2201 Capstone

Brody Kretz - Write up

For this project, I worked with the FBI Most Wanted API, the world_country_and_usa_states_latitude_and_longitude_values.csv file, and the FBI Most Wanted Wiki. My goal was to examine both current and former criminals on the list to identify trends and assess whether there's any predictive value in the FBI's success rate at capturing fugitives.

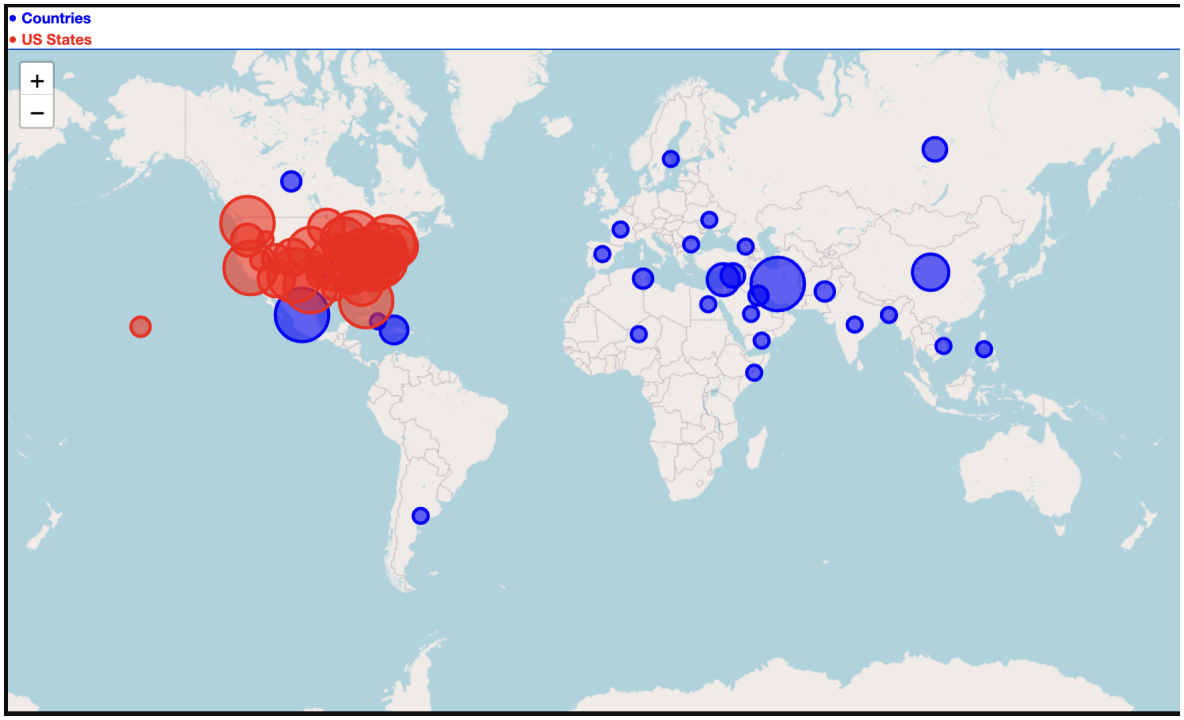
I took a different approach to this project than maybe was expected of me. I wanted my project to be interesting, not just random data. This required me to teach myself alot of new things and read up on many python libraries. When doing this I used a lot of debugging, printing names and information too screen, to see why data isnt coming out how I wanted. I had to make a decision if I wanted to keep this in because it really shows what was happening behind the scenes, but I needed to focus on what the goal of this project was and not have distractions/unnecessary outputs. I left some minor things so the user knows when data is done processing.

