

# Tracking the Streets: NYC Taxi Patterns Over Time

## SC475 Time Series Aalysis

Vyom Narsana(202203026)  
Krushi Sutariya(202203044)  
Ayush Popshetwar(202201412)

Mentor: Professor Mukesh Tiwari

July 4, 2025

# Yellow Taxi Trips Data Fields

---

- VendorID
- tpep\_pickup\_datetime
- tpep\_dropoff\_datetime
- passenger\_count
- trip\_distance
- RatecodeID
- store\_and\_fwd\_flag
- PULocationID
- DOLocationID
- payment\_type
- fare\_amount
- extra
- mta\_tax
- tip\_amount
- tolls\_amount
- improvement\_surcharge
- total\_amount
- congestion\_surcharge
- airport\_fee
- cbd\_congestion\_fee

# Features to Study

---

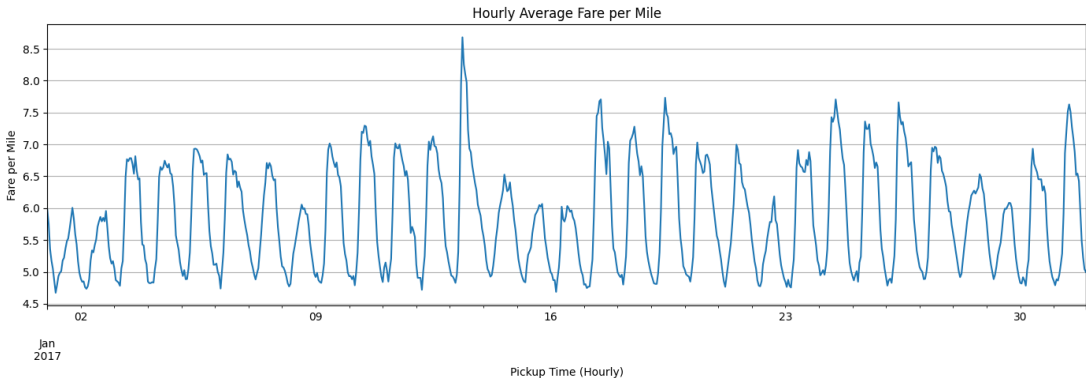
- **Fare per Mile over Time(January 2017)**
  - Shows price changes
  - Helps with fare rules
- **Volume Density of Trips over Time(Year 2017)**
  - Shows how busy taxis are
  - Helps plan traffic and transport

# Exploratory Data Analysis on Fare per Mile for January 2017

---

- Rows with dates not falling in January 2017 were removed.
- A new feature,  $\text{fare\_per\_mile} = \frac{\text{fare\_amount}}{\text{trip\_distance}}$ , was created.
- Trips with negative fare amounts were removed.
- Trips with zero distance were excluded.
- Trips exceeding 3 standard deviations from the fare per mile were removed.

# Fare Per Mile of January 2017



Fare per Mile over Time

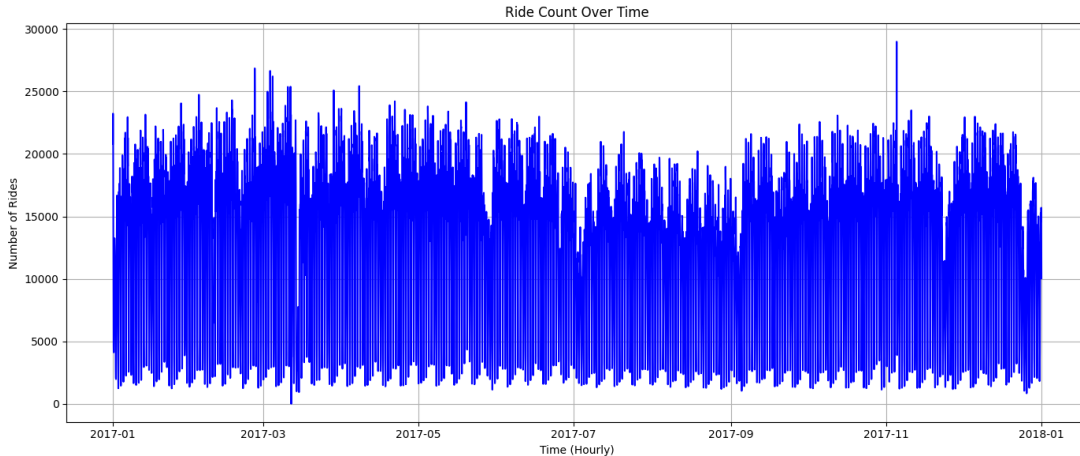
# Exploratory Data Analysis on Volume of Trips for 2017

