Modeling Crime in Los Angeles

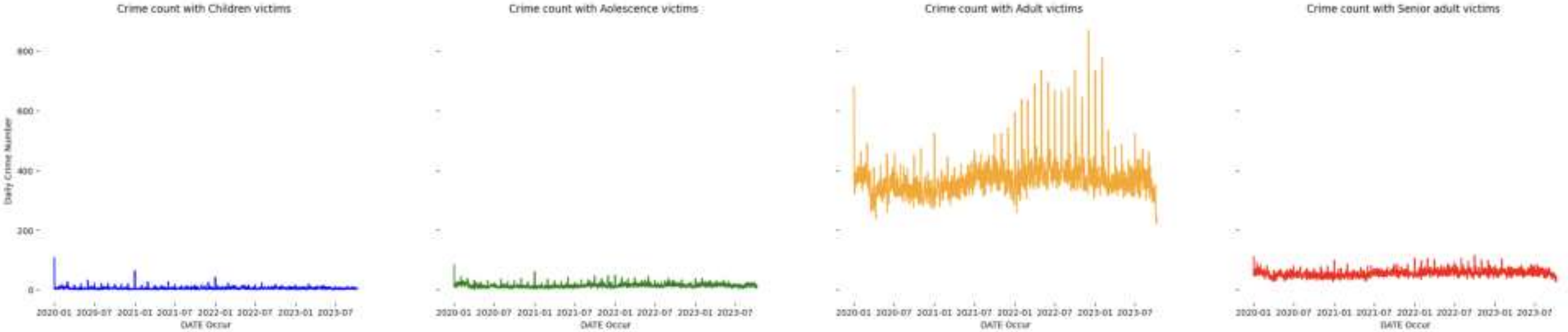
Julia McNerney, Yue Wan & Xukan Zhang ACMS40842

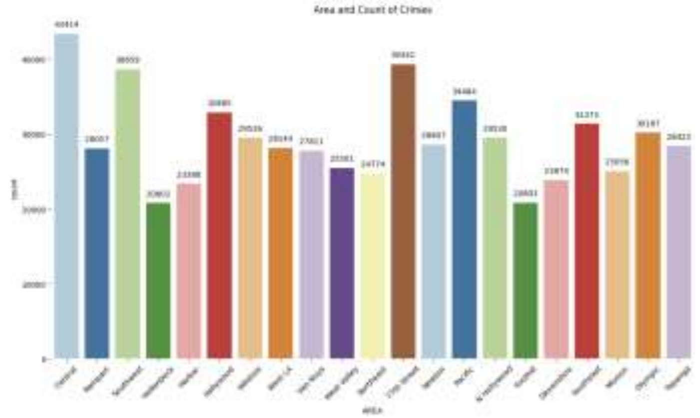
Background

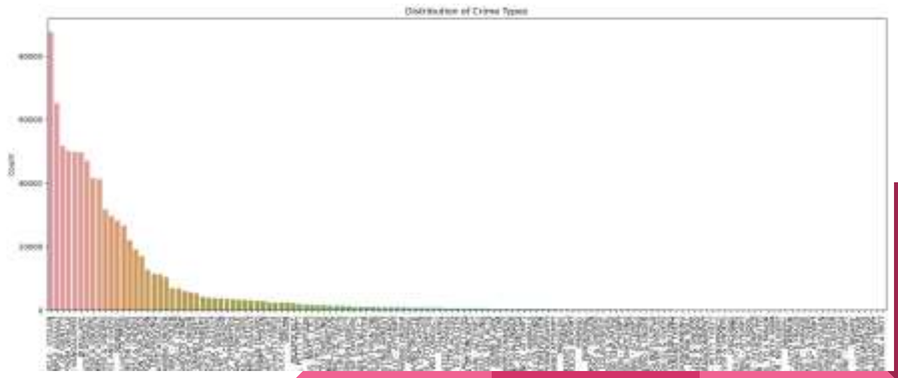
- Los Angeles ranks 17th in crime rate for US cities with population >100,000
  - Highest among the biggest cities in the US (NYC, Chicago, Houston, Phoenix, Philadelphia)
- 3rd largest police department in US
  - $1.74 billion budge
  - $420 per resident
- One of the most diverse cities in the US
  - 50% latino, 11% asian, 10% black

Dataset Overview: 01/01/2020 - 10/08/2023   Over 700k Crime Case

Selected Features: Area, Crime Type, Victim Sex, Age & Descent, Exact Time





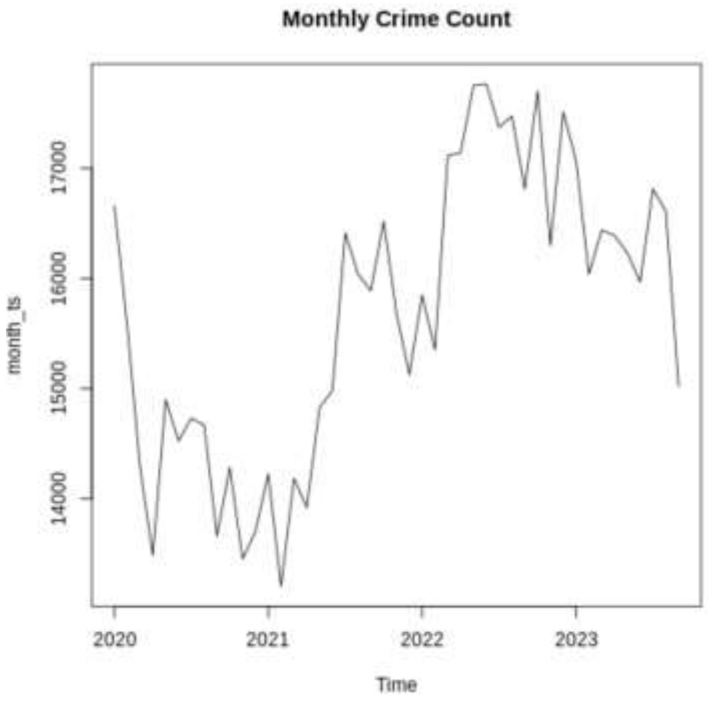


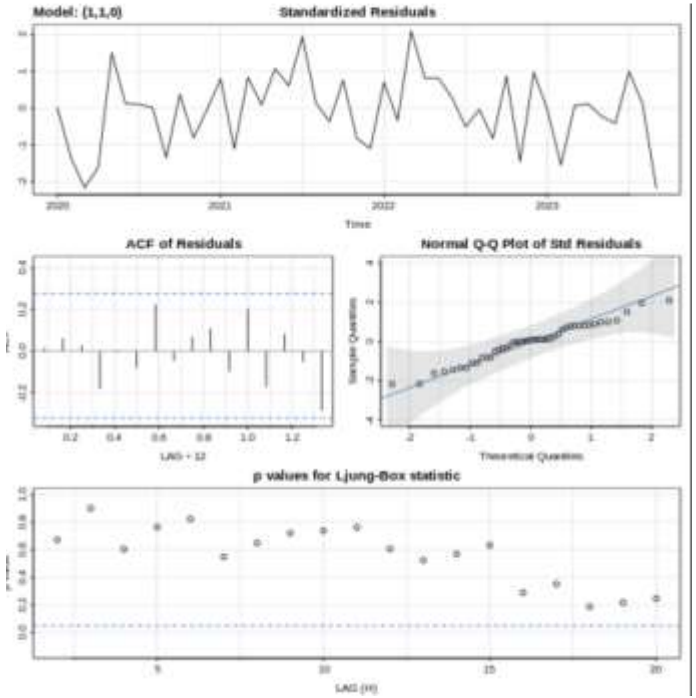
Research Questions

Examine how crime trends differ for different groups of victims.

- Enable policy makers to create more appropriate solutions to fighting crime
- Deeper understanding of who is most affected by crime
- Help the general public become aware of the possible crimes

Overall crime count model (monthly)

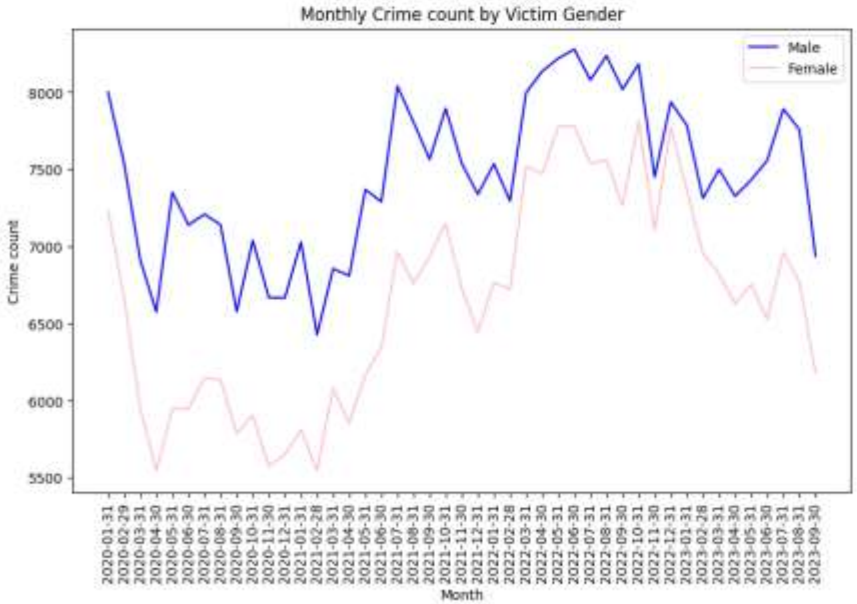


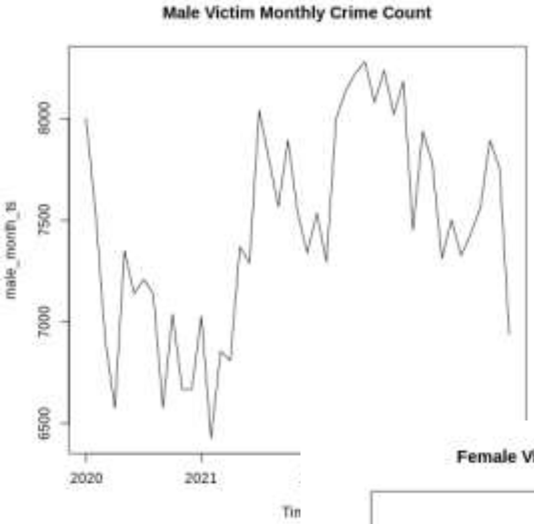


ARIMA(1,1,0)

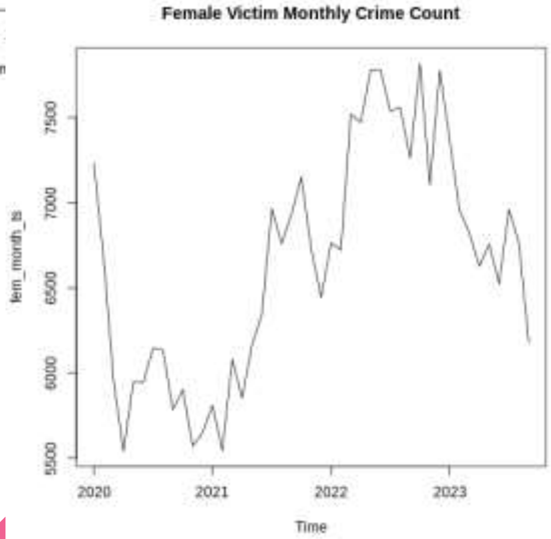
No seasonality

Segmented by Victims' Gender (Monthly): EDA



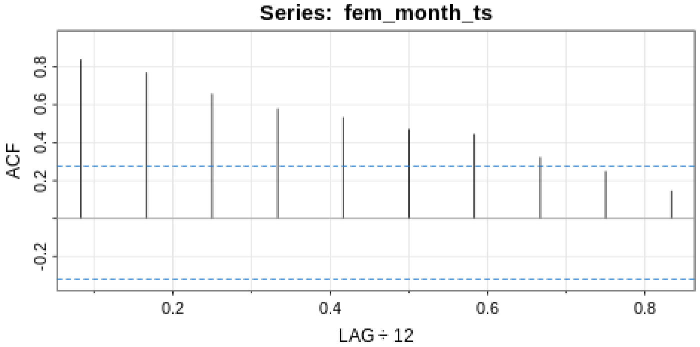


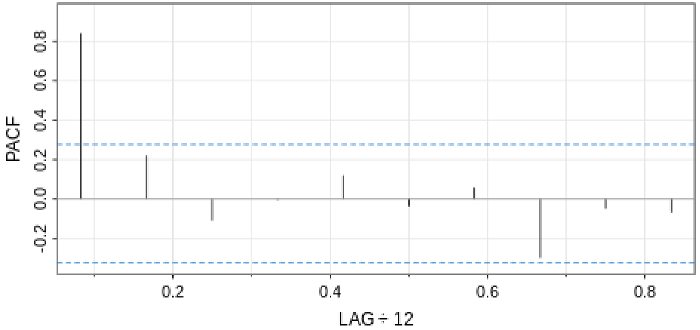
Largely similar, but enough differences to explore further