One challenge I faced was how to plot each criminal accurately on the map. The FBI's API has two versions: /v1 and /v1/list. Initially, I used /v1, which only returned the nationality of each criminal. This was a problem because my CSV file only included country names not nationalities with corresponding latitude and longitude values. I attempted an overly complex solution using several Python libraries to convert nationalities into countries. Later, while reviewing past coursework, I remebered the /v1/list endpoint. Although I assumed it returned the same data, I tried it anyway and discovered it included the country names directly solving the problem entirely and was able to create the following dataframes

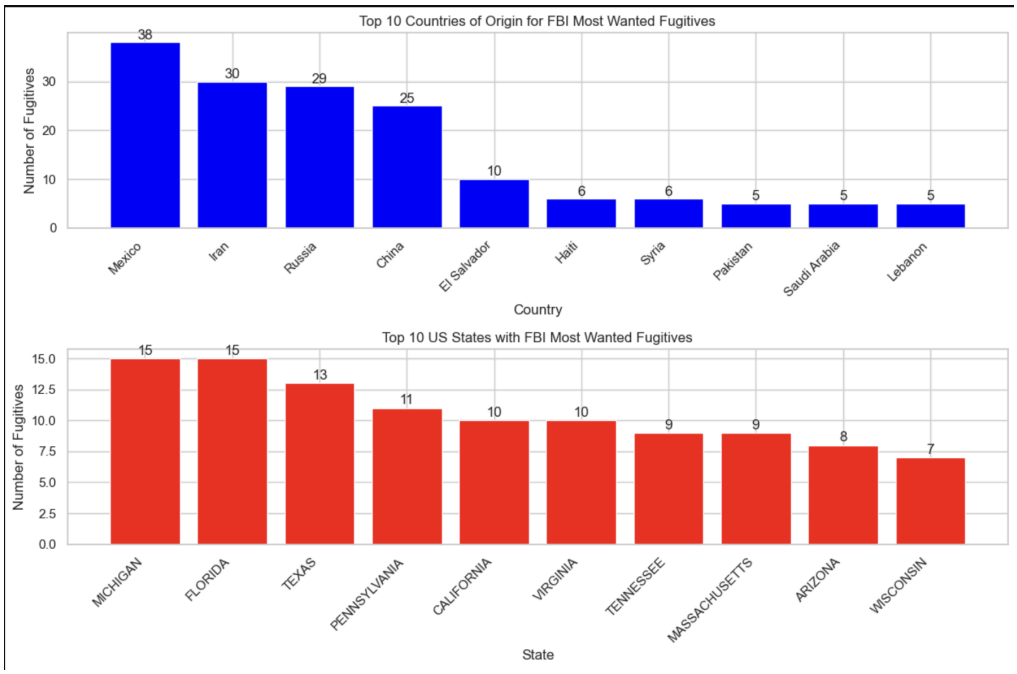
	country	latitude	longitude	fugitive_count
0	United States	37.0902	-95.7129	2
1	Mexico	23.6345	-102.5528	38
2	Honduras	15.2000	-86.2419	4
3	Syria	34.8021	38.9968	6
4	Pakistan	30.3753	69.3451	5
5	China	35.8617	104.1954	25
6	Afghanistan	33.9391	67.7100	1
7	El Salvador	13.7942	-88.8965	10
8	Colombia	4.5709	-74.2973	1
9	Bulgaria	42.7339	25.4858	1
10	Haiti	18.9712	-72.2852	6
11	Saudi Arabia	23.8859	45.0792	5
12	France	46.2276	2.2137	1
13	Iraq	33.2232	43.6793	1
14	Lebanon	33.8547	35.8623	5
15	Somalia	5.1521	46.1996	1
16	Kuwait	29.3117	47.4818	3
17	Yemen	15.5527	48.5164	2
18	India	20.5937	78.9629	4
19	Palestine	31.9522	35.2332	1
20	Jordan	30.5852	36.2384	1
21	Philippines	12.8797	121.7740	2
22	Vietnam	14.0583	108.2772	4
23	Russia	61.5240	105.3188	29

	state	latitude	longitude	fugitive_count
0	TEXAS	31.054487	-97.563461	13
1	NEW MEXICO	34.840515	-106.248482	4
2	FLORIDA	27.766279	-81.686783	15
3	ARIZONA	33.729759	-111.431221	8
4	WISCONSIN	44.268543	-89.616508	7
5	OKLAHOMA	35.565342	-96.928917	3
6	CALIFORNIA	36.116203	-119.681564	10
7	TENNESSEE	35.747845	-86.692345	9
8	GEORGIA	33.040619	-83.643074	4
9	COLORADO	39.059811	-105.311104	7
10	PENNSYLVANIA	40.590752	-77.209755	11
11	INDIANA	39.849426	-86.258278	4
12	MASSACHUSETTS	42.230171	-71.530106	9
13	MICHIGAN	43.326618	-84.536095	15
14	MARYLAND	39.063946	-76.802101	3
15	LOUISIANA	31.169546	-91.867805	2
16	NEW YORK	42.165726	-74.948051	5
17	ILLINOIS	40.349457	-88.986137	6
18	MISSOURI	38.456085	-92.288368	4