---

- Rows with dates not falling in 2017 were removed.
- The dataset was reduced to the sum of trips per hour.

# Volume of Trips of Year 2017

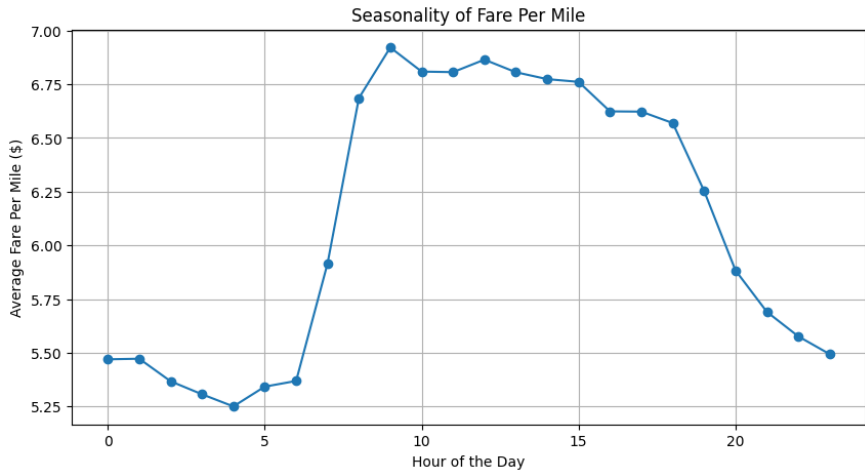
---



Volume Density of Trips over Time

## Avg. Hourly Fare per Mile (Jan 2017)

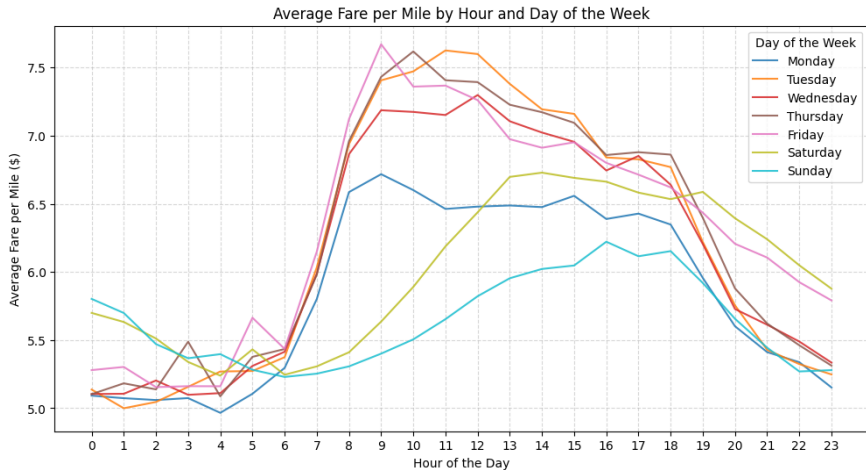
---



This visualization presents the average fare per mile over each hour of the day in January 2017.

