

Capstone Project - The Battle of Neighborhoods

Table of Contents

1. Introduction	1
1.1 Overview and Analytical Approach	1
2. Data Description	1
2.1 Data Requirements.....	1
2.2 Data Collection.....	2
2.3 Data Understanding and Preparation	2

1. Introduction

1.1 Overview and Analytical Approach

In a city like San Francisco with population of 3.3 million¹ people and home to Silicon Valley, home evictions are getting very common. San Francisco has many areas, including the Bay Area, which are the among the most densely populated areas in United States. In these difficult times, through my project I wanted to explore and analyze the best neighborhoods for renting a house in SF based on the probability of not getting evicted.

As mentioned, the population density in San Francisco makes is very challenging to rent a house in a good neighborhood. Understanding the details of a good neighborhood is key to having a peaceful shelter, especially when the occurrences of home evictions carried out by landlords are increasing. One of the primary reason for increased evictions can be factored to easy availability of tenants, willing to pay high rents. In a real estate market where buyer power is less than seller power, selecting the right neighborhood, property and landlord will reduce the chances of becoming homeless.

Analyzing the above discussed scenarios by creating density maps and information tables of the past evictions in San Francisco will provide good insights into the underlying neighborhoods. By segmenting and clustering the neighborhoods based on the data, one can select the best fit neighborhood to rent a house/apartment in San Francisco.

2. Data Description

2.1 Data Requirements

¹ https://en.wikipedia.org/wiki/San_Francisco

For data requirements, I identified websites that can provide me details of San Francisco in a specified format (CSV, Excel, .json etc.). The data should include details of the neighborhoods and boroughs in San Francisco, CA. Also, optional data regarding postal/zip code will also be useful for analyzing.

2.2 Data Collection

In the data collection stage, I found the details of websites or articles that provide details of San Francisco real estate. The data I am particularly interested was the renter's marketplace.

I have used the following data sets to analyze the problem statement discussed in the above section:

- From data.world² > Make Over Monday data sets, I downloaded the CSV file for the San Francisco Eviction Notices. During the data cleansing, I grouped the neighborhoods to Supervisor Districts³ and mapped the probability of an eviction in San Francisco using a choropleth map.
- I used Foursquare API to analyze the most common venues around the San Francisco county (in San Francisco there is no concept of Borough, instead I used county to segment).
- I did not have to use csv for Geospatial data since the data set included latitude and longitude coordinates corresponding to each data entry.
- I used the JSON file **san-francisco.geojson** to create the choropleth map. This JSON file was used while analyzing the San Francisco crime related data during the data visualization lab⁴.

2.3 Data Understanding and Preparation

In the data understanding stage, I checked if the details provided in the data sets collected are representative of San Francisco tenant market. I prepared data by cleansing and removing duplicates, addressing missing values, and ensuring it is properly formatted.

² <https://data.world/makeovermonday/2019w39>

³ https://en.wikipedia.org/wiki/San_Francisco_Board_of_Supervisors

⁴ <https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/DV0101EN/DV0101EN-3-5-1-Generating-Maps-in-Python-py-v2.0.ipynb?lti=true>