Early on, I created an interactive Folium map displaying the current FBI Most Wanted list. I also generated bar charts comparing the distribution of fugitives by country and U.S. state to identify geographic trends. I spent considerable time learning how to plot long/lat points on a map and found Folium to be the most effective tool. I was pleased with how the visualizations turned out

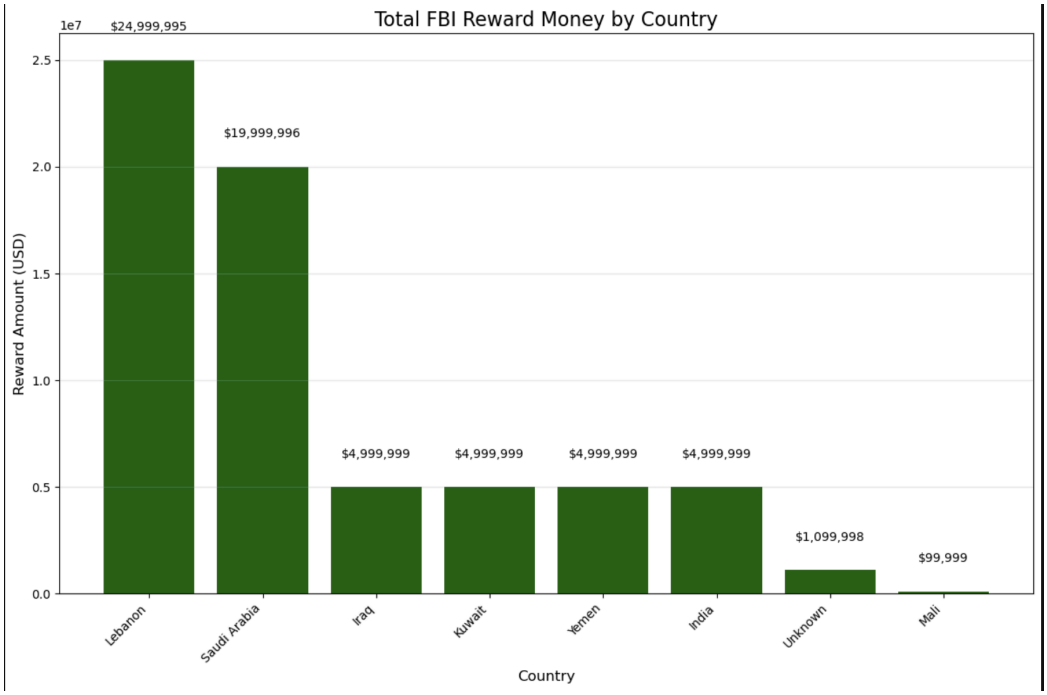


Then I wanted to visualize this data as a bar chart with each state and country keeping their respective colors. Each country's criminals were summed, and using my data frame I created and matplotlib, I was able to get the final product as shown below.



There are many countries that cause problems as seen on the map, but which country causes the most trouble overall? I saw the each person is given a bounty and figured that would be a