Segmented by Victims' Gender (Monthly): EDA

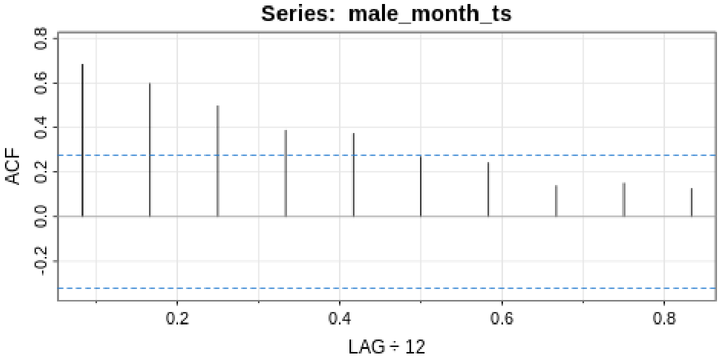
Female Victims

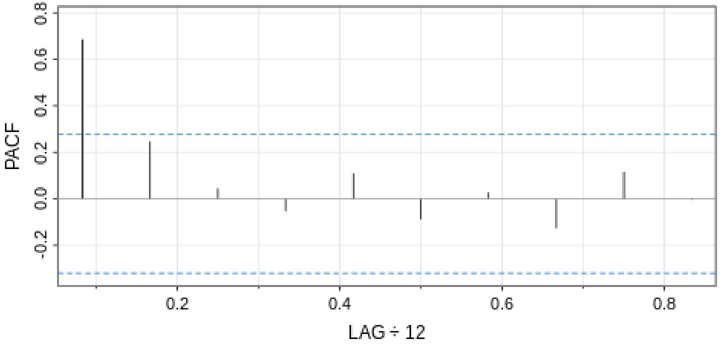




Indicate AR(1) Model

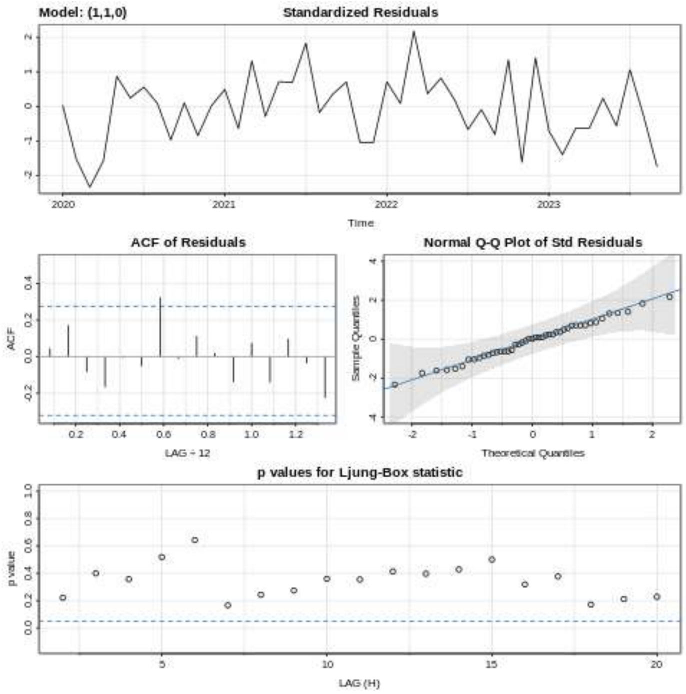
Male Victims





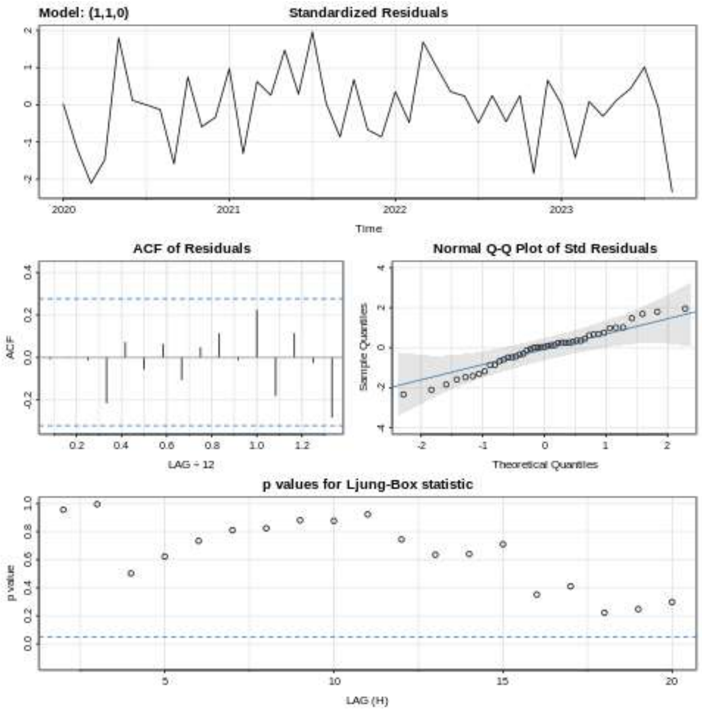
Segmented by Victims’ Gender (Monthly): Model Selection

Female Victims: ARIMA(1,1,0)



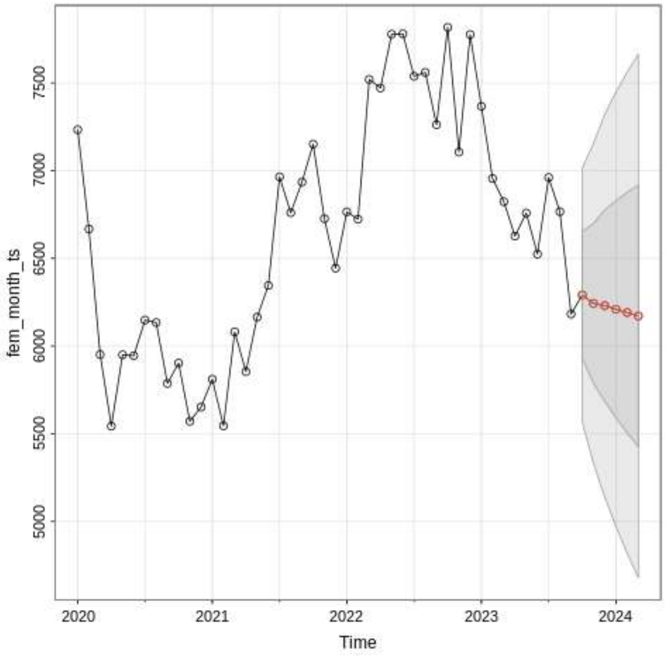
All diagnostic plots look good

Male Victims: ARIMA(1,1,0)



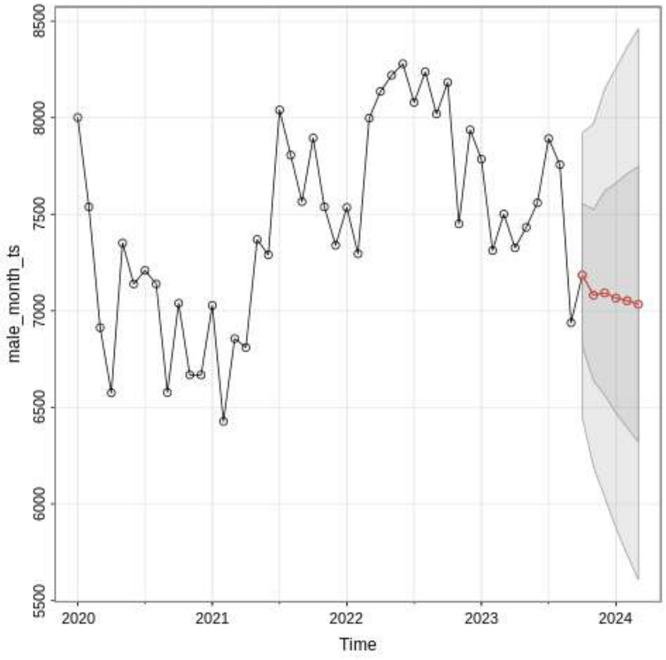
Segmented by Victims' Gender (Monthly): Forecasting

Female Victims:



Decreasing prediction

Male Victims



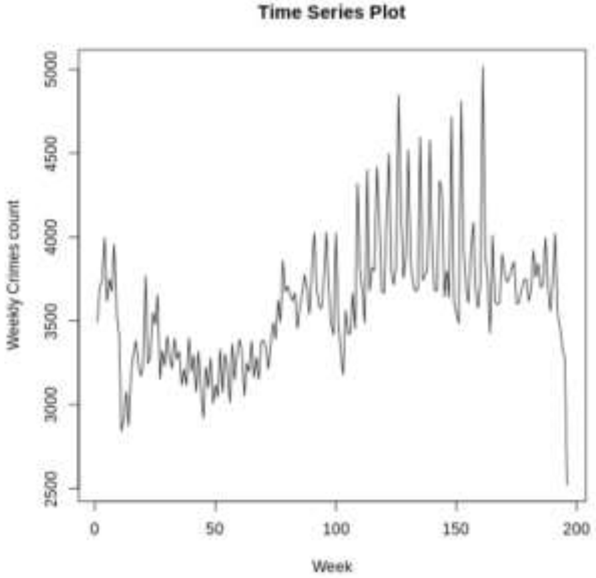
Summary of monthly trend

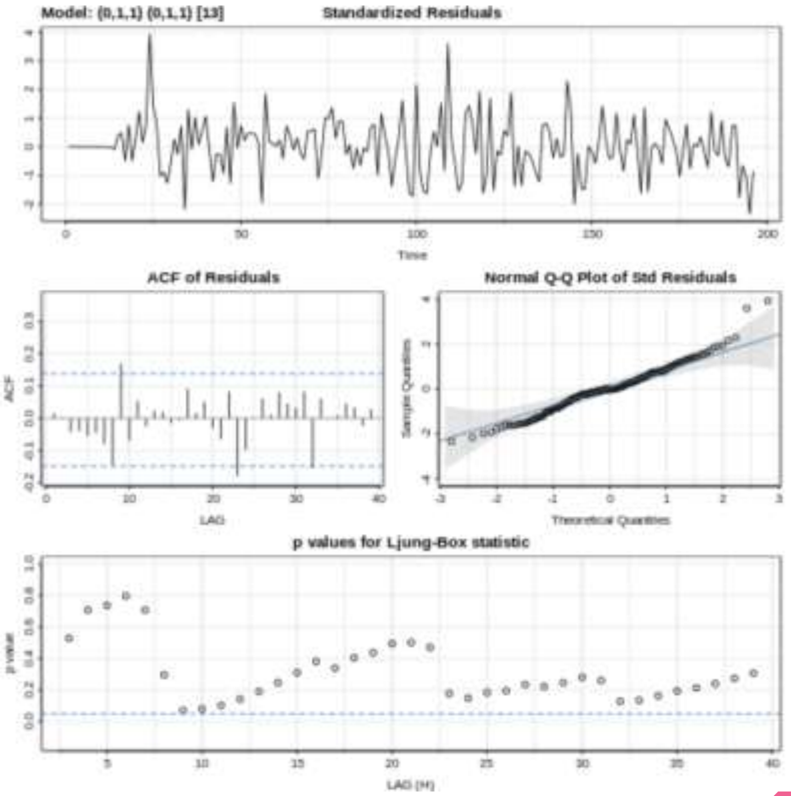
No difference in crime count trends toward male and female victims

- Efforts aimed to reduce crime against males or females may be effective in reducing crime overall

- Analysis into type of crime, area, etc may be more insightful when dealing with gender of victim

Overall crime count model (weekly)

