SD Police Vehicle Stop Analysis

Group 14
Raul Pegan
Sikai Liu
Guangjun Xue
Lingfeng Chen

Motivation

- San Diego is a highly populated city with a lot of vehicle traffic
- Law enforcement is frequently monitoring this traffic
- We wish to provide the common drivers with insight as to how to better avoid encounters with the police

Motivation

- Frequently, it has been observed that law enforcement might have a negative bias towards certain demographic groups and behaviors
- We will examine San Diego's police data to observe any trends regarding this topic

Methodology

Data Source:

City of San Diego Open Data Portal (https://data.sandiego.gov)

Visualization Method

Geographic information system (GIS) - geopandas

Stacked bars, line plot, Pi charts - pandas, matplotlib

Part 1: Map

1. **Which regions in San Diego have the most stops?**
2. **Which regions have the highest probability of arrest?**

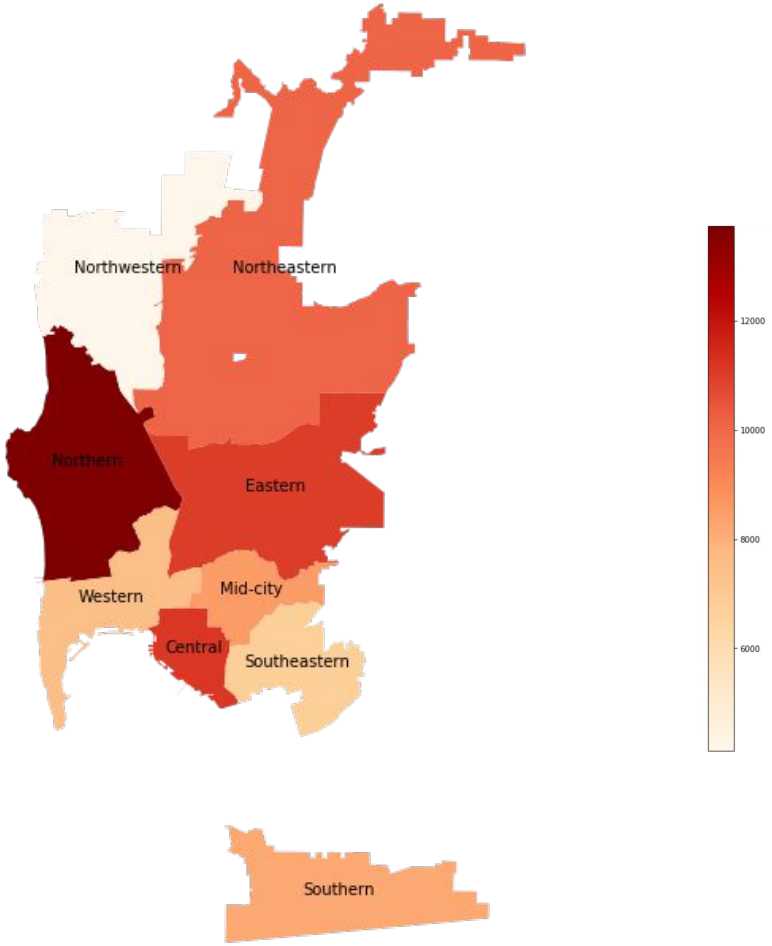
San Diego county is subdivided into 9 different areas:

Southern, Southeastern, Central, Mid-city, Western, Eastern, Northern, Northeastern, Northwestern.

Part 1: Map

Number of Stops

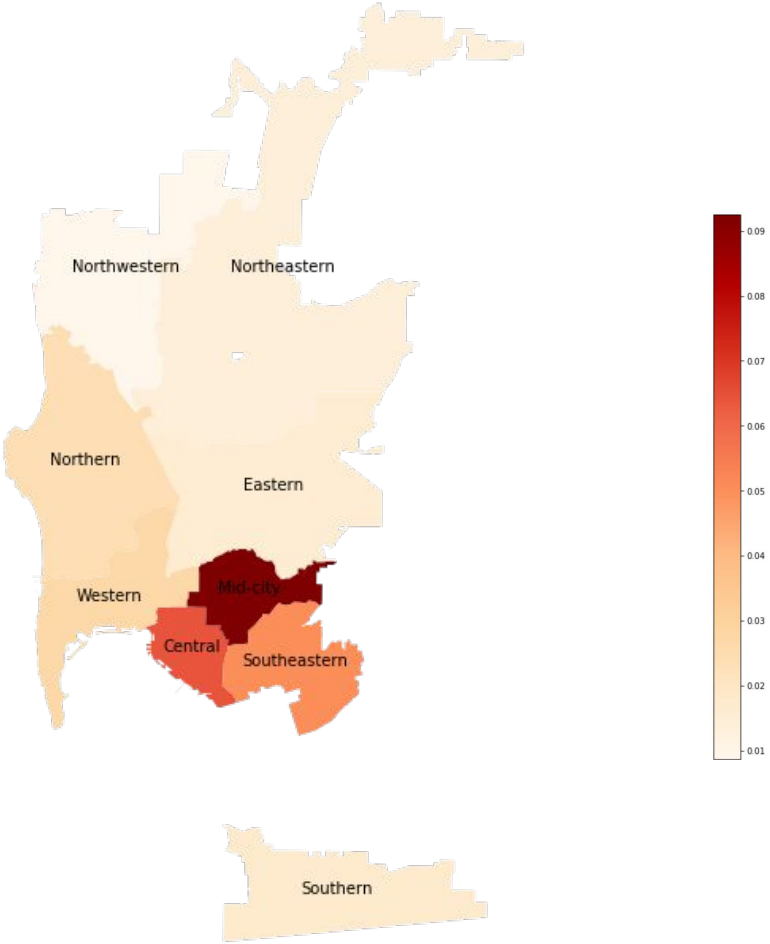
- Northern San Diego has the highest amount of traffic stops



