IS445 : Data Visualization

Police Department Incident Reports: San Francisco

Group 7

Shrey Shah, Sanmil Ralkar, Ketaki Gujar, Kairav Pandya, Rohit Doraiswamy

ILLINOIS

Dataset

- The dataset is about the **Police Department Incident Reports: 2018 to Present** in the city of San Francisco.
- It includes incident reports, filed from January 1, 2018 to present.
- Has around 668k rows and 26 columns (Incident Date, Incident Time, Report Datetime, Incident category, Latitude, Longitude, etc)
- Data is updated automatically daily by 10:00 am Pacific.
- We are fetching the data using the SODA API and we do not need any tokens as it is a public dataset.

Dataset Link:
https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

Data Cleaning, Preprocessing & EDA

Data Cleaning & Preprocessing

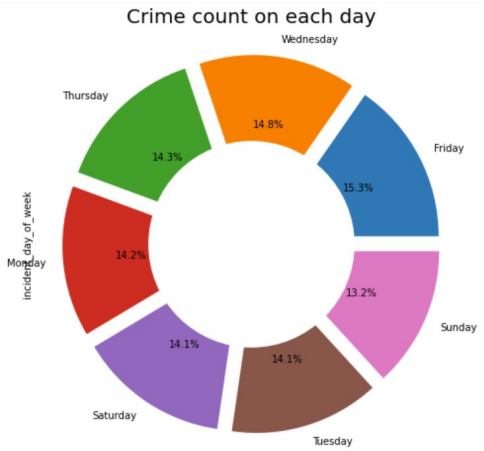
1. Removed columns irrelevant to the dataset and its analysis. Reducing the number of columns also increased the processing speed.

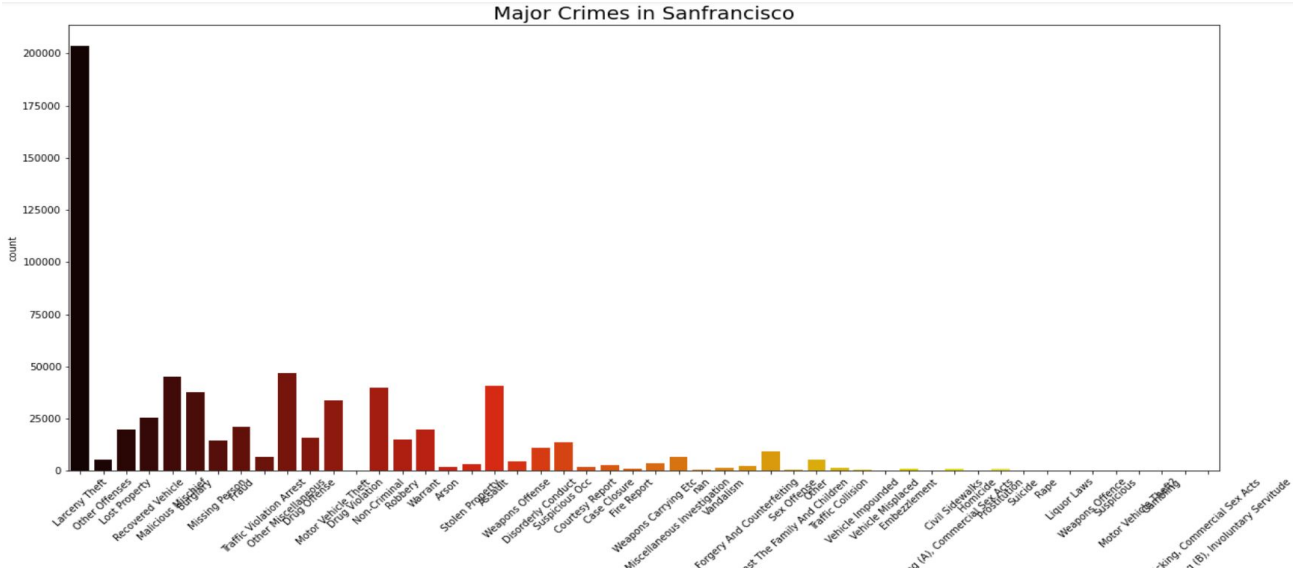
2. Checked for null values in the dataset. Removed rows with null values for some columns and added zero value for other columns to make dataset uniform and clean.

3. Changed the datatypes of columns based on the data in it. All columns were object data type. Changed the columns to datetime, float and string respective to the column data.