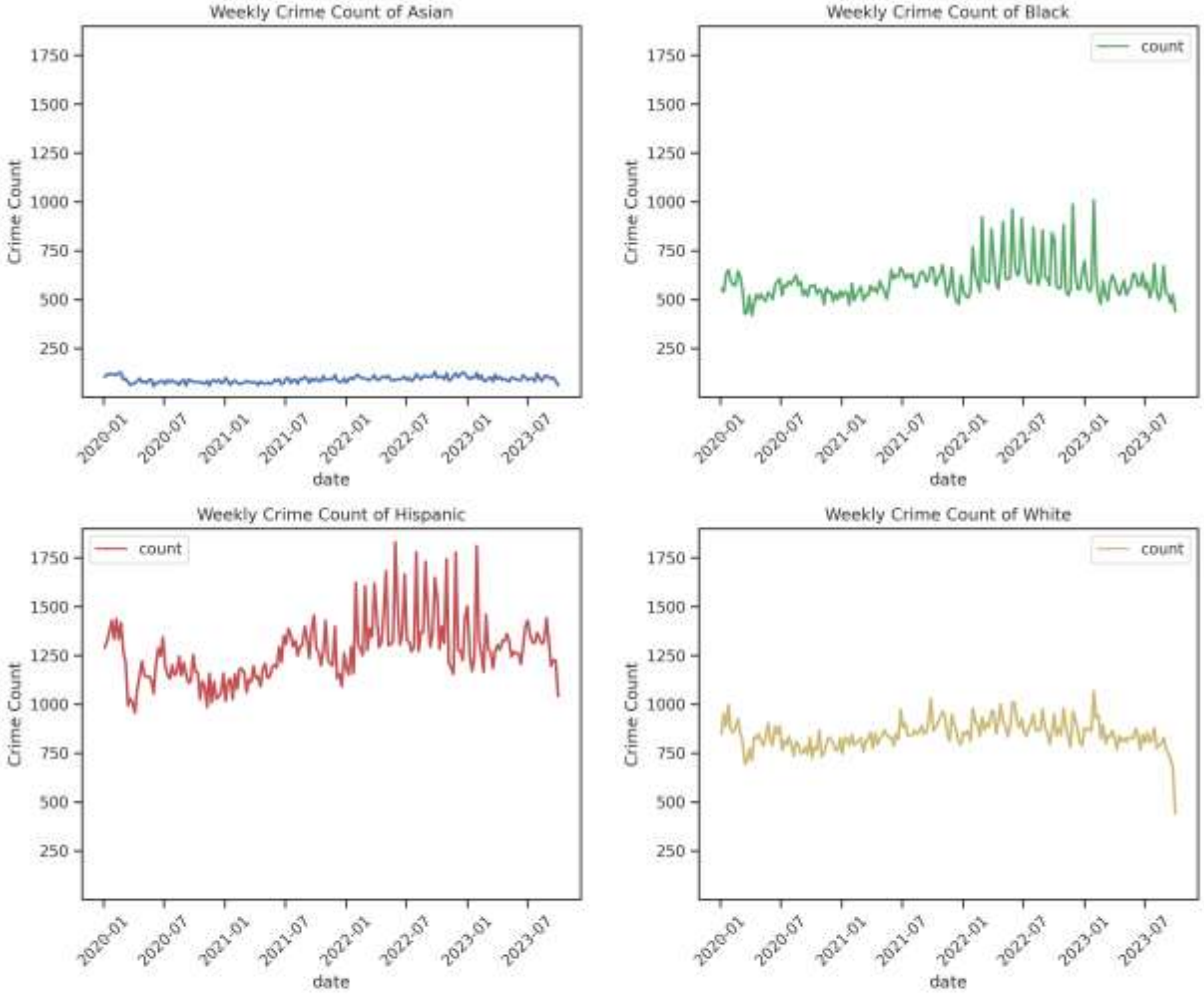




3 months seasonality

ARIMA(0,1,1) x (0,1,1) [S=13]

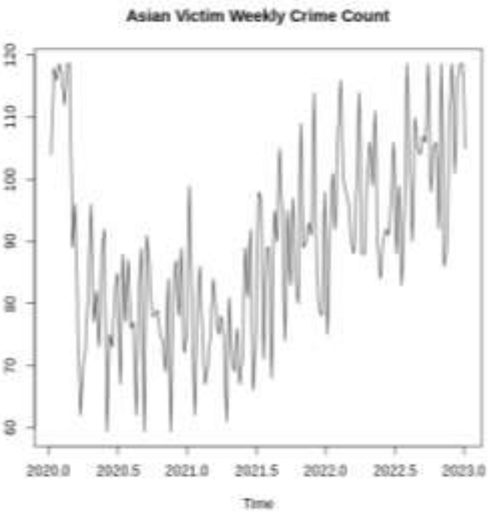
01/06/2020- 10/08/2023

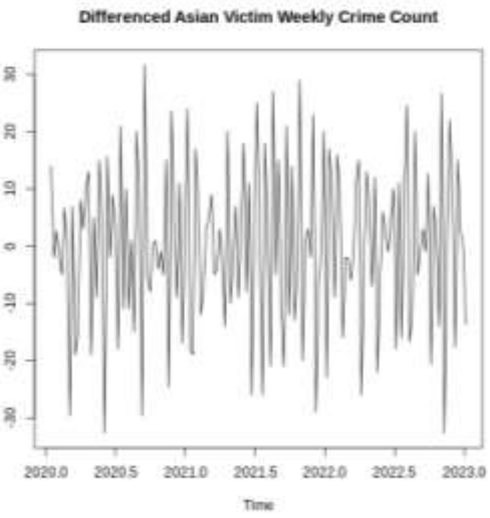


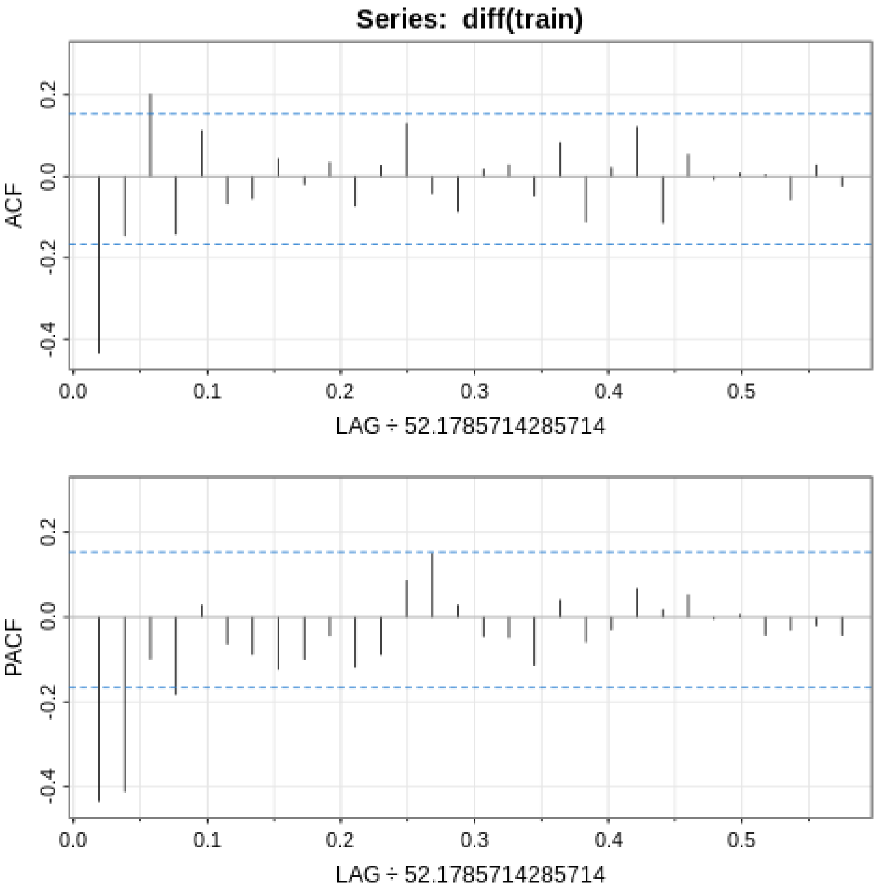
Victim Descent

- Obvious differences in average counts
- Possibly different trends and seasonality

Weekly Asian Victim Crime Count



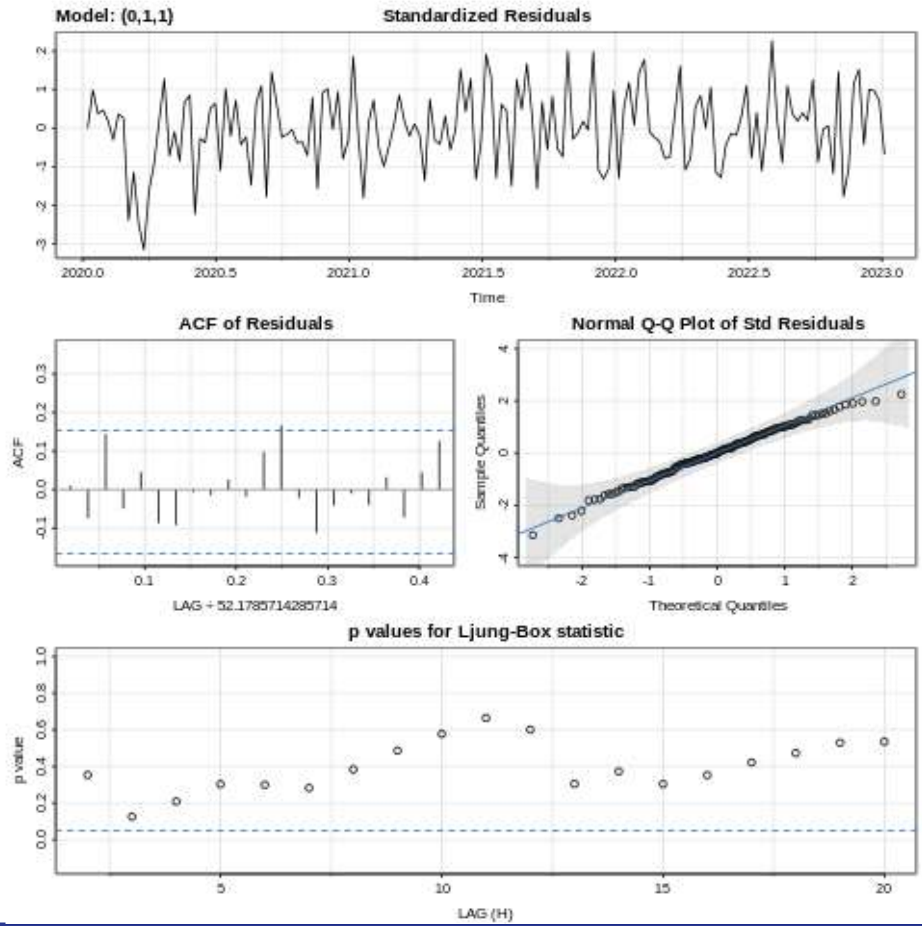




No seasonality

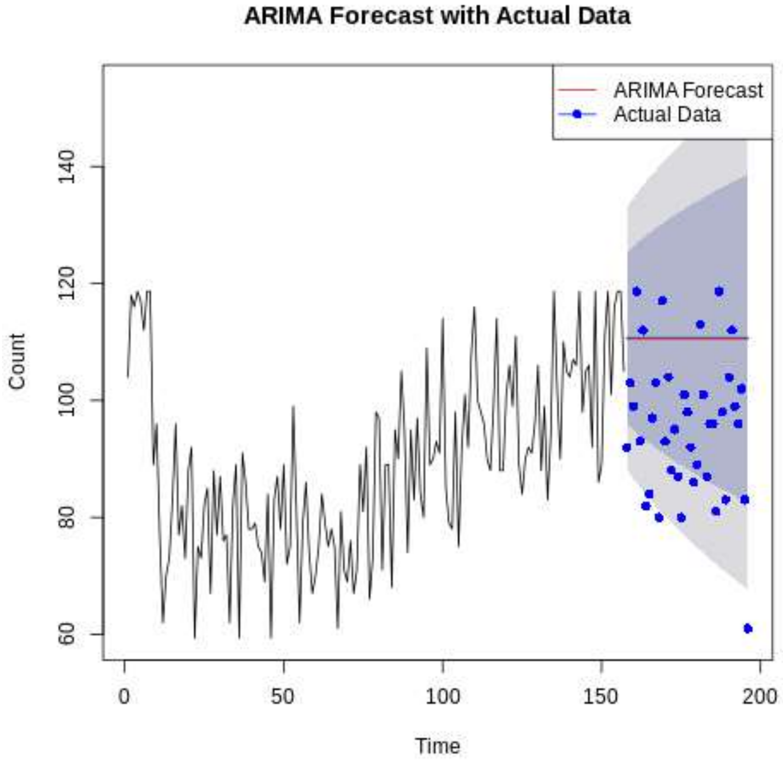
Possible Model: ARIMA(0,1,1)

