



Finding Locations for Setting Up Restaurant in Singapore

Thomas Chua
31 Oct 2020

IBM Applied Data Science Capstone
Battle of Neighborhoods



Business Problem

- The study aims to achieve a two fold objective:
 - Where to set up a restaurant?
 - What cuisine to focus on for a selected location



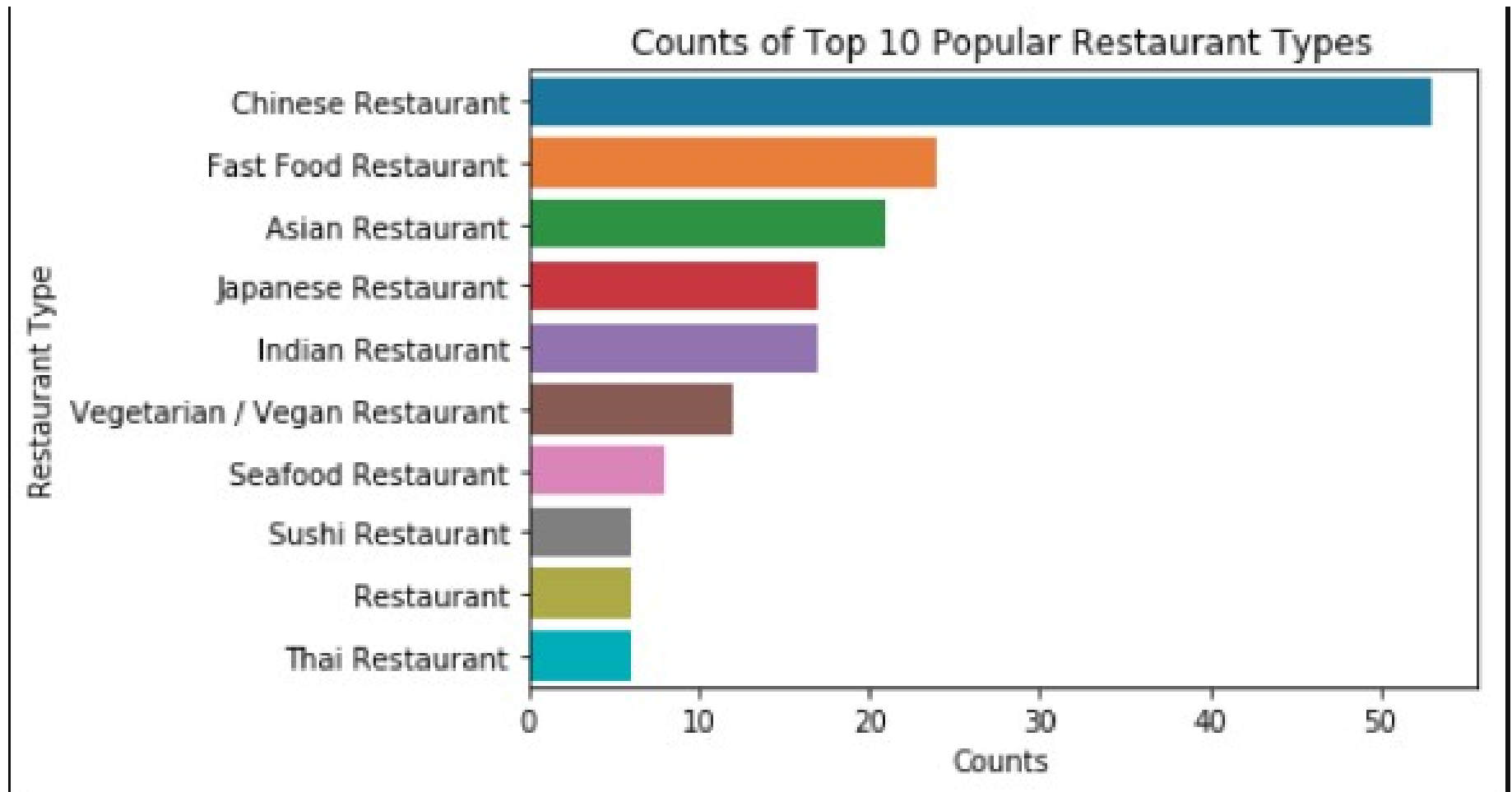
Data Sources

- List of Singapore Planning Areas, which are referred to as Towns in this study
- List of MRT Stations with the Towns they are located in.
- Latitude and Longitude coordinates of MRT stations