Part 1: Map

Arrest probability

- Mid-city San Diego has the highest probability of arrest
- This is one of the most diverse places in the US



Part 2: Time VS Stops

1. Which periods in time have more stops?

    Which month?

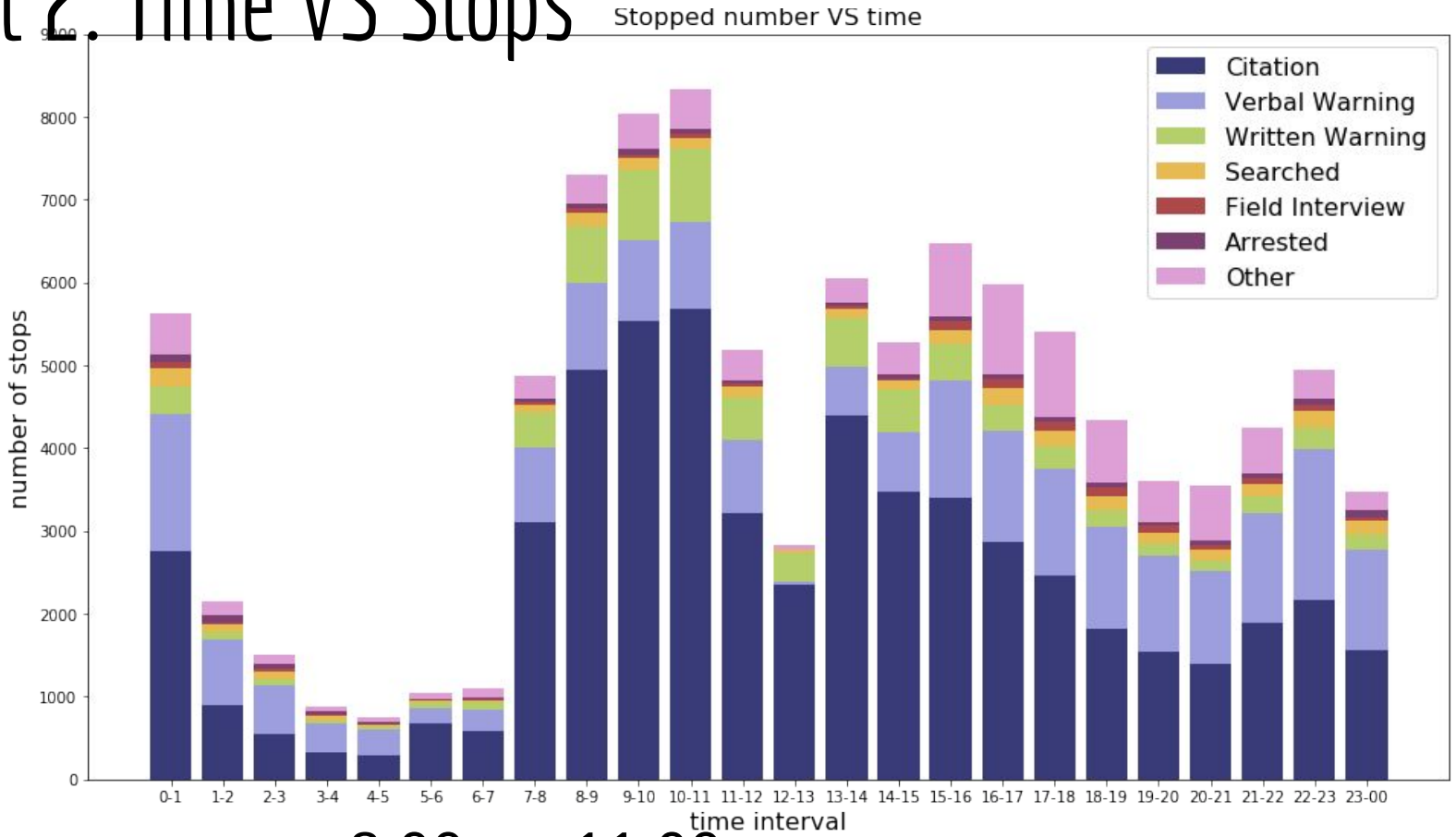
    Which weekday?

