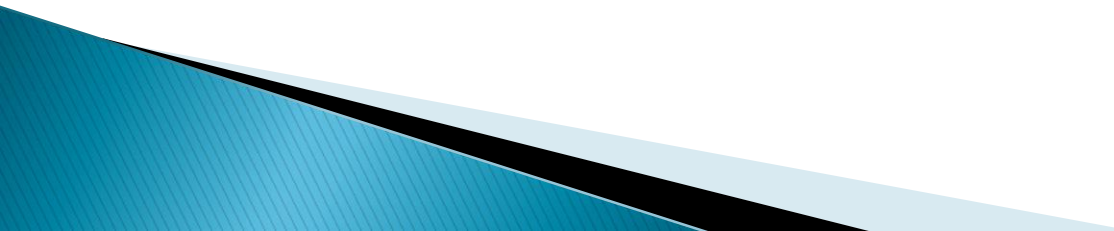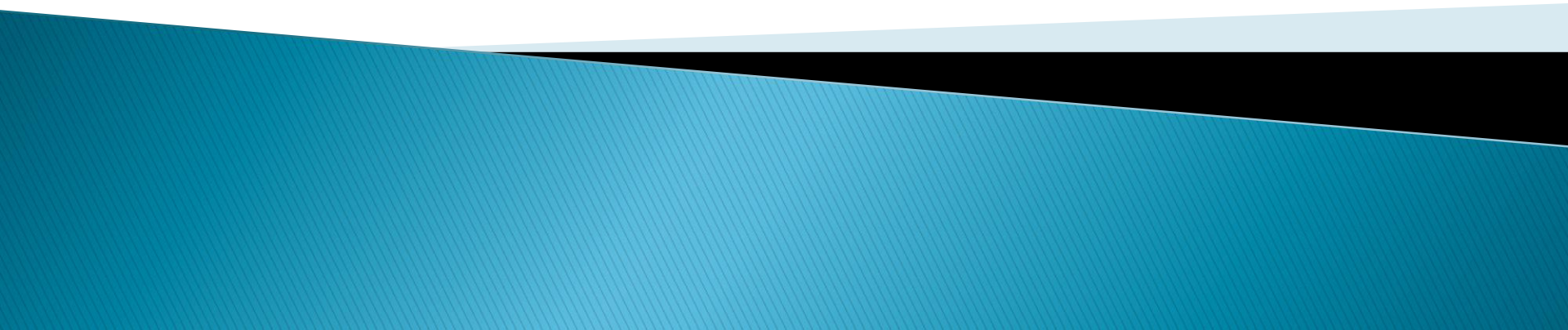# Capstone Project – The Battle of Neighborhoods

Salvador Roca

INTRODUCTION

- Background: fast food business in Sacramento expanding to San Francisco

- Problem: where should we establish the first restaurant?

