Crime data in San Francisco for businesses and tourists

Augusto González Bonorino

June 14, 2020

Introduction

Often, both entrepreneurs and tourists move to a different city or country seeking better business opportunities or new experiences. It is common that those who move to, or visit, a new location is not aware of the dangerous areas of such new place. Not like those who grew there know. Therefore, those individuals might be driven towards opening a store or visit a new area that looks interesting or shows potential without having any information about the downsides of such area or neighborhood. With this project I have decided to create a tool for such individuals. First, I will be plotting information about crimes in San Francisco divided by neighborhoods in a choropleth map. Second, I will extract from Foursquare.com information about the location of the following types of businesses: Coffee shops, liquor stores, nightclubs, and restaurants. I decided to pick these four types because they are the most common places tourists look for.

Moreover, this map (which potentially could be used by foursquare or a third-party app) also provides the entrepreneur looking to open a new business in San Francisco with the location of most of the businesses around the area. Such information could be used to locate the area with the least amount of businesses similar to his or hers and conduct further research on the potential of such area regarding target audience and several other factors to be determined by the entrepreneur.

Problem

Tourists or entrepreneurs are often victims of robberies or different types of crime, most of which located on historic dangerous areas. In San Francisco, the Northern and Mission neighborhoods are the two areas with the higher crime rates (around twenty thousand cases), followed by the Central neighborhood located downtown (with seventeen thousand cases). Ironically, these are also the neighborhoods where businesses tend to cluster together, driving, in consequence, the most amount of human traffic around those areas. Tourists, in particular, are exposed to such risks for a couple of reasons 1) they often use apps like Foursquare's city guide, yelp or google maps which provides no information about the dangers of the areas the user might be looking at and, 2) tourists are the preferred target of local criminals because they often find them distracted either by the landscapes or taking pictures. In addition, entrepreneurs that move to San Francisco might not be aware of the specific characteristics of each neighborhood and often do not perform research beforehand on such topics. Therefore, tempted by low prices or high human traffic such incoming entrepreneurs might decide to open their business in a dangerous neighborhood exposing themselves and their business to unnecessary risks.

Interest

I strongly believe that providing this kind of information will help businesses and tourists be better prepared before going deciding on the location to open their new business or where to go visit. No app of the kind of Foursquare city guide, yelp, and others now provide information on such issues.

They provide the information of nearby stores or trending places but no background information of the dangers of the neighborhood where such businesses are located. Therefore, the interest of this project is that already established apps, like the ones previously mentions, adopt this simple yet significant upgrade on their platforms to provide better information to their customers, both clients and businesses owners.

Data acquisition

The information regarding crime data of the different neighborhoods in San Francisco was obtained from the following address: https://cocl.us/sanfran_crime_dataset. To plot the boundaries of each neighborhood on the choropleth map we have used information gathered from the following geojson file: https://cocl.us/sanfran_geojson. Finally, to be able to plot the location of the businesses in San Francisco I used the Foursquare places API. I have downloaded information about the name, category and location of all the coffee shops, restaurants, nightclubs and liquor stores around a 10km radius from downtown San Francisco. For better understanding on what Foursquare is I attach the link to the product page Foursquare_places and to its main page Foursquare.com.

