Capstone Project
The Battle of Neighbourhoods

Yiteng Zhang

August 2020

Problem Statement

The Bay Area has a lot of restaurants already with different cuisines.

Starting a new restaurant without a proper business justification is surely not ideal.

This capstone tries to use data science technique to analyse the restaurants and in particular the cuisine of the restaurant near the Bay Area.

Find out the best possible place/neighbourhood to open a new business.

Objectives

The main objective is to analyse the restaurant chains in the Bay Area near the technology companies and to look for the areas where a successful restaurant business can be started.

The second objective is to pay special attention to the restaurant cuisine. It aims to analyse the restaurants and in particular the cuisine of the restaurant near the Bay Area and to find out the best possible neighbourhood for opening a new business.

Data Acquisition

▶ The Bay Area companies list is an open source list of San Francisco Bay Area. According to the author, Connor Leech, the focus in creating this list is to make a resource for candidates to find tech jobs in the Bay Area. The .csv file has been uploaded to Github and will be used for the analysis.

data.head()

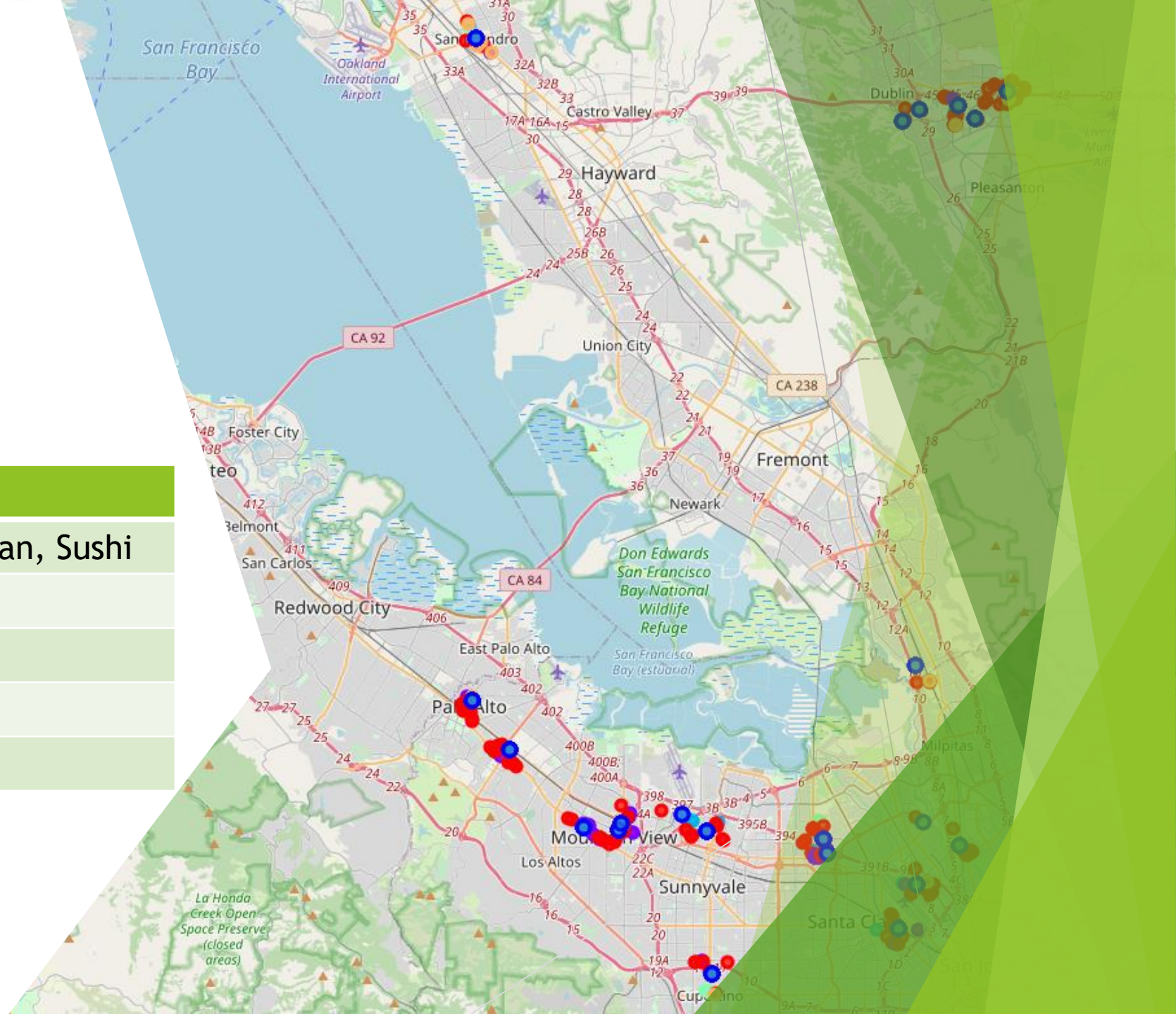|	Company Name	Tags	Location	Investors	Description	Website	Foun Y
0	3scale	B2B Software and Services,Developer Tool	San Francisco	NaN	Unlock the power of your APIs (owned by RedHat)	https://www.3scale.net/	20
1	8tracks	Music,Consumer Goods and Services	San Francisco	NaN	8tracks is the best place to make and discover...	https://8tracks.com/	20
2	10 by 10	B2B Software and Services,Talent Acquisition	San Francisco	Y Combinator	We help independent recruiters and agencies ea...	https://www.10by10.io/	20
3	15Five	Employee Engagement	San Francisco	NaN	15Five software elevates the performance and e...	15five.com	20
4	21Tech	Real Estate and Construction,Financial Technol...	East Bay	NaN	Enterprise Asset Management: 21Tech has broad ...	21tech.com	19

Data Methodology

- ▶ The data/list is extracted from Github.

- ▶ The data consists 750+ rows. The first 200 rows were extracted from the data source. • Feature selection and cleasning were carried out.

- ▶ Map for all the offices in the Bay Area was plotted.

- ▶ Foursquare API was used to find the restaurants near the offices.

- ▶ The restaurants were classified using k-means algorithm.

- ▶ The plot of all the classified restaurants and the offices was made on the map of the Bay Area.

Clustering

▶ Five clusters have been constructed using k-means algorithm.

Cluster	Colour	Cuisine
0	Red	Japanese, Indian, Sushi
1	Purple	Mexican
2	Light Blue	American
3	Green	Chinese
4	Orange	Vietnamese

Observations

- In East Bay area near Art.com, there is only one restaurant nearby. Several other restaurants in Cluster 0 are a bit far away. So opening a new restaurant as long as a Mexican restaurant shall be a good idea.

- In Chinatown near Okaland City Center, there are quite many Chinese restaurants and few Far East restaurants. This is a good at least. As other cuisines might not be successful in Chinatown area.

- Around Oakland City Center, there are no Mexican restaurants.

- In the San Lendro region, there is no American restaurant, so opening one there might lead to a good business.

- In San Ramon Marketplace Shopping Center, there is no Mexican restaurant in their food court.

Conclusion

The capstone project documents how to use data technique to analyse the restaurants and in particular the cuisine of the restaurant near the Bay Area to find out the best possible place for opening a new business. The restaurants have been clustered into five groups and the cluster map can offer insights to advise where might be a good place to start a restaurant business in the Bay Area.