    Which time of a day?

2. How did the stop end up?

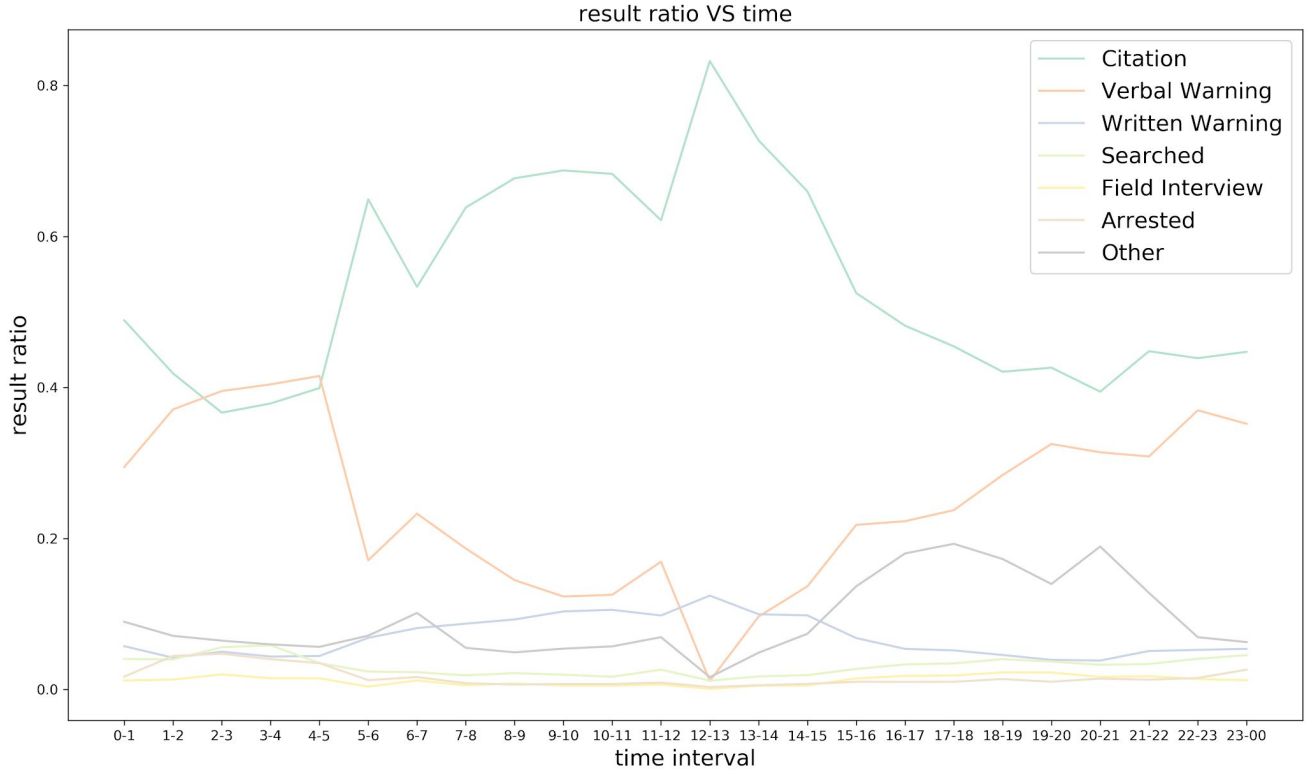
    Arrested, Searched, Citation, Verbal Warning, Writing Warning, Field Interview...

