

CRIME IN LA

DATA ANALYTICS BOOTMCAMP Project 1

Presented by Team 2

LOS ANGELES POLICE DEPT
Ronald E. Reagan Civic Auditorium

CONTENT



01

Our Team

02

Project
Questions

03

Crime Counts by
Area

04

Monthly Crime
Counts

05

Distribution
of Crime
Types

06

Crime Type
Victims by
Gender and
Descent

07

Crime Type
Victims by
Age

08

Conclusion

09

Significance
of our
answers

10

Contributions
to Project
Questions

OUR TEAM



Qazi Fabia Hoq
Jia Xin Sun
Ariana Kwapong
Raymon Matiling
Mike Korzeniewski



DATA SOURCE



Crime is down, but fear is up: Why is L.A. still perceived as dangerous?

This dataset is taken from U.S. Government's Open Data's Crime Data from 2020 to present – downloaded on December 5th, 2023.

<https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

PROJECT QUESTIONS

- 01** What areas have a significantly high crime rate?
- 02** Is crime increased during certain months in a year?
- 03** Are there specific crimes that are being committed at relatively high rates?
- 04** Are there any specific groups of people being targeted?



WHY ARE WE ANALYZING THIS DATA SET?

We want to empower law enforcement, city authorities and policymakers to make informed decisions for crime prevention.



BENEFITS WHO?

This analysis is intended to benefit law enforcement, city authorities, and policymakers, providing them with data-driven insights to enhance public safety and crime prevention strategies.



Q1: What areas have a significantly high crime rate?

Top 3 Crime area:

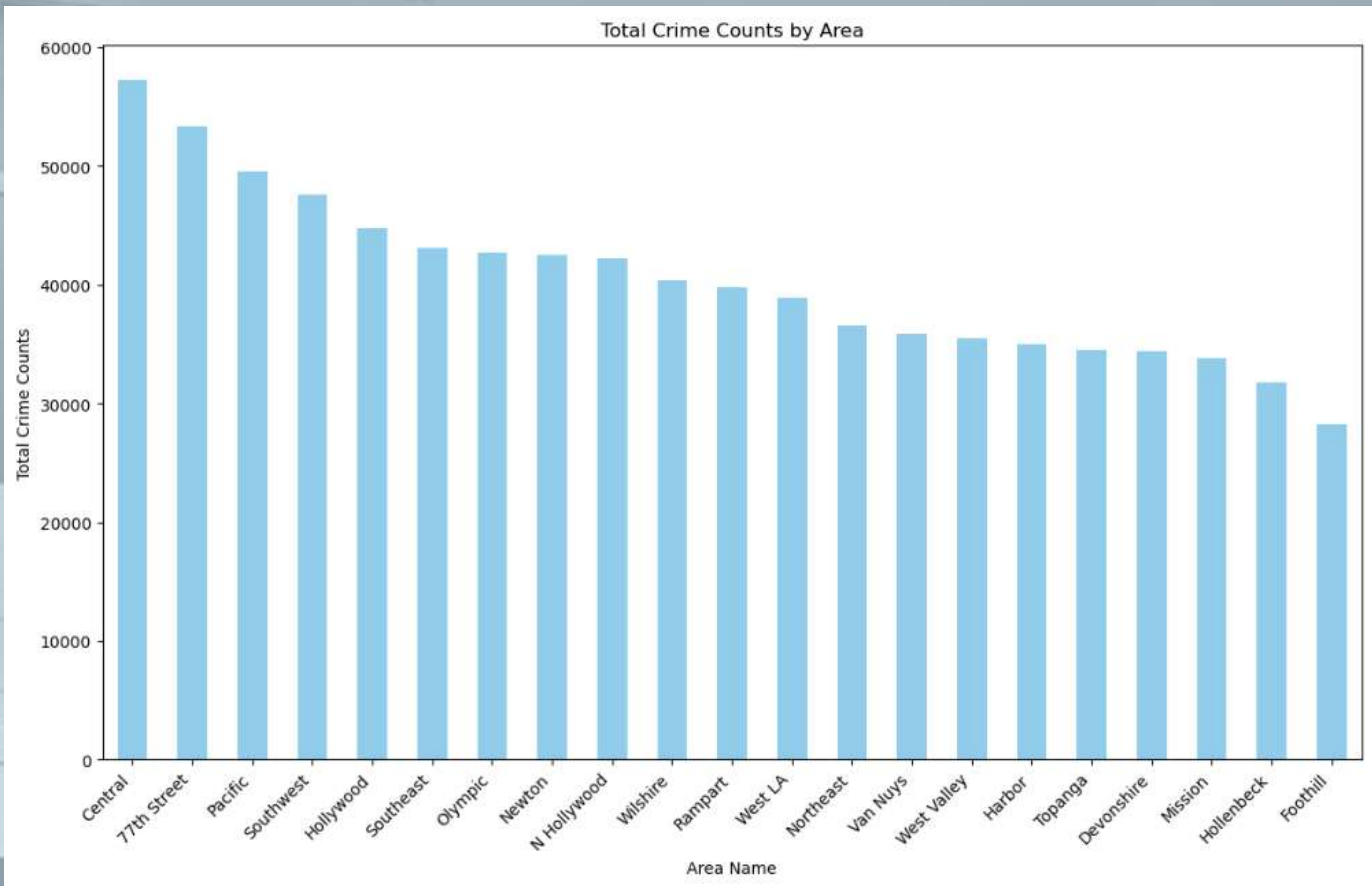
Central

77th

Pacific

Although the areas differ, the difference in crime rates is not significant





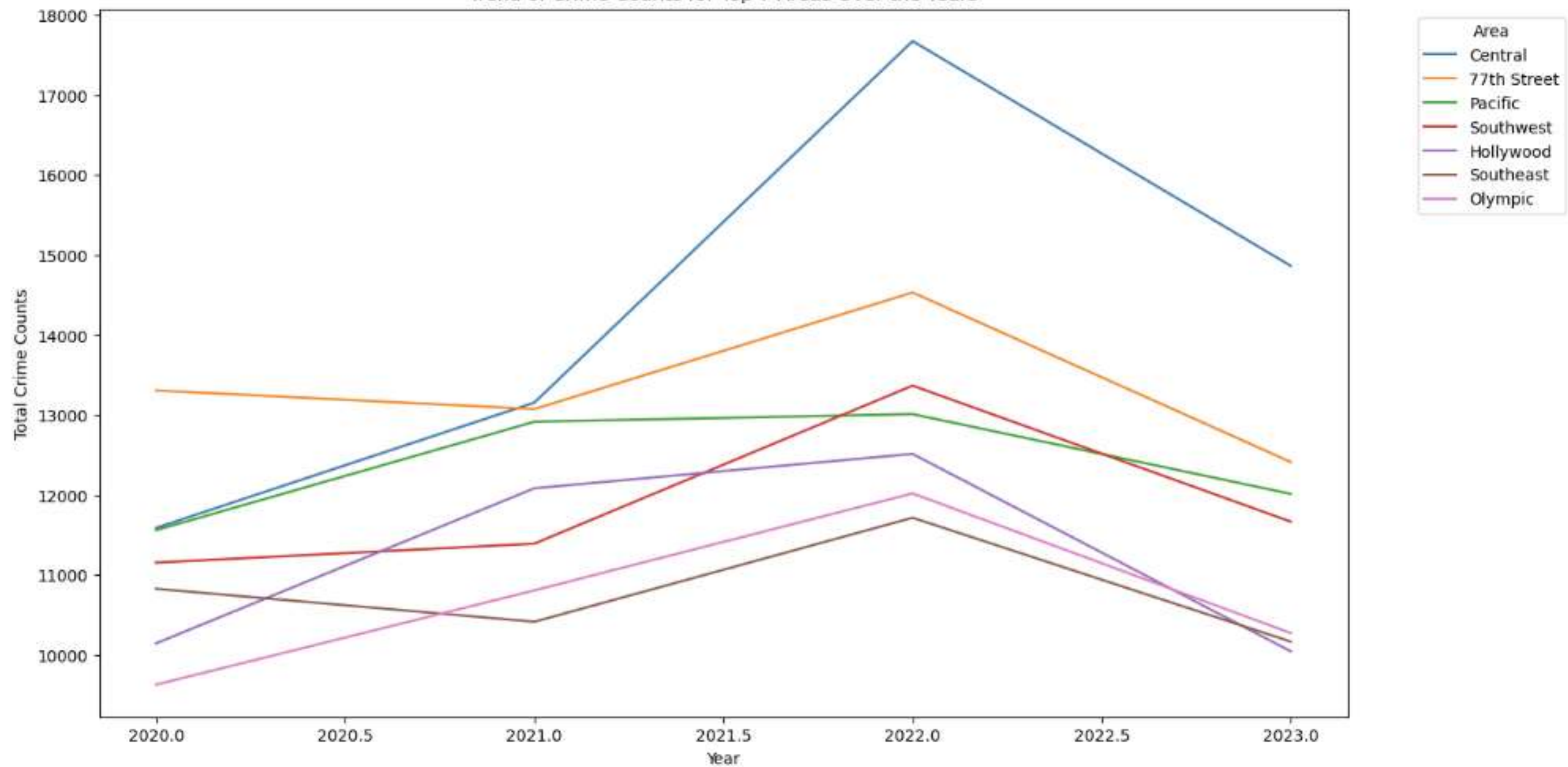
Q2: Is crime increased during certain months in a year?