Methodology

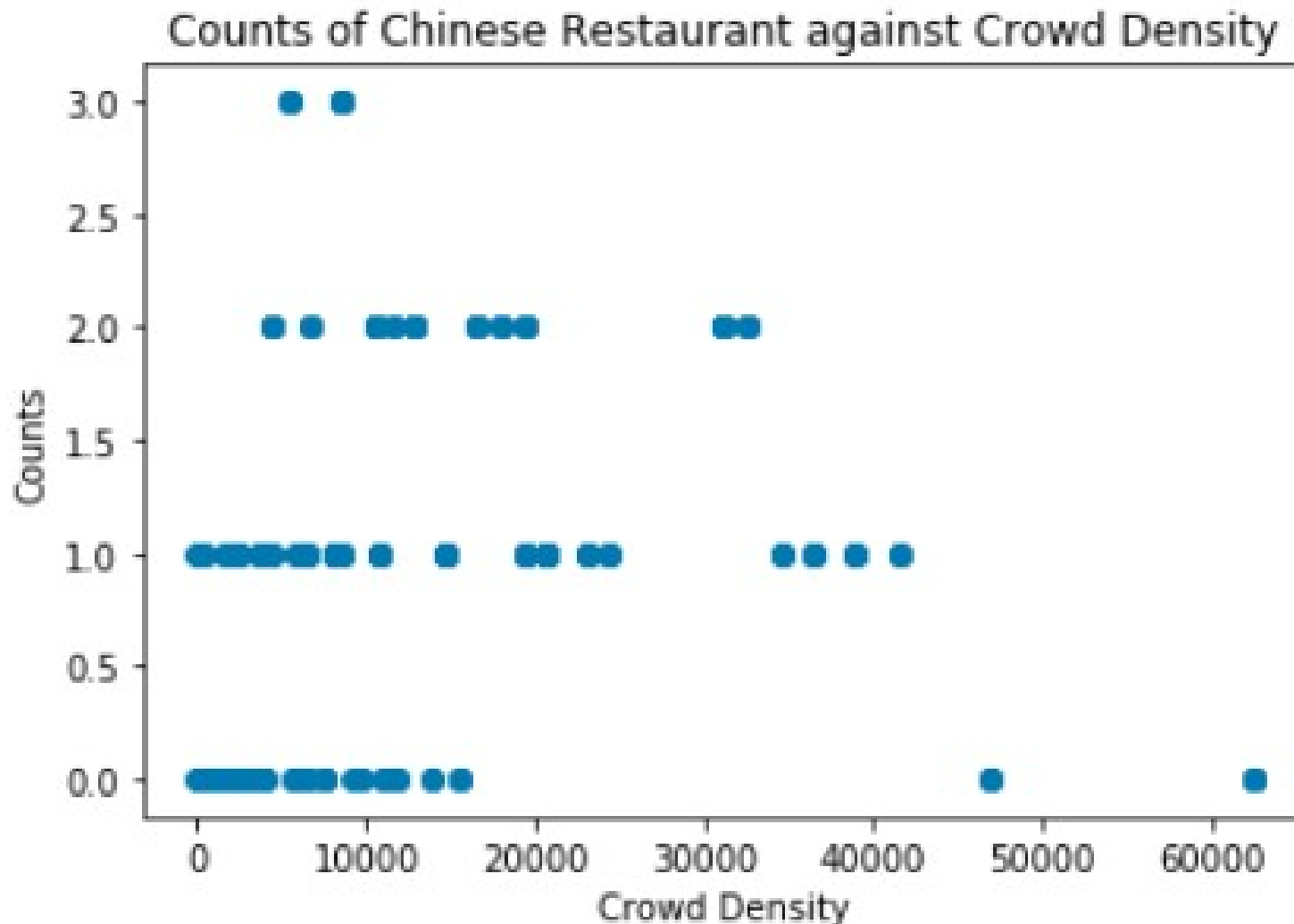
- Data Acquisition
 - List of MRT stations and Towns are obtained by web scraping from Wikipedia,
 - Coordinates of MRT Stations are obtained from OneMapSG website using a python wrapper for client OneMapSG API.
- Data Wrangling
 - To obtain crowd density, which is population density of town divided by number of stations in the town
 - To create dataframes suitable for performing exploratory analysis and K-means clustering
- Exploratory Analysis
 - to investigate relationship between crowd density and counts of restaurant types
- Perform K-means clustering
- Data visualization using bar plots, scatter and plots, folium map

Top 10 Common Restaurant Types



Scatter Plot Analysis

- Scatter plot is used to explore the relationship between crowd density and counts of Chinese Restaurant