- Interest: find zones with as less competence as possible, with highest population density
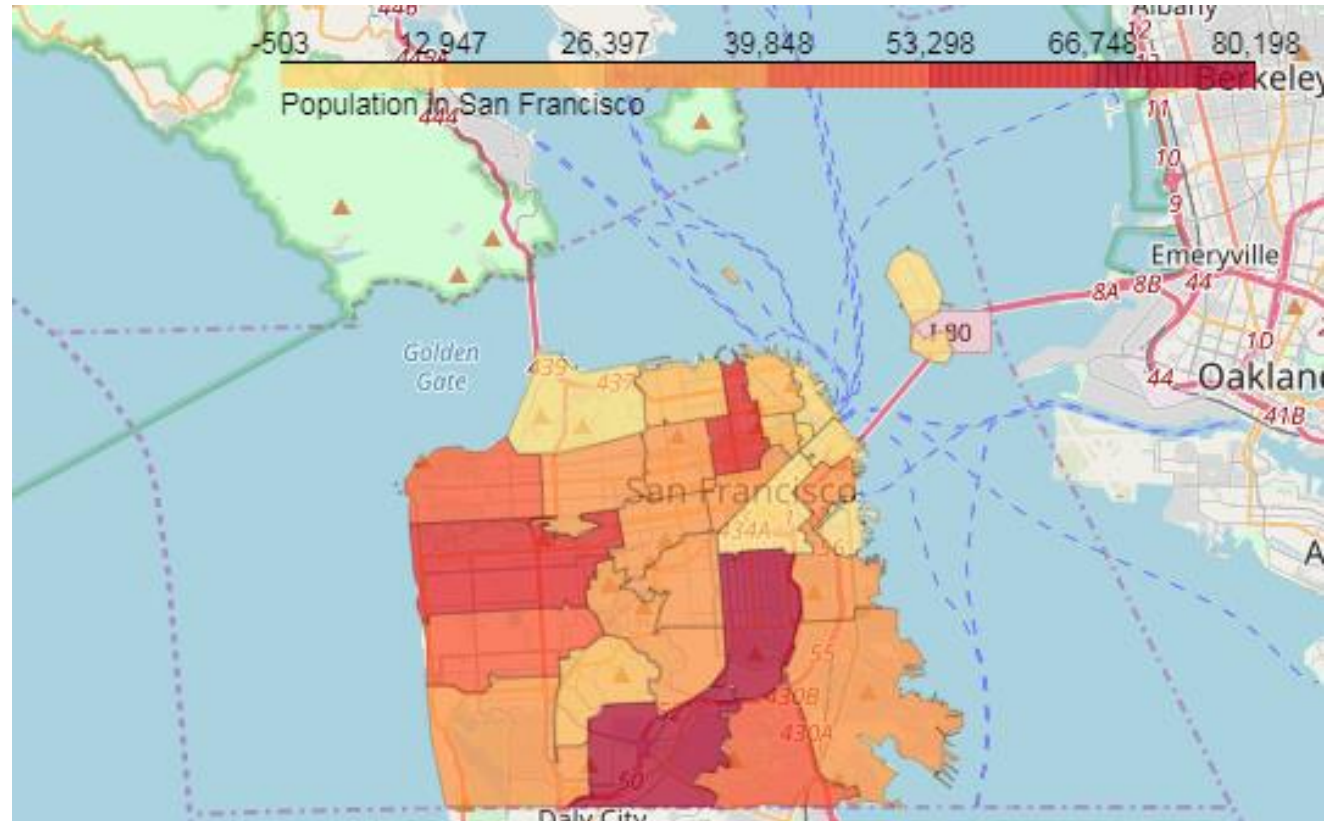
DATA ACQUISITION AND CLEANING

▸ San Francisco geographical data → ZIP codes, Choropleth maps

▸ Population by ZIP code

▸ Data of all different businesses in each zone
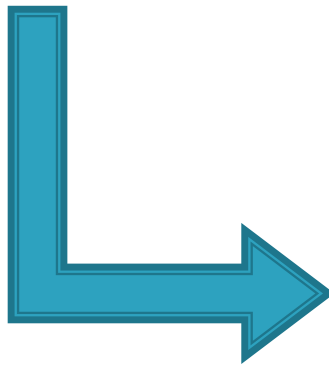
FOURSQUARE

EXPLORATORY DATA ANALYSIS

ZIP Code	Population
94112	79407
94110	69333
94122	56023
94109	55984
94116	43698

ANALYSIS: K-MEANS CLUSTERING

▸ 5 different clusters, fitting all venues retrieved by calling Foursquare database in terms:
  ◦ Number of restaurants
  ◦ Population density

▸ Summary:

Cluster	Restaurants Count mean	Population mean	Ratio
1	20.125	26565.375	1320.018
2	19.000	4339.125	228.375
3	28.000	56003.500	2000.125
4	23.000	38600.571	1678.286
5	24.000	74370.000	3098.750

DISCUSSION AND CONCLUSION

▸ Best option: <u>cluster 5</u> → highest Population vs Restaurants ratio.

▸ Cluster with highest population density, but not highest in restaurant number.

▸ First restaurant will open in ZIP code 94110 or 94112.

ZIP Code	Restaurants Count	Population	Cluster Labels
94110	23	69333	4
94112	25	79407	4

THANK YOU FOR REVIEWING MY ASSIGNMENT!

SALVADOR ROCA