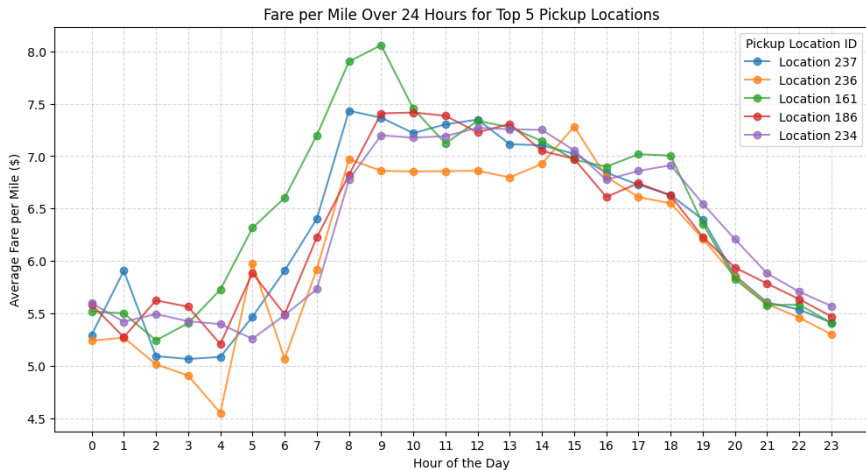


# Avg. Fare per Mile by Hour and Day of Week (Jan 2017)



This plot shows how the average fare per mile varies by hour and day of the week in January 2017.

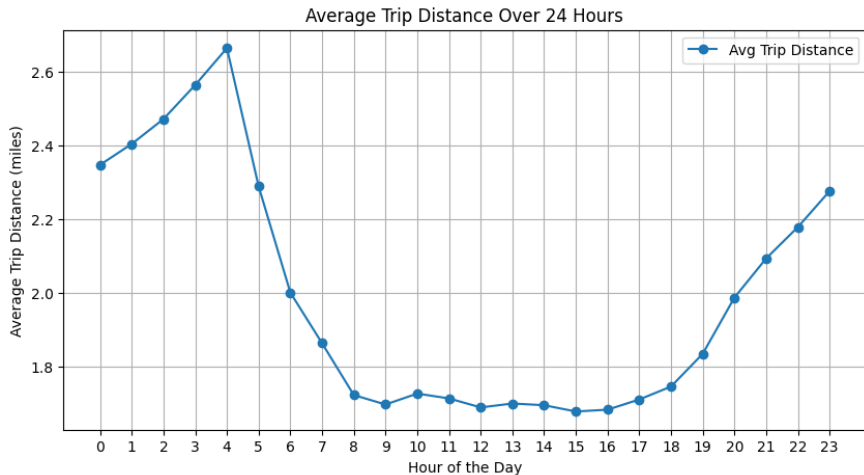
# Fare per Mile Over 24 Hours for Top 5 Pickup Locations



