San Francisco Venues VS Crime Incidents

The Battle of Neighborhoods - Coursera Capstone Project

Final Report

Author: Angelo Di Marco

1. Introduction

1.1 Background

San Francisco is one of the main cities of United States of America, located in the so-called San Francisco Bay in the state of California. It is the cultural, commercial and financial center of Northern California and in 2019 it reached 881,549 inhabitants, distributed on a surface of "just" 121.4 square kilometers. The population of San Francisco is strongly multiethnic and heterogeneous. As a result this city is the meeting point of a large variety of cultures which contribute to its peculiar identity in different ways. In particular, this diversity accounts for the presence of many different kind of venues including, e.g., restaurants with kitchens from every part of the world. These venues are spread over the 36 different neighborhoods in which the city can be divided and the category they belong to can depend on their location.

Besides all the interesting, beautiful and "delicious" venues that San Francisco can offer to its inhabitants and especially to tourists from all over the world, this city is also, unfortunately, the scene of many criminal events. Every year, indeed, several thousands of crime incidents are reported by the Police such as, for instance, murders, rapes, robberies, assaults, and gangs fights. Every neighborhood of the city is affected by these unpleasant happenings.

1.2 Business Problem

In this context, understanding how type of venues and type of crimes are spatially distributed in San Francisco and finding out how they are correlated may have a relevant impact in the economy and safety of this American city. This might be a key point for every entrepreneur who wants to start an economical activity in the city and for the Police Department in reaching a solution for the criminality problem.

In this Data Science project, we want to use machine learning techniques to determine (1) how the different venues and crimes are located in San Francisco neighborhoods and (2) what is the correlation between them. The ultimate questions we want to answer are: Which location is more convenient for a new business activity? Did a particular crime happen in a certain location because there are certain types of venues around there? Given a crime with no "certain" location, is it possible to predict where it occurred based on the venues spatial distribution?

1.3 Data Description

To answer the questions listed above we will use several different datasets. First of all we will access to all the venues in San Francisco available at the locations platform called Foursquare. By means of API calls to foursquare.com, we will have access to datasets containing name, category, latitude and longitude of the venues within a certain radius from a specified location. We will choose the latter to be the "center" of one of the 36 neighborhoods in which the city can be divided. Each neighborhood is listed together with the coordinates of its "center" in a dataset created by the author of this report with the help of Google Maps. Finally the "Police Department

Incident Reports" will provide us the data about the criminal events reported in San Francisco from 2018 to present. Specifically, this dataset contains information like the type of incident (category, subcategory and description), the name of the neighborhood where the crime happened as well as the exact point on the map (latitude and longitude). Eventually more datasets will be included in the analysis to achieve the goals of this project.