Weekly Asian Victim Crime Count

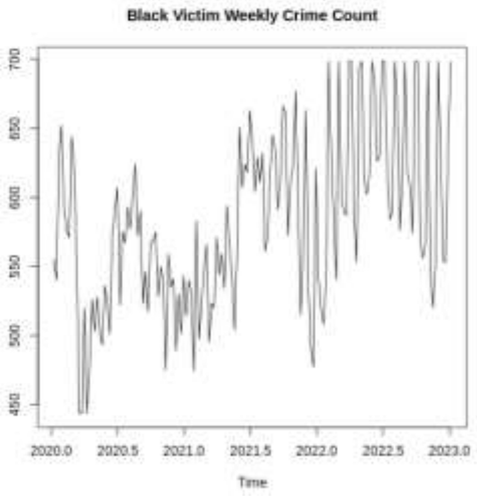


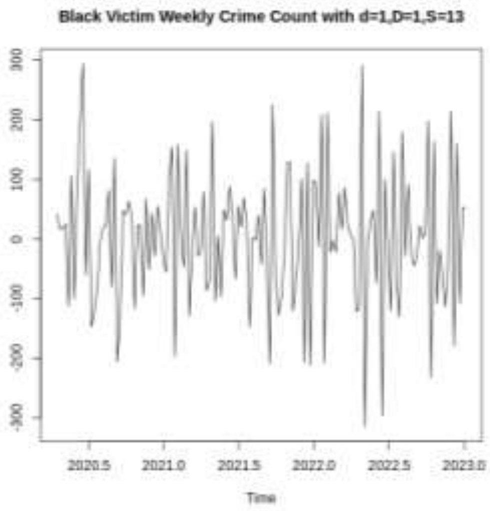
$$\boxed{ARIMA(0,1,1)}$$

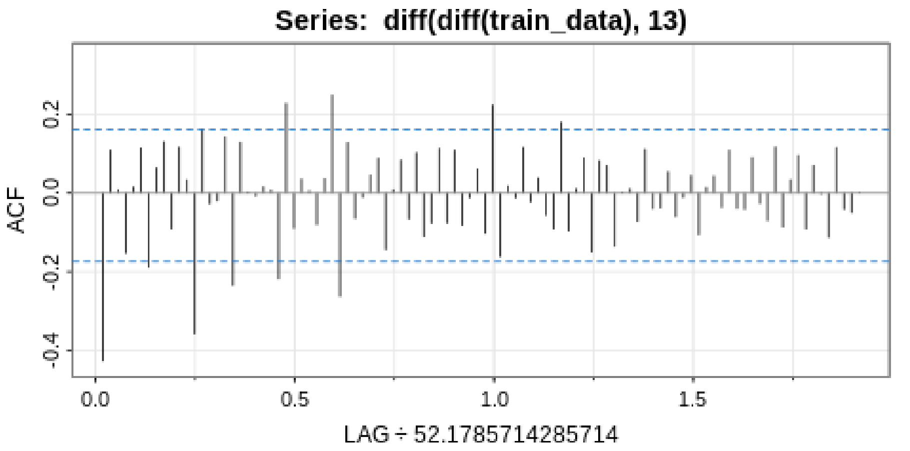
Prediction: constant trend



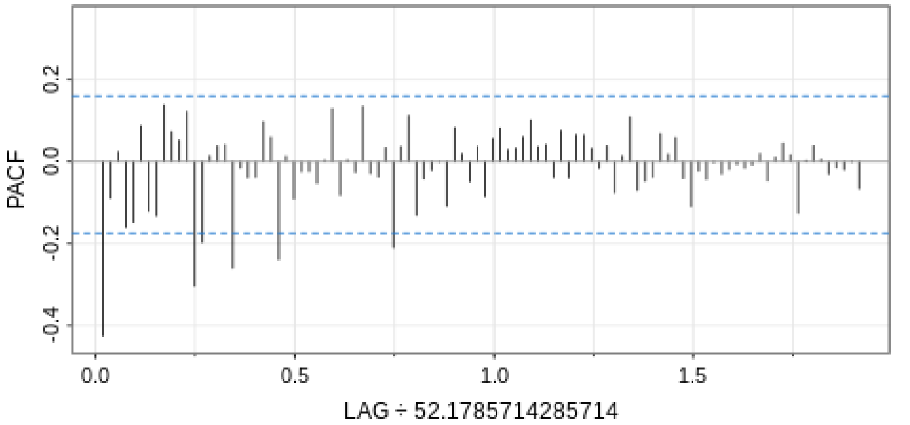
Weekly Black Victim Crime Count





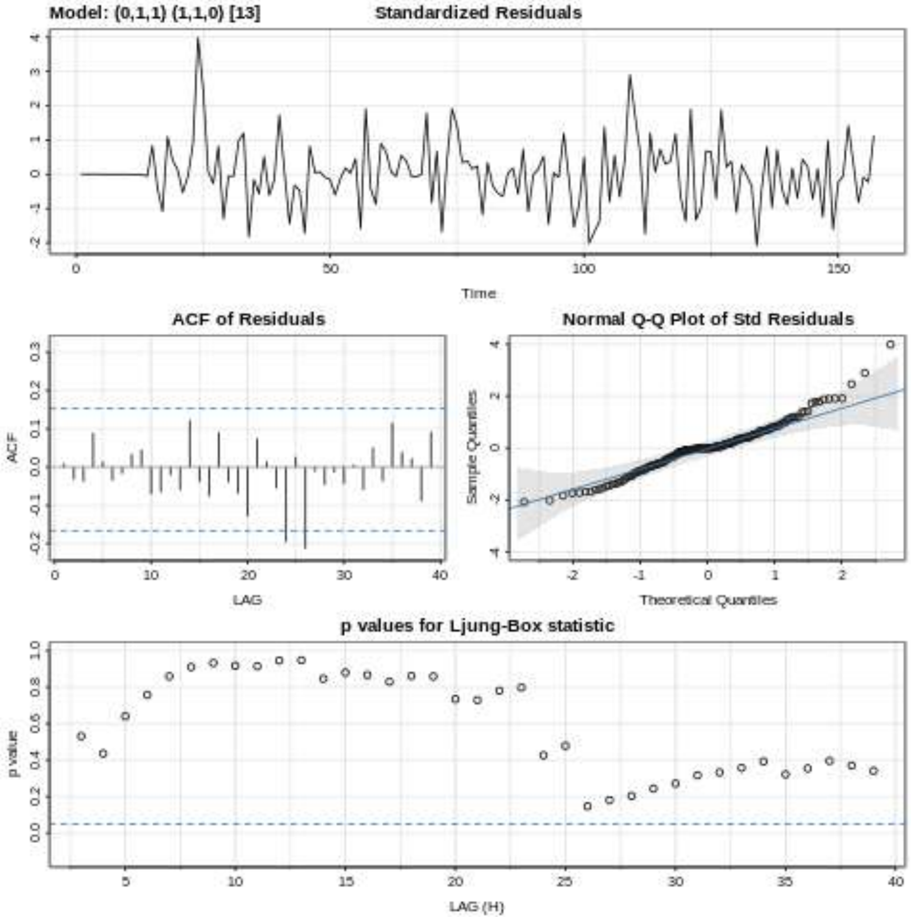


Seasonality with S=13
Possible Model: MA(1), undetermined for seasonal component

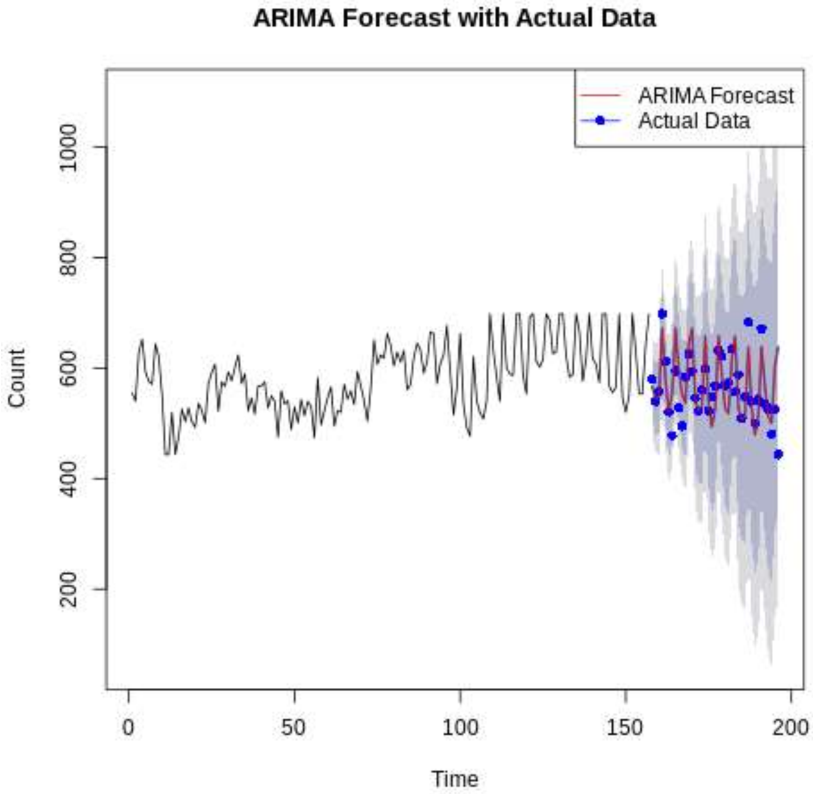


Weekly Black Victim Crime Count

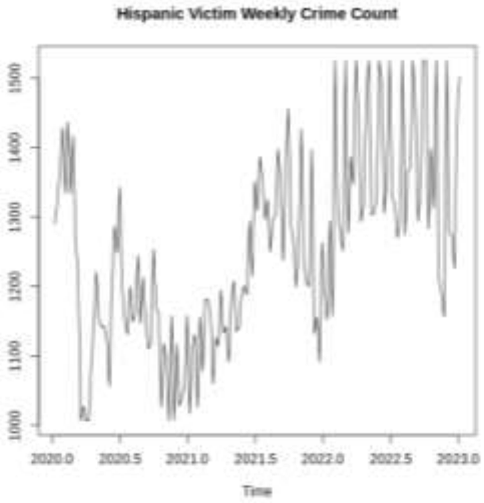
$$ARIMA(0,1,1) \times (1,1,0)_{13}$$

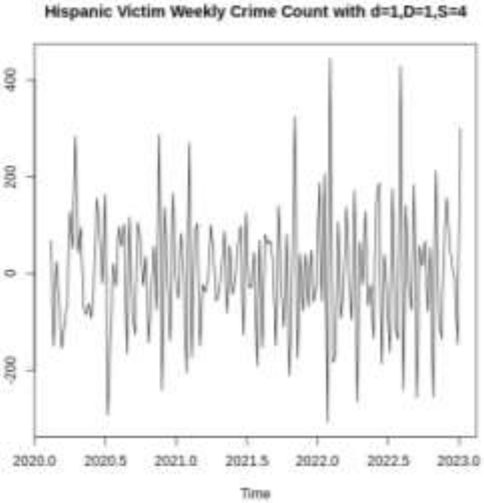


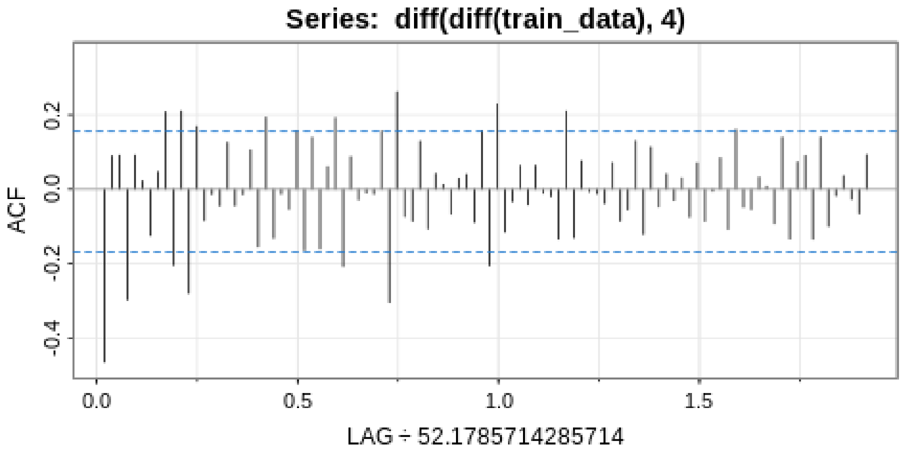
Prediction:
decreasing trend with seasonality