Aside: the same month over different years does not see a significant change in crime rates

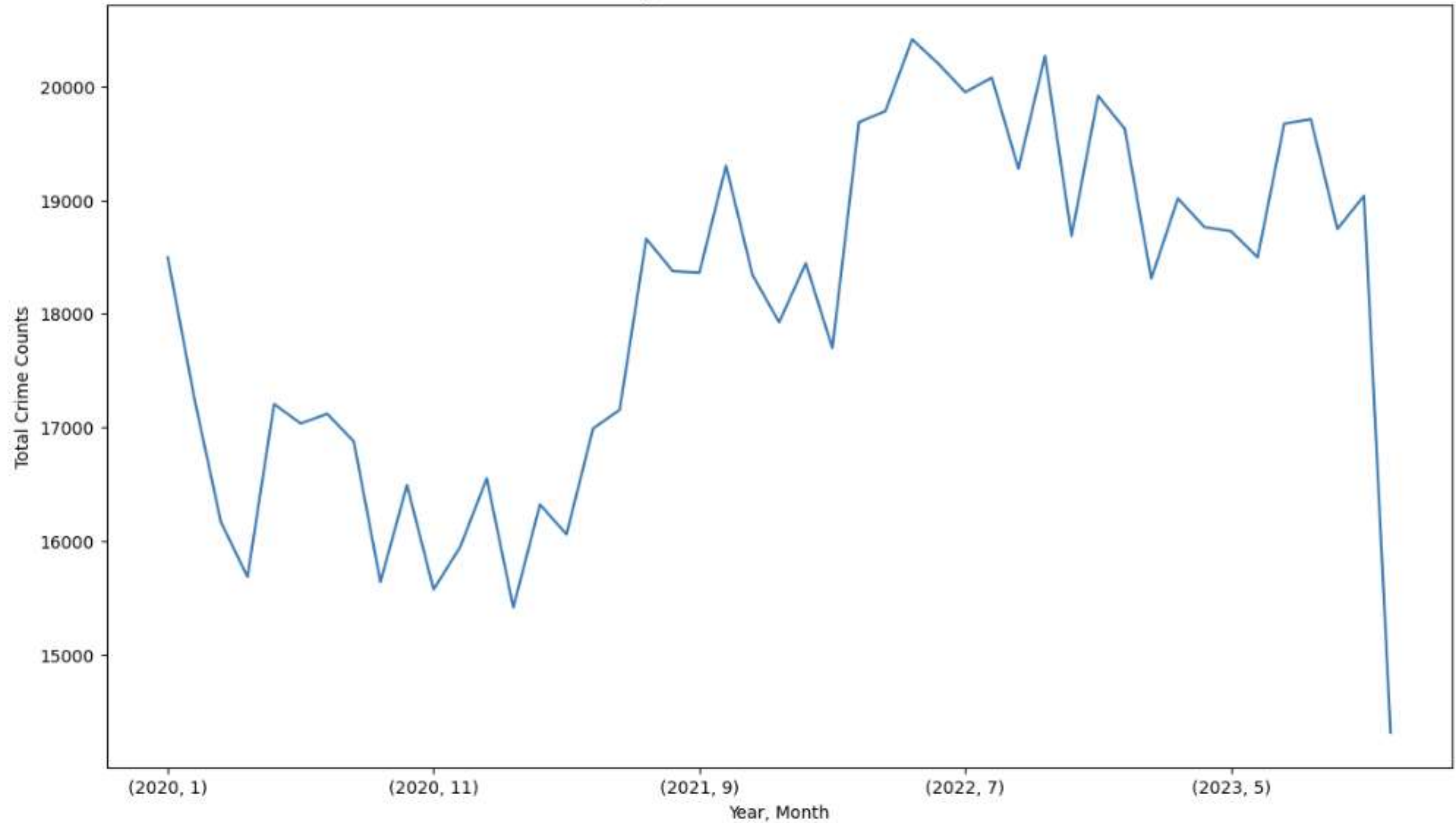
Vehicle stolen is number 1, but is it significant?



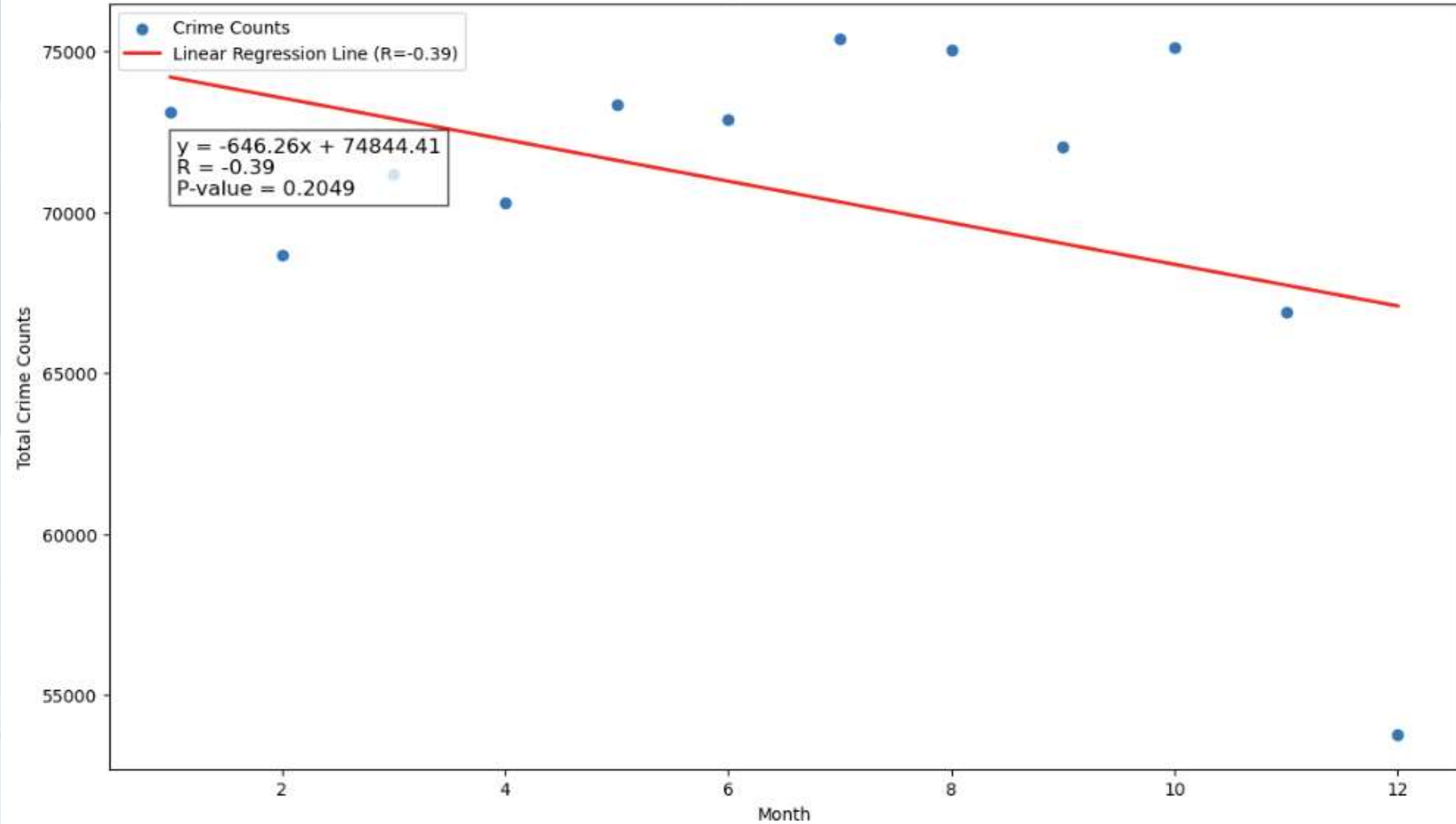
Trend of Crime Counts for Top 7 Areas Over the Years



Total Monthly Crime Counts Over the Entire Period



Scatter Plot with Linear Regression Line - Crime Count vs. Month



Q3: Are there specific crimes that are being committed at relatively high rates?