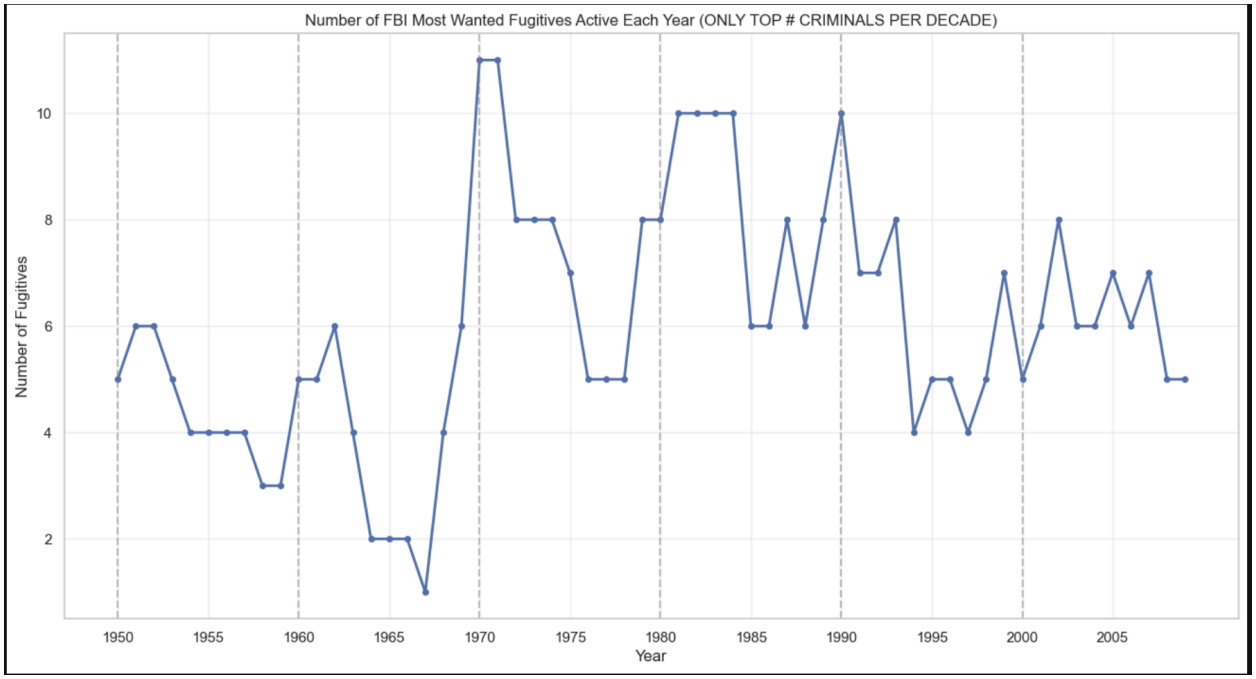
great way to determine how much trouble each individual person is causing, and by grouping them by country and summing their total worth. You can find as seen in my visualization that we have a lot of money put towards catching people in Lebanon as seen below.