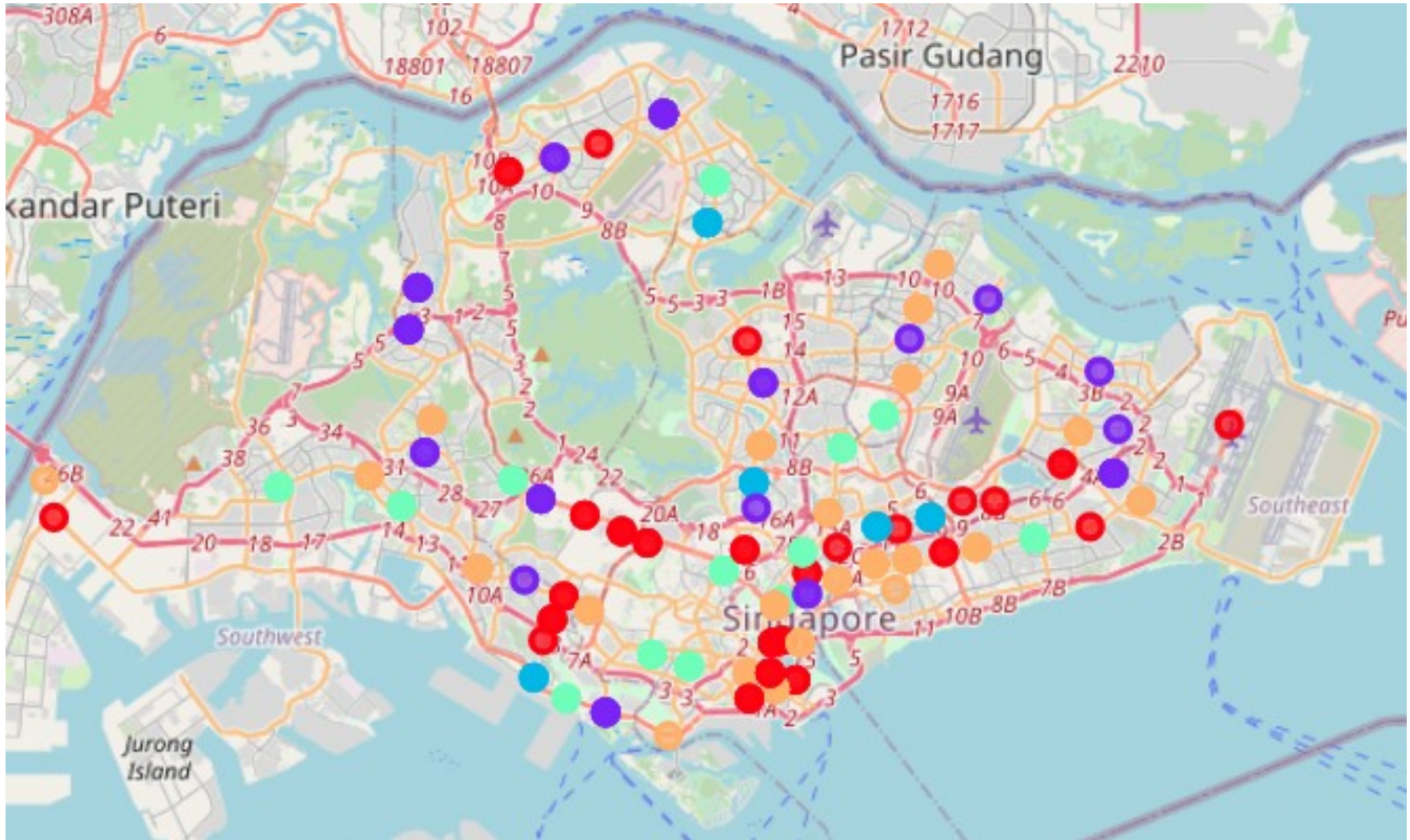
Pearson Correlation

- Pearson p-values are calculated correlating counts of each of top 10 most common restaurant types to crowd density

