Crime Analysis in the City of San Francisco

Data Visualization Project

Group 8:
Maddie Bowman, Jinlu Wang, Brian Lee, Justin Bein, Sakurako Kikuchi

Project Proposal

Following the 2020 pandemic, San Francisco has seen a steady rise in crime throughout the city.

We proposed a project to analyze crime patterns within San Francisco neighborhoods, prior to, and following the pandemic. Through the data cleaning process, we analyzed the ‘safest’ and ‘least safe’ neighborhoods within San Francisco, recording common crimes committed throughout the city.

Our goal was to create a city map that toggles through available crime data (2018 - current), containing a base layer of markers specified by incident-type, and a drop down allowing you to display markers for the various crime categories. The remaining visualizations further explain these data trends. The Flask App will host the outlined visualizations and analysis.

Research Questions

1. Which are the safest neighborhoods in San Francisco? Which are the least safe?
2. What types of crime are most common throughout the city?
3. How has overall crime in San Francisco changed since 2018?
    - Are there different crime types more common now, than they were prior to 2020?
4. What time of day is crime most prevalent? What day of week is crime most prevalent?

Data Source

DataSF - Police Department Incident Reports: 2018 to Present

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

* Data is compiled from the department's **Crime Data Warehouse (CDW)**

→ Provides information on incident reports filed by the SFPD in CDW, or filed by the public with the SFPD

Data Privacy & Ethics

Guide provided by DataSF was used as reference to ensure we appropriately use identifiers, interpret different kinds of records, and limitations of analysis related to active **privacy controls** ↓

https://datasf.gitbook.io/datasf-dataset-explainers/sfpd-incident-report-2018-to-present

Data Cleaning Process

Use `pd.read_csv` to read in 'Police_Department_Incident_Reports.csv' and create our DataFrame: `crime_df`

→ `crime_df` returns a total of: **850,895 rows** and **35 columns**

1. Drop the following columns (not necessary for crime map or analysis) →
2. Remove any rows that contain null values with `.dropna( )`
3. Using our cleaned '`crime_new_df`' - return a list of incident categories

   → Select necessary incident categories to keep for final dataset
4. Create a final cleaned '`crime_new02_df`' with the selected categories

→ `crime_new02_df` returns a total of: **503,415 rows** and **13 columns**

```
columns_to_drop = ['Report Datetime',
                   'Row ID',
                   'Incident ID',
                   'Incident Number',
                   'CAD Number',
                   'Report Type Code',
                   'Report Type Description',
                   'Filed Online',
                   'Incident Code',
                   'Intersection',
                   'CNN',
                   'Police District',
                   'Supervisor District',
                   'Supervisor District 2012',
                   'Point',
                   'Neighborhoods',
                   'ESNCAG - Boundary File',
                   'Central Market/Tenderloin Boundary Polygon - Updated',
                   'Civic Center Harm Reduction Project Boundary',
                   'HSOC Zones as of 2018-06-05',
                   'Invest In Neighborhoods (IIN) Areas',
                   'Current Supervisor Districts',
                   'Current Police Districts']

crime_new_df = crime_df.drop(columns=columns_to_drop)
```

MongoDB - Database Creation

1. Connect to MongoClient
2. Create a new database: `crime_db`
3. Create a collection storing our sample data: `incidents`
4. Create a collection storing all of our data: `incidents_full`
5. Import our previously cleaned sample data into the: '`incidents`' collection
   Import our previously cleaned data into the: '`incidents_full`' collection

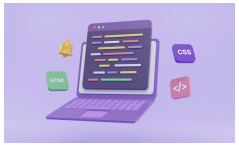
```
client = MongoClient("mongodb://localhost:27017/")
db = client.crime_db

collection = db[collection_name]

# Load data into incidents_full collection
load_csv_to_mongodb('data/sf_crime_data.csv', 'incidents_full')

# Load data into incidents collection
load_csv_to_mongodb('data/sample_data_by_year.csv', 'incidents')
```

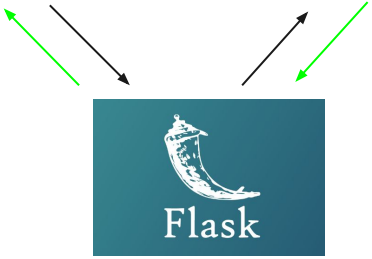
Flask App - Design

Objective: Develop a web application for visualizing and serving data end points for crime data analysis.

