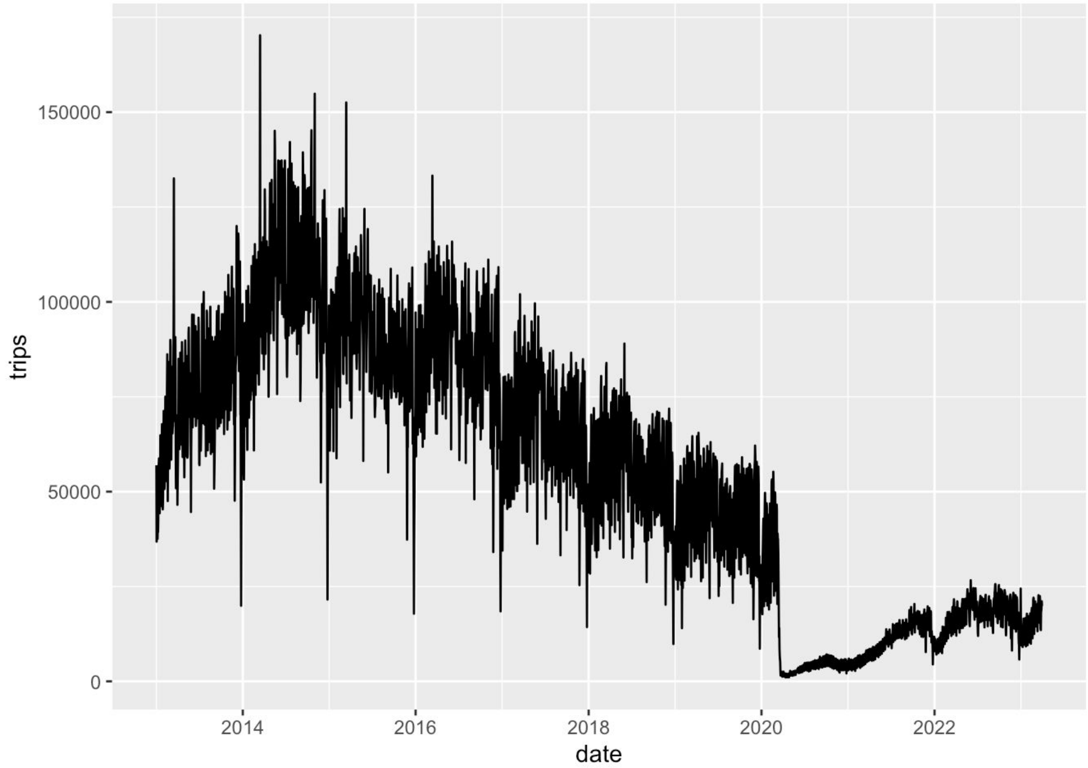
Analyzing and Forecasting Taxi Demand Patterns in Chicago Over Time

Team 4: Maggie Chuang, Kelsey Liu, Shirley Liu, Evelyn Wu, Vivian Yeh

Date: Tuesday, May 23, 2023

Table of Contents

Business Problem - Downward Taxi Trips



Chicago Taxi Trips by Date (2013/01-2023/03)