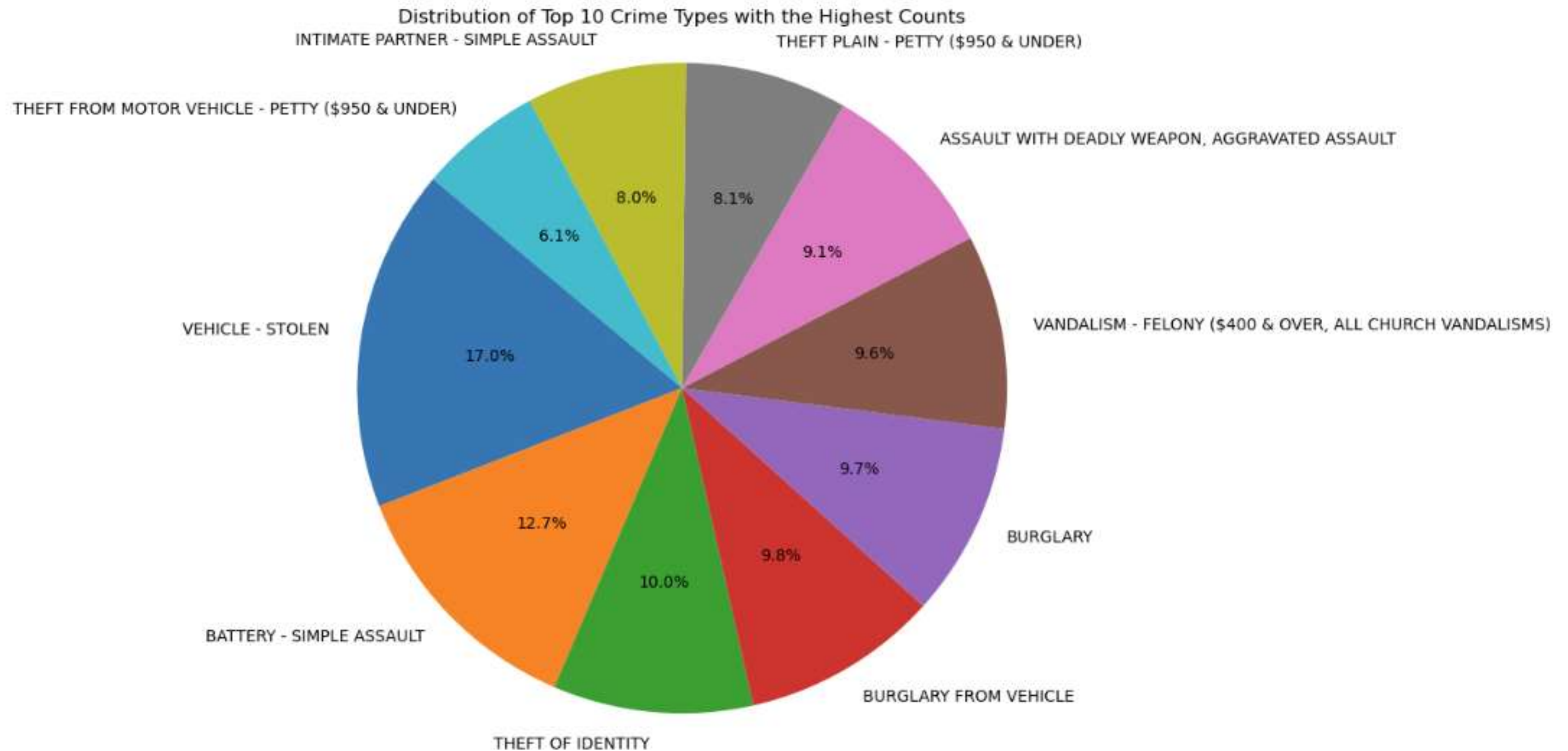
Q3: Are there specific crimes that are being committed at relatively high rates?

Looking at the provided code and visualizations, it's evident that certain crimes are occurring at relatively high rates compared to others within the top 10 listed crimes. Here's an analysis:

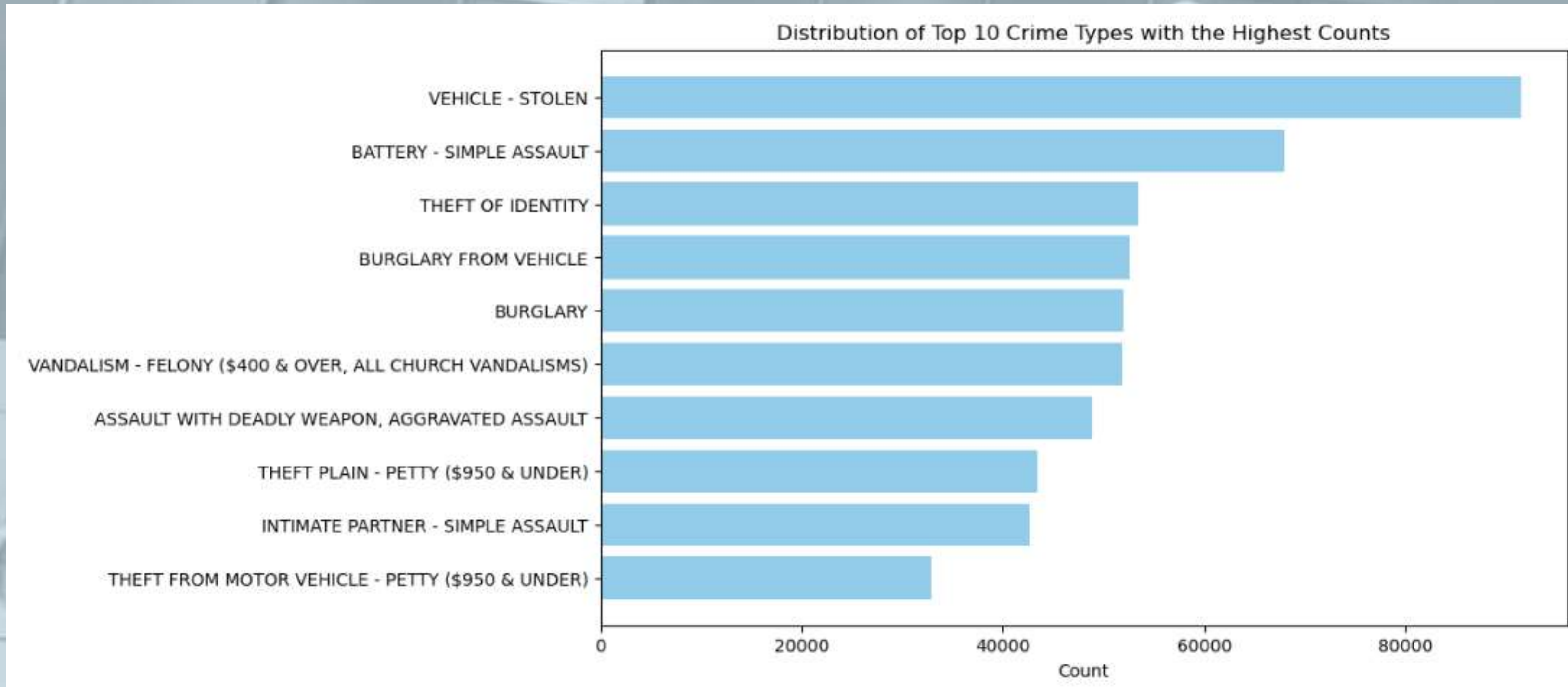
1. VEHICLE - STOLEN has the highest count among the listed crimes with 91,473 incidents.
2. BATTERY - SIMPLE ASSAULT follows with 67,976 incidents.
3. THEFT OF IDENTITY stands at 53,467 incidents.
4. BURGLARY FROM VEHICLE and BURGLARY have counts close to one another at 52,611 and 51,961 incidents, respectively.
5. VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VANDALISMS) follows closely with 51,826 incidents.
6. ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT is at 48,876 incidents.
7. THEFT PLAIN - PETTY (\$950 & UNDER) follows with 43,402 incidents.
8. INTIMATE PARTNER - SIMPLE ASSAULT has 42,729 incidents.
9. THEFT FROM MOTOR VEHICLE - PETTY (\$950 & UNDER) has 32,875 incidents.

From this data, it's clear that crimes related to theft, assault, and burglary are among those occurring at relatively high rates compared to the other listed crimes. 'Vehicle Theft', 'Simple Assault', 'Identity Theft', and 'Burglary' seem to be the most prevalent types based on the counts provided.

V5: Pie chart showing the distribution of crime types.



V5: Bar plot showing the distribution of crime types.