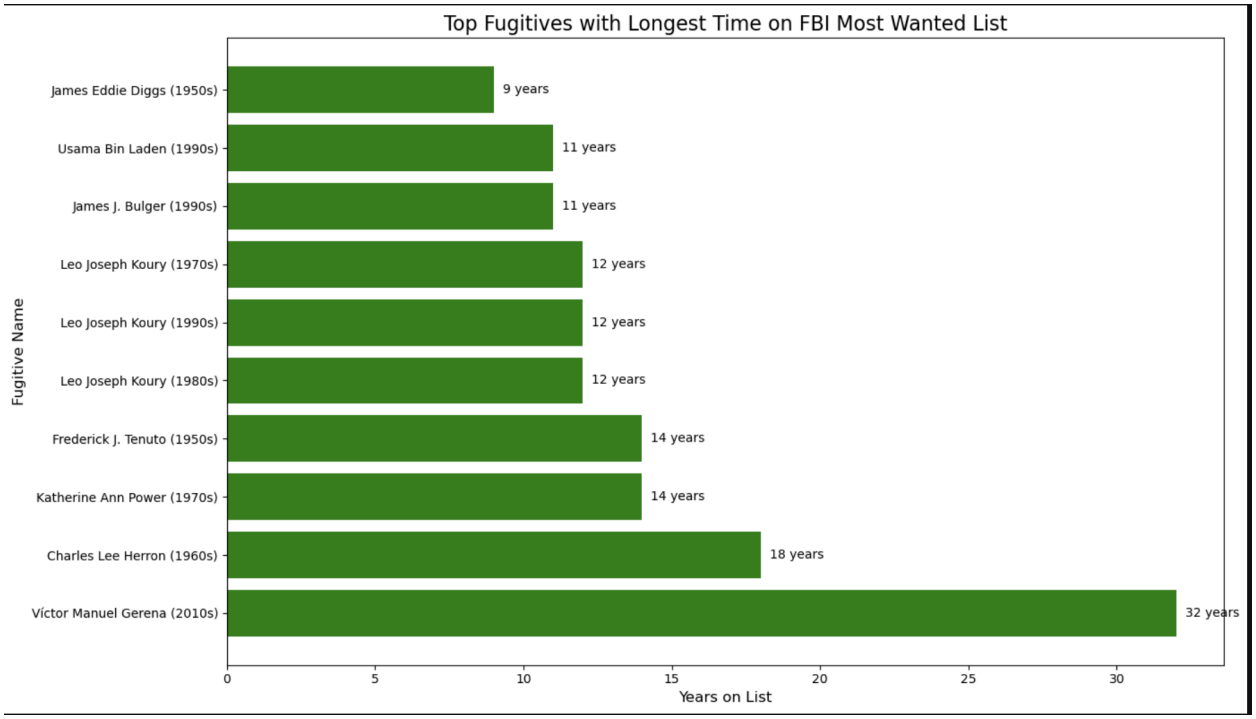
Once I completed the current data analysis, I wanted to explore the history of the FBI's Most Wanted list. However, the FBI does not publicly preserve a record of past fugitives only the current list is available. My first solution was to use the Wayback Machine, which archives snapshots of websites. While Python has a library for accessing this data, there weren't enough archived API snapshots to be useful.

Eventually, I found that Wikipedia had compiled sufficient historical data from 1950 to the present. I spent several days writing regex patterns and functions to extract and graph the historical fugitive data. This part of the project consumed the majority of my time, but I was ultimately happy with the results as seen below.

At the end of my project, I created a graph representing the number of individuals wanted in a given year. For each year, I selected the three individuals who had been on the list the longest. I used the years they were added to the list and the years they were captured to fill in data points across the timeline. In this graph, a higher y-axis value reflects a worse year for the FBI, indicating that more fugitives remained at large.



You can see the people that were wanted the longest and how many years they were on the list in the following list.



Due to the limitations of the Wiki data (including gaps that restricted my sample size), I couldn't make many strong inferences. However, I did see a slight decline in the number of wanted individuals over time. There was a spike in 1970 and after further research, I learned this was part of a broader crime wave spanning the 1970s and early 1990s, driven by factors like social unrest, economic instability, and the emergence of infamous serial killers such as the Zodiac Killer.

To improve this project, I would focus on creating more visually appealing charts. In some ways, I wish there had been more criminals listed in the past, as that would have allowed for a more thorough analysis. I noticed that certain decades lacked sufficient data, so to ensure consistency, I chose the top three fugitives with the longest time on the list to create Figure 6. It's difficult to make a definitive judgment on whether the FBI has improved at capturing criminals over time which was my original goal.

In the future, I'd be interested in applying this same methodology to a dataset with more consistent historical records, such as sports data, which would provide enough data points to draw more reliable conclusions. I believe my approach in this project is unique, as I couldn't find much prior research that attempted what I did. Of course, individuals within the FBI likely have access to better data than the public.

I'd like to share this project with future employers because it highlights my problem solving skills, particularly in plotting country data by integrating a CSV file with API outputs, and in using web scraping to track historical data. Initially, I tried using the Wayback Machine to retrieve old FBI Most Wanted snapshots, but due to limited archives, I turned to Wikipedia, which ultimately provided exactly what I needed. In the end I spent around 7 hours on the FBI API, 2 Hours On the Folium Map, 1 Hour on the Rewards Figure, 3 Hours on the States/Countries Bar Chart, and close to 10 hours on the Wiki portion. I spent 2 hours a day for 2 weeks on this and I had so much fun doing it. I enjoy the challenge of having no idea how to do something and needing to research and trial an error. Yes, I wanted to pull my hair out many times, but now that I look at it and am proud that it works even if I didn't get the outcome I wanted.