Part 2: Time VS Stops



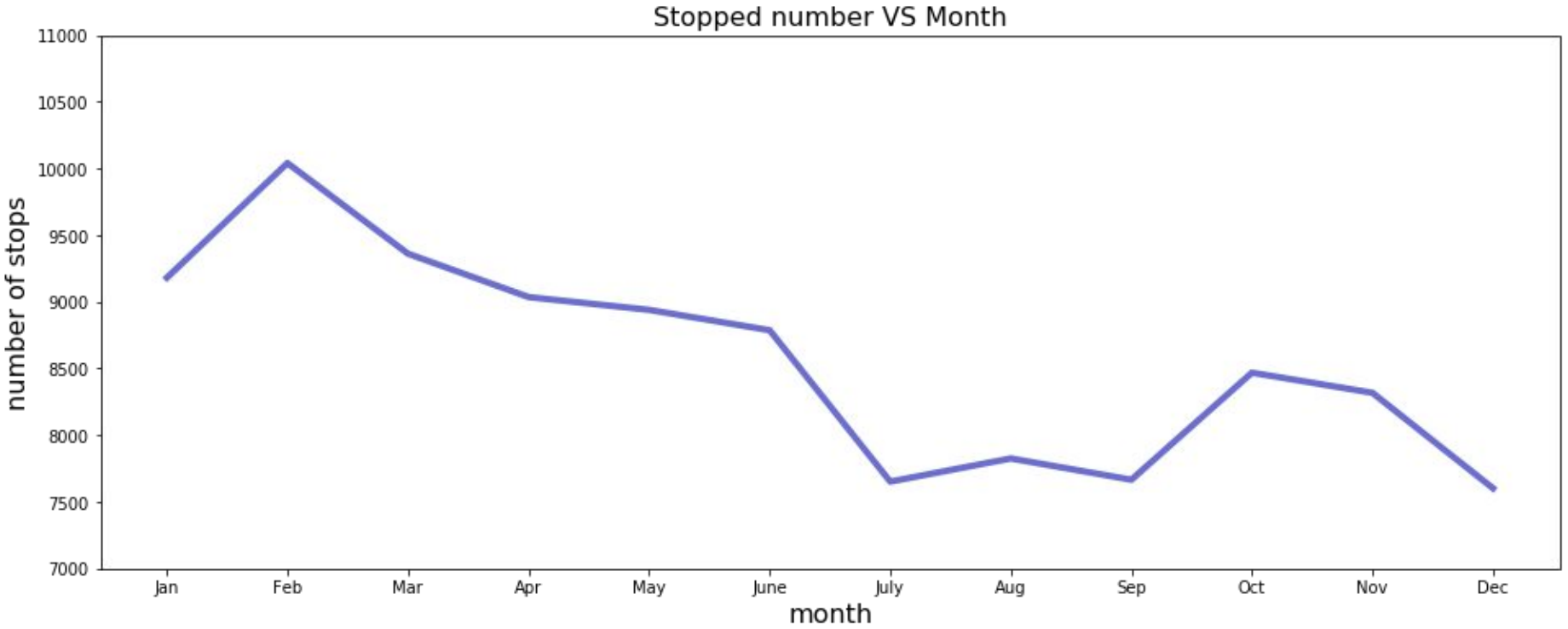
8:00am-11:00am: most likely to be stopped

Part 2: Time VS Stops



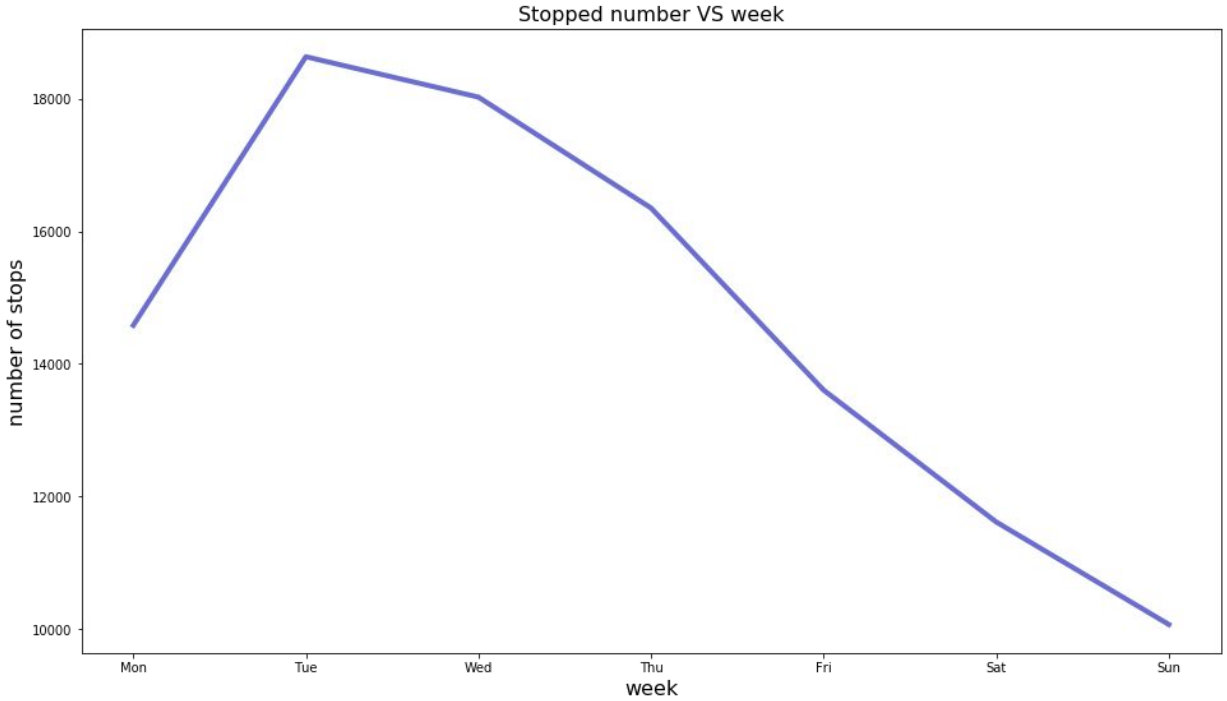
Citation: most common result after being stopped