Framework

Backend → Flask
Database → MongoDB
Frontend (Client side) → HTML, CSS, JavaScript (Leaflet, D3, Apexcharts, etc.)





Key Features

→ Multiple data endpoints
→ Interactive data visualization
→ User-friendly interface with various data views

Flask Demo

SF Crime Map

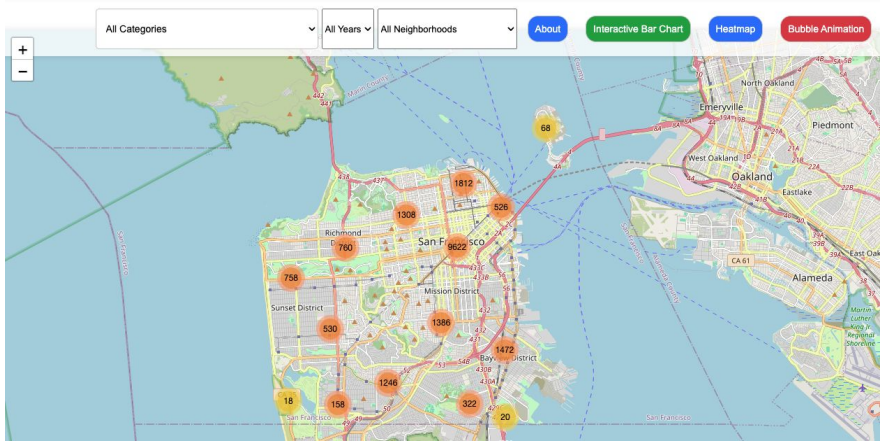
Marker Clusters

Filtered by Category, Year & Neighborhood

index.html

Leaflet Plugins:

https://unpkg.com/leaflet/dist/leaflet.css
https://unpkg.com/leaflet.markercluster/dist/MarkerCluster.css
https://unpkg.com/leaflet.markercluster/dist/MarkerCluster.Default.css
https://unpkg.com/leaflet/dist/leaflet.js
https://unpkg.com/leaflet.markercluster/dist/leaflet.markercluster.js



app_map.js

1. Using **Leaflet**, create a city map centered on San Francisco coordinates

2. Define API URLs for incident and neighborhood data

```
// Store the API query variables
    let incidentUrl = "http://127.0.0.1:5000/reduced_data";
    let neighborhoodUrl = "https://data.sfgov.org/resource/gfpk-269f.json";
```

3. Create a group for **marker clusters**

4. Use `d3.json` to fetch incident and neighborhood data
   → Fetch incident data and create the markers for each incident
   → Fetch neighborhood data to add GeoJSON layer for neighborhood boundaries

5. Add marker clusters to the city map

6. Add **event listener** to create drop down menus for filtering markers

Data Visualizations

Neighborhood Trends

Question 1:

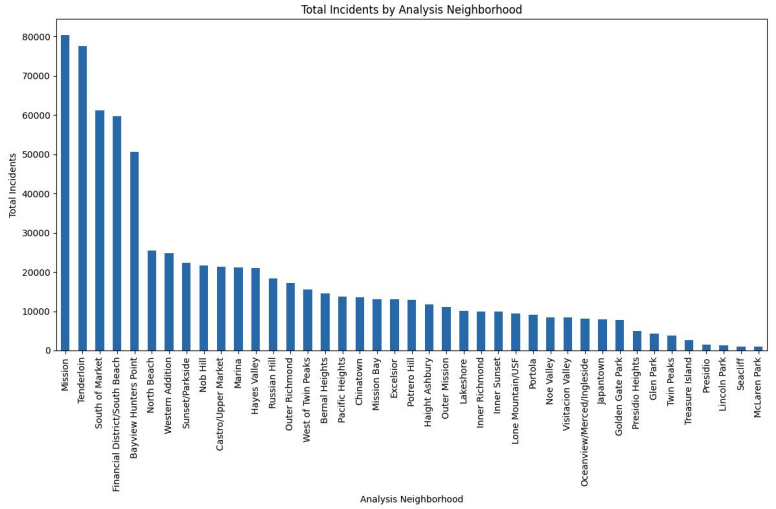
What are the safest neighborhoods in Downtown San Francisco?
What are the least safe?