- Downward trend overall
- Sudden drop due to COVID-19

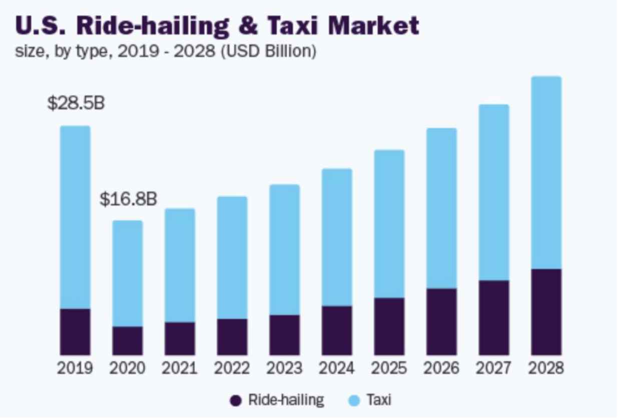
Business Problems - Market Expansion

01 → **02** → **03**

The growth of urban areas

The increasing popularity of ride-hailing service

The aging population



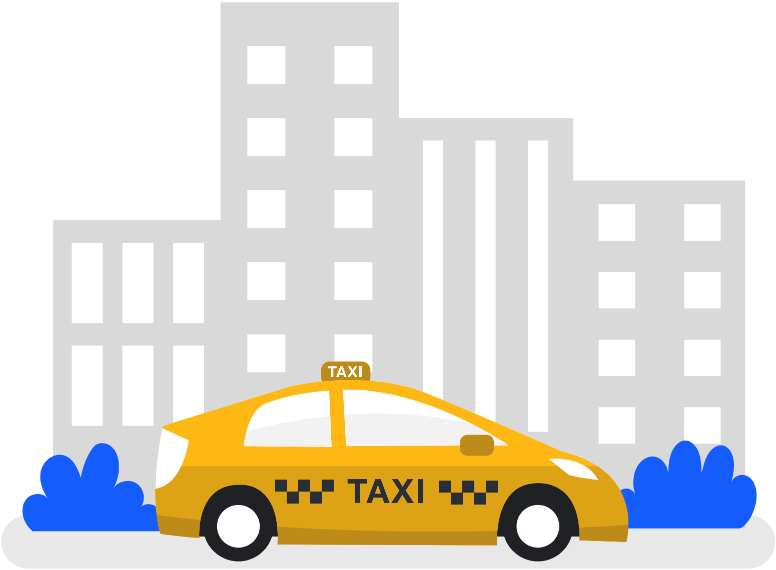
Project Goals

Analysis

Identify temporal patterns and trends in taxi trips

Explore factors influencing passenger demand

Develop predictive models to forecast future demand



Benefits

Optimize taxi companies fleet size

Reduce wait times for passengers

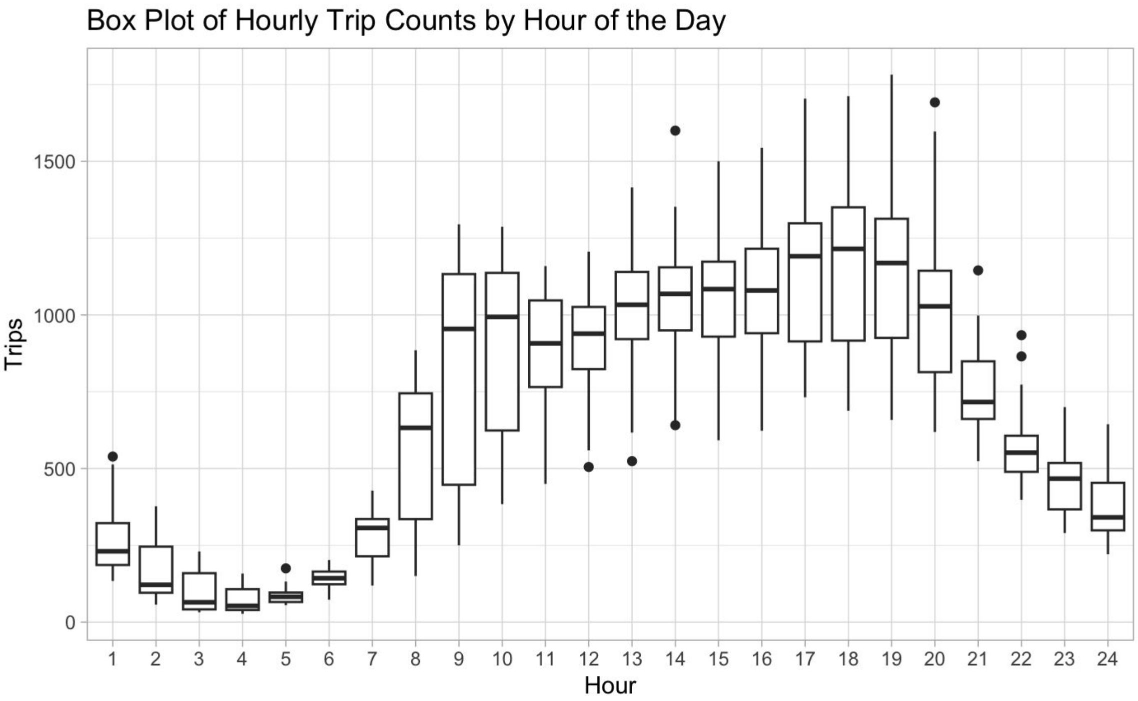
Improve profitability and resource utilization

Dataset

Data source:
Chicago Data Portal

Time range:
February 2023

Valuables:
trip ID, pickup/drop-off locations, distance, fares, and timestamps



Data - Seasonality

ACF Test

Outputs

Positive Autocorrelation:

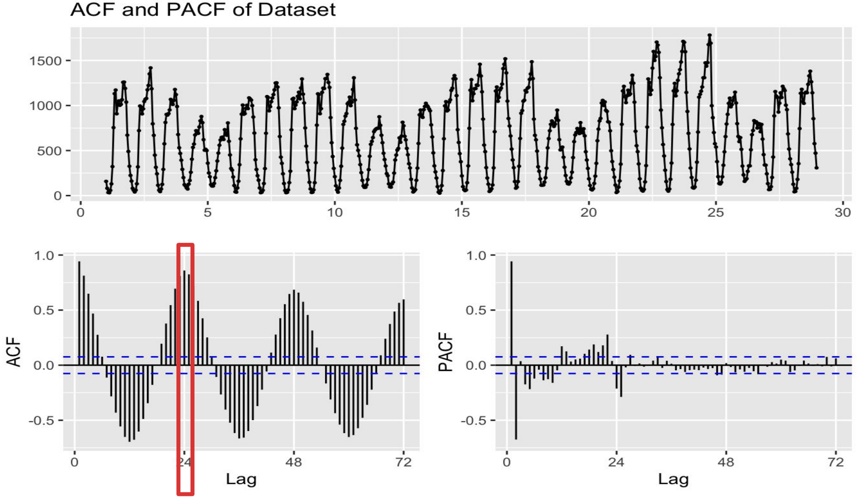
- Daily pattern observed at lag 24, 48, and 72
- Consistent relationship between taxi trip counts at the same hour across consecutive days

Negative Autocorrelation:

- Alternating pattern observed at lag 12, 36, and 60
- Tendency for high and low trip counts at specific intervals within a day

Conclusions

- Significant correlation exists in the hourly taxi trip count data



Data - Normality

Shapiro-Wilk Normality Test

Null Hypothesis

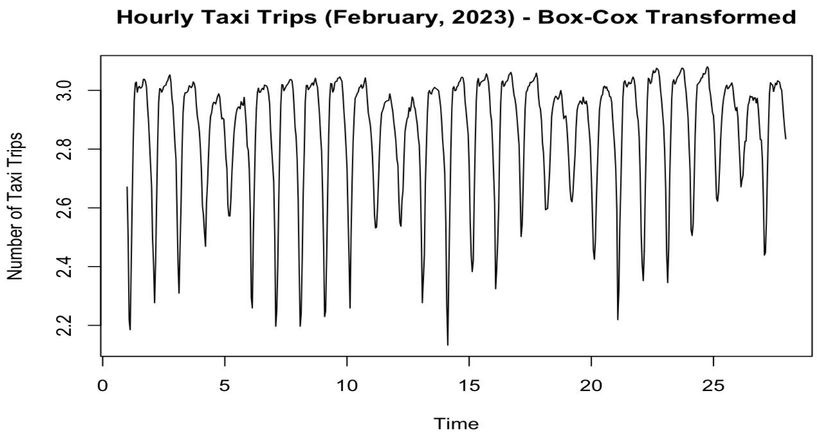
- Data follows a normal distribution
- Significance Level: 0.05

Outputs

- p-value: 0.000000000000002999
- Rejection of the null hypothesis
- Box-Cox transformation lambda: -0.2866391

Conclusions

- Not follow a normal distribution
- Applying Box-Cox transformation may help approximate normality in the data