Part 2: Time VS Stops



No obvious differences among months

Part 2: Time VS Stops



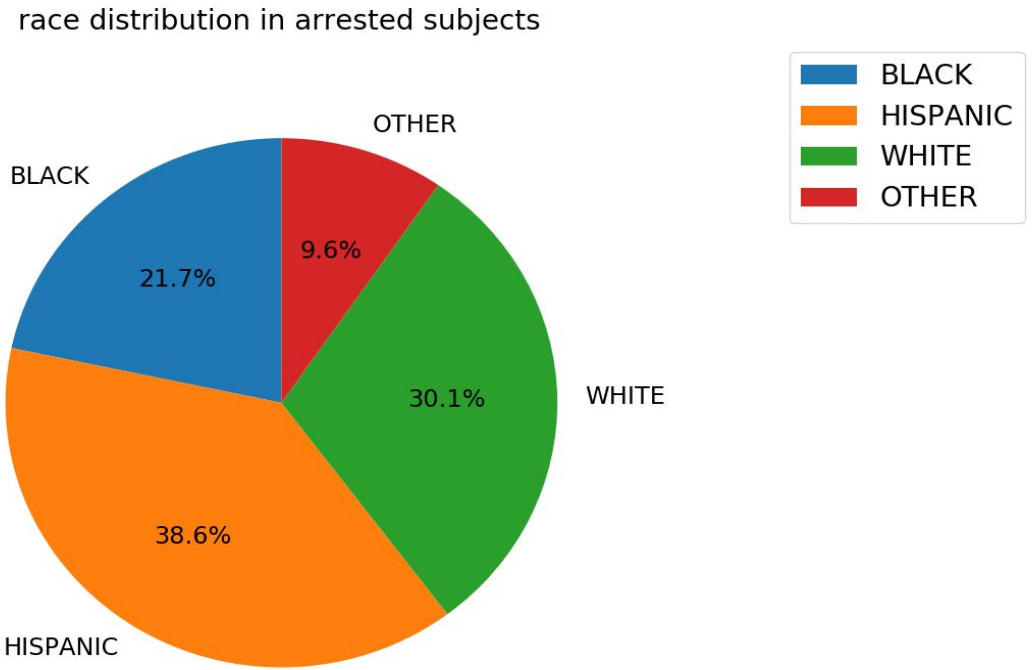
30%-80% more likely to be stopped on **weekdays** than on weekends（Rush Hour!）

Part 3: Race and Arrests

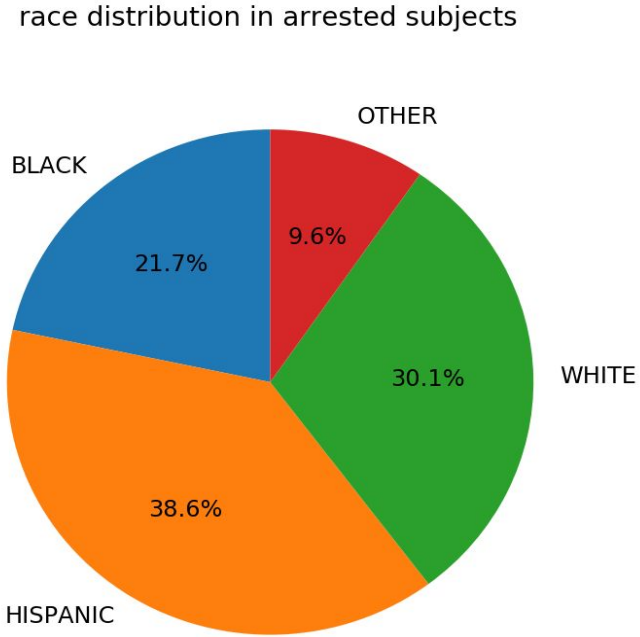
1. What is the race distribution of the arrested subjects?

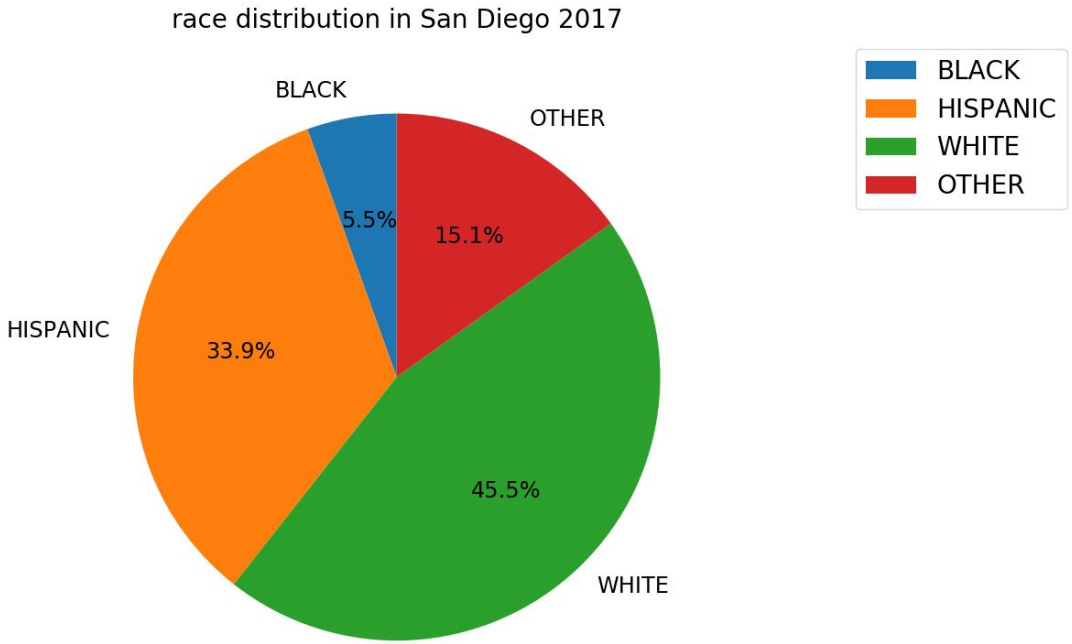
2. Is there any relationship between the race distribution of arrested subjects and San Diego race demographics?