Fare per Mile Over 24 Hours for Top 5 Pickup Locations

# Average Trip Distance Over 24 Hours

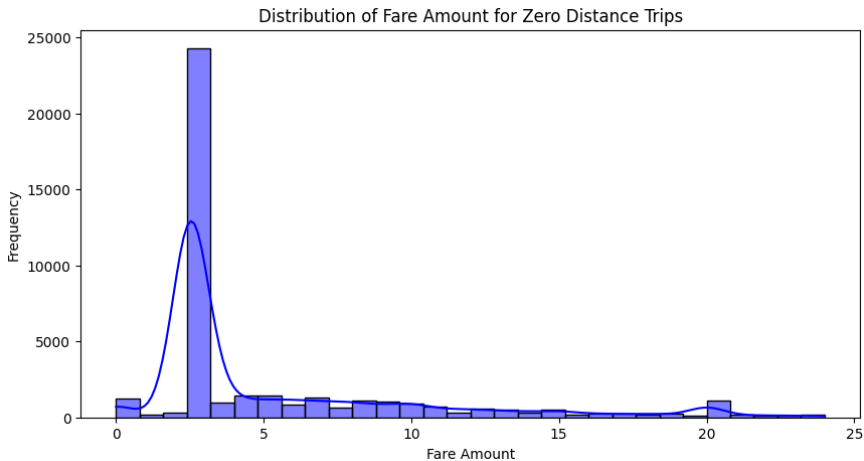
---



Average Trip Distance Over 24 Hours

# Distribution of Fare Amount of Zero Distance Trips

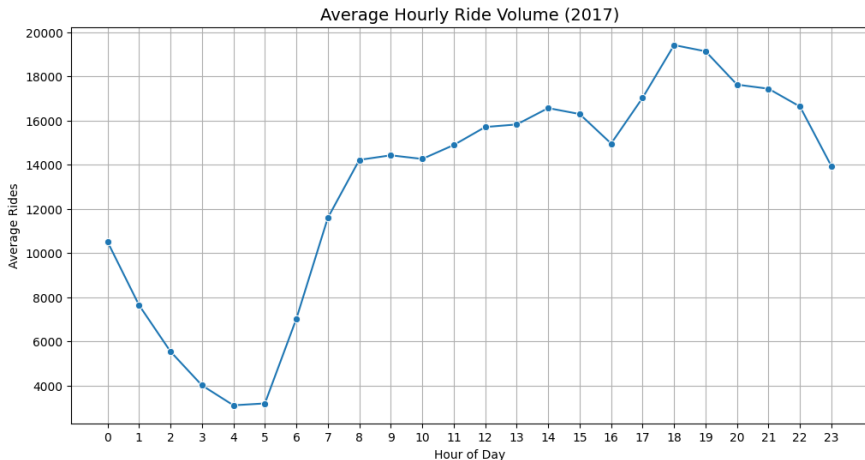
---



Distribution of Fare Amount of Zero Distance Trips

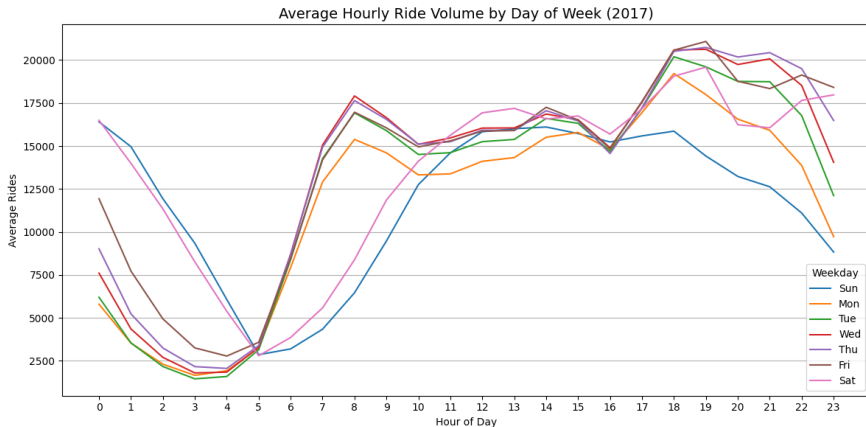
# Average Hourly Ride Volume (Overall – 2017)

---



This visualization displays the overall average hourly ride volume throughout the year 2017.

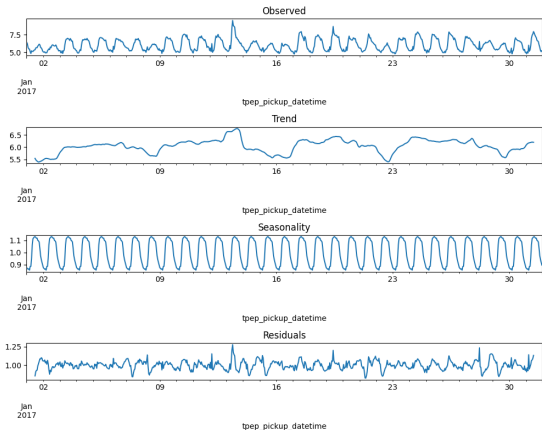
# Average Hourly Ride Volume by Day of Week (2017)



This plot shows how ride volume varies across different days of the week, averaged by hour.