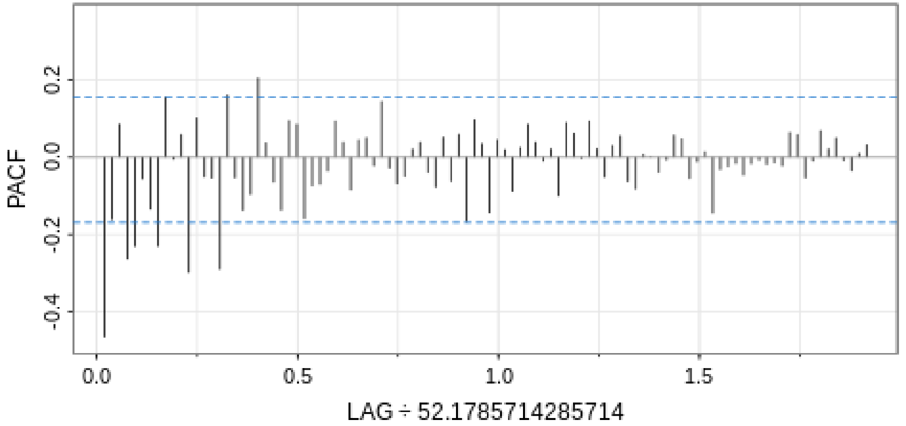


Weekly Hispanic Victim Crime Count







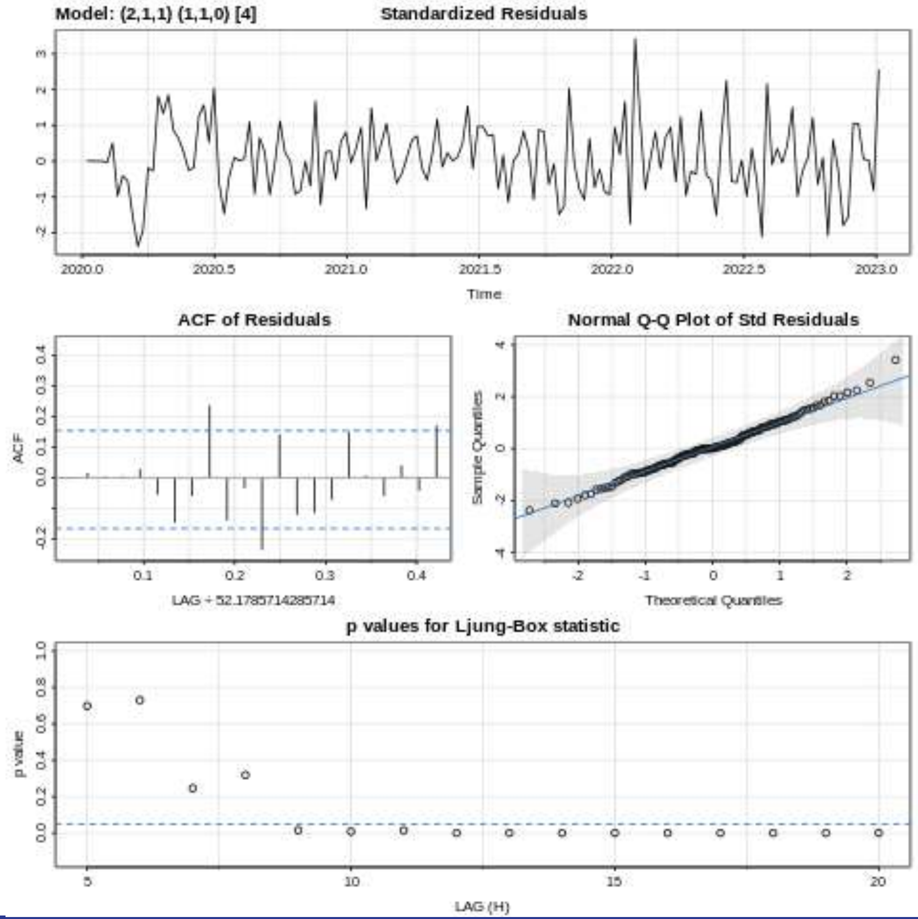


Seasonality with S=4

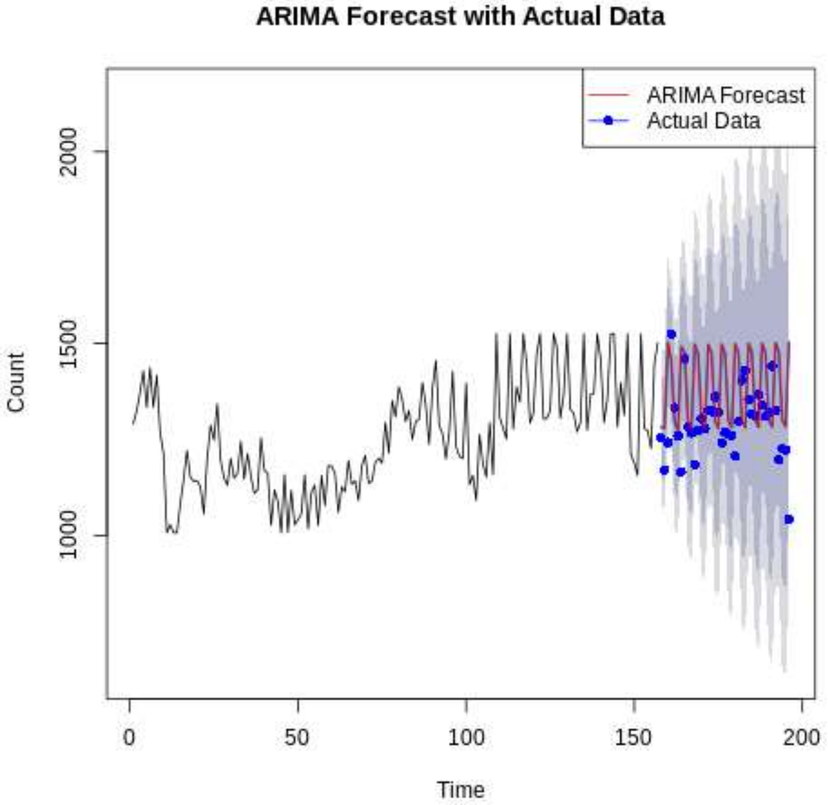
Possible Model: both seasonal and non-seasonal components cannot be determined from the ACF and PACF plot

Weekly Hispanic Victim Crime Count

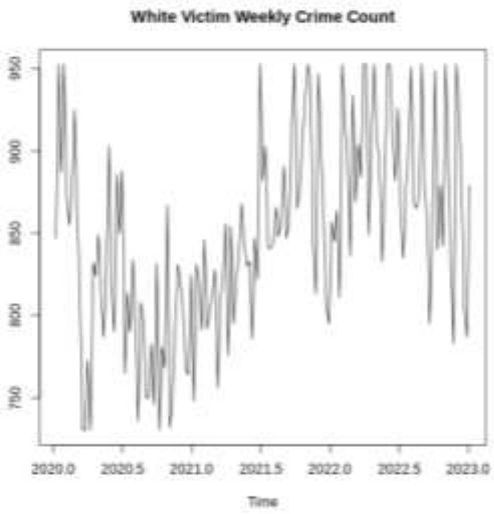
$$\boxed{ARIMA(2,1,1) \times (1,1,0)_4}$$

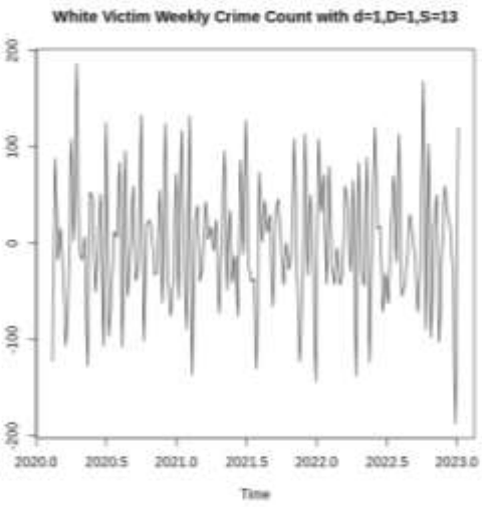


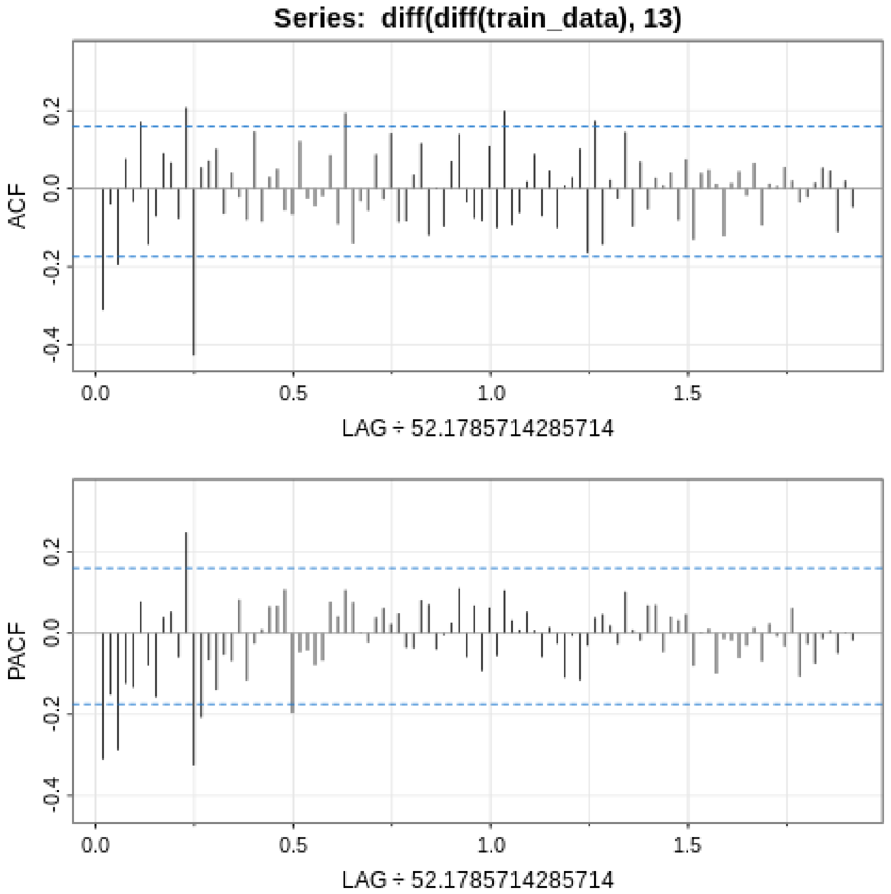
Prediction:
constant trend with seasonality