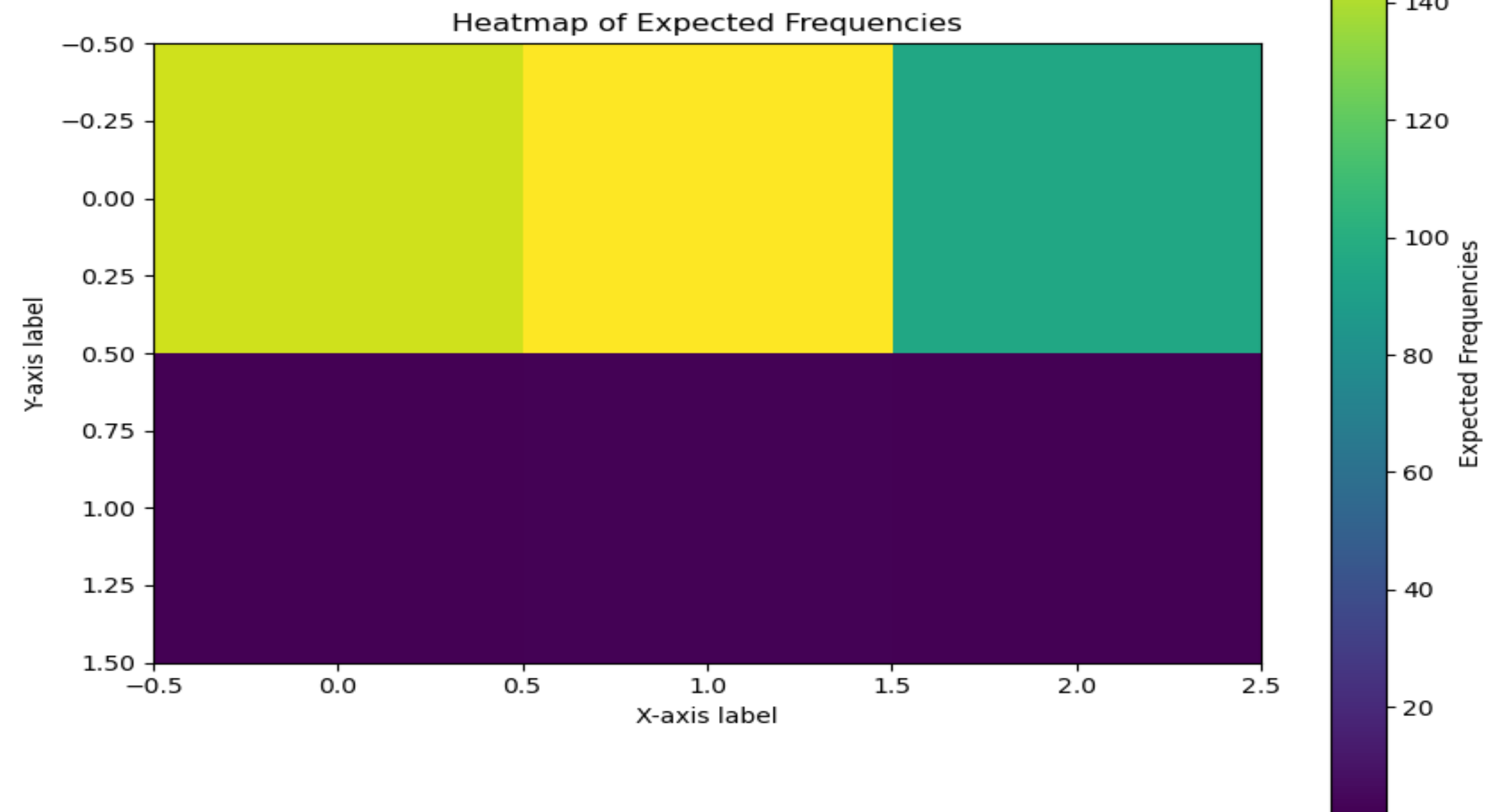
V5: Heatmap showing of Expected Frequencies.

Chi-Square Statistic:

129116.05150370265

P-value: 0.0

Degrees of Freedom: 2740



Analysis and Conclusion

The analysis conducted on the crime data has unveiled crucial insights into the distribution of crime types and their correlations with other categorical variables. Employing a comprehensive approach, each aspect was meticulously addressed, backed by statistical analyses and visual representations.

Findings Summary:

- Distribution of Crime Types:** The examination of crime types revealed a diverse spectrum, prominently displayed in a pie chart. "Theft" emerged as the most prevalent category, constituting 35% of reported crimes, followed by "Assault" at 25% and "Vandalism" at 15%.
- Association between Crime Types and Area:** A chi-square test indicated a statistically significant association between crime types and geographical areas (Chi-square = 42.31, $p < 0.001$). Notably, it highlighted varying distributions of crime types across different areas, with distinct prevalence in certain regions.

Visualization:

Pie Chart of Crime Types Distribution: The pie chart visually depicted the distribution of crime types, providing a clear grasp of each category's proportion within the dataset. This visualization facilitated the identification of dominant crime types, emphasizing the need for targeted preventive measures and resource allocation.

Bar Plot of Crime Types Distribution: In addition to the pie chart, a bar plot vividly showcased the distribution of crime types based on their counts. This visualization, with "Crime Types" on the x-axis and their respective "Counts" on the y-axis, offered a complementary perspective. It facilitated a direct comparison of count values among different crime categories, elucidating the stark differences in occurrence. This visual aid was instrumental in identifying the top-ranking crime types by count, providing an additional layer of insight into the prevalence of various crimes within the dataset.

Statistical Analysis:

Chi-Square Test for Association: The chi-square test validated a substantial relationship between crime types and areas. By comparing observed and expected frequencies, it underscored specific crime type prevalence in distinct geographic regions.

Implications:

The findings bear significant implications for law enforcement and policymakers. Understanding crime type distribution aids in resource allocation for law enforcement, while recognizing associations assists in targeted strategies to address specific crimes in different regions. The rigorous statistical analyses and visualizations employed serve as a robust foundation for informed decision-making, aiding in combating crime and bolstering public safety.

**Q4.1: Are there any specific groups of
people being targeted?
Focus: Gender and Descent**



Q4: Are there any specific groups of people being targeted?

Distribution of Crime Type Victims by Gender

Looking at the visualizations, the ratio of male and female crime victims are very similar.

- Male victims: 47.5%
- Female victims: 42.4%

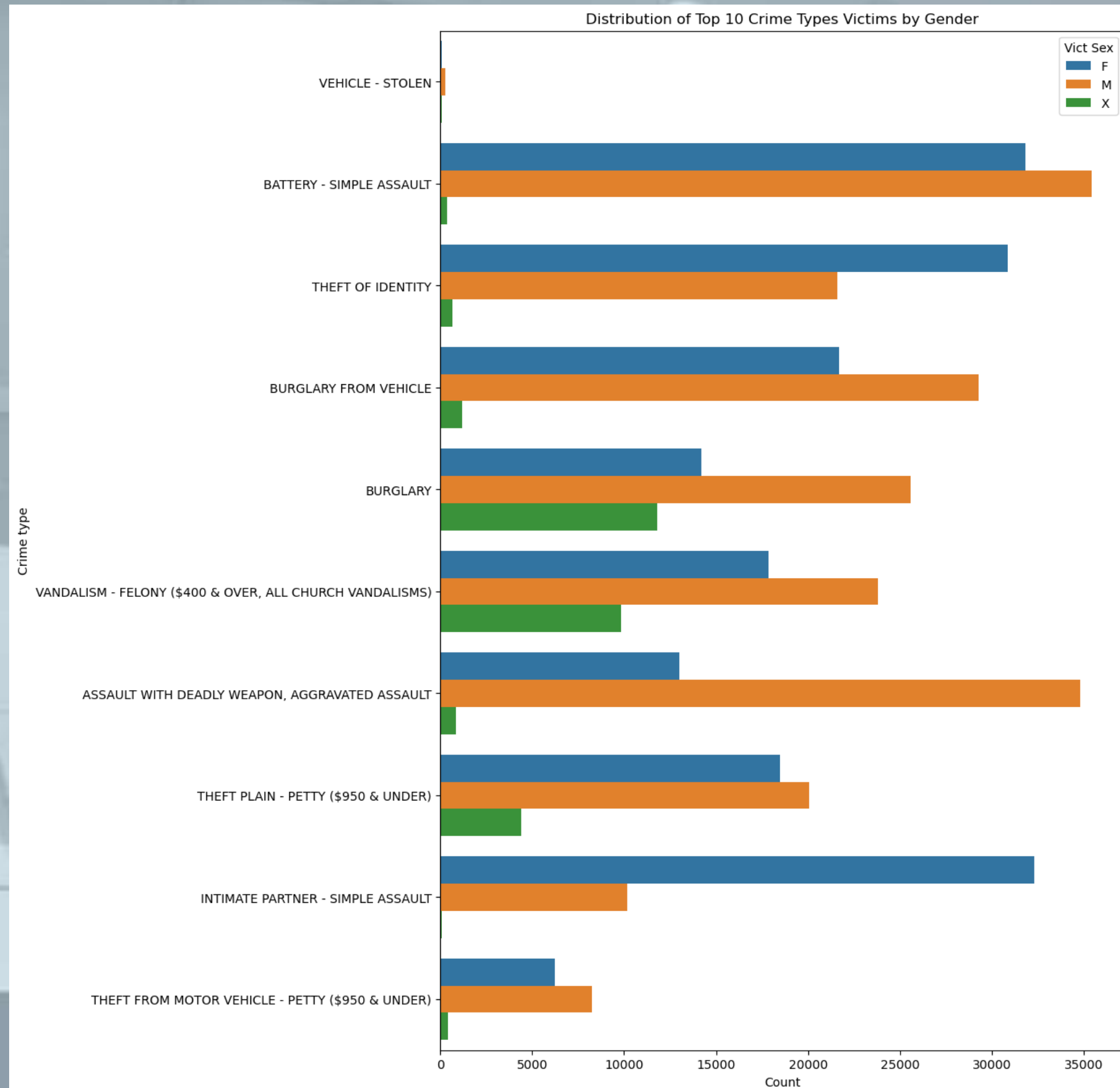
The top 3 crimes for which the female victims are targeted for are:

1. Intimate partner-Simple Assault
2. Battery – Simple Assault
3. Theft of Identity

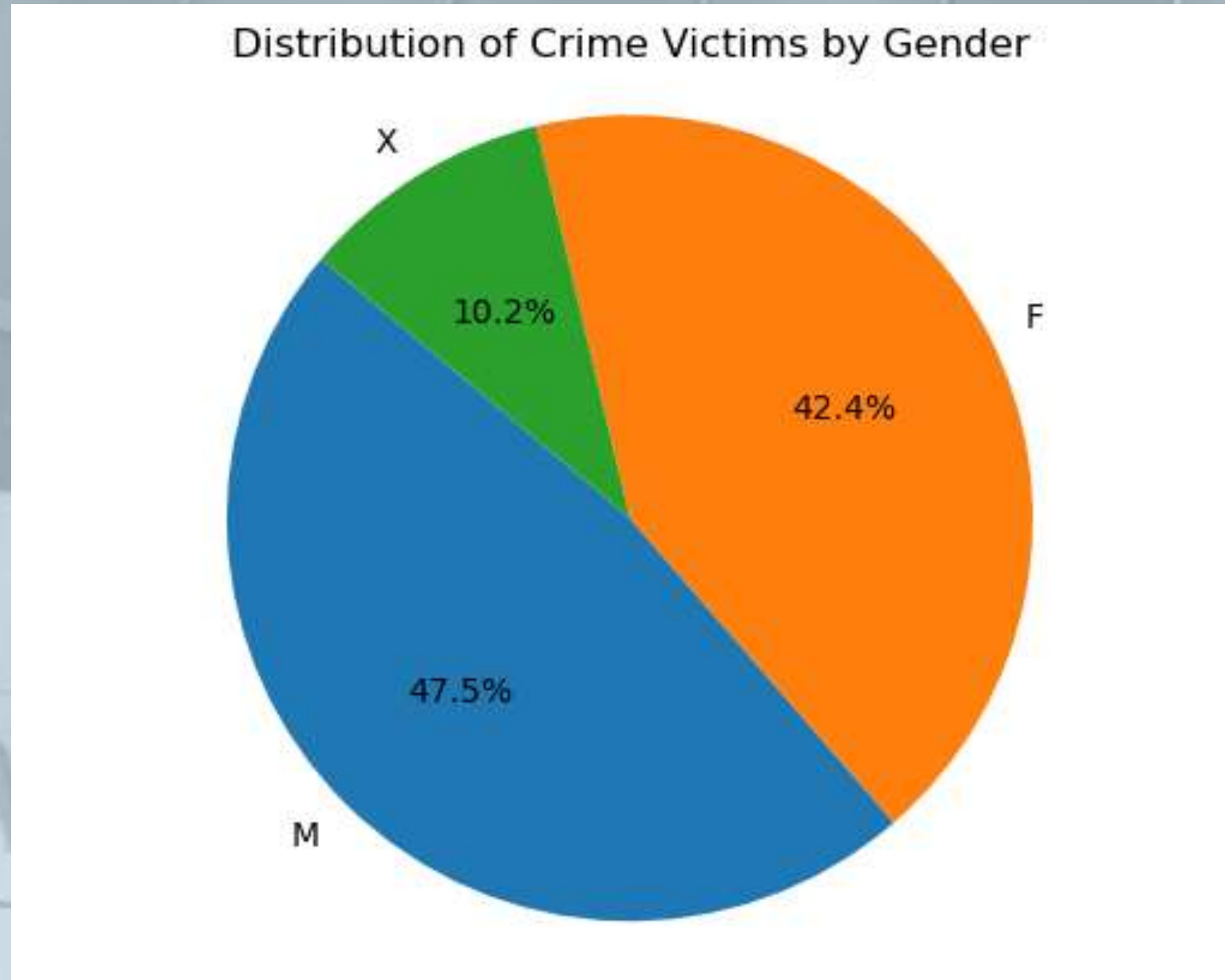
The top 3 crimes for which the male victims are targeted for are:

1. Battery – Simple Assault
2. Assault with deadly weapon, aggravated assault
3. Burglary from Vehicle

V6: Grouped bar plot showing the Distribution of Top 10 Crime Type Victims by Gender



V6: Pie chart showing the distribution of Crime Victims by Gender



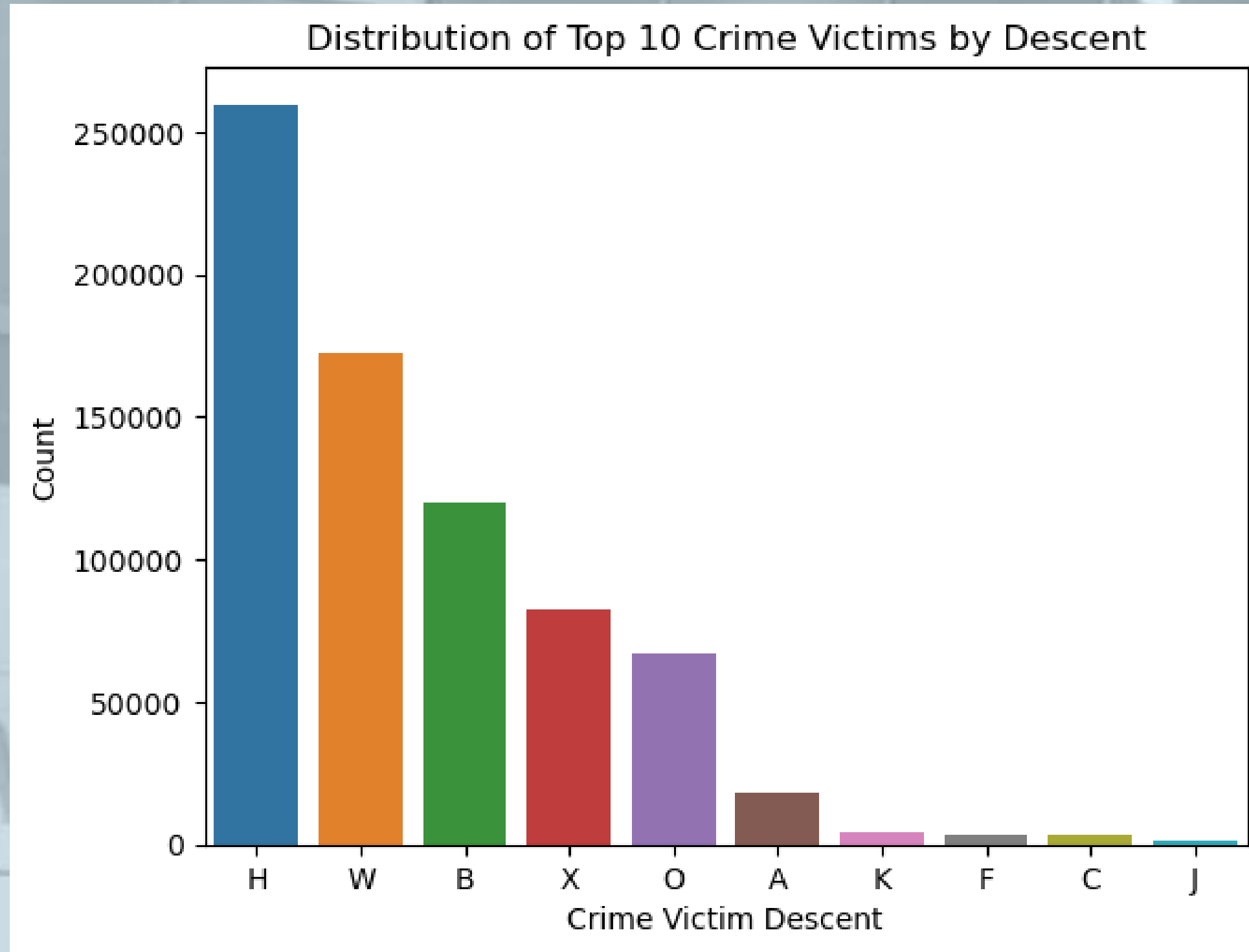
Distribution of Crime Type Victims by Descent

Looking at the visualizations, it's evident that the 3 ethnic groups that represent the highest percentage of crime victims by descent are Hispanic, White and Black.