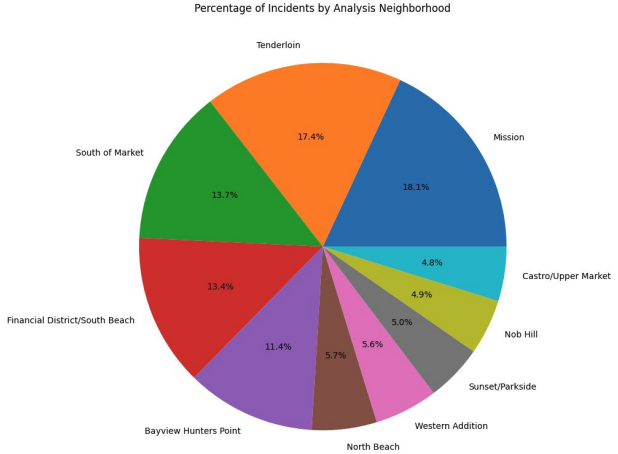
Safest Neighborhoods

1. McLaren Park
2. Lincoln Park
3. Presidio
4. Seacliff
5. Treasure Island

Least Safe Neighborhoods

1. Tenderloin
2. Mission
3. South of Market
4. Financial District/South Beach
5. Bayview Hunters Point





Data Visualizations

Neighborhood Trends

Question 1:

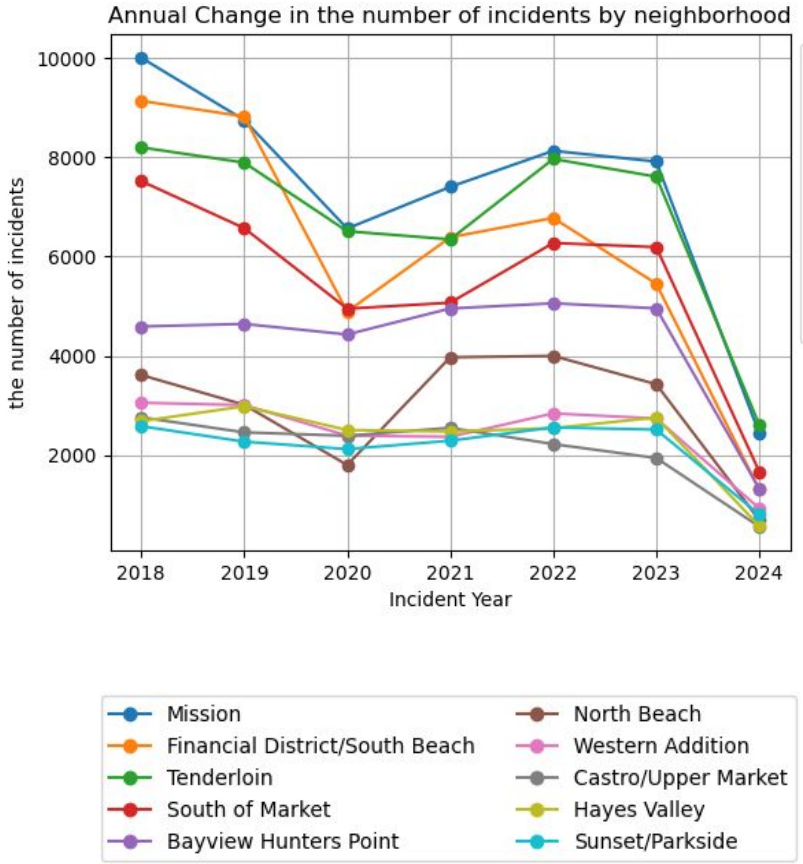
What are the safest neighborhoods in Downtown San Francisco?
What are the least safe?