Part 3: Race and Arrests



Arrested subjects mainly consist of Hispanic, White, and Black.





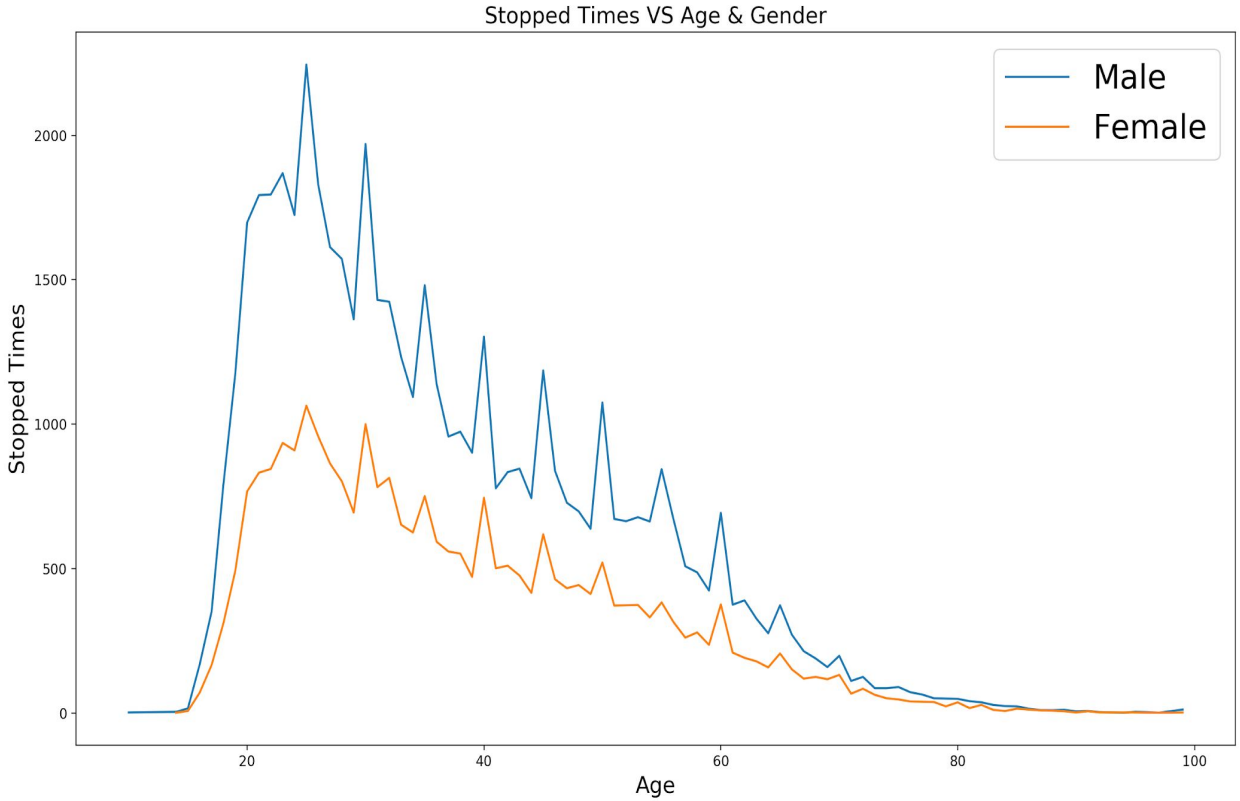
The amount of arrests are not proportional to the population distribution

Part 4: Age, Gender and Searched

1. **At what age are people more likely to be searched when stopped?**
2. **People of which gender are more likely to be searched when stopped?**

We have the dataset for people at ages ranging from 14~93. Using the number of stops and searches, we derive the probability of being searched.