# Decomposition of Fare per Mile Time Series

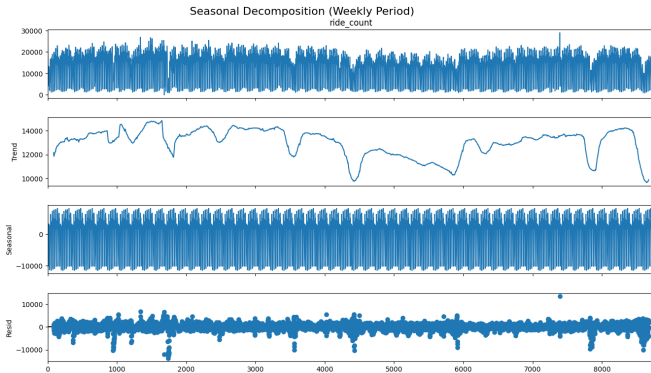
---



Shows decomposition of the fare per mile time series with a period of 24 hours.

# Decomposition of Volume of Trips Time Series

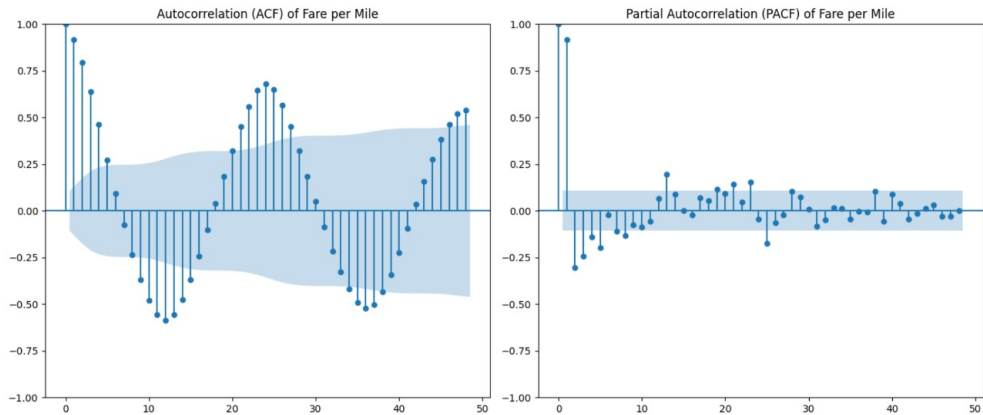
---



Time series decomposition of ride volume into trend, seasonality, and residual components