Data - Trend Stationarity

KPSS Test

Null Hypothesis

- The data is stationary
- Significance Level: 0.05

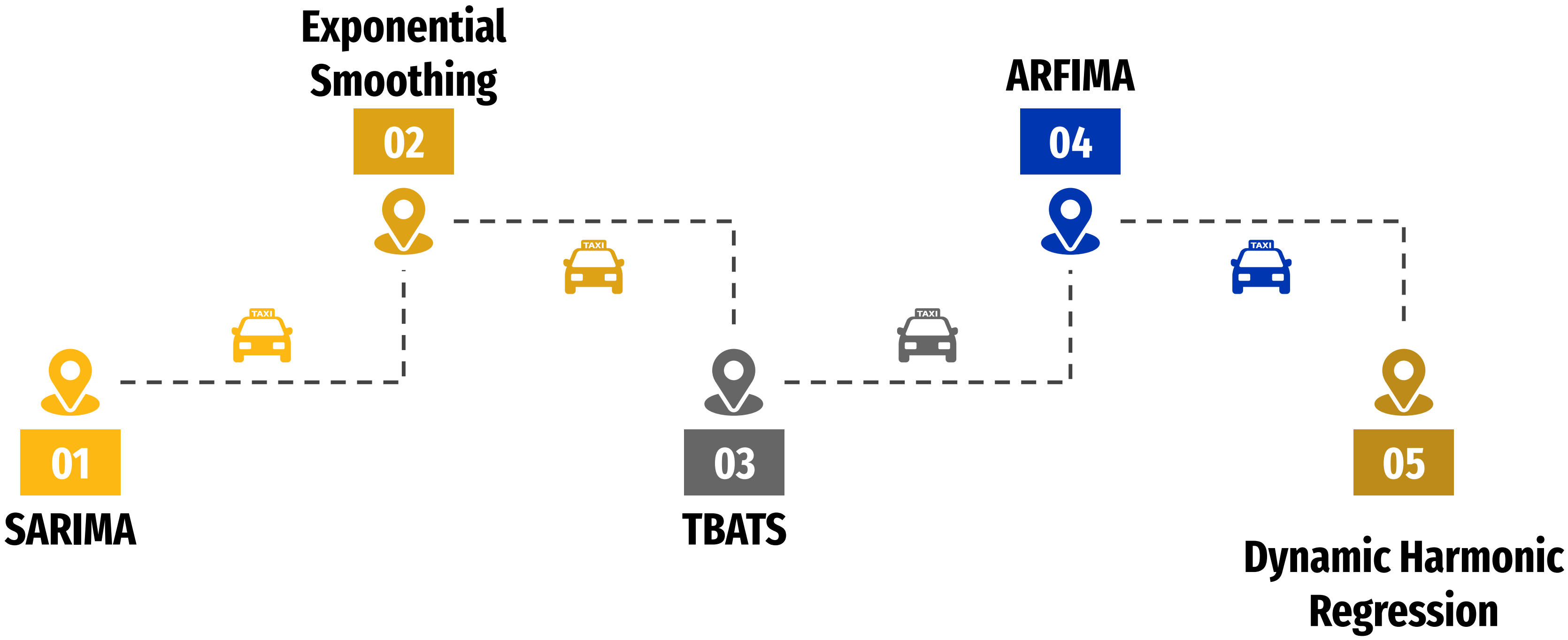
Outputs

- p-value: 0.1
- Not reject the null hypothesis

Conclusions

- The data is stationary
- No further differencing or transformation is required.

MODELS


Exponential Smoothing
02
ARFIMA
04
TAXI
01
SARIMA
03
TBATS
05
Dynamic Harmonic Regression

SARIMA: Seasonal ARIMA

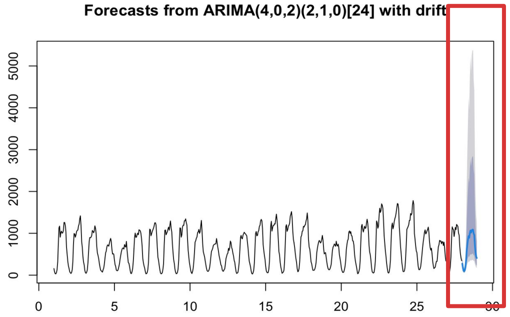
Incorporates both non-seasonal and seasonal factors in a multiplicative model

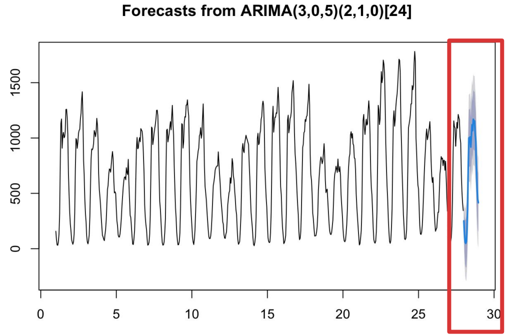
Using auto arima(): ARIMA(4,0,2)(2,1,0) with drift

- Better account for the regular fluctuations in taxi demand throughout the day
- Capture the **temporal dependencies** in taxi demand using **AR & MA** components
- Capture both **short-term fluctuations** and **long-term trends**

Fine tuning using eacf(): ARIMA(3,0,5)(2,1,0)

- Non-seasonal component
  - Negative relationship between the current and lagged value
- Seasonal component
  - Negative relationship between current value & past lags at the same seasonal periods
- 1 seasonal differencing is applied to remove seasonality



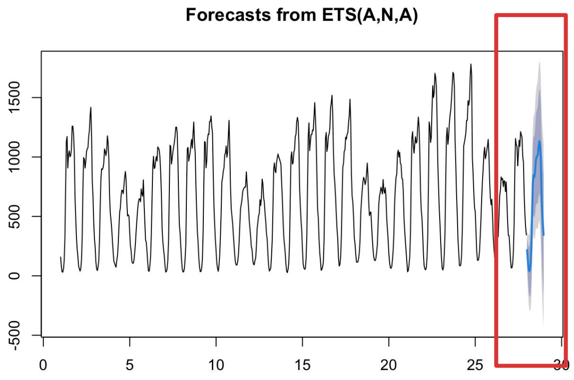


Exponential Smoothing

Capturing seasonal fluctuations but large prediction intervals

Model Interpretation

- The errors are additive, there is no trend in the data, and the seasonality is additive
- The prediction intervals show that **there is considerable uncertainty in the future values** of hourly taxi demand for next day



ETS(A, N, A)