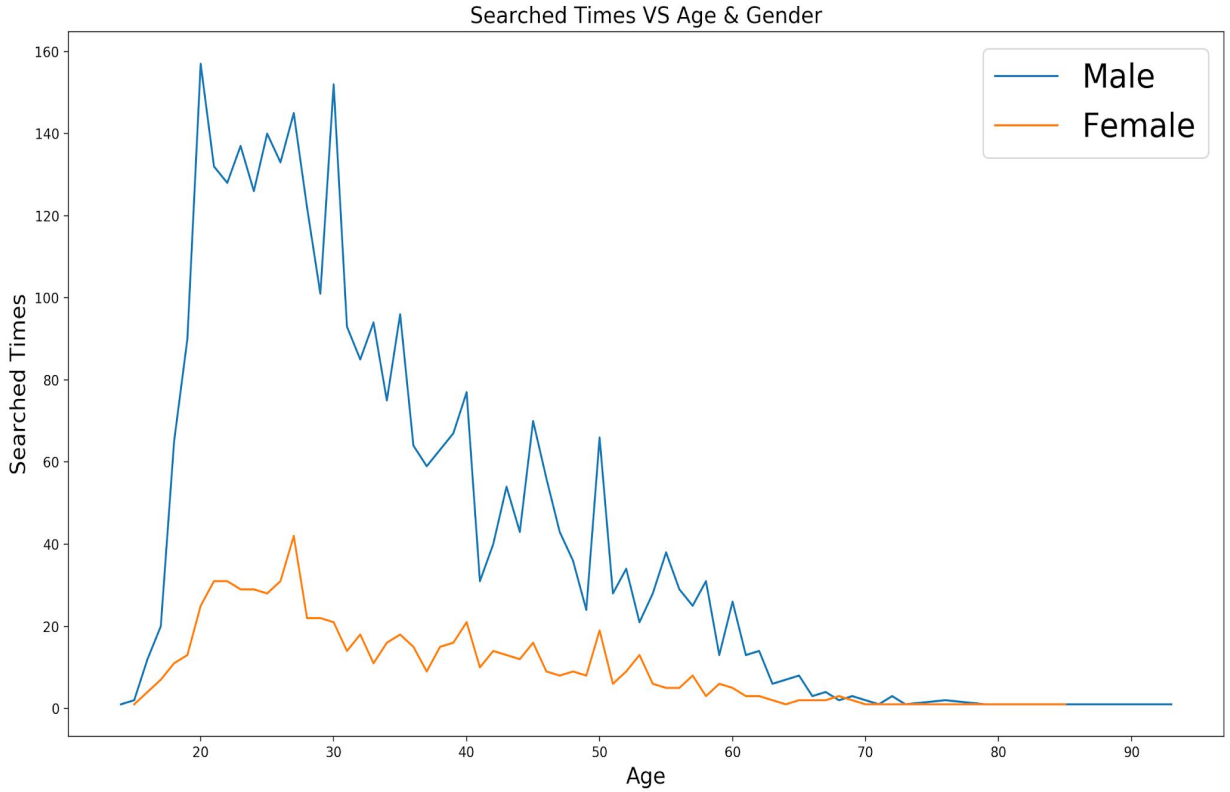
Part 4: Age, Gender and Searched - Stopped Times



- Stopped Times of Men are greater than Women.

- Most drivers are 20~60 years old.

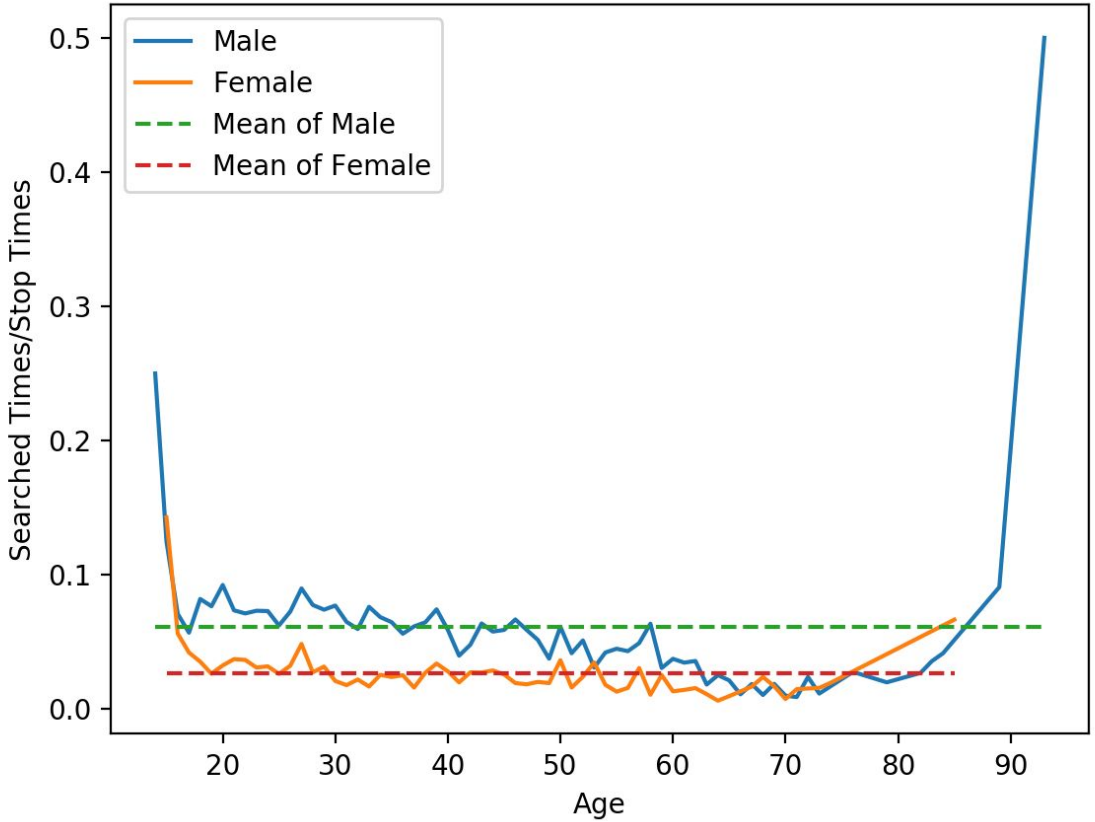
Part 4: Age, Gender and Searched - Searched Times