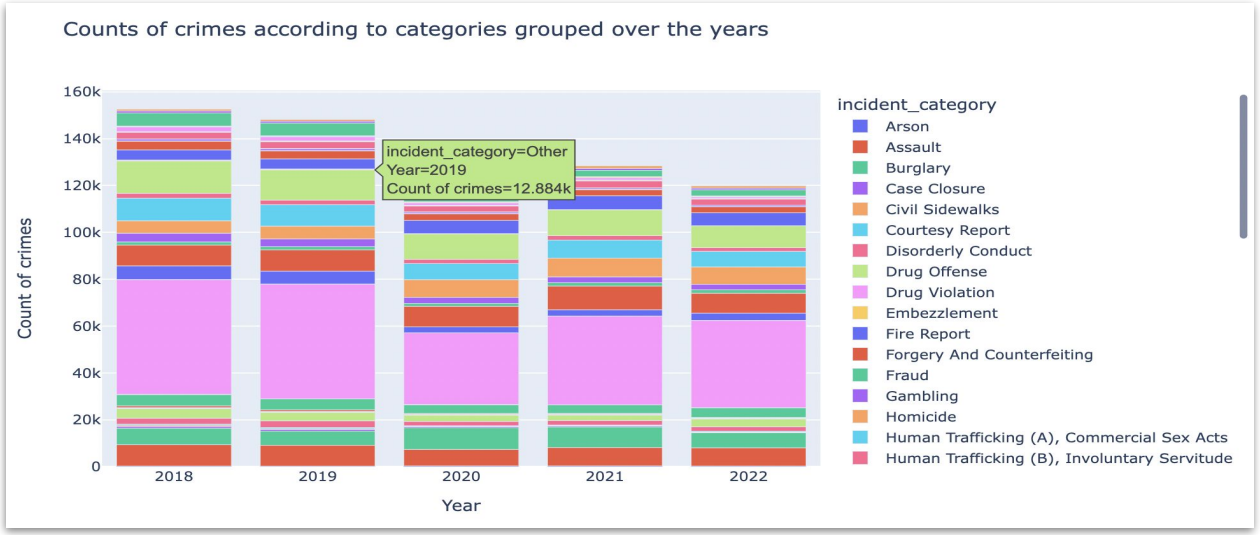
Exploratory Data Analysis

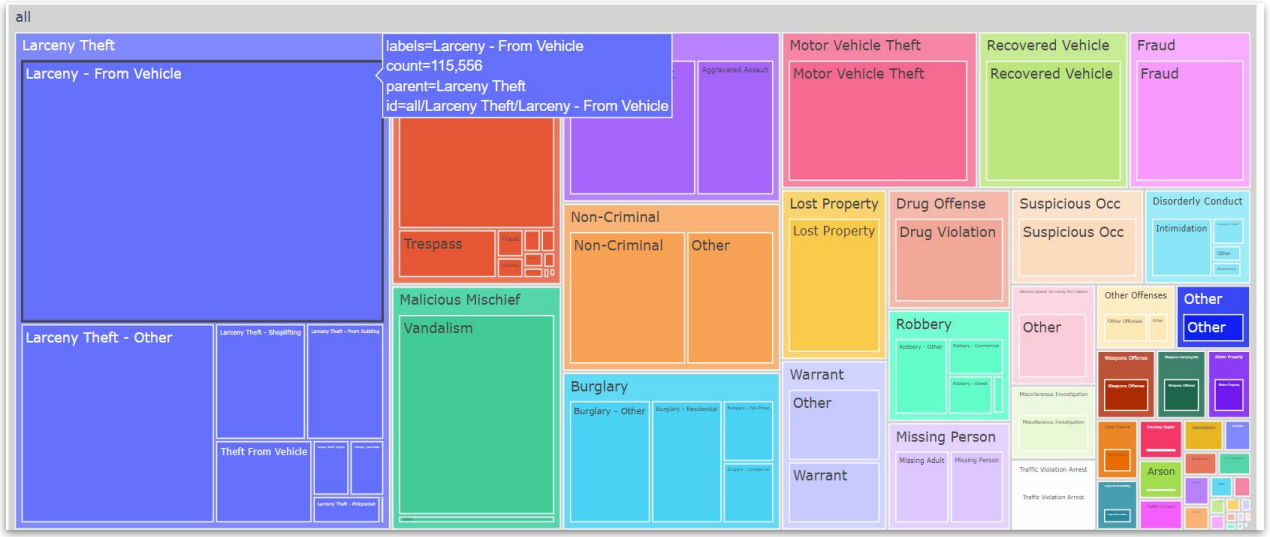
We created a donut pie chart and a bar chart as a part of the exploratory data analysis. The donut pie chart shows crime incidents reported based on the day of the week. The bar chart shows the type of crimes and their count since 2018.