A: additive error term

N: the trend is assumed to be non-existent or a flat line

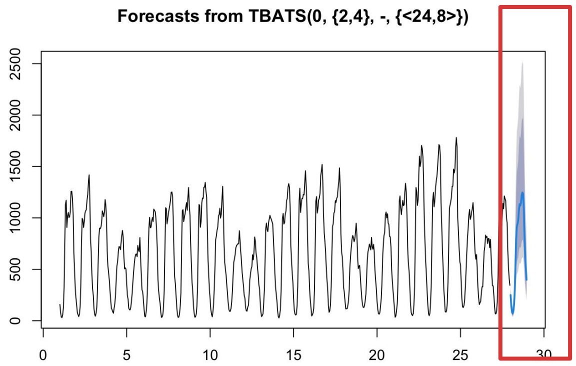
A: additive seasonal component

TBATS: suitable for capturing multiple cyclic patterns

Trigonometric Seasonal, Box-Cox Transformation, ARMA Errors, Trend, and Seasonal

Model Interpretation

- Hourly data has various types of seasonalities.
- Suitable for higher frequency time series data that exhibit complicated seasonally pattern.
- Allows automatic box-cox transformation and ARMA errors.
- Allows for dynamic seasonality: the seasonality is allowed to change slowly over time.



TBATS(0, {2, 4}, -, {<24, 8>})

1. Natural log transformation
2. {2, 4}: ARMA error
3. -: no trend, no need to adapt damping parameter
4. <24, 8>: seasonal period & fourier terms

ARFIMA: capturing long term memories

Autoregressive Fractionally Integrated Moving Average

Deal with 'long memory' processes

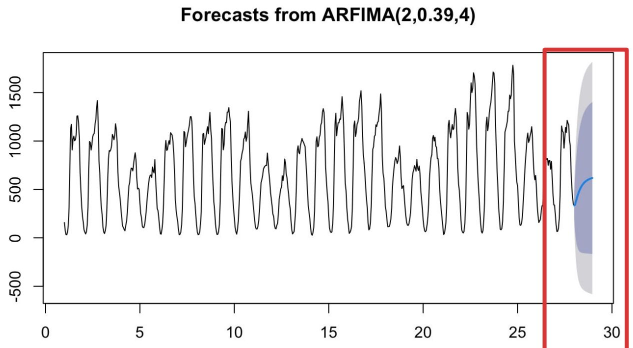
If a big event (like a seasonal sports game) led to a spike in taxi demand one week, a similar spike may be expected when a similar event occurs in the future.

ARFIMA(2, 0.39, 4)

AR: p = 2 - the number of trips depends on previous 2 hours

I: d = 0.3889 - fractionally differencing

MA: q = 4 - error term depends on previous 4 error terms



Dynamic Harmonic Regression: forecasting with 7 possible seasonal periods and short-term fluctuations

Multiple Seasonal Periods

7 Fourier terms capture up to 7 cycles of seasonality present in the train data.

Short-term Dynamics

AutoRegressive Moving Average (ARMA) error handles short-term dynamics that are not seasonality.

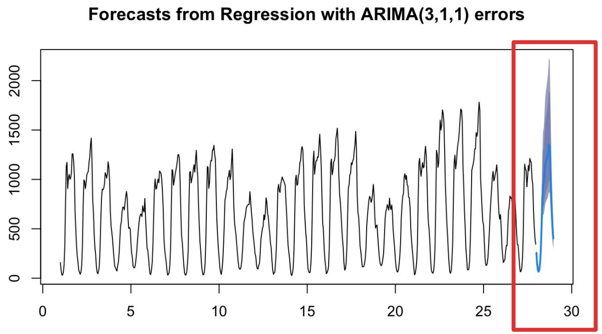
Regression with ARIMA(3,1,1) Errors

AR: p = 3 - the # of trips depends on previous 3 hour

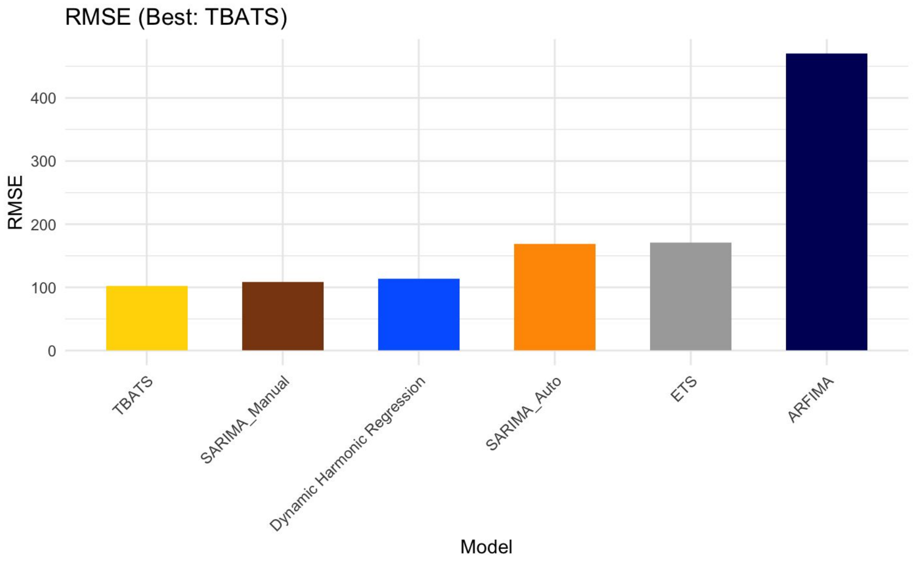
I: d = 1 - 1st differencing

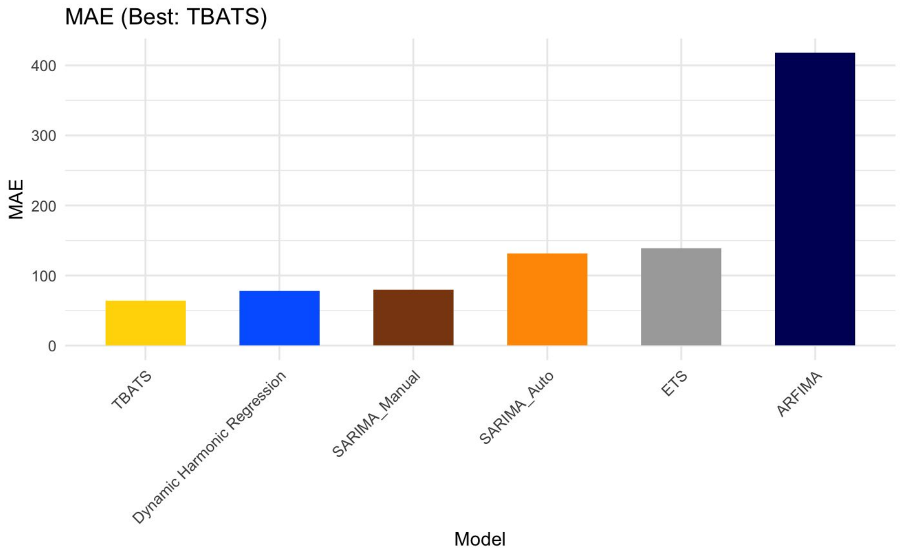
MA: q = 1 - error term depends on previous 1 error term

