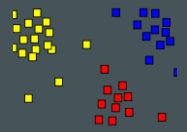
APPLIED DATA SCIENCE CAPSTONE

JATIN SELMOKAR

BUSINESS PROBLEM

A business client wishes to open a restaurant in the San Francisco area and has reached out for recommendations



Our objective is to build a clustering model on San Francisco neighborhood data to gain insights on the restaurant businesses and recommend best locations to the client

DATA

We will need

- San Francisco neighborhood data
- Geo-Coordinates for neighborhoods
- Surrounding restaurant venues data for each neighborhood

METHODOLOGY

Extract neighborhood data using Beautiful Soup & Requests python packages

Populate geo-coordinates for neighborhood using Google Maps API

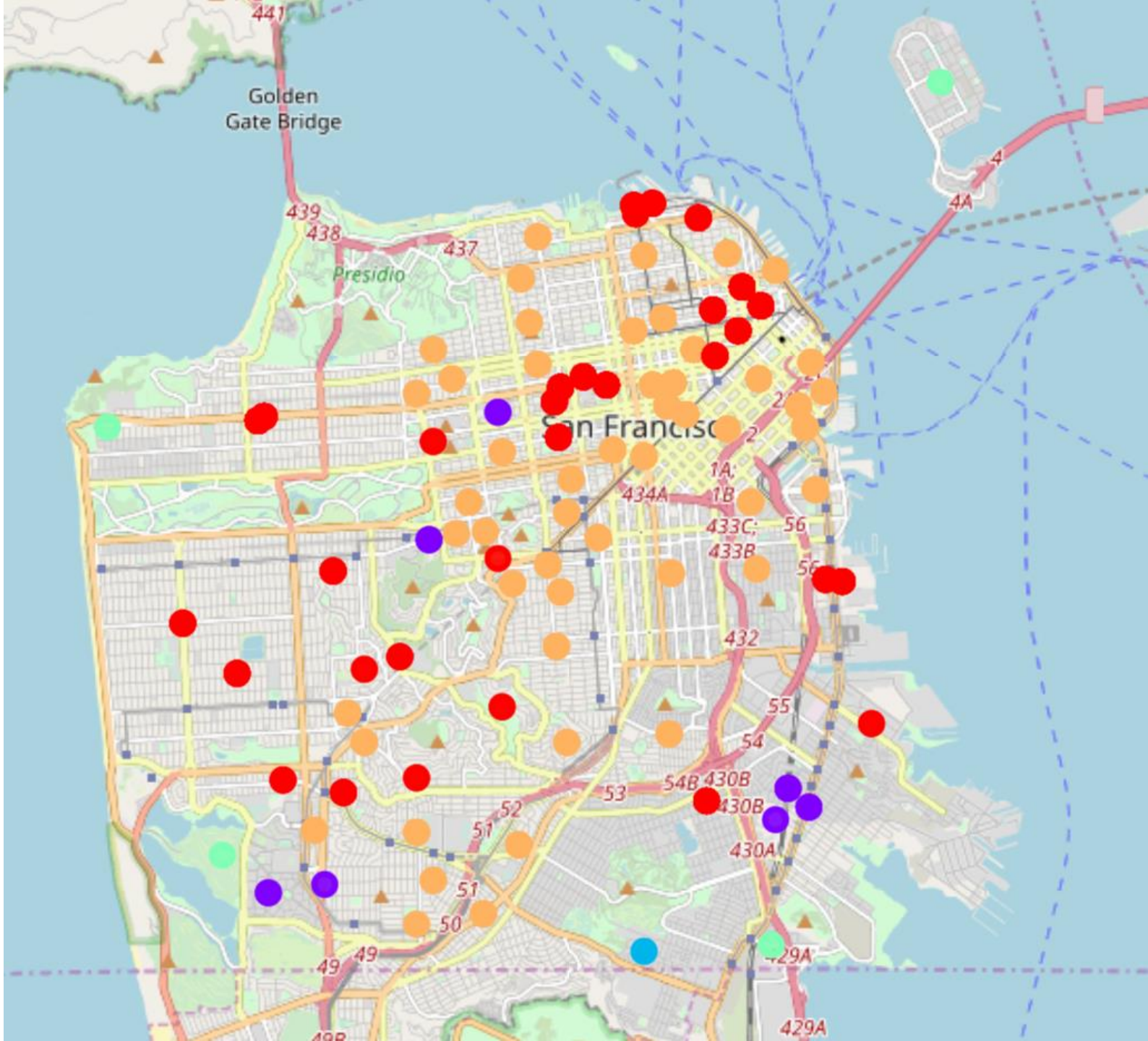
Get restaurant venues using Foursquare API

Use KMeans clustering to get clusters

Define each cluster

RESULTS

- Cluster 1 (Red) – Japanese, Sushi, and Chinese restaurants
- Cluster 2 (Purple) – Mexican & Southern/Soul restaurants
- Cluster 3 (Blue) – Asian restaurants
- Cluster 4 (Green) – American restaurants
- Cluster 5 (Orange) – Vietnamese, Italian, and New American restaurants



DISCUSSION

- Majority of the neighborhoods fall in the Orange cluster where Vietnamese, Italian, and New American are the popular cuisines
- Wide variety of cuisines available throughout San Francisco
- Asian food is popular in cluster 3 (Sunnydale area)

CONCLUSION

- Based on client's choice of cuisine, clusters can be studied to recommend locations
- For instance, if the business owner had to open an Indian restaurant, neighborhoods in clusters 2,3, and 4 would be the ideal locations