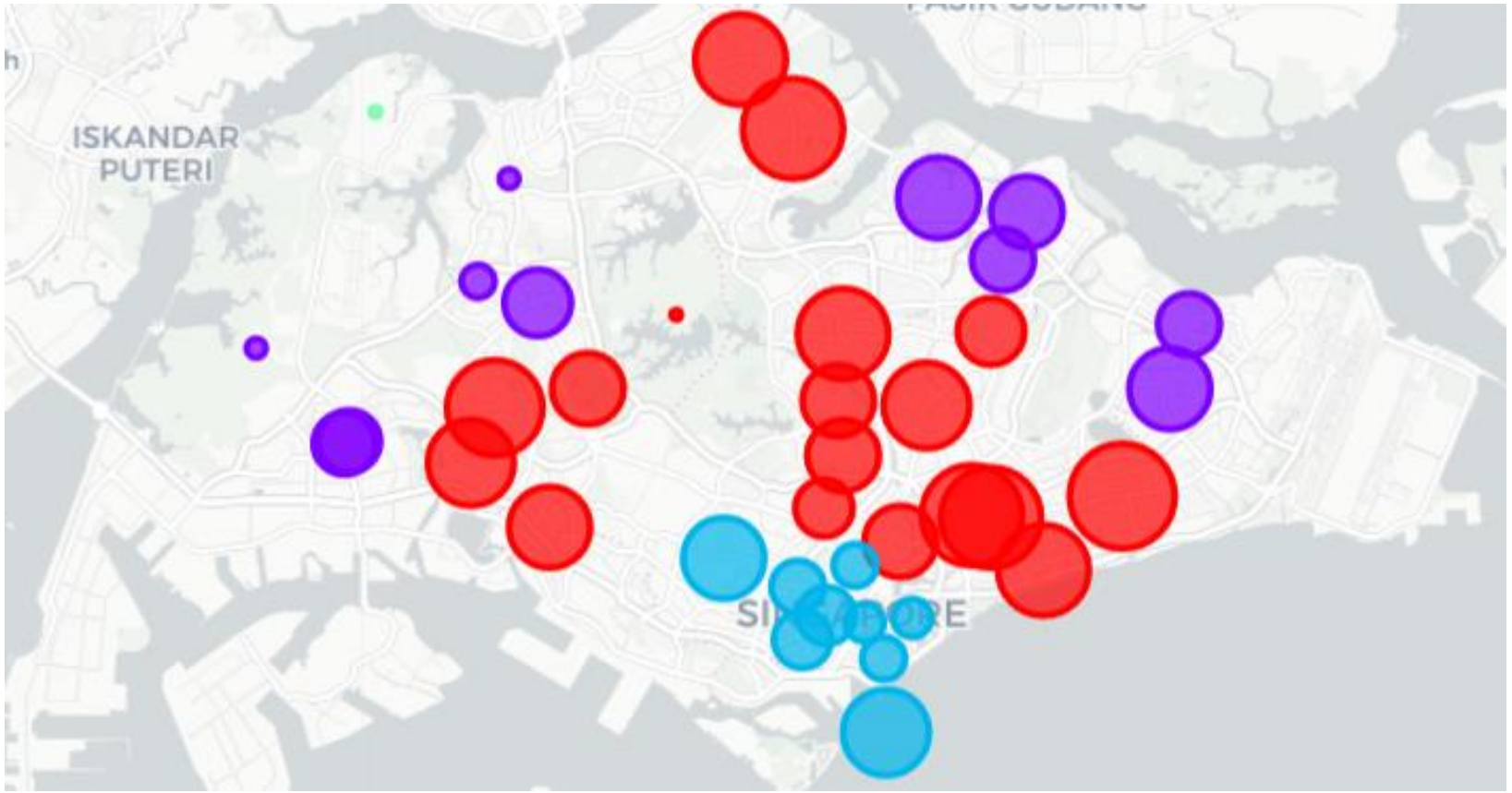
Out[100]:

	Neighborhood	Region	Area (km2)	Population[7]	Density (/km2)	Latitude	Longitude	Cluster Labels_x	1st Most Common Venue_x	2nd Most Common Venue_x	3rd Most Common Venue_x	4th Most Common Venue_x	5th Most Common Venue_x	6th Most Common Venue_x	7th Most Common Venue_x
0	Ang Mo Kio	North-East	13.94	163950	13400	1.370080	103.849523	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant
1	Bedok	East	21.69	279380	13000	1.323976	103.930216	0.0	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Fast Food Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant
2	Bishan	Central	7.62	88010	12000	1.350986	103.848255	0.0	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Indian Restaurant
3	Boon Lay	West	8.23	30	3.6	1.338550	103.705812	1.0	Fast Food Restaurant	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Indian Restaurant	Japanese Curry Restaurant	Sushi Restaurant
4	Bukit Batok	West	11.13	153740	14000	1.349057	103.749591	0.0	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Thai Restaurant

Results

Results

Used folium to put the information into a map



Results

Description of Clusters

- Cluster 1 – Chinese Restaurants
- Cluster 2 – Fast Food Restaurants
- Cluster 3 – Japanese Restaurants
- Cluster 4 – Vegetarian / Vegan Restaurants
- Cluster 5 – Chinese Restaurants

*Please refer to Github for the code used

Discussion

Discussion and Insights

- Chinese Restaurants top the list as the most common type of restaurants, and it is quite spaced out across Singapore
- Fast Food Restaurants are popular along the outskirts of Singapore's Central area
- Japanese Restaurants seem to concentrate in the Central area, particularly close to where Singapore's Business Districts are
- There do not seem to be many Vegetarian/Vegan Restaurants around

Other Factors for Consideration

It is important to note that there are other factors that determine the location of the restaurants, such as the distance from the venue to closest train stations, and type of housing or offices in each area. Hence, from the above, the businessman can decide on the type of restaurant he wishes to open and narrow down the geographical location he wishes to explore for his restaurant.

Conclusion

Conclusion

It is interesting how data can be pulled from various sources to gather insights about a problem that one wishes to solve. For example, Foursquare's data is most helpful in giving us an overview about restaurants in Singapore without having to do any legwork. It can give the businessman some good understanding about the types of restaurants and cuisines that he can introduce in Singapore.

From this project, the data is malleable can be used to solve other problems. For example, it can also be pivoted to produce food recommendations for tourists.