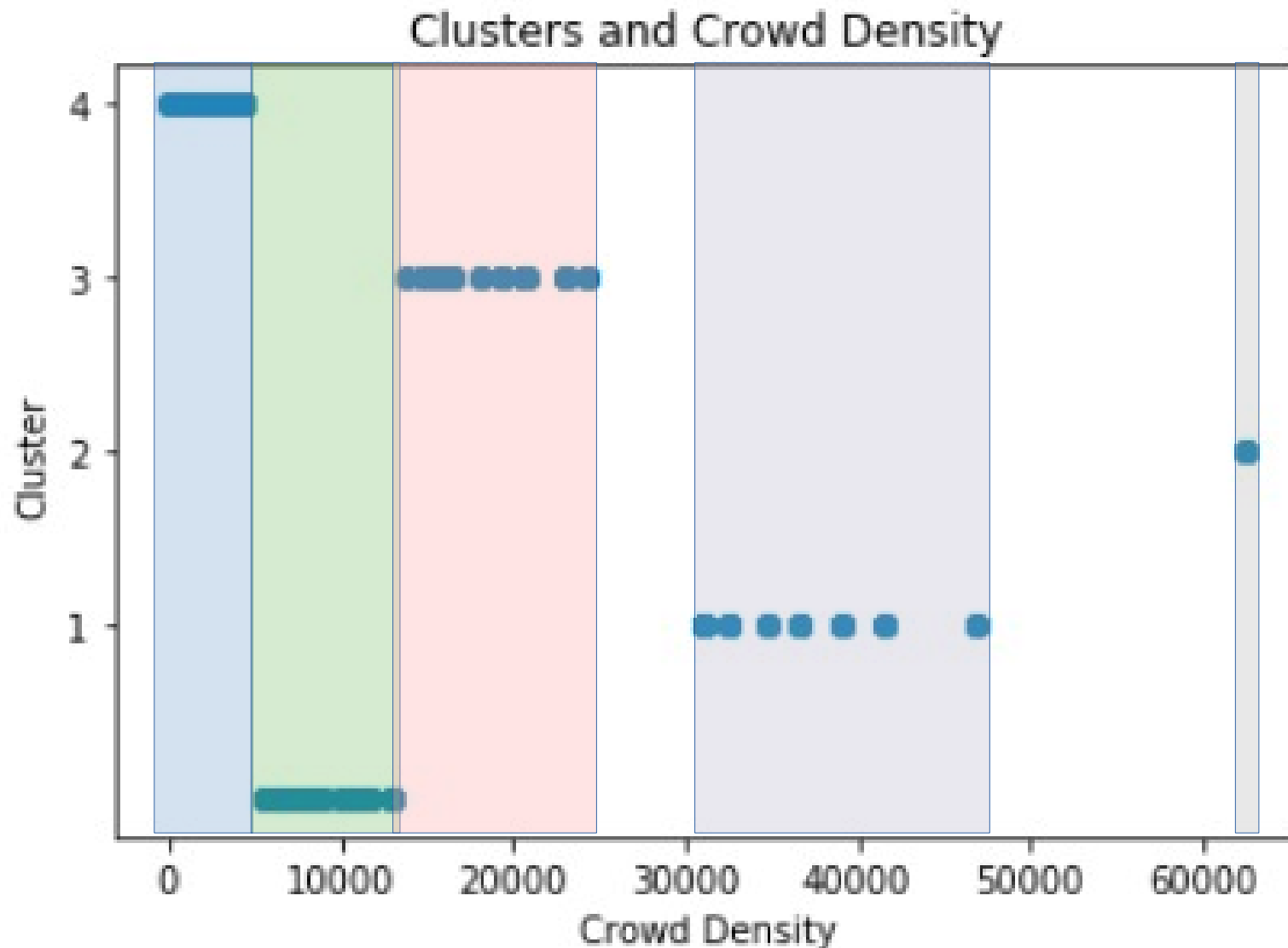
	Restaurant Type	p_value
29	Thai Restaurant	0.360426
28	Sushi Restaurant	0.291626
6	Fast Food Restaurant	0.271617
2	Chinese Restaurant	0.181349
25	Seafood Restaurant	0.056789
1	Asian Restaurant	0.011385
24	Restaurant	0.001572
13	Indian Restaurant	-0.000178
30	Vegetarian / Vegan Restaurant	-0.078464
15	Japanese Restaurant	-0.105755

- P-values range above -0.5 and below 0.5.
- No linear relationship between crowd density and counts**

Map with K-Means Clustering



K-means Cluster and Crowd Density



Discussions

- The MRT stations are clustered according to Crowd Density
- Chinese Restaurant is the 1st most common restaurant type for all clusters except Cluster 3.
- Vietnamese Restaurant appears as 2nd most common restaurant type in Cluster 1 and 5 although it is not among the overall top 10 most common restaurant types. This suggests an emerging market for vietnamese cuisine.
- Data for Crowd Density may not be truly reflective as it's based on residential population. Crowd Density for central business districts and shopping district may be understated.
- Certain Restaurant Type data obtained from Foursquare API are unclear and potentially duplicative. For e.g. "Restaurant", "Dim Sum Restaurant", "Hong Kong Cuisine Restaurant".



Conclusion

- Information from the internet, get coordinates for every MRT Station, and used Foursquare API to get venues surrounding every MRT Station.
- Data is wrangled, correctly formatted, and normalised before further data analysis was done. Exploratory analysis and visualisations are done to gain a better understanding of the data.
- Finally, machine learning algorithms are used to cluster data. Result suggested clustering to be done in accordance to Crowd Density.
- The clustering results was discussed to illustrate how restaurateurs could decide the type of restaurants to open and where to open.