- Searched Times of Men are greater than Women.

- The Line of Women is much smoother.

Part 4: Age, Gender and Searched - Probability



- The right and left side data is much higher than the average of the rest. Why?

-Drivers of under 16 and over 90 highly likely to be searched.

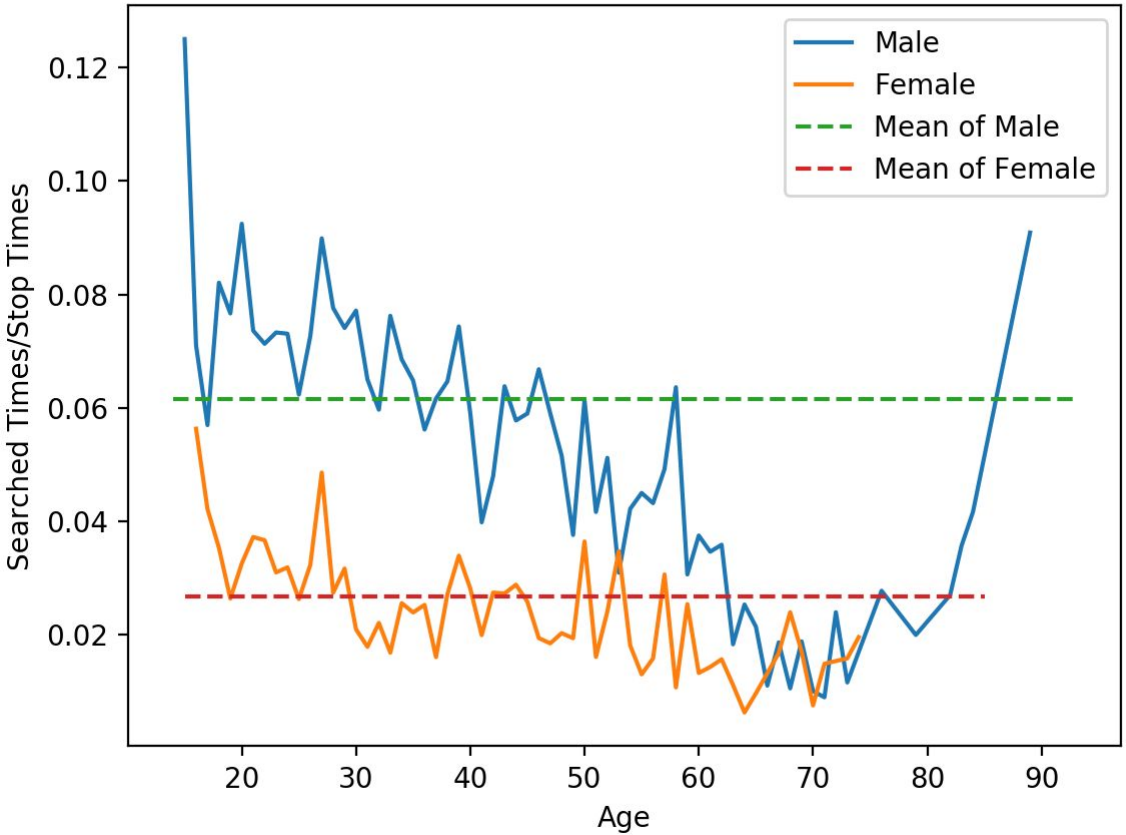
- Let's focus on the central part.

Part 4: Age, Gender and Searched - Probability

ZOOMED Chart for Probability

- The average probability of being searched of Men is twice as much as Women.

- The Young are more likely to be searched.



Conclusion

- **Rush Hour:** Higher probability to be stopped
- **Citation:** Most common outcome
- **Mid-city San Diego:** Most arrests per stops
- **Men**: Twice more likely to be searched
- **Arrests:** Not proportional to the population distribution

Outliers in datasets tend to skew our perception

Don’t break the law