Box Cox transformation: lambda= -0.2866



Evaluating Forecast Accuracy: Scale-Dependent Measures





RMSE (Root Mean Squared Error):

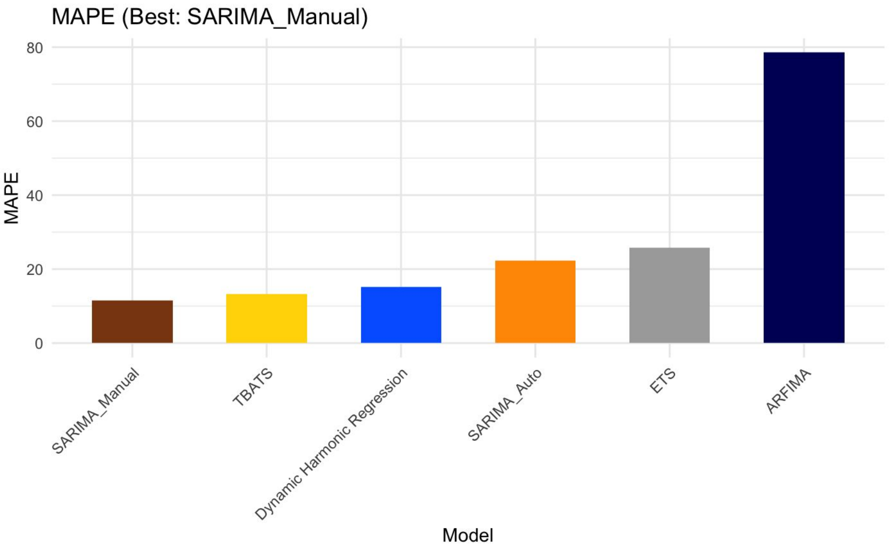
- The RMSE value for **TBATS** is 102.5660.
- On average, the **TBATS** model's predictions have an error of approximately **103 taxi trips**.

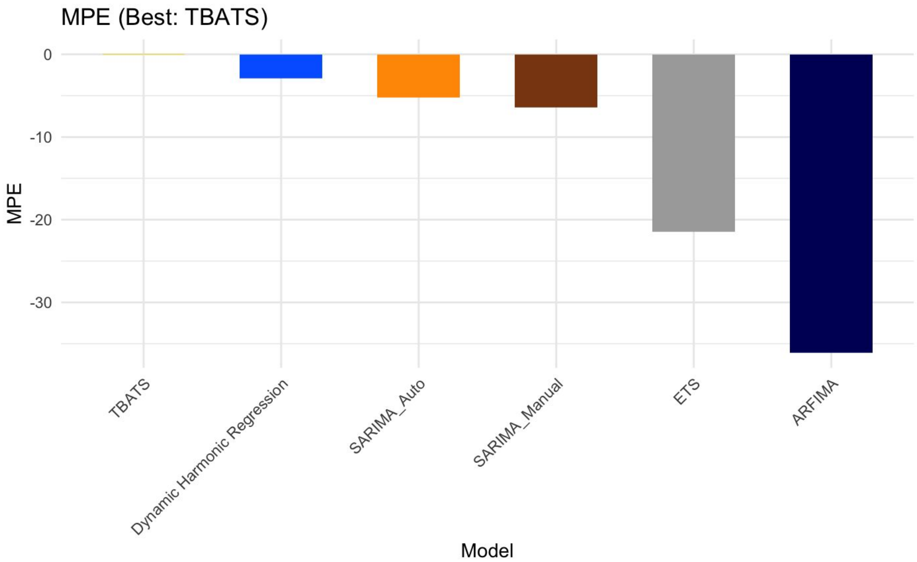
MAE (Mean Absolute Error):

- The MAE value for **TBATS** is 64.18240.
- On average, the **TBATS** model's predictions deviate by approximately **64 taxi trips** from the actual values.

Please consult Appendix E, which provides detailed information on the forecasting metrics and accuracy values for evaluating forecast accuracy.

Evaluating Forecast Accuracy: Scale-Independent Measures





MAPE (Mean Absolute Percentage Error):

- The MAPE value for **TBATS** is 13.35608.
- On average, the **TBATS** model's predictions have a deviation of approximately **13.36%** from the actual values.