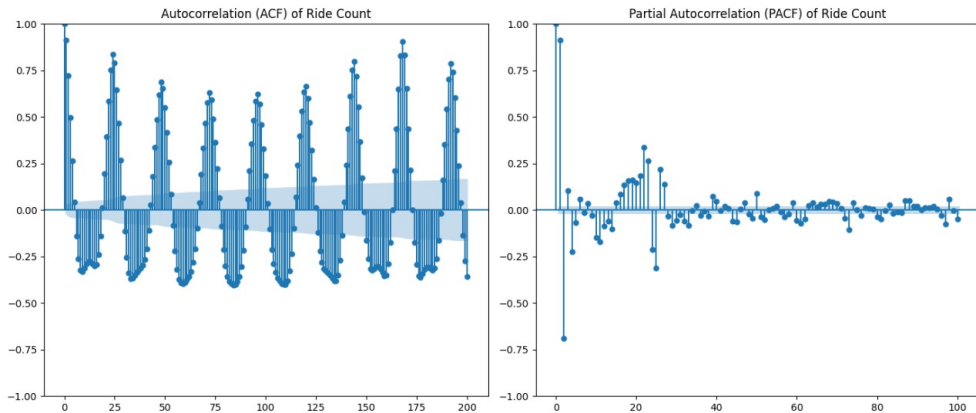


# ACF and PACF for Fare per Mile Time Series



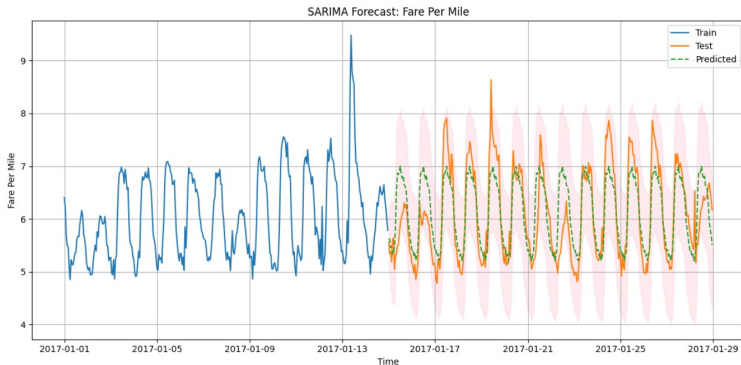
ACF and PACF for fare per mile

# ACF and PACF for Volume of Trips Time Series



ACF and PACF for Volume of trips

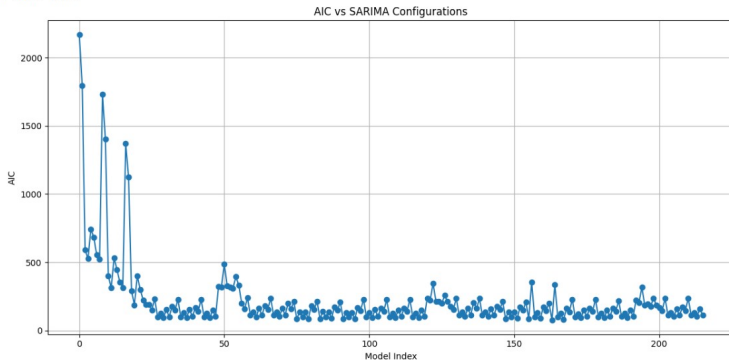
# SARIMA Model for Fare Per Mile



SARIMA (1,0,2)(1,1,1,24) forecast of fare per mile. The model captures daily seasonality and performs better than the AR(1) model, explaining about 65% of the variance with low prediction errors (MAE = 0.3890, RMSE = 0.4999, MAPE = 6.38%).

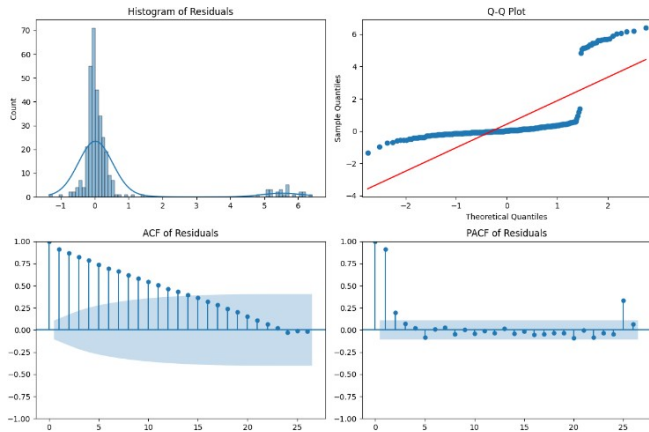
# AIC Scores for different SARIMA Configurations

Best SARIMA: (2,0,2) x (0,1,1,24) with AIC: 76.93886441601828  
MAE: 0.3863  
MSE: 0.2495  
RMSE: 0.4995  
MAPE: 6.36%  
R<sup>2</sup> Score: 0.6523



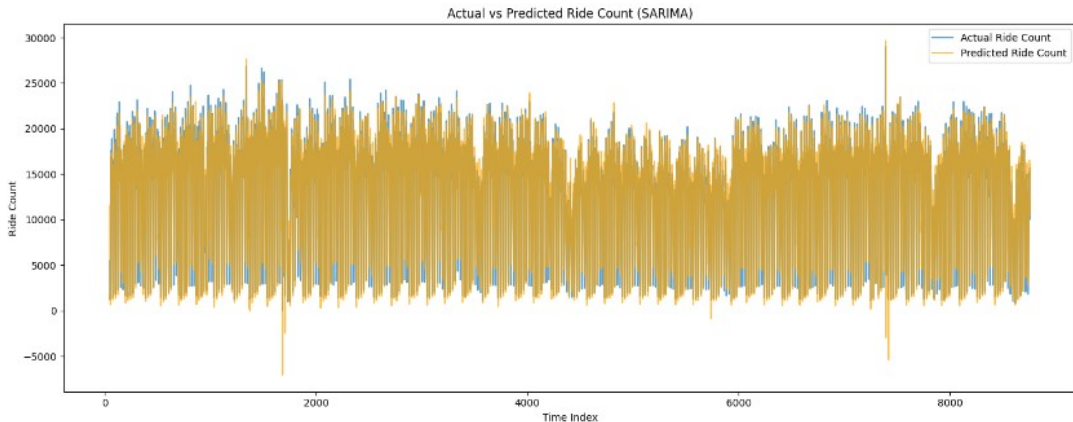
AIC scores for different SARIMA configurations for fare per mile time series