Weekly White Victim Crime Count





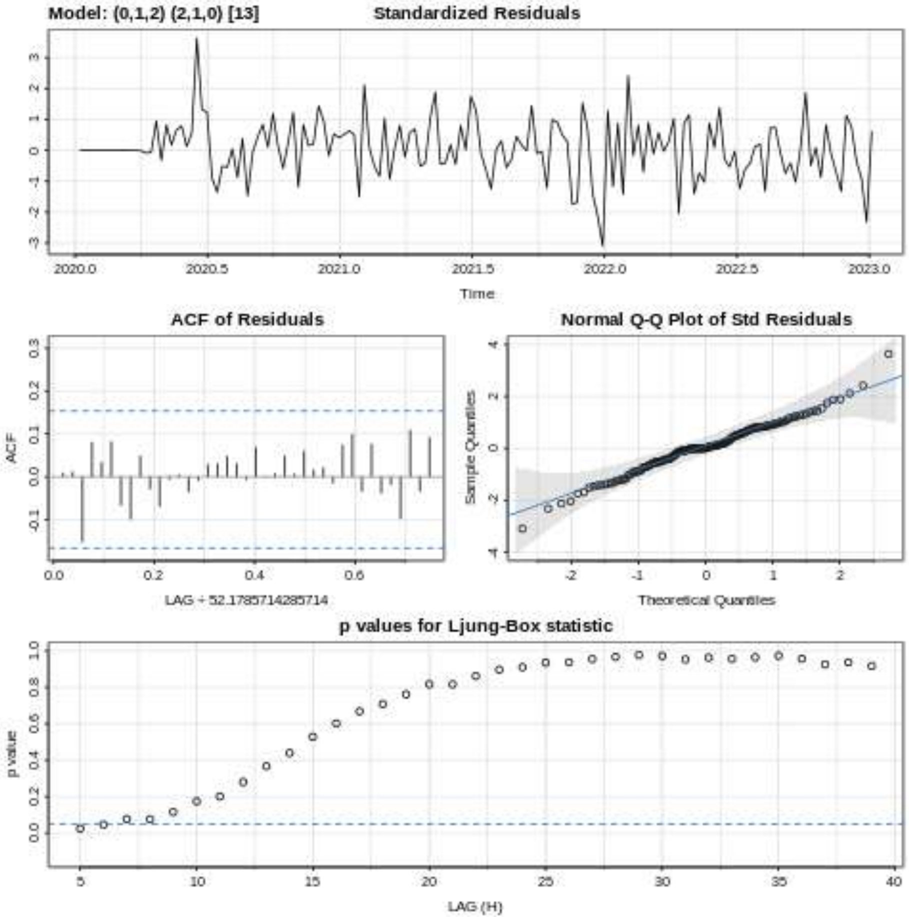


Seasonality with S=13

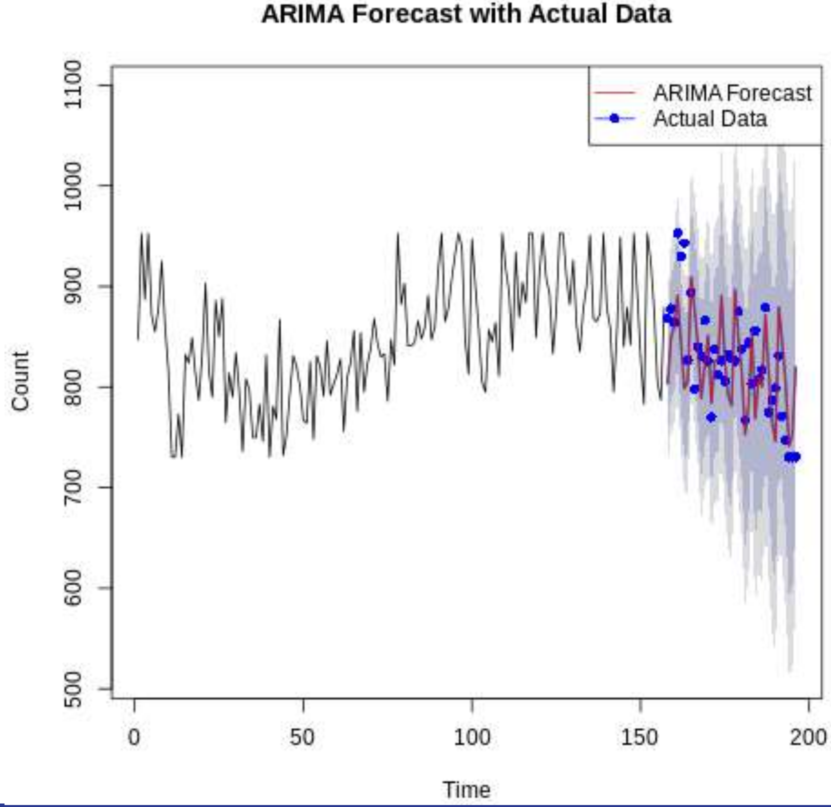
Possible Model: both seasonal and non-seasonal components cannot be determined from the ACF and PACF plot

Weekly White Victim Crime Count

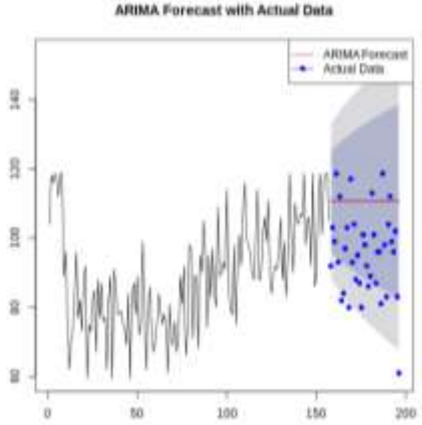
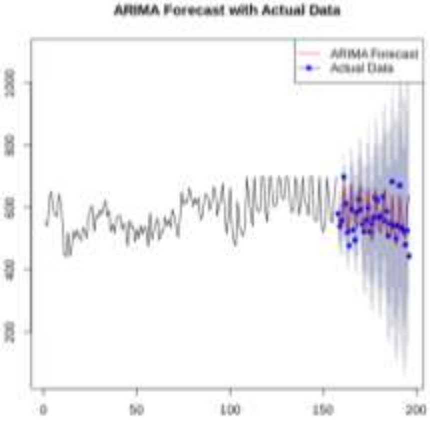
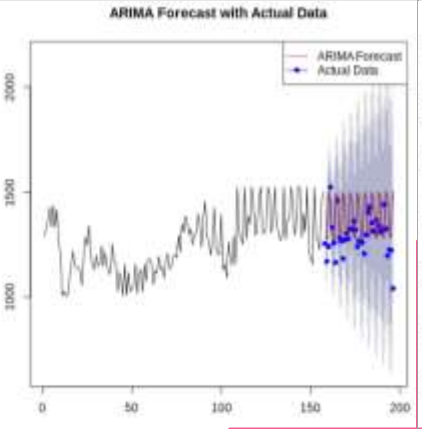
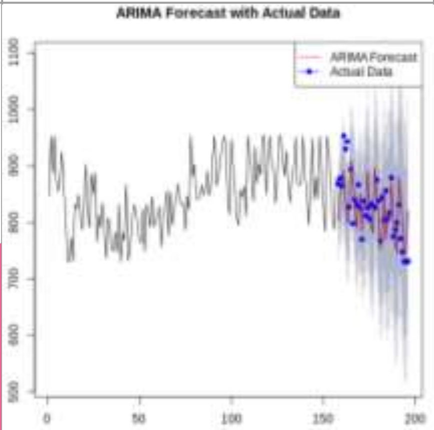
$$\boxed{ARIMA(0,1,2) \times (2,1,0)_{13}}$$



Prediction:
decreasing trend with seasonality

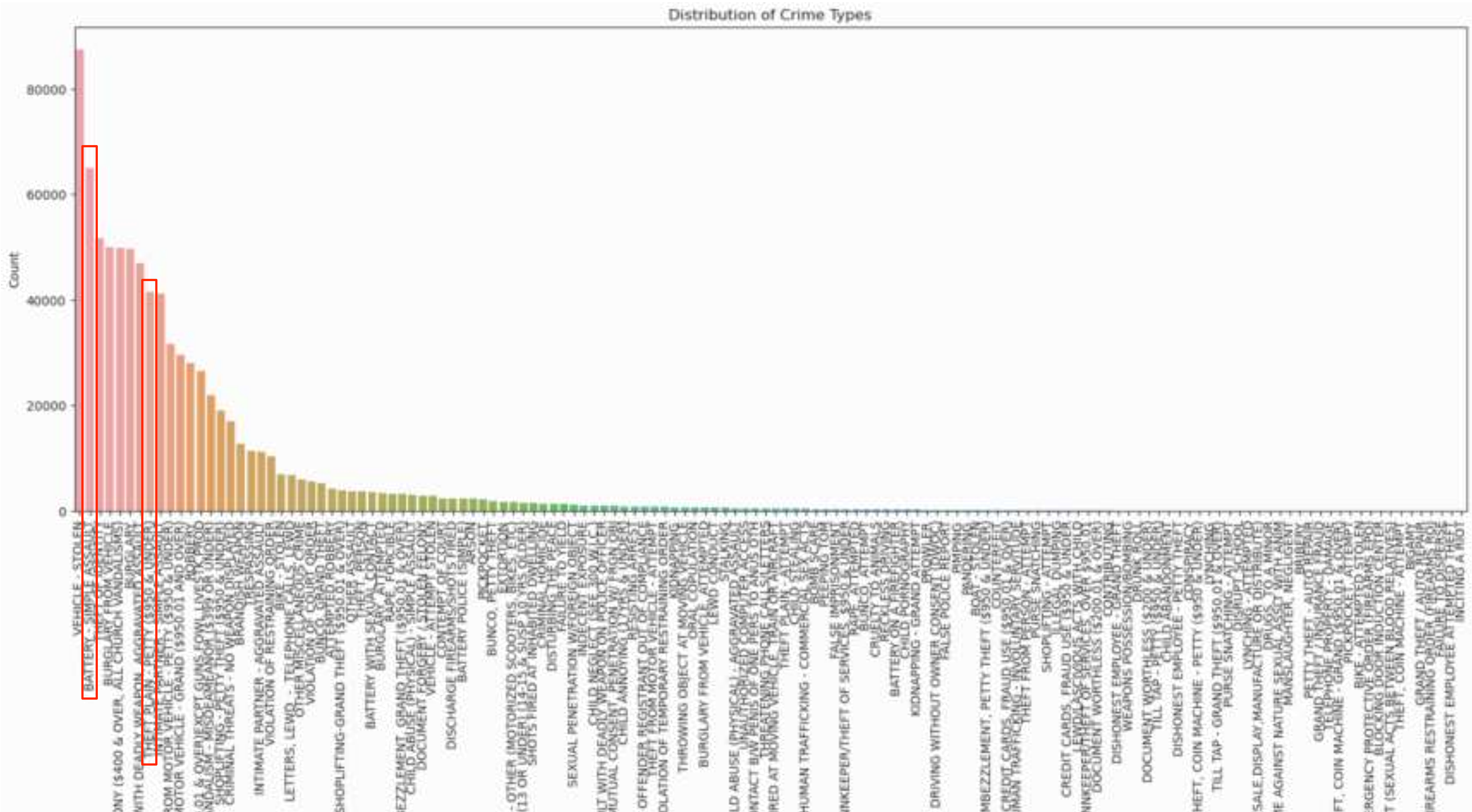


Models for different Victim Descent

Victim Descent	Asian	Black	Hispanic	White
Model Fitted	$ARIMA(0,1,1)$	$ARIMA(0,1,1)\times(1,1,0)_{13}$	$ARIMA(2,1,1)\times(1,1,0)_4$	$ARIMA(0,1,2)\times(2,1,0)_{13}$
Equation of Model	$x_t = x_{t-1} + w_t - 0.7364w_{t-1}$ $w_t \sim N(0, 131.8)$	$z_t = x_t - x_{t-1} - x_{t-13} + x_{t-14}$ $z_t = 0.474x_{t-13} + w_t - 0.520w_{t-1}$ $w_t \sim N(0, 1996)$	$z_t = x_t - x_{t-1} - x_{t-4} + x_{t-5}$ $z_t = 0.2325x_{t-1} + 0.2727x_{t-2} - 0.3278x_{t-4} + 0.0762x_{t-5} + 0.0894x_{t-6} + w_t - w_{t-1}$ $w_t \sim N(0, 10415)$	$z_t = x_t - x_{t-1} - x_{t-13} + x_{t-14}$ $z_t = -0.7113x_{t-13} - 0.3954x_{t-26} + w_t - 0.5175w_{t-1} - 0.1922w_{t-2}$ $w_t \sim N(0, 1850)$
39-week Prediction				