Question 3:

How has overall crime in San Francisco changed since 2018?
Are different crime types more common now, than prior to 2020?

The annual crime count for San Francisco neighborhoods are relatively consistent.

We see a noticeable and expected drop in crime activity in 2020. While we do note an increase in crime since the pandemic, SF crime incident count in recent years is seemingly lower than prior to.



* 2024 can be ignored in the graph above - this visual reflects count
* Only 4.5 months of data for this current year

Data Visualizations

Crime Trends

Question 2:

What types of crime are most common throughout the city?

No matter the year, the types of crime reported in Downtown San Francisco remain fairly consistent.

However, there are few select categories that have become more common in recent (post-pandemic) years ↓

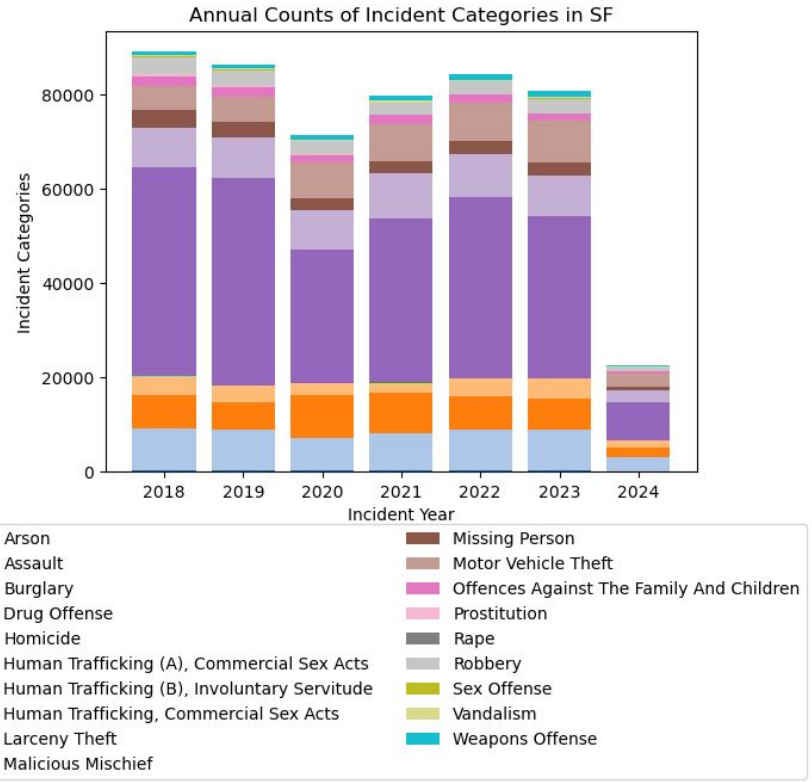
Crime Categories Increasing After COVID-19 Pandemic:

- Drug offense
- Motor Vehicle Theft
- Arson
- Weapons offense
- Vandalism
- Homicide

*The dynamic bar chart helps to better visualize the annual trends for each incident category.

Overall Most Common Crimes in San Francisco:

1. Larceny Theft
2. Malicious Mischief
3. Assault
4. Burglary

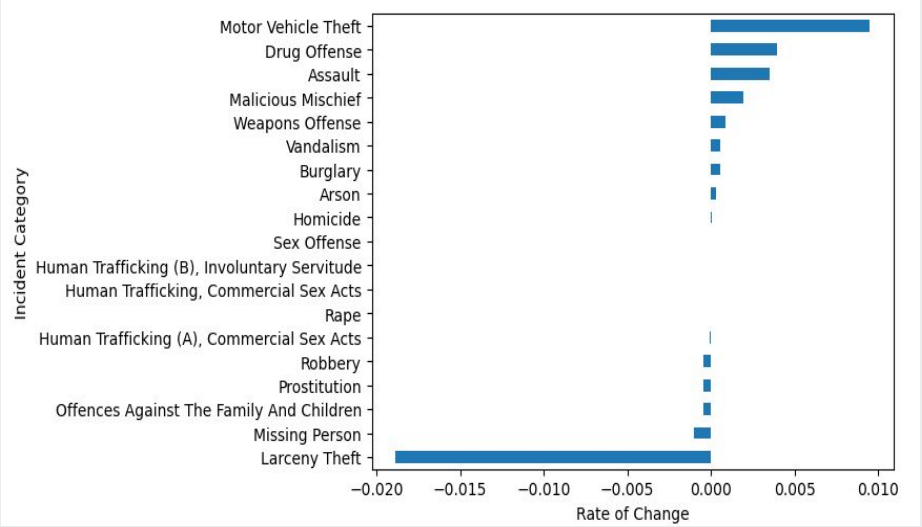


Data Visualizations

Incident Category - Rate of change

Question 3:

What types of crime are most common throughout the city?



Slope trend

Incident Category

Incident Category	
Larceny Theft	-0.0189
Missing Person	-0.0010
Offences Against The Family And Children	-0.0004
Prostitution	-0.0004
Robbery	-0.0004
Human Trafficking (A), Commercial Sex Acts	-0.0001
Rape	-0.0000
Human Trafficking, Commercial Sex Acts	-0.0000
Human Trafficking (B), Involuntary Servitude	-0.0000
Sex Offense	0.0000
Homicide	0.0001
Arson	0.0003
Burglary	0.0005
Vandalism	0.0006
Weapons Offense	0.0009
Malicious Mischief	0.0020
Assault	0.0035
Drug Offense	0.0039
Motor Vehicle Theft	0.0095

Visual Analysis

Heat Maps & Box Plots

Could religious activities on Sunday be a factor?

Day of Week:

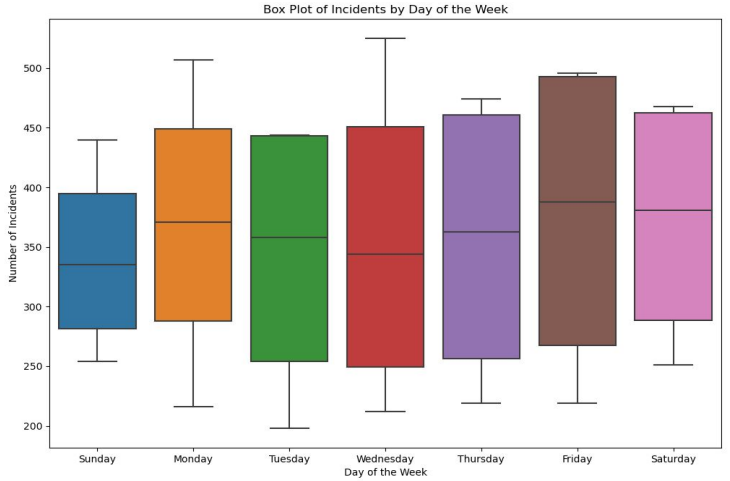
Friday's have the **highest** crime count, **Sunday's** have the **least.**

→ Likely due to lifestyle factors, it is fairly common for city activity to increase over weekends, as people's social lives are more active then.

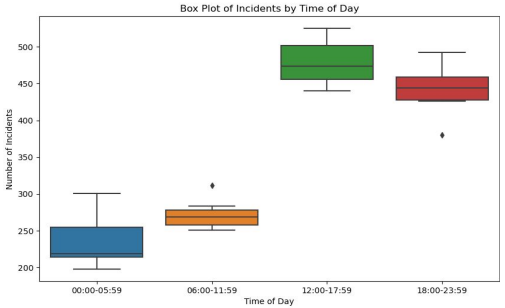
→ An increase in overall activity gives more opportunities for crime to occur. **Sundays are rest days, *even for criminals.***

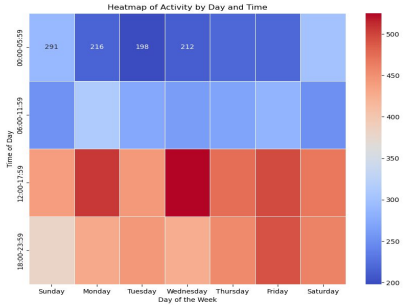
Time of Day:

Between the hours of **noon** and **5:59pm** we see the most incidents, **6:00pm - 11:59pm** trail shortly behind in total incident count.