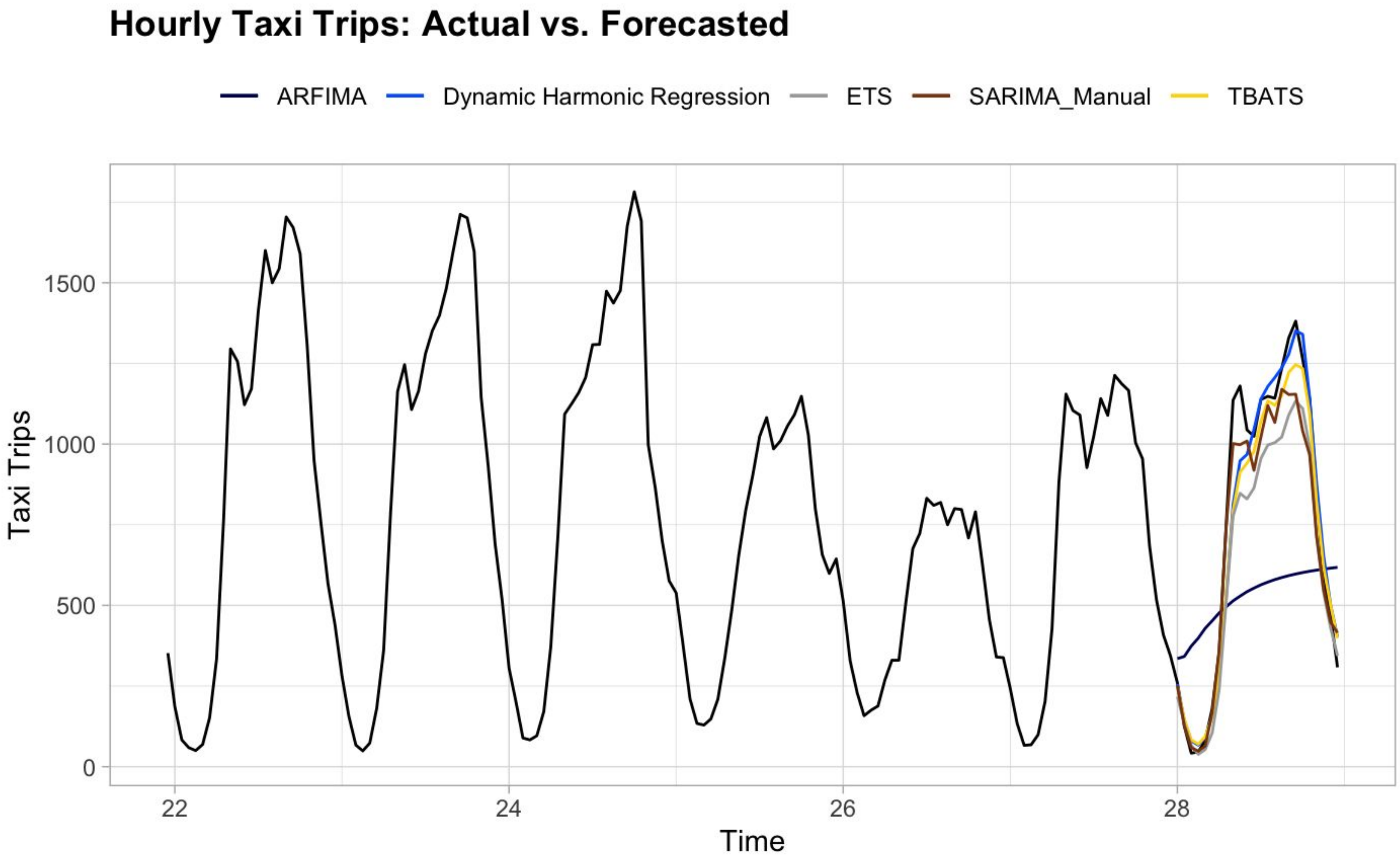
MPE (Mean Percentage Error):

- The MPE value for **TBATS** is 0.02603666.
- On average, the model's predictions have a slight positive bias, **overestimating** the taxi trip counts by approximately **0.03%**.

Please consult Appendix E, which provides detailed information on the forecasting metrics and accuracy values for evaluating forecast accuracy.

Model Comparison

	SARIMA_Manual	TBATS	Dynamic Harmonic Regression	ETS	ARFIMA
RMSE	★★★★	★★★★★	★★★	★★	★
MAE	★★★	★★★★★	★★★★	★★	★
MAPE	★★★★★	★★★★	★★★	★★	★
MPE	★★★	★★★★★	★★★★	★★	★


Hourly Taxi Trips: Actual vs. Forecasted
ARFIMA
Dynamic Harmonic Regression
ETS
SARIMA_Manual
TBATS
Taxi Trips
1500
1000
500
0
22
24
26
28
Time

Conclusion

01

Data Characteristics

Contain a strong seasonality in hourly taxi trip counts

02

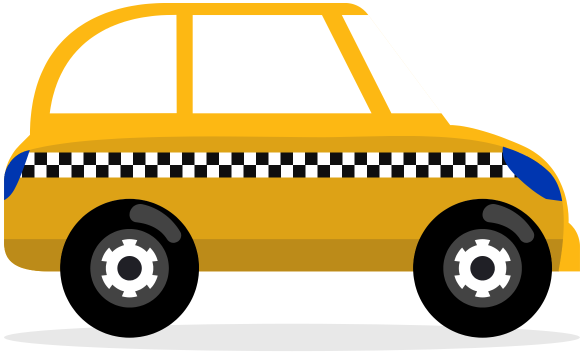
Data Assumption

Made stationary after transformation

03

Model Performance

TBATS and Dynamic Harmonic Regression models work best due to their ability to capture multi-seasonality



04

Model Residual

Haven't adequately captured all the patterns and information

05

Model Prediction

Mostly capture the seasonality despite the scale difference in prediction

06

Improvement

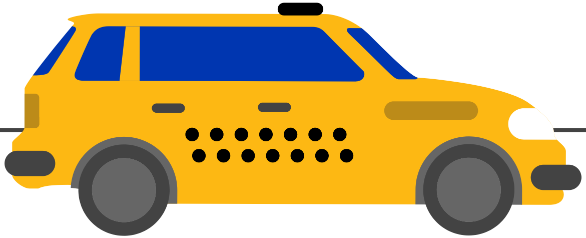
Incorporate additional variables or features

Future work

01

02

03



Model Improvement

- Multivariate Time Series and corresponding model structure
- Refine model specification

Timely Update

- Rerun model everyday to get better prediction
- Make necessary changes to model structures if needed

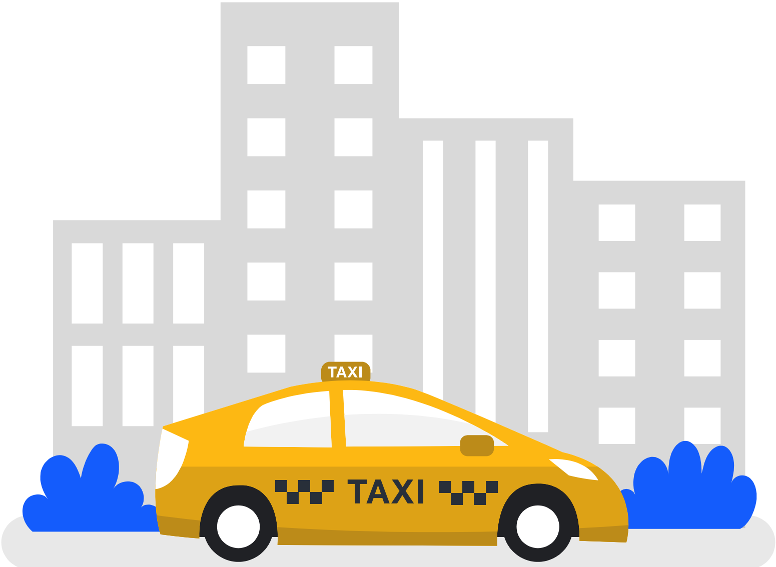
Supply Optimization

- Utilize the prediction for better supply allocation
- Build model based on different Taxi company

Thank you!