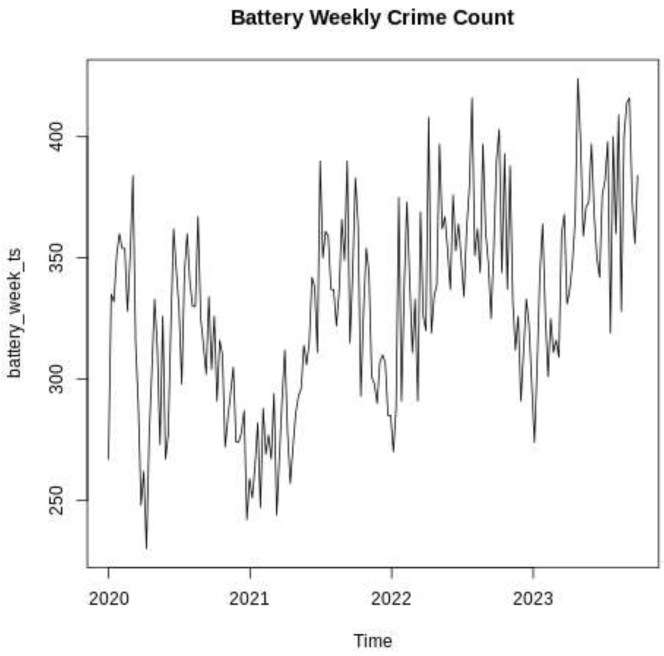
Segmented by Crime Type: EDA

210 Crime Types Overall

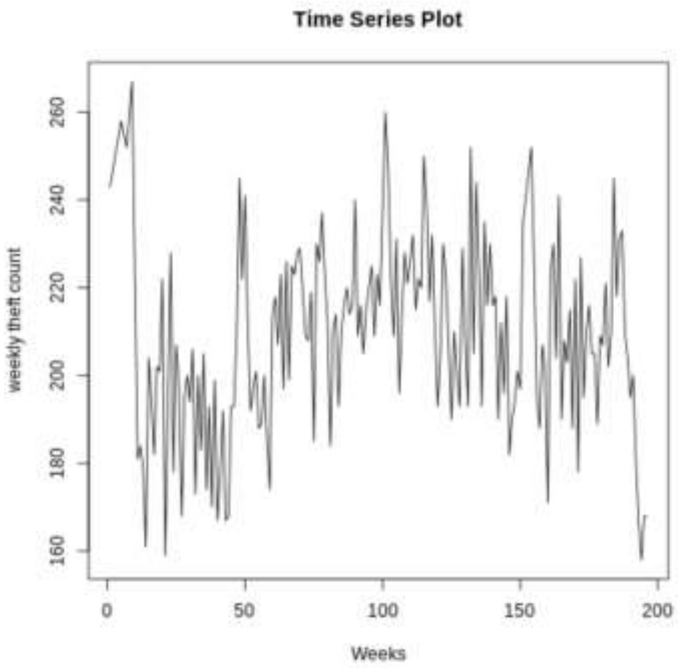


Segmented by Crime Type: EDA

Battery - Simple Assault

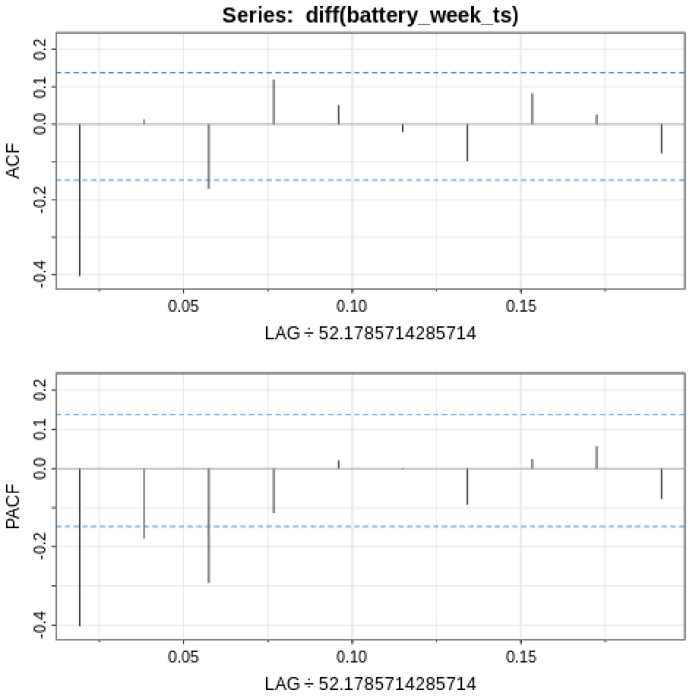


Theft - Plain (Under 950$)

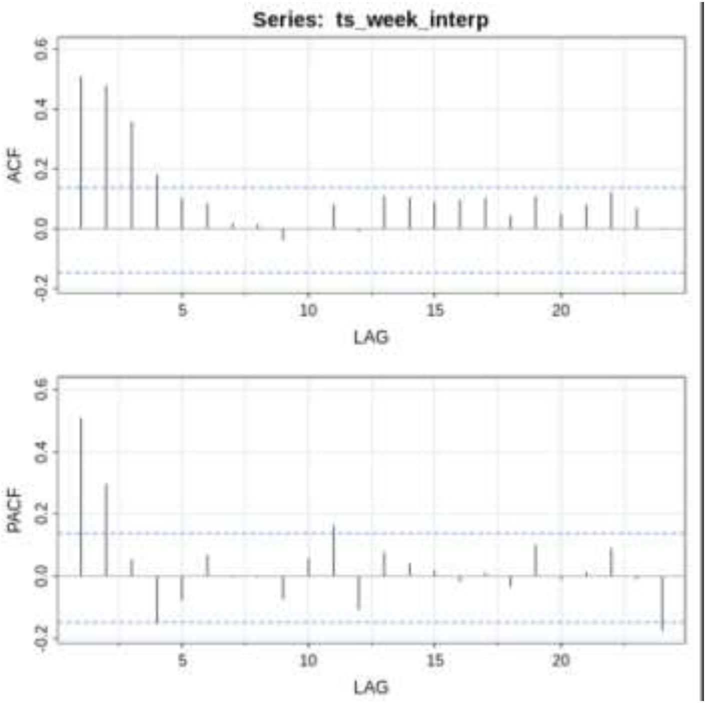


Segmented by Crime Type: EDA

Battery: possible MA(1) model

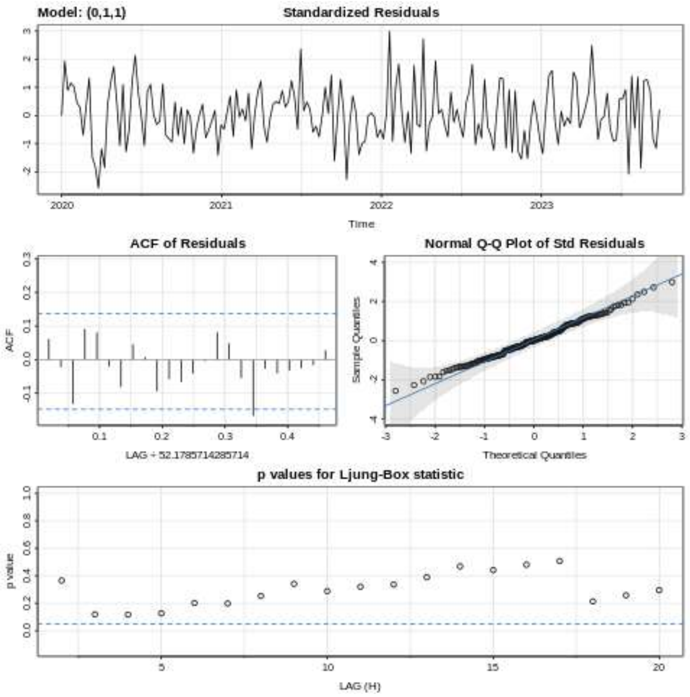


Theft：possible AR(2) model

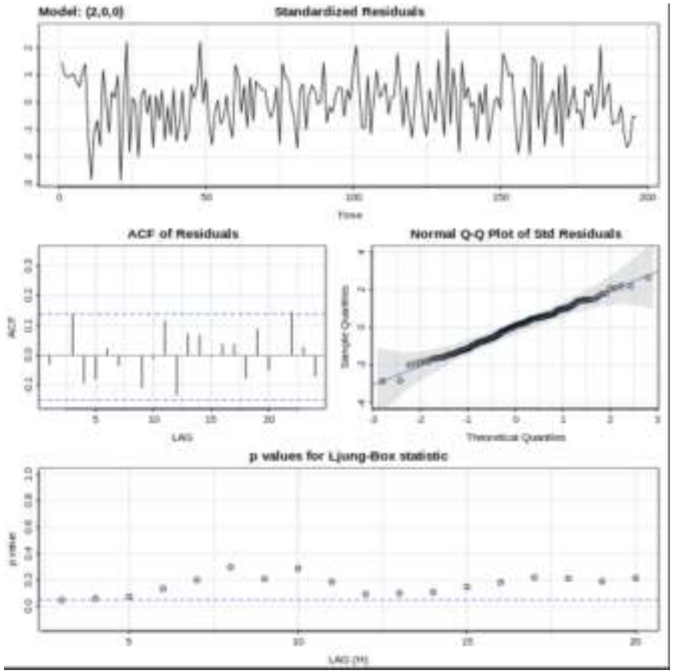


Segmented by Crime Type: Model Diagnostics

Battery: ARIMA(0,1,1)

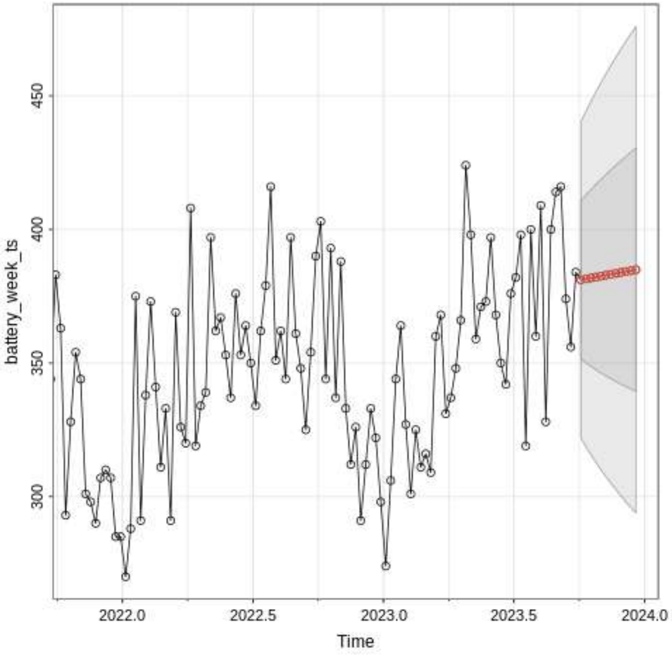


Theft: MA(2)

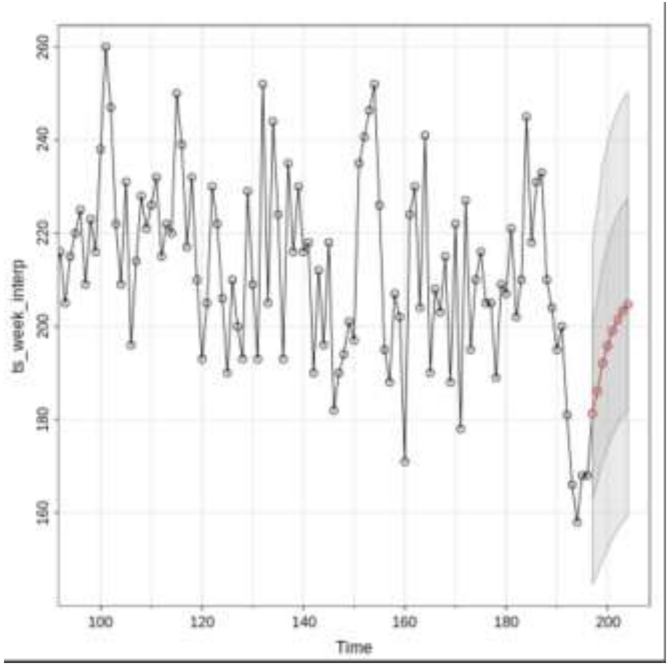


Segmented by Crime Type: Forecasting

Battery: slight increase 12 weeks ahead



Theft



Summary of Crime Type Trend

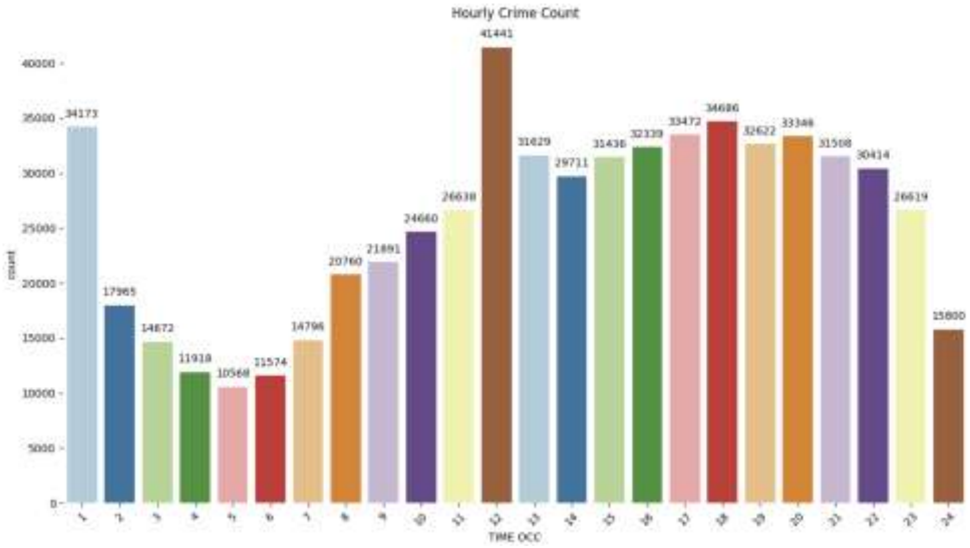
Difference in modeling the type of crime occurred