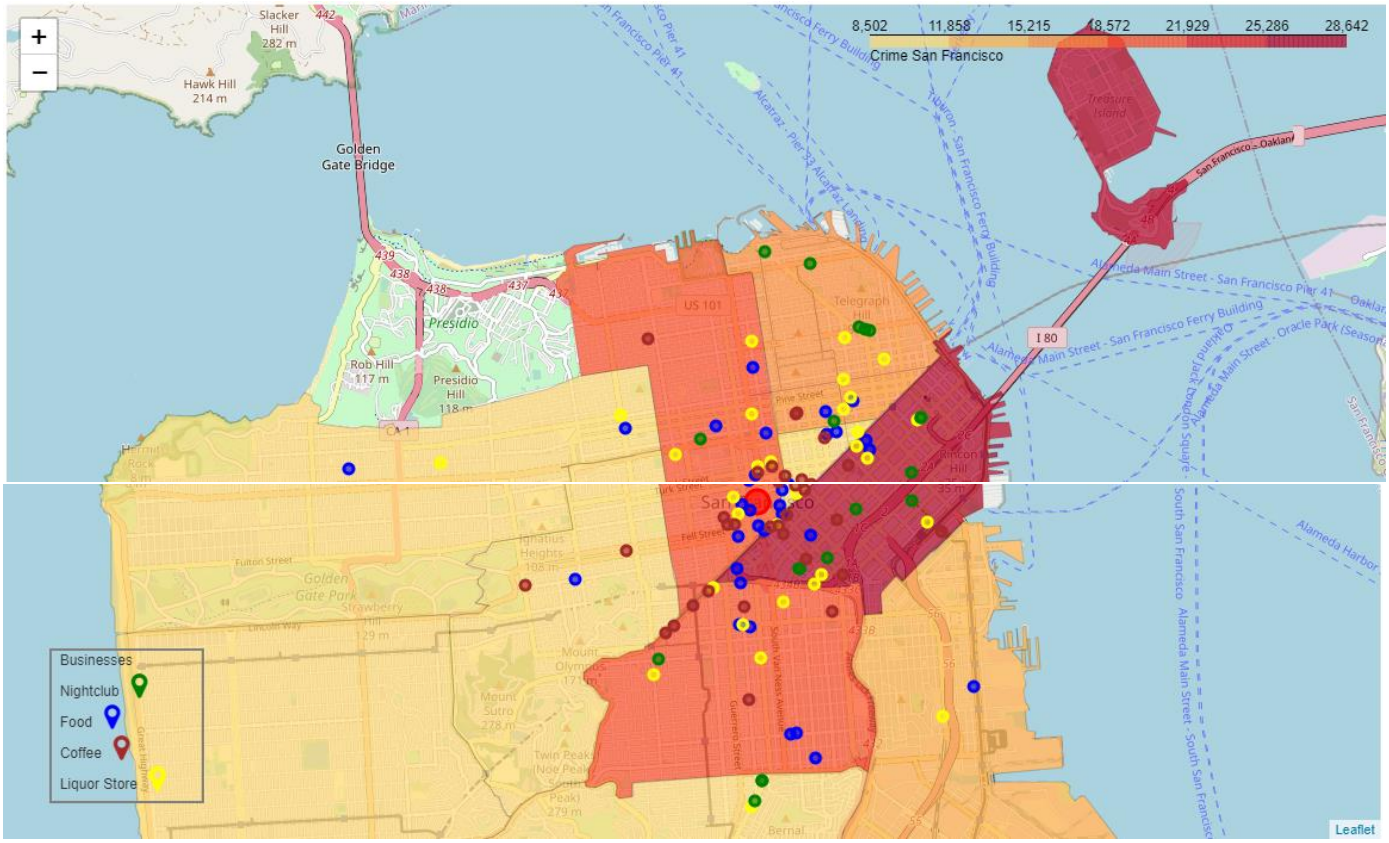
Data cleaning

The process of cleaning the data consisted on three simple steps. First, I imported the csv file containing the San Francisco crime data to the IBM Watson Studio platform to be able to manipulate it using python 3.6. Second, after a short visual analysis of the dataset I dropped the columns that were unnecessary for the project. The following columns from the original data set were dropped: *IncidntNum, Category, Descript, DayOfWeek, Date, Time, Resolution, Address, X, Y, Location, PdId*

It is worth noting that some of the data found under such columns, like *Time* and *DayOfWeek,* might be of future interest to further improve the accuracy of the current project. Third, a new column called 'Count' was generated by grouping the crimes by neighborhoods. This way I was left with the total amount of cases per neighborhood. Such data was used as the scale to plot the choropleth map and assign the proper color tonality to each neighborhood according to the total quantity of cases.

Graph

The graph shown below was generated using the all the data previously mentioned. A choropleth map was chosen because it facilitates the observation of the data and it is a great tool to visualize data such like the one used in this project where we differentiated among the various neighborhoods in San Francisco. The data extracted from Foursquare was plotted on top of the choropleth map to provide a clear visualization of the location of each business and their respective distance from the high crime rate neighborhoods. As it can be observed in the legend, the brown points correspond to the coffee shops, the green dots to nightclubs, the blue ones to restaurants (those under the "Food" category on Foursquare's city guide and the yellow dots correspond to the liquor stores.



Conclusions

There is tremendous potential to this project, which has been simplified for the purposes of meeting the IBM capstone course deadlines and requirements. We have found that, as we thought, most businesses are located inside or near those neighborhoods that present high numbers of crime cases, therefore proving our initial suggestion that tourists were exposed to high risks in the city of San Francisco due to the proximity of the most attended categories of businesses (cafes, restaurants, nightclubs and liquor stores) to dangerous areas of the city. Regarding the benefits for entrepreneur and prospective business owners, we have provided information, in addition to the potential risks each area conveys, of locations where certain businesses might be allowed to thrive. In conclusion, most of the businesses in San Francisco are clustered near high risk neighborhoods which increases the risks for tourists.

Further thoughts on the potential of this data

We believe much more can be added to what it is been currently presented, more detailed data from crimes and businesses as well as geo spatial data to incorporate other cities in the US and maybe even countries in the future. Some of the potential applications of the data presented in this project include the implementation of new policies issued by the department of tourism to protect tourists; development of new apps like yelp, google maps and foursquare city guide, or the improvement of those mentioned, where this information is provided to the user; this information could also be included by businesses such like Airbnb or travel agencies and many other possibilities.