The box plot and each heat map, collectively show a clear distribution in terms of the time a crime has occurred ↓



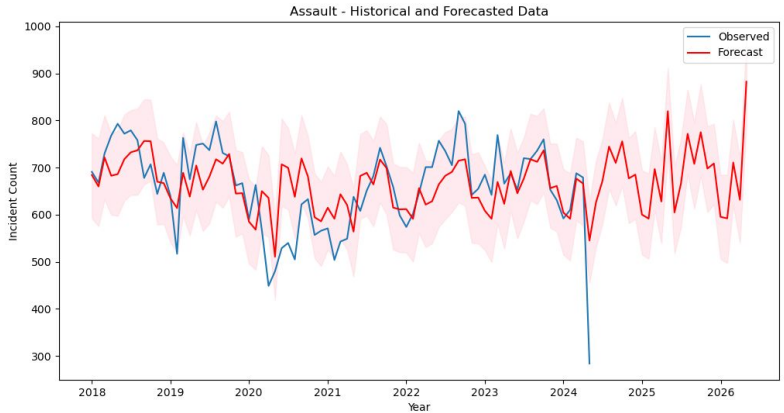


Major Category Crime Forecast into 2026

After evaluating different models, we chose Meta's **prophet** to model the crime data forecast:

→ Modeling for the next 2 years
→ Handling seasonality and missing/incomplete data
→ Evaluate every 180 days

Drawbacks: Accuracy drops significantly after 90 days

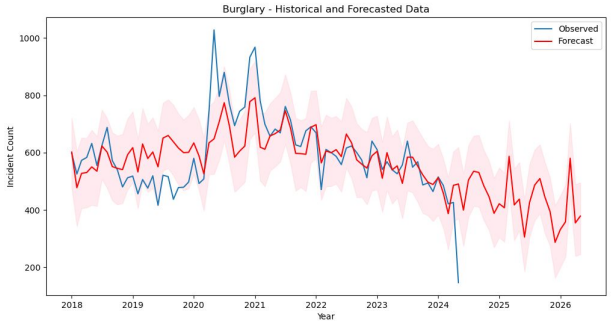


	horizon	mae	mse	mape
0	79 days	139.568105	33788.718178	0.[illegible]
1	81 days	147.362915	35171.092922	0.24582
2	85 days	146.507341	35021.510128	0.246983
3	86 days	126.997625	23005.895174	0.206078
4	90 days	131.796127	23849.162389	0.213722
..	...	...	...	...
148	719 days	200.857314	54890.238946	0.281569
149	723 days	210.127680	58719.234591	0.297745
150	724 days	224.252378	62819.145651	0.318086
151	728 days	209.746283	54378.994404	0.303804
152	730 days	214.307007	57889.666887	0.368219

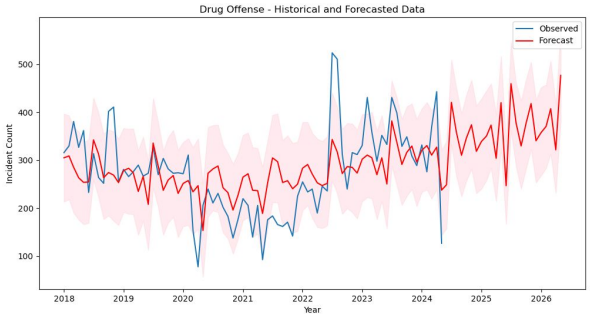
Mean Absolute Percentage Error (MAPE):

The lower the MAPE, the higher the forecast accuracy.