# Residual Analysis Plots



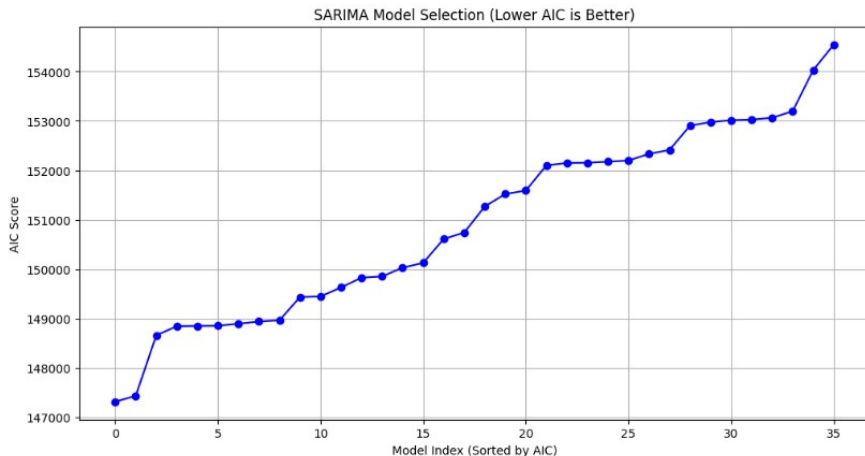
Residual analysis plots: Histogram, Q-Q plot, ACF, and PACF of residuals from the SARIMA model of Fare per mile time series

# SARIMA Model for Volume of Trips



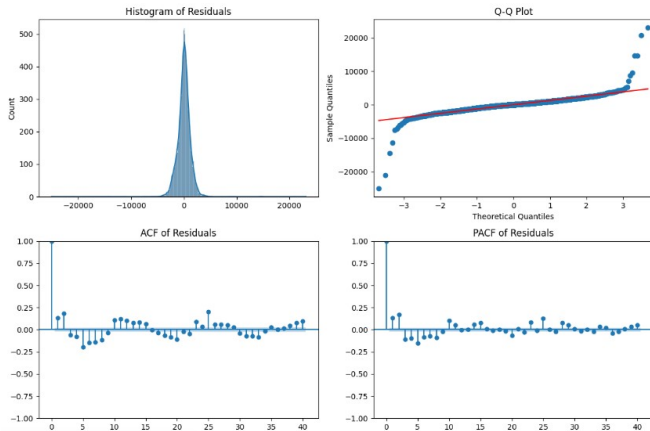
Actual vs. predicted ride count using the SARIMA model

# AIC Scores for different SARIMA Configurations



SARIMA model selection (Lower AIC is better)

# Residual Analysis Plots



Residual analysis plots: Histogram, Q-Q plot, ACF, and PACF of residuals from the SARIMA model of Volume density of trips



# Applications

---

This study shows a link between fare per mile, trip volume, and time. Practical applications include:

- **Better Pricing:** Adjust prices using real-time demand.
- **Smarter Fleet Use:** Deploy more taxis during peak hours.
- **Better Policies:** Support urban mobility planning.
- **Rider Benefits:** Help passengers avoid high fares and delays.

## Future Work

---

- Add features like weather, special events, and location data.
- The current model only uses hourly average data. For more detailed insights, we can compute averages at the second level using high-performance machines, enabling deeper analysis and richer information extraction.
- Including pickup and drop-off locations can improve prediction accuracy.
- A detailed analysis based on weather, location, and specific time periods can help optimize public transport arrangements, which will also benefit the environment.

**Thank You!**