1. Hispanic – Victims of Hispanic descent represent the highest percentage of crime victims: 35.8%.
2. White – White victims represent the second highest percentage of crime victims: 23.8%.
3. Black – Black victims represent the third highest percentage of crime victims: 16.6%.
4. Unknown – 9.3% of Victims are of unknown descent.

Since the start of the COVID-19 pandemic, there has been an increase in the number of Hispanic and Black victims has played out almost entirely among Latino and Black victims of homicides in LOS (Rector, 2021, paras. 1),

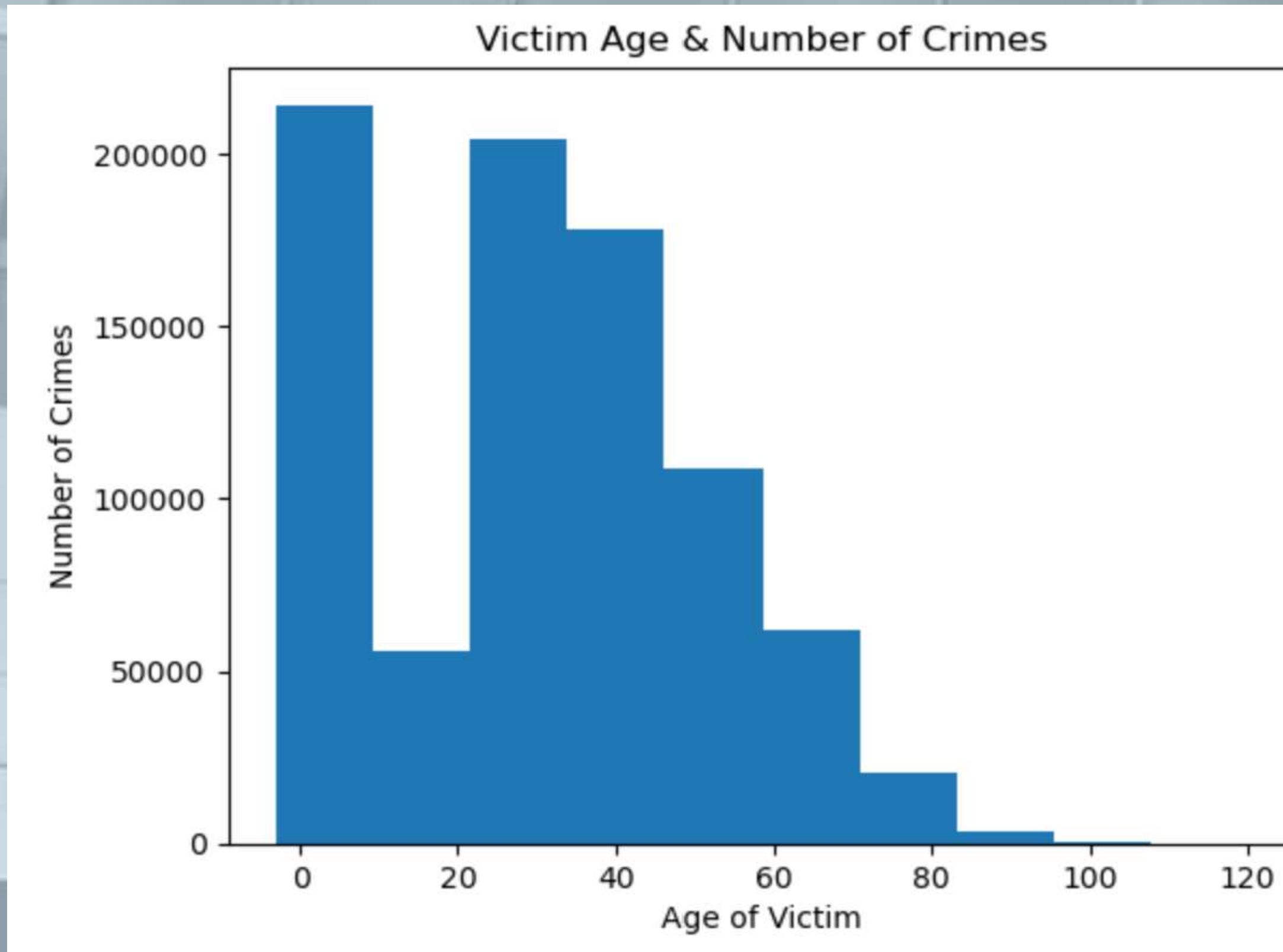
Barplot showing the distribution of Top 10 Crime Victims by Descent



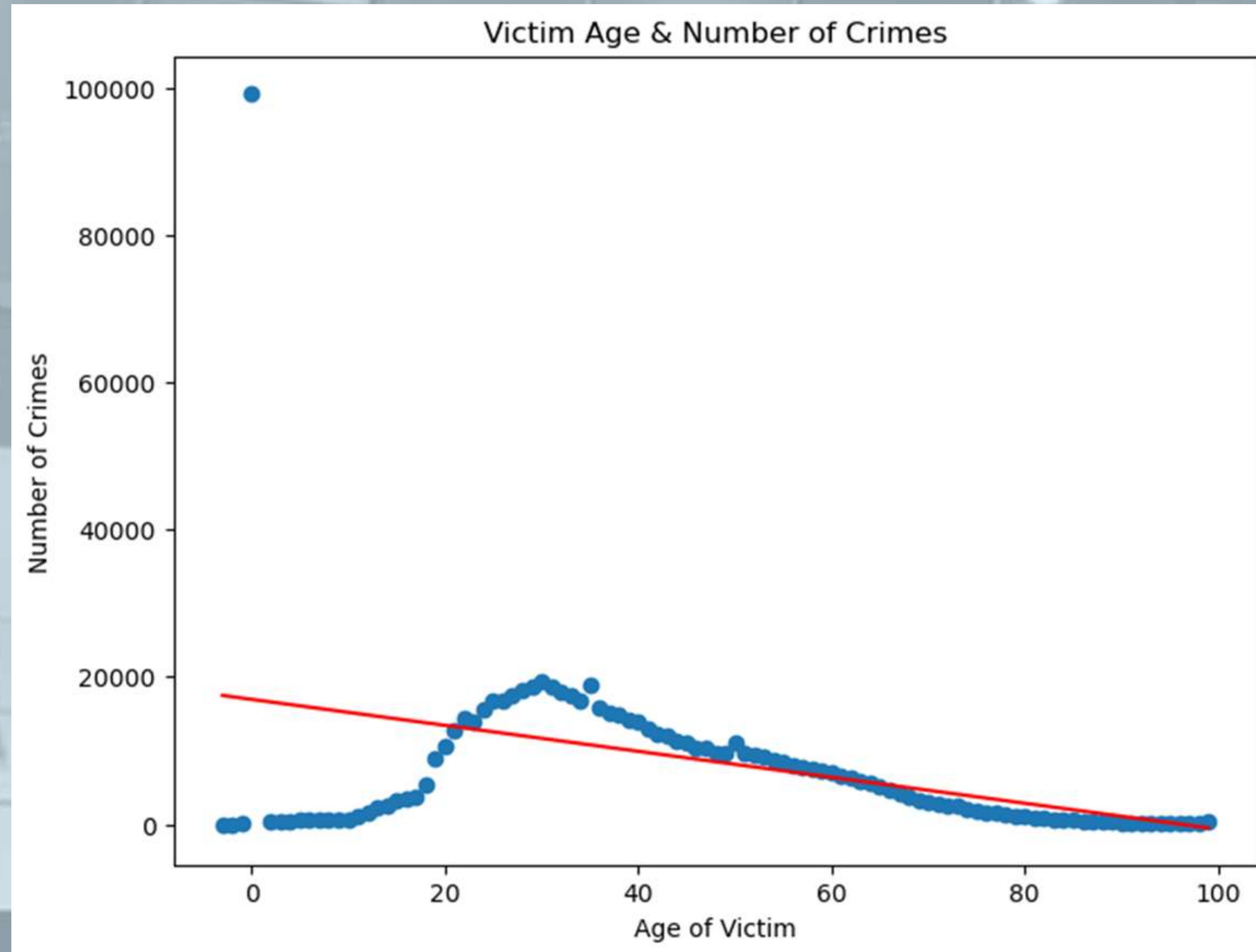
Q4.2: Are There Any Specific Groups of People Being Targeted? Focus: Age