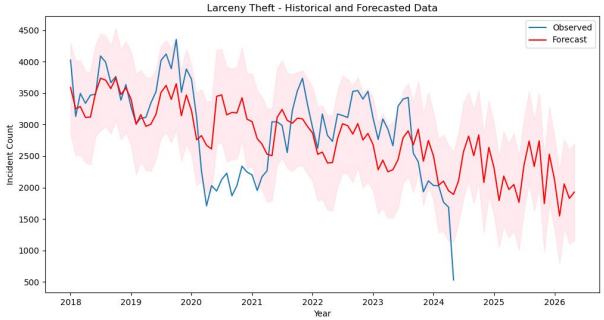
Crime Forecasts



	horizon	mae	mse	mape
0	79 days	203.074280	57669.298234	0.326226
1	81 days	184.810800	51979.046842	0.281656
2	85 days	165.772418	44273.437294	0.241065
3	86 days	141.540477	29913.578763	0.211547
4	90 days	152.985357	34131.202964	0.241937
..	...	...	...	...
148	719 days	393.044686	208864.777854	0.687730
149	723 days	415.877329	239030.803068	0.743025
150	724 days	431.263735	248069.976533	0.775098
151	728 days	408.464288	222872.582406	0.745373
152	730 days	431.238624	252060.130059	1.047220



	horizon	mae	mse	mape
0	79 days	94.633320	16628.350504	0.339961
1	81 days	94.674716	16637.179942	0.344566
2	85 days	79.915723	13087.070673	0.290866
3	86 days	76.852110	12868.003345	0.279407
4	90 days	75.693886	12649.657479	0.265345
..	...	...	...	...
148	719 days	237.229259	73897.446212	1.022865
149	723 days	242.250437	75462.991050	1.016996
150	724 days	253.959341	80628.635644	1.039744
151	728 days	254.148601	80712.939074	1.040787
152	730 days	238.576782	75085.804078	1.015256



	horizon	mae	mse	mape
0	79 days	1104.539629	1.832727e+06	0.414404
1	81 days	1113.589087	1.862199e+06	0.434519
2	85 days	1142.035874	1.901972e+06	0.456584
3	86 days	986.152834	1.442786e+06	0.374704
4	90 days	923.117802	1.275866e+06	0.343352
..	...	...	...	...
148	719 days	1890.504093	8.035551e+06	0.654840
149	723 days	1919.264517	8.160768e+06	0.661303
150	724 days	2006.225662	8.352935e+06	0.695474
151	728 days	1883.885961	8.001415e+06	0.667491
152	730 days	1825.778699	7.776055e+06	0.792054

Conclusions

San Francisco Crime Rate (2019 - 2024)**:**

We were surprised that our initial expectation of seeing a stark rise in crime since the COVID-19 pandemic was not in fact the reality of crime in San Francisco post-pandemic. Crime has seen a steady rise, climbing back up to normal incident count levels, as compared to previous years, however, it has not yet reached the heights that were noted in 'high crime' years in the past (2018).

Neighborhood Analysis:

→ Overall, neighborhoods isolated from SF's downtown city center tend to have less incidents of crime
→ As expected, the neighborhoods that were located in busier parts of the city saw more crime activity

Crime-Category Commonality:

→ Larceny theft is by far the most common crime that occurs within SF neighborhoods
→ Post-pandemic saw a rise in Drug Offenses, Weapon Offenses, Motor Vehicle Theft, Vandalism,Arson, and Homicide

Incident Date Time Occurance:

→ Crime occurs at both the day of week and time of day when you would expect a city to most active
→ Afternoon-evening, along with weekends see the overall highest counts of crime

Limitations

Data Size:

The volume of data precluded us from delving deep into all crime categories. With more time, we could evaluate more models for projection -aiming at improving accuracy.

Data Cleaning:

In the cleaning process we dropped many incident categories that didn't necessarily match the interest of our project's outline. There is a possibility that in doing so we may have dropped categories due to their generic descriptions that might have included subcategory and descriptions of significant interest.

Time Restraints:

Given the probability for error, we found ourselves spending a lot of time fixing small problems, given more time we could comfortably accomplish adding further style and functionality to our flask app and the related visualizations.