- ARIMA(0,1,1) vs MA(2) indicate a more complex relationship with previous theft incidents compared to battery crimes

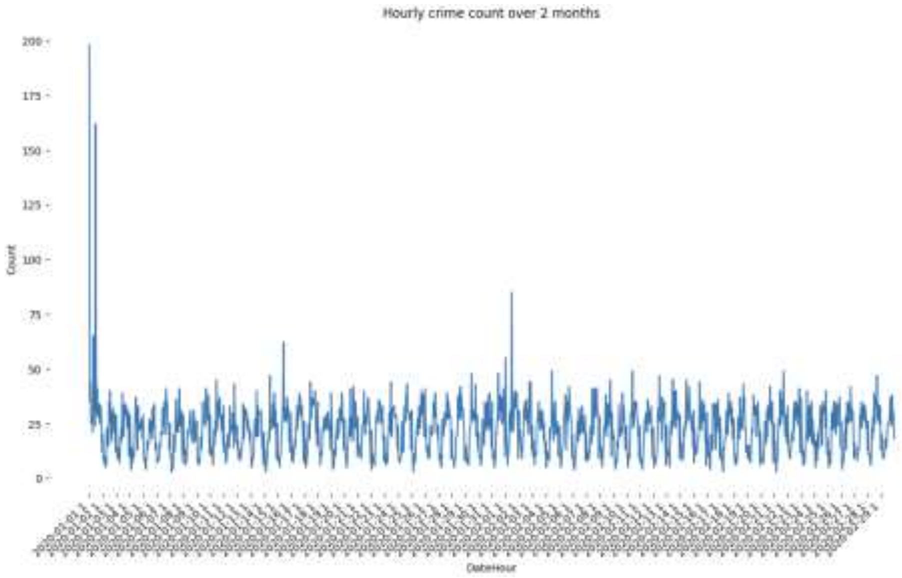
- Measures to prevent theft may be more difficult to predict the success of compared to battery

Segmented crime count model EDA (hourly)

01/01/2020 - 02/29/2020

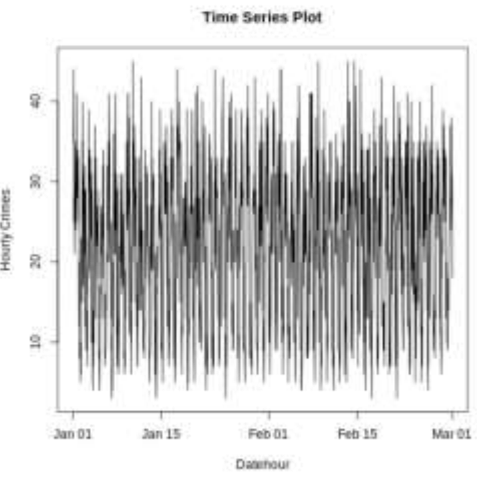


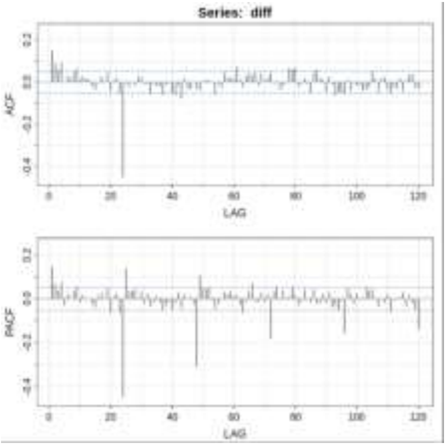
Overall Hourly Crime Count

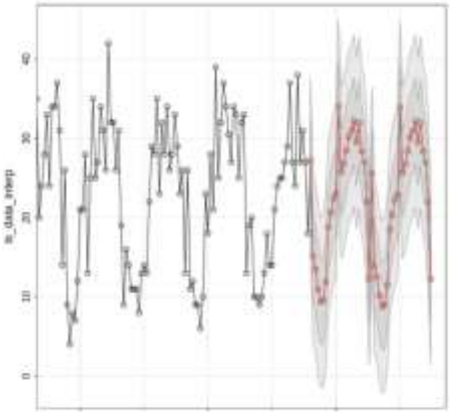


Hourly Crime Count over 2 months

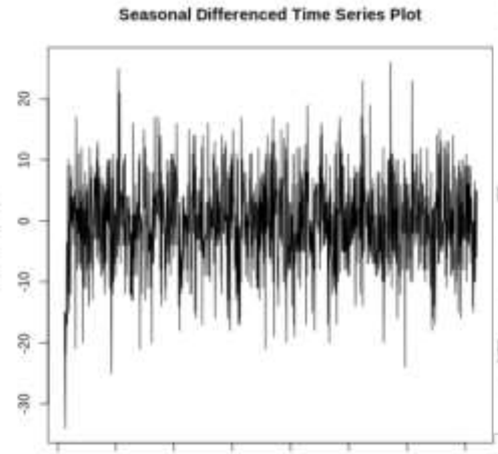
Segmented crime count model fitting (hourly)

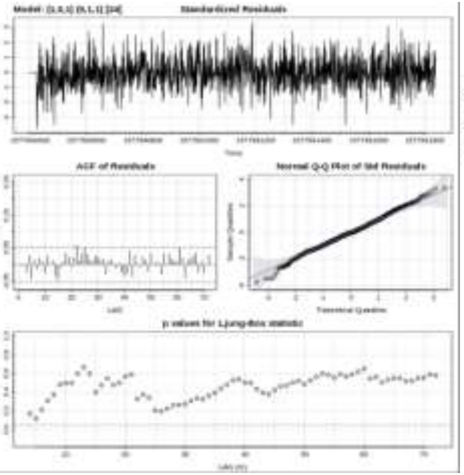


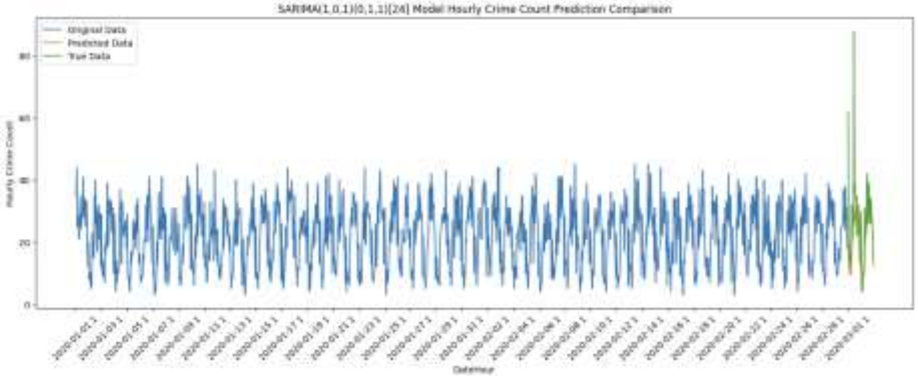




Apparent Seasonality







SARIMA(1,0,1)x(0,1,1) S=24

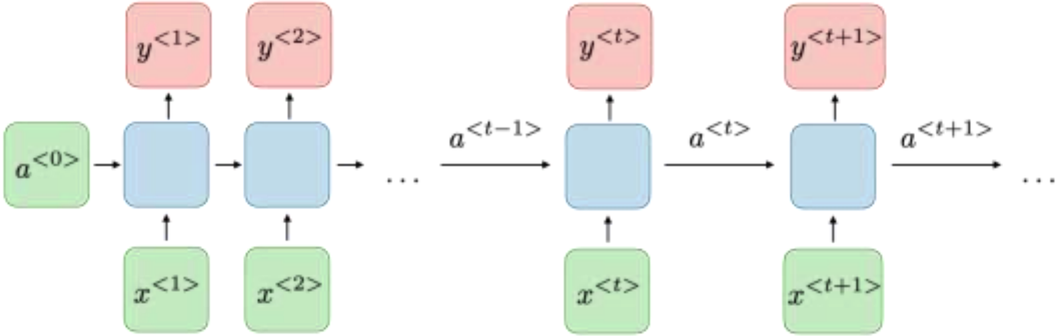
Question: - Is there a relationship between victim profile and crime info?
- Instead of overall crime count trend, can we predict a future crime within one hour time precision?

Proposed Solution: RNN model (Capture Sequential Dependencies)

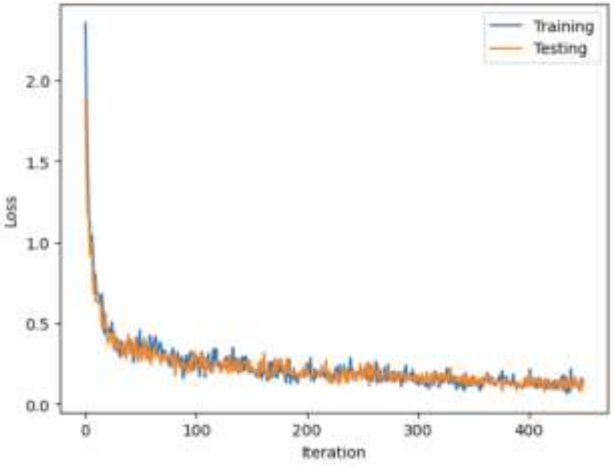
- Using victim’s profile (sex, age & descent) to predict crime area and type

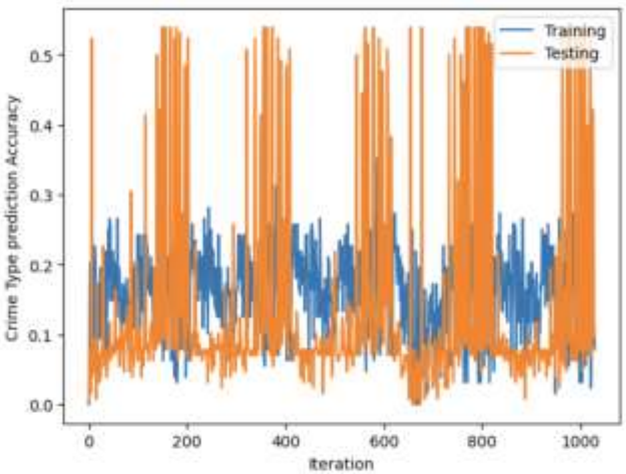
Data Pre-processing: 700k crime => 4k crime

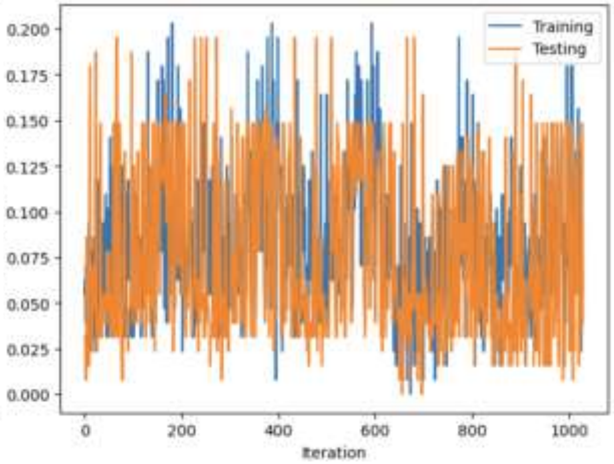
- Rounding each crime time to the closest integer hour
- Rounding each age to multiples of 5
- Use mode for each feature at the same rounding hour:
  - Do not choose real crime, choose the most possible crime pattern



Training and Prediction Results







Take aways:

- Training is effective, best accuracies are good compared to untrained accuracies
- Model converges quickly (within each epoch), may be due to lack of data, and model is relatively complex (500 hidden neurons)
- Unbalanced class (crime type)
- Future analysis and models can be developed to predict crime more precisely

Q&A