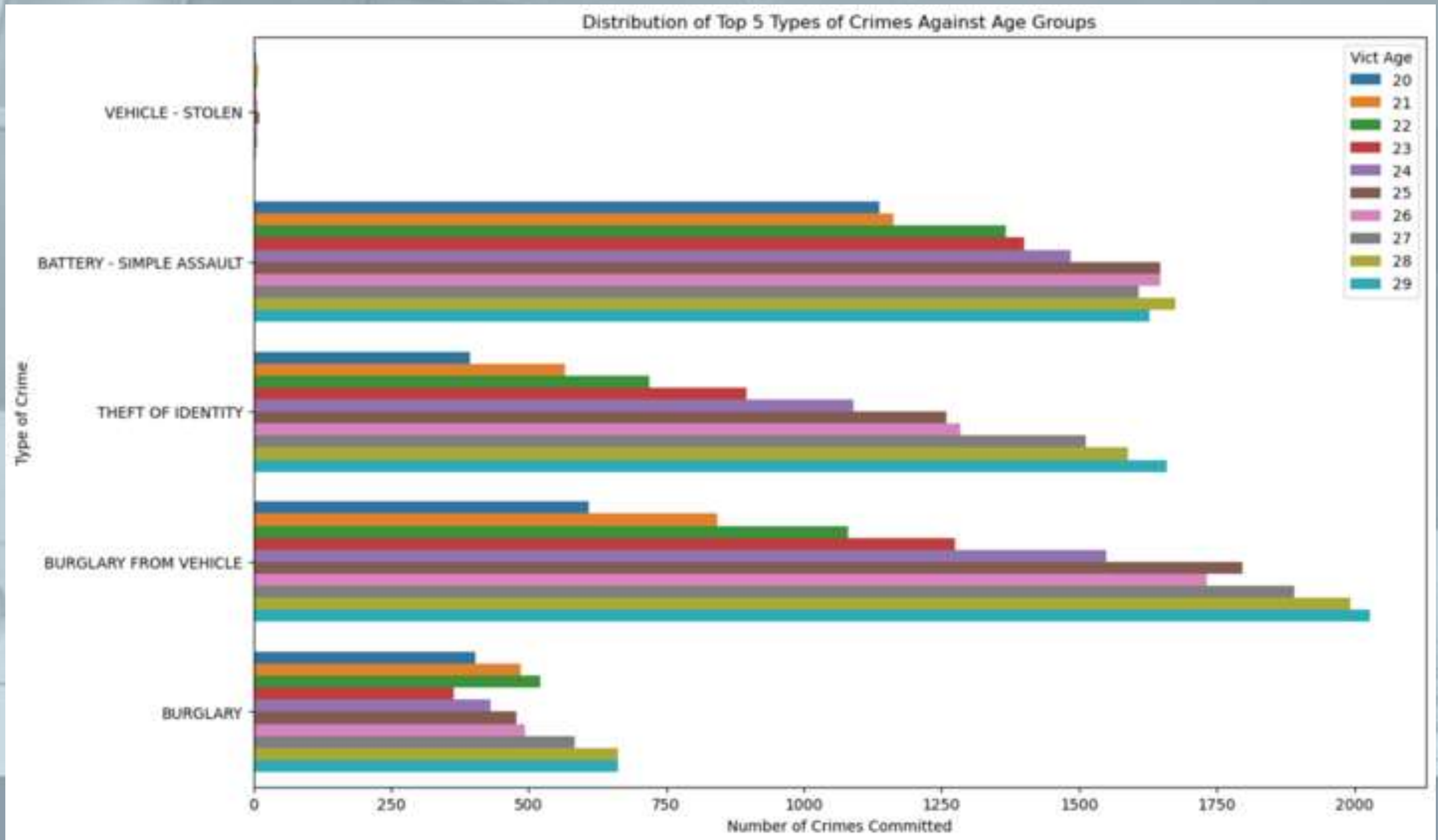
Barplot showing the distribution of Victims by Age



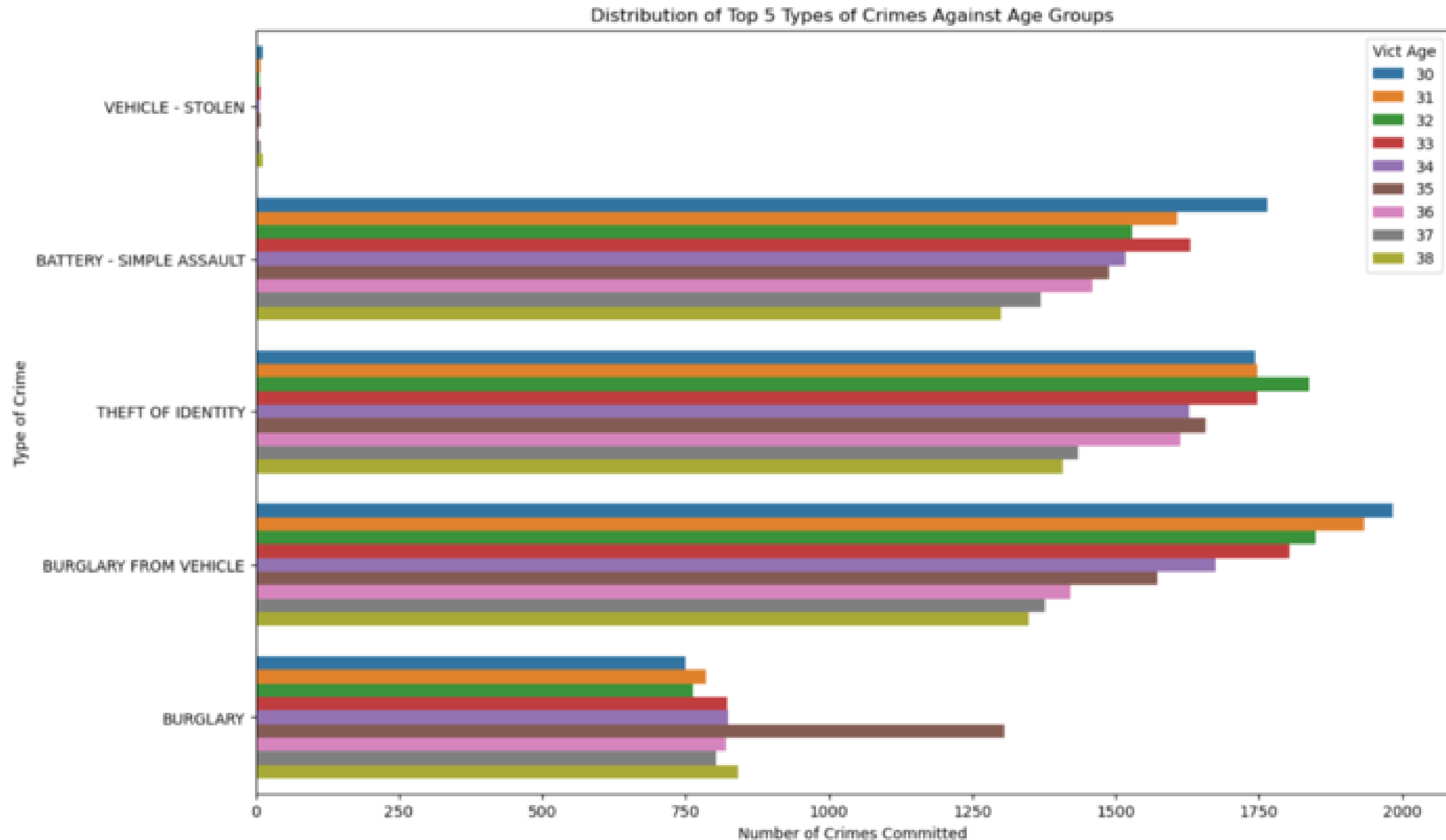
V7: Scatterplot displaying the Number of Crimes Committed Against Victims at Different Ages



Victims at Ages 20-29



Victims at Ages 30-39



Conclusion

Age and Crime Relationship:

- Crimes exhibit a clear age-related pattern, peaking around 30 and declining as age increases.

Peak Age Group:

- The highest crime frequency occurs in the 25 to 30 age group.

Negative Trend:

- A negative trend line emphasizes a consistent decrease in crimes as age advances.

Baseline and Endpoint:

- The trend line starts at 15,000 crimes, showcasing a baseline, and gradually approaches zero for older age groups.