Visualizations and Insights

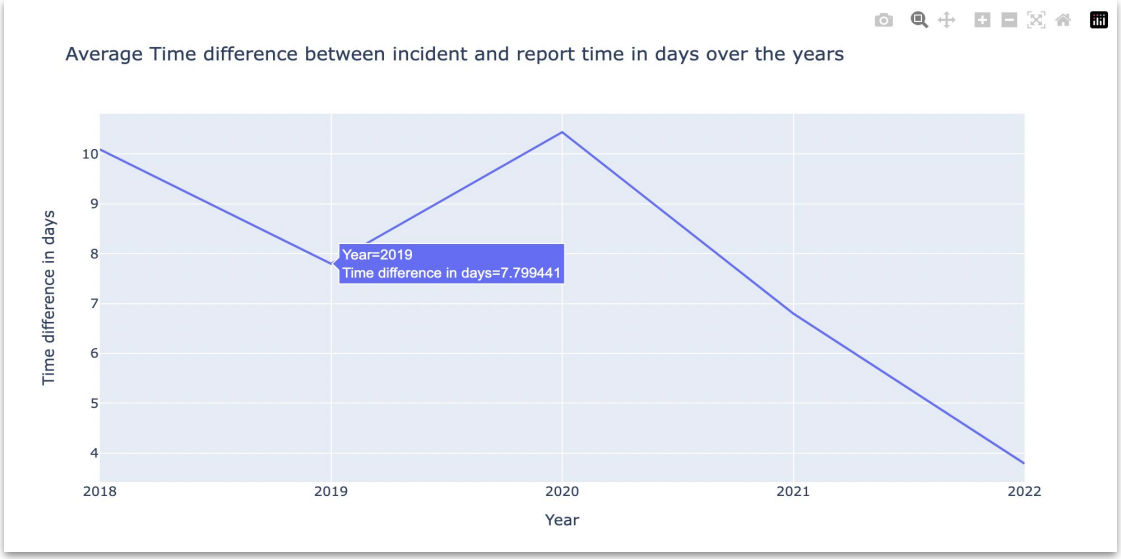




Number of crimes based on category over the years

- Firstly, we checked for the unique crime categories. There were a few which were overlapping and we updated the values to make them consistent.
- We initially plotted a stacked bar chart showing the number of crimes per year with each bar segregated by the incident type/category.
- Since the bars appeared cluttered and we decided to use a tree-map instead for better visualization which adds hierarchy based on category and subcategory.

Visualizations and Insights

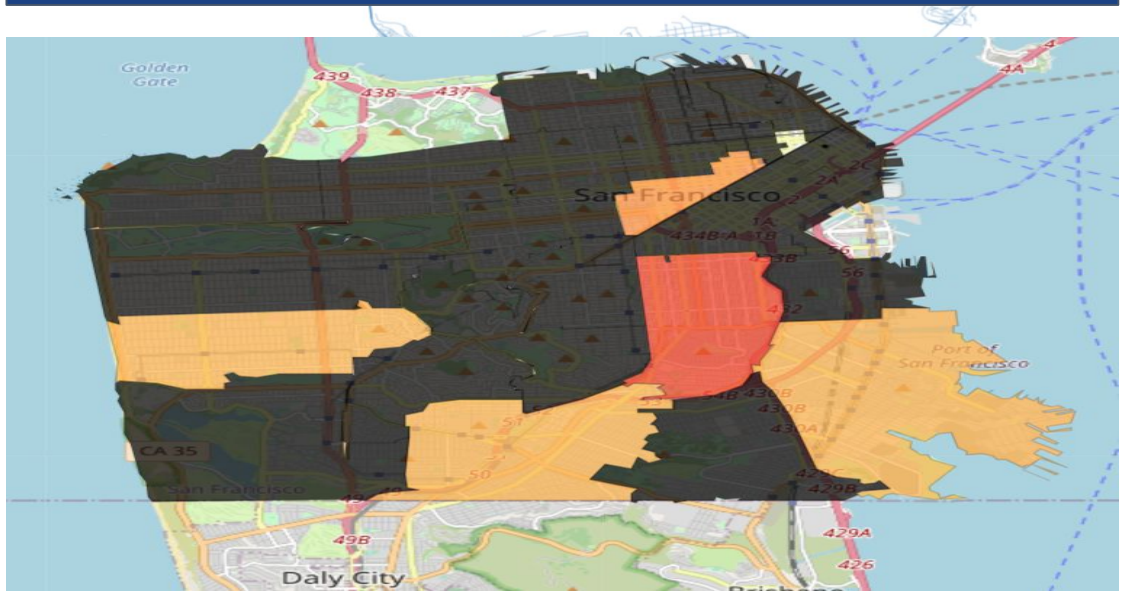


Average time difference between incident & report time over the years.

- Converted incident & report time to datetime datatype.
- Found the difference between them and converted it to show the value in ***number of days.***
- Grouped the time difference by year and found the average.

Crime rates in different regions of the city filtered by incident type

- The heatmap is an indicator of the number of crimes that has occured in each of the main regions of SFO city.
- Darker shade indicates more crimes have occured.
- We see the black region on this map because the data that we had for those regions were not identifiable on the map.



Future Work

- We plan on analyzing the number of incidents as per the time of their occurrence. Time of the incident will be categorized into 4 groups: Early Morning (12.01 am - 6.00 am), Morning (6.01 am - 12.00 pm), Afternoon (12.01 pm - 6.00 pm) and Evening (6.01 pm - 12.00 am) to get insights about the time at which most incidents occur in San Francisco.

- Initially, we have plotted the rough outline of the map of San Francisco using the cartopy and plotted the hotspots of crimes. We have used folium to plot the map with extra features. To improve it further, we will be adding a filter based on the incident category and based on it the hotspots of that specific crime will be plotted on the map.

- Create more relevant visualizations to answer the questions that we have on hand.

Thank You