Any questions?

Appendix



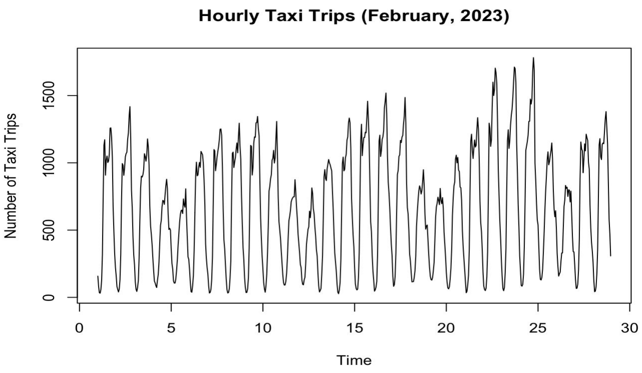
Team 4 Members and Task Distribution

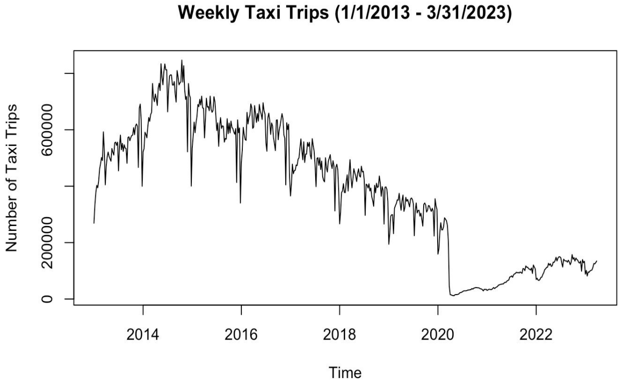


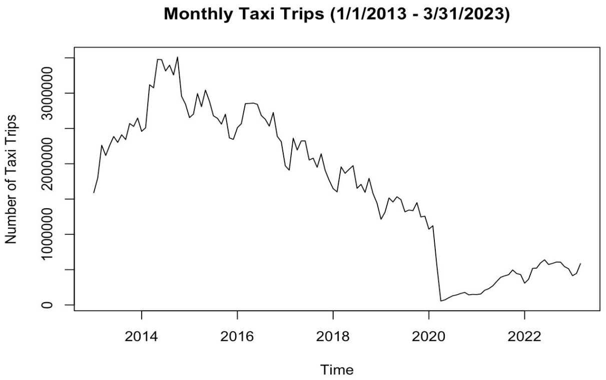


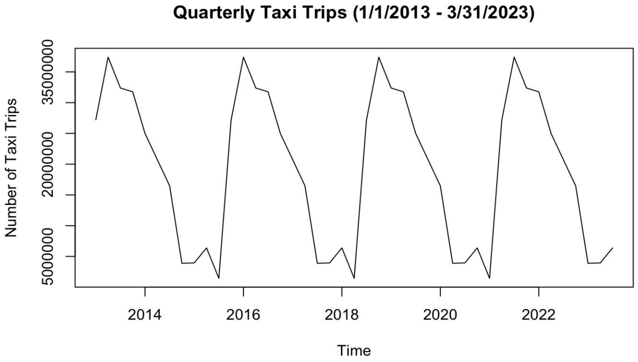
	Maggie Chuang	Kelsey Liu	Shirley Liu	Evelyn Wu	Vivian Yeh
Model Analysis	★	★	★	★	★
Code Integration			★	★	
Business Analysis	★	★	★	★	★
Slide Creation	★	★	★	★	★

Appendix A. Hourly/Weekly/Monthly/Quarterly Comparison









Appendix B. Residual Plots

SARIMA

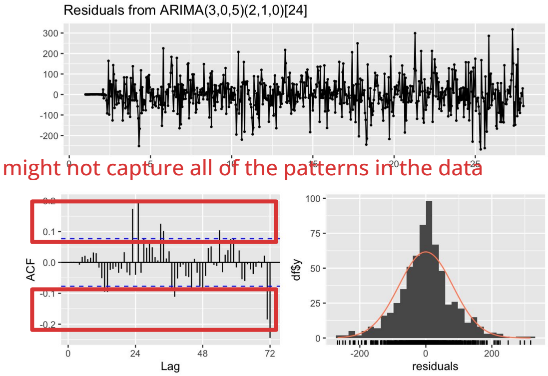
- <u>ACF residual plots</u>
  - not white noise
- <u>Residual Histogram</u>
  - normally distributed

Exponential Smoothing

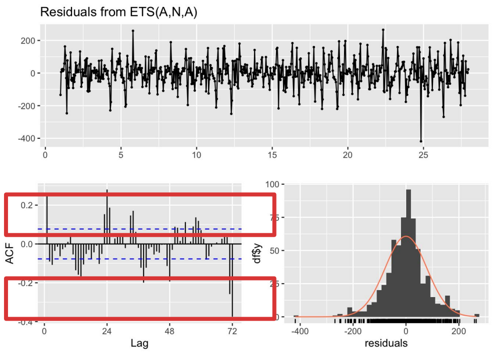
- <u>ACF residual plots</u>
  - not white noise
- <u>Residual Histogram</u>
  - normally distributed

