

The Battle of Neighbourhoods

Yiteng Zhang

August 2020

1 Introduction

The San Francisco Bay Area, popularly referred to as the Bay Area or simply the Bay, is a populous region surrounding the San Francisco, San Pablo, and Suisun Bay estuaries in Northern California.

The Bay Area has the second-most Fortune 500 companies in the United States, after the New York metropolitan area, and is known for its natural beauty, liberal politics, entrepreneurship, and diversity.

Among the Fortune 500 companies located in the region include technology companies Google, Facebook, Apple Inc., Hewlett Packard, Intel, Adobe Inc., Applied Materials, eBay, Cisco Systems, Symantec, Oracle, Netflix, Sony Interactive Entertainment, Electronic Arts, and Salesforce; energy companies Chevron and PG&E; financial service companies Charles Schwab Corporation, Visa Inc., and Wells Fargo; apparel retailers Gap Inc., Levi Strauss & Co., and Ross Stores; aerospace and defense contractor Lockheed Martin; local grocer Safeway; pharmaceutical company McKesson; and biotechnology companies Genentech and Gilead Sciences. The largest manufacturers include Tesla Inc., Lam Research, Bayer, Chevron, and Coca-Cola. Oakland is the site of the fifth-largest container shipping port in the United States and is also a major rail terminus. In research, NASA's Ames Research Center and the federal research facility Lawrence Livermore National Laboratory are based in Mountain View and Livermore respectively. In the North Bay, Napa and Sonoma counties are known for their famous wineries, including Fantesca Estate & Winery, Domaine Chandon California, and D'Agostini Winery.

1.1 Problem Statement

Nearly all those technology companies have headquarters located in the Bay Area. During the lunch time, most of the staffs will either grab a quick lunch in the nearby restaurants or have a lunch at company's canteen. It turns out there are already many restaurant chains in the valley and near the offices of almost every technology company. However, this still attracts some adventurers to start a new restaurant business in the Bay Area. This capstone tries to use data science technique to analyse the restaurants and in particular the cuisine

of the restaurant near the Bay Area to find out the best possible neighbourhood for opening a new business.

1.2 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.

1.3 Beneficiaries

This capstone is beneficial to anyone who is looking to start a new restaurant business in the Bay Area. The restaurants already in operations can utilise the analysis to adjust their business prospects also.

2 Data Integration

The data requirements of this capstone include:

- List of all the technology companies;
- The coordinates of the offices;
- The list of all restaurants near the offices using Foursquare API.

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.

2.1 Methodology

This subsection explains how the data will be used to solve the problem. Some highlights are given here:

- 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 cleansing 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.