Interpretation:

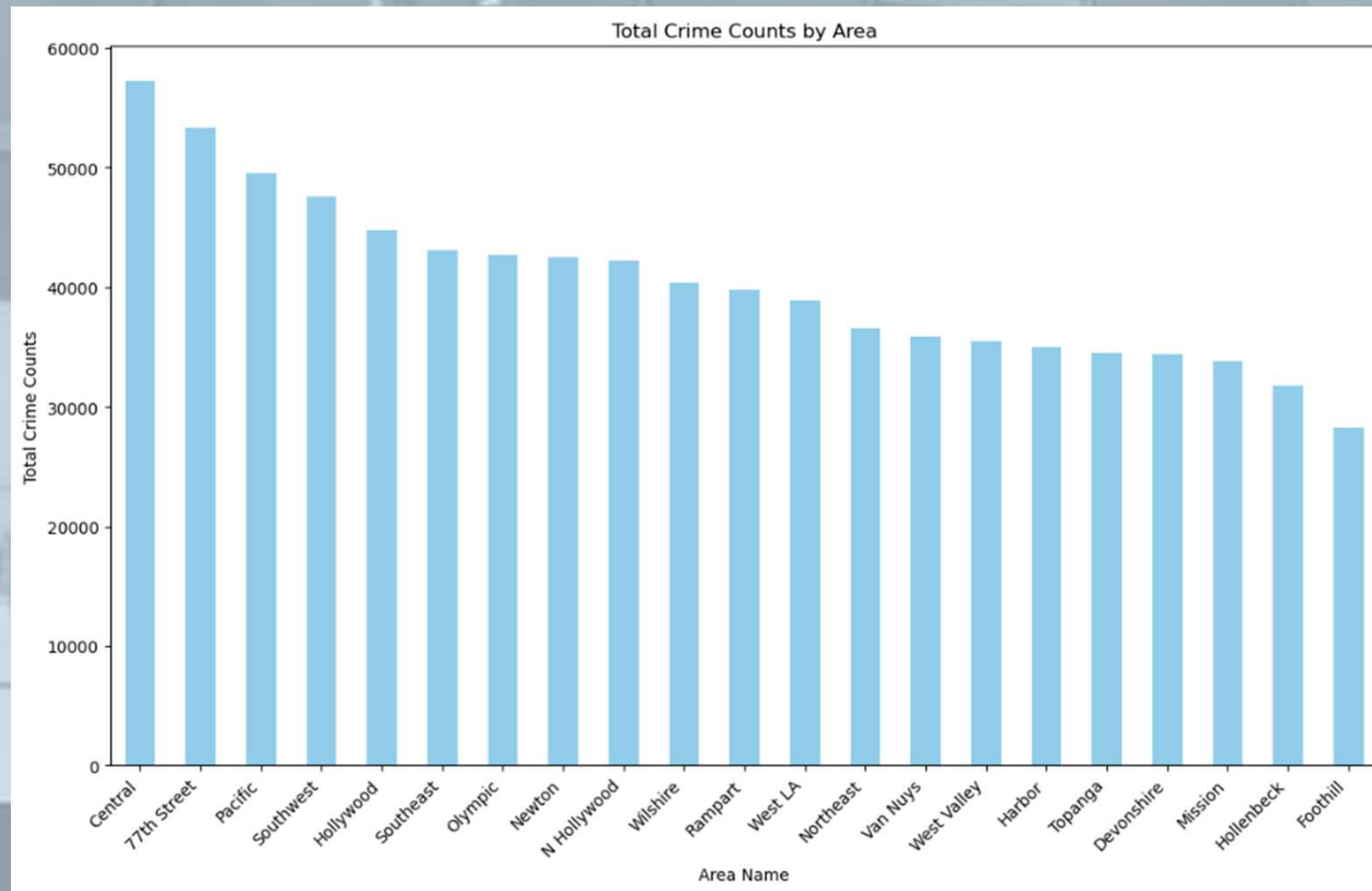
- The data underscores a distinct age-related trend, indicating a significant peak around 30, followed by a decline. The negative trend line suggests a diminishing victimization likelihood among older individuals, necessitating deeper analysis for a comprehensive understanding.

Additional Insights - Histogram Analysis:

- The histogram delves deeper into the age distribution of crimes, revealing nuanced insights. The 25 to 30 age group stands out with the highest crime frequency, warranting focused attention for crime prevention initiatives. Notably, age groups such as 58 to 73, 73 to 78, 78 to 90, and 90 to 100 exhibit lower crime frequencies, presenting opportunities for targeted interventions in these segments. The visual representation of the histogram aids in identifying concentration and dispersion of crime occurrences across different age brackets. This expanded summary provides a more detailed exploration of the histogram analysis, emphasizing its role in uncovering age-specific crime patterns and guiding strategic preventive measures.

Significance of our answers

- Are the variations in the crime rate across the different areas significant?
 - Central and 77th are 1 and 2 for highest crime
 - there is not enough evidence to conclude that there exist significant differences in crime count between the 21 areas (p-value=2.53)



Is crime increased during certain months in a year?

- There is not enough evidence to conclude that there exist significant differences in crime count between the months

year	2020	2021	2022	2023
month				
1	18496	16552	18448	19633
2	17255	15411	17700	18313
3	16165	16320	19687	19024
4	15683	16057	19784	18775
5	17204	16989	20418	18738
6	17034	17156	20201	18500
7	17121	18659	19952	19697
8	16874	18374	20081	19743
9	15639	18362	19279	18861
10	16491	19305	20274	19394
11	15574	18340	18688	17358
12	15946	17926	19924	1545

Is crime consistent for each month over the years?

- Weak to medium correlations across the board, meaning there is no particular month thats sees a consistent trend

year	2020	2021	2022	2023
year				
2020	1.000000	-0.266965	-0.196548	0.304204
2021	-0.266965	1.000000	0.461569	-0.091707
2022	-0.196548	0.461569	1.000000	-0.093743
2023	0.304204	-0.091707	-0.093743	1.000000

Vehicle Theft is the most common crime. Is the contribution specifically from one area?

- There is not enough evidence to conclude that one area significantly contributes to vehicle theft

AREA	
1	3755
2	4366
3	5204
4	5029
5	4885
6	3211
7	3326
8	2831
9	3636
10	3393
11	4337
12	7046
13	6693
14	5222
15	3960
16	3585
17	3151
18	5825
19	4662
20	4652
21	2704
Name: VEHICLE - STOLEN,	

Are there any specific groups of people being targeted?

- There is not enough evidence to conclude that one age group is being specifically targeted, significantly (p-value = 3.83)
- Men are victims significantly more than the other (p-value = 0.003)

```
F-statistic: 307.8055140858127  
p-value: 0.003233057840408918  
{ 'F': 313468, 'M': 351362 }
```

Contributions to Project Questions

Question 1 (Crime distribution by area):

Fabia's expertise in data collection and preprocessing, combined with Raymon's prowess in data visualization, resulted in a bar plot displaying total crime counts for each area. Additionally, statistical analysis, particularly the t-test comparing crime counts between different areas, was conducted.

Question 2 (Monthly crime trends over the years):

Mike's leadership in statistical analysis and Jia's contributions to interpreting findings led to comprehensive insights into crime trends. Visualizations, such as line plots for crime trends and correlation analysis using scatter plots, were key aspects of this investigation.

Question 3 (Distribution of crime types):

Raymon's expertise in data visualization contributed to the creation of pie charts or bar plots illustrating the distribution of crime types. Statistical analysis was conducted using the chi-square test to explore associations between crime types and other categorical variables.

Question 4 (Demographics of crime victims):

Visualizations depicting demographic distributions were a result of Ariana's and Jia's contributions. Additionally, statistical tests, including t-tests, were conducted to compare different demographic groups based on victim demographics.

BIBLIOGRAPHY

ASSOCIATED PRESS. (2021, July 1). California homicides jump 31% in 2020, reaching highest total in 13 years. *Los Angeles Times*.

<https://www.latimes.com/california/story/2021-07-01/california-homicides-jump-31-in-2020-the-most-in-13-years>

Deliso, M and Stone, A. (2023, February 13). Alleged gunman in custody after 2 Jewish men shot in suspected hate crime: Police. *Abc News*.

<https://abcnews.go.com/US/2-jewish-men-reportedly-shot-religious-services-la/story?id=97269248>

Los Angeles Police Department. (n.d.). *The Official Site of the Los Angeles Police Department*. <https://www.lapdonline.org/#>

Rector, K. (2021, August 13). Latino and Black victims account for nearly all of L.A.'s surge in homicides. *Los Angeles Times*.

<https://www.latimes.com/california/story/2021-08-13/la-homicides-since-covid-take-mostly-black-latino-victims>

U.S. Government's Open Data. (2023). *Crime Data from 2020 to Present* [Data set]. <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>