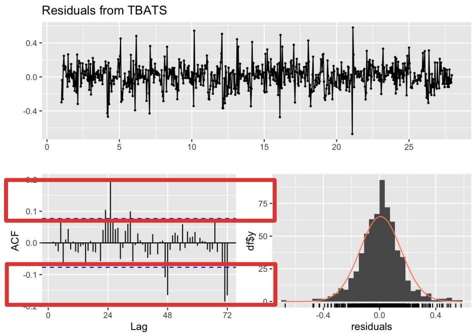
TBATS

- <u>ACF residual plots</u>
  - not white noise
- <u>Residual Histogram</u>
  - normally distributed



model might not capture all of the patterns in the data

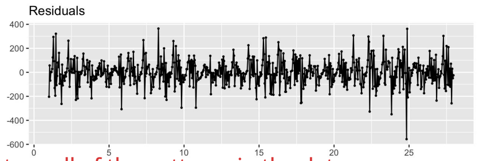




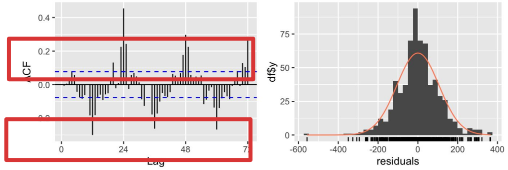
Appendix C. Residual Plots

ARFIMA

- <u>ACF residual plots</u>
  - not white noise
- <u>Residual Histogram</u>
  - normally distributed

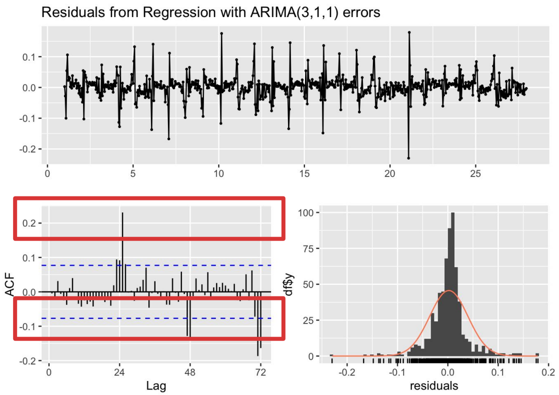


model might not capture all of the patterns in the data

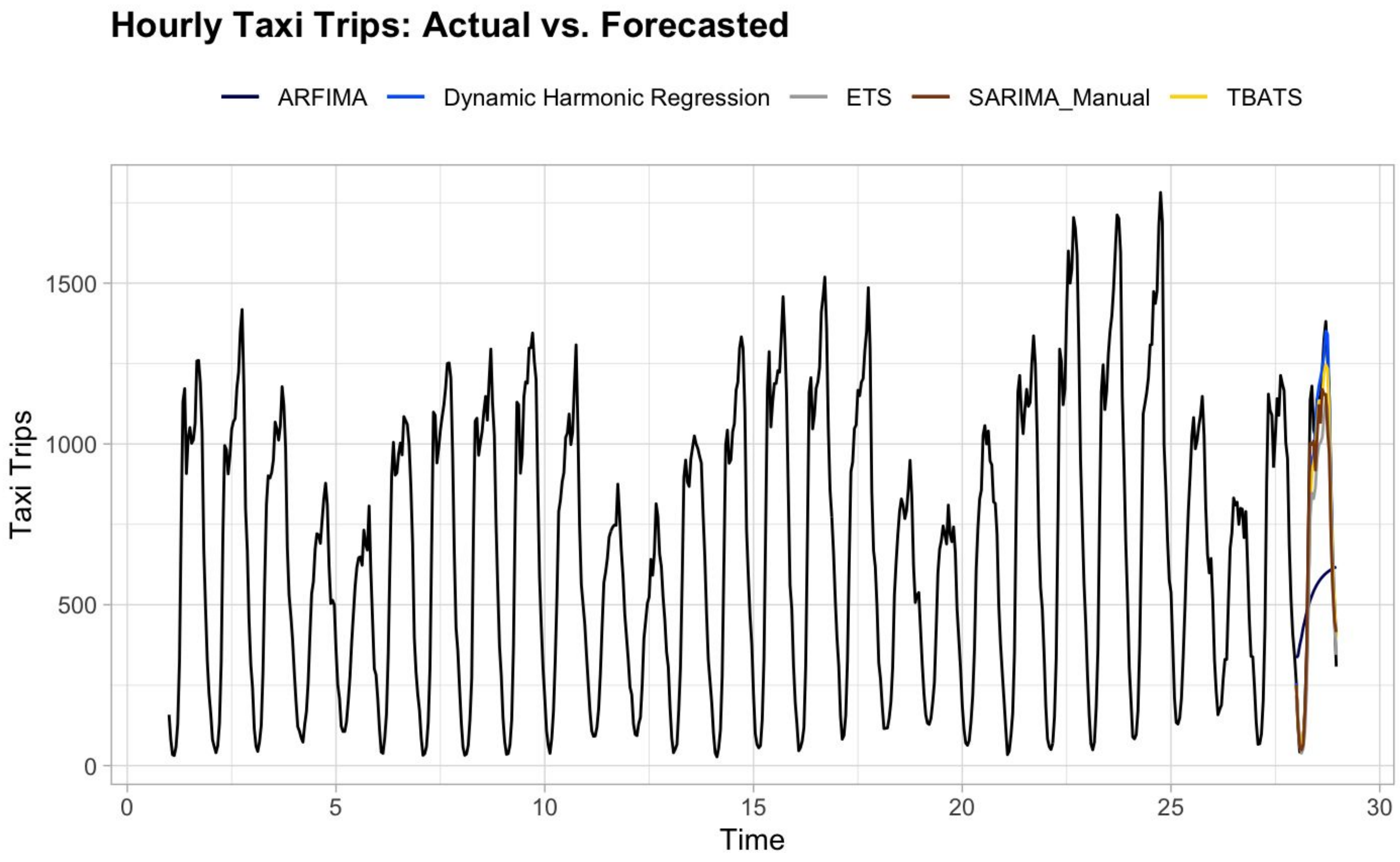


Dynamic Harmonic Regression

- <u>ACF residual plots</u>
  - not white noise
- <u>Residual Histogram</u>
  - normally distributed



Appendix D. Monthly Taxi Trips: Actual vs. Forecasted



Appendix E. Forecast Accuracy

Model <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
ARIMA	-357.64449	497.4271	410.41048	-75.22301746	99.70367
SARIMA_Auto	-105.16264	169.1483	131.88928	-5.19455729	22.25136
SARIMA_Manual	-68.43938	108.6076	79.41854	-6.45875717	11.55319
ETS	-133.29504	170.7190	138.95600	-21.46020955	25.85445
ARFIMA	-233.74790	469.8839	417.61335	-36.12321805	78.53504
Dynamic Harmonic Regression	-56.03499	113.7951	77.86034	-2.92832838	15.16795
TBATS	-21.47085	102.5660	64.18240	0.02603666	13.35608