Restaurant in San Francisco REPORT

S E L E C T I N G T H E S A F E S T A R E A F O R A R E S TAU R A N T I N S A N F R A N C I S C O

INTRODUCTION

The purpose of this project is to find a safe placefor restaurants to open in San Francisco, United States.

First I have to choose the safest neighborhood by analyzing the crime data for the opening of the restaurant and a list of a neighborhood, where the restaurants are located as close as possible to the city of San Francisco, but in any case in unusual places.

I will use data science tools to analyze the data and focus on the safest district and explore its neighborhoods and the most common locations in each neighborhood so that I can select the best neighborhood where the restaurant is not among the most common places.

WHICH DATA? WHICH TECH_TOOL?

I need very well correlated data to the scope of this project. Here follow the best sourced: - KAGGLE, specifically all the crimes of any kind are described one-by-one from 2007 till 2015 - WIKIPEDIA, have helped me to identify the neighborhoods and the districts - OPENCAGE GEOCODER joined with the power of Folium helped me to explor the gathered data - FOURSQUARE API helped me to analyze the neighborhoods locations plotting them on maps

SAN FRANCISCO CRIMES

The real data from Kaggle traced as followed:

Dates - timestamp of the crime incident

Category - category of the crime incident (only in train.csv).

Descript - detailed description of the crime incident (only in train.csv)

DayOfWeek - the day of the week PdDistrict -name of the Police

Department District Resolution - how the crime incident was resolved (only in train.csv)

Address - the approximate street address of the crime incident

X - Longitude

Y – Latitude

METHODOLOGY

Categorized the methodology section into two parts:

Exploratory Data Analysis: Visualise the crime repots in different Vancouver boroughs to idenity the safest borough and normalise the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.

Modelling: To help stakeholders choose the right neighborhood within a borough we will be clustering similar neighborhoods using K-means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use K-Means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.

DATA ANALYSIS

Neighbourhoods in San Francisco with the Highest crimes
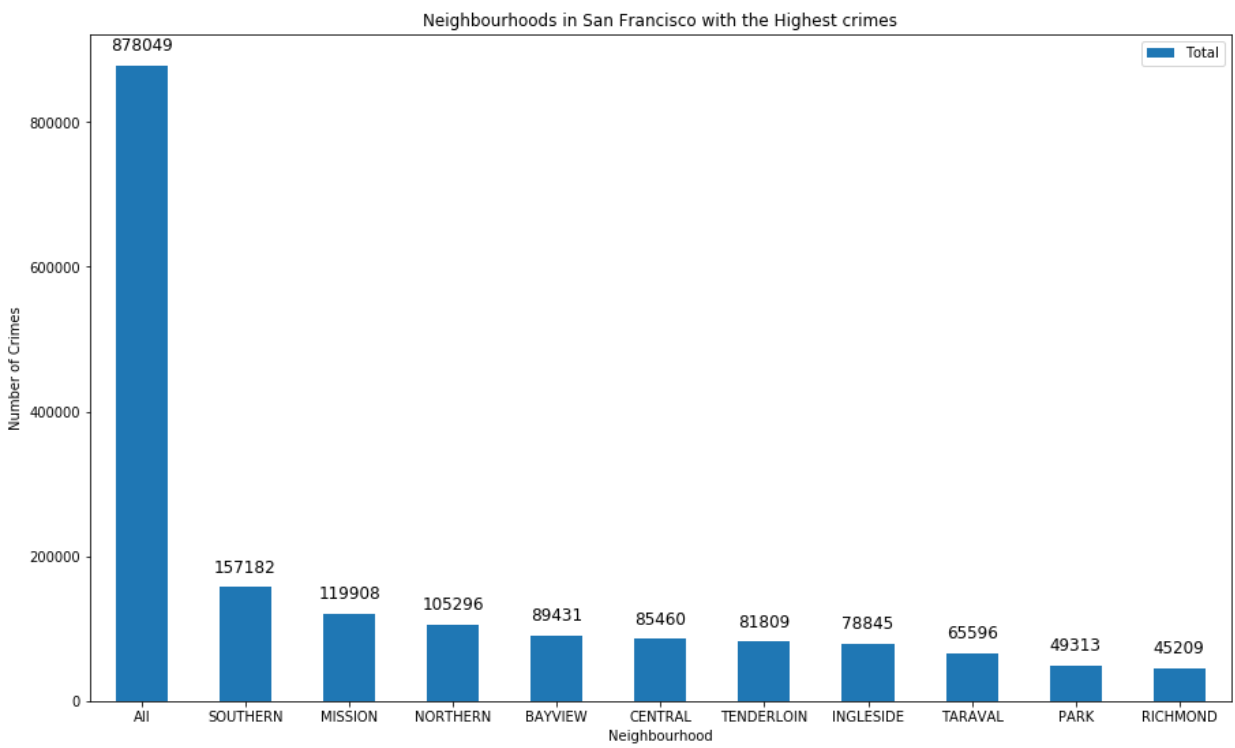
RESULTS

I set myself the goal of identifying the safest neighborhoodin San Francisco where to open a restaurant. To do this, I first collected the data on crime of any kind, large, small, important or not. The goal was to find a safeplace so as to make customers and even restaurant managers safe. I achieved this by grouping the neighborhoods in clusters by providing the potential entrepreneur and manager of the restaurant with places and greater protections to be taken.

CONCLUSIONS

I analyzed the data on the crimes that tookplace in the city of San Francisco to understand what happened. So I divided them by neighborhoodand districtsand grouped in such a way as to identify the safest neighborhoods. I focused on districtswith fewer crimes to choose where to place the restaurant with greater customer safety.