Divvy Bike Usage Forecasting

Presented by: Snigda G

Chapters

1. Project Overview
2. Data Description
3. Exploratory Data Analysis
4. Exploring Time Series Models
5. Model Evaluation
6. Future Work

Project Overview

Divvy

A bike-sharing program that operates in Chicago, Illinois, United States.

Objective

Predicting the number of trips for Divvy bikes on a monthly basis

Outcome

1. Resource Optimization: Strategically allocate bikes and docking stations to high-demand areas, reducing instances of unavailability or overcrowding.

2. Expansion and Infrastructure Planning: Installing additional docking stations or expanding Divvy's bike fleet during months with high demand.

Data Description

Our original dataset consists of approximately 4GB of data with Divvy bike trips between 2013 and 2023, with 9 features.

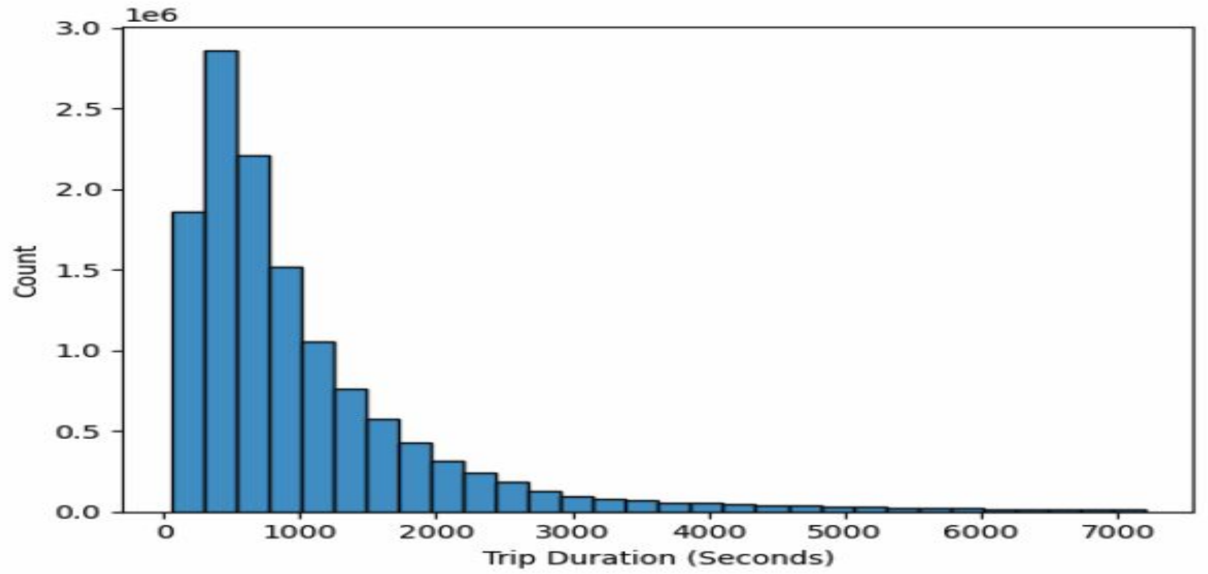
Due to the large size, which may cause the kernel to crash, we have finalized our data for modeling to include 3 main features, which are:

1. Year_month: from 06/2013 and 04/2023
2. Monthly Trip Count: number of trips in each month (in thousands)
3. Average Monthly Duration: average length of time for each trip in minutes (formula = total duration/total trip*60)

Train period = July 2013 - Dec 2021 & Test period = Jan 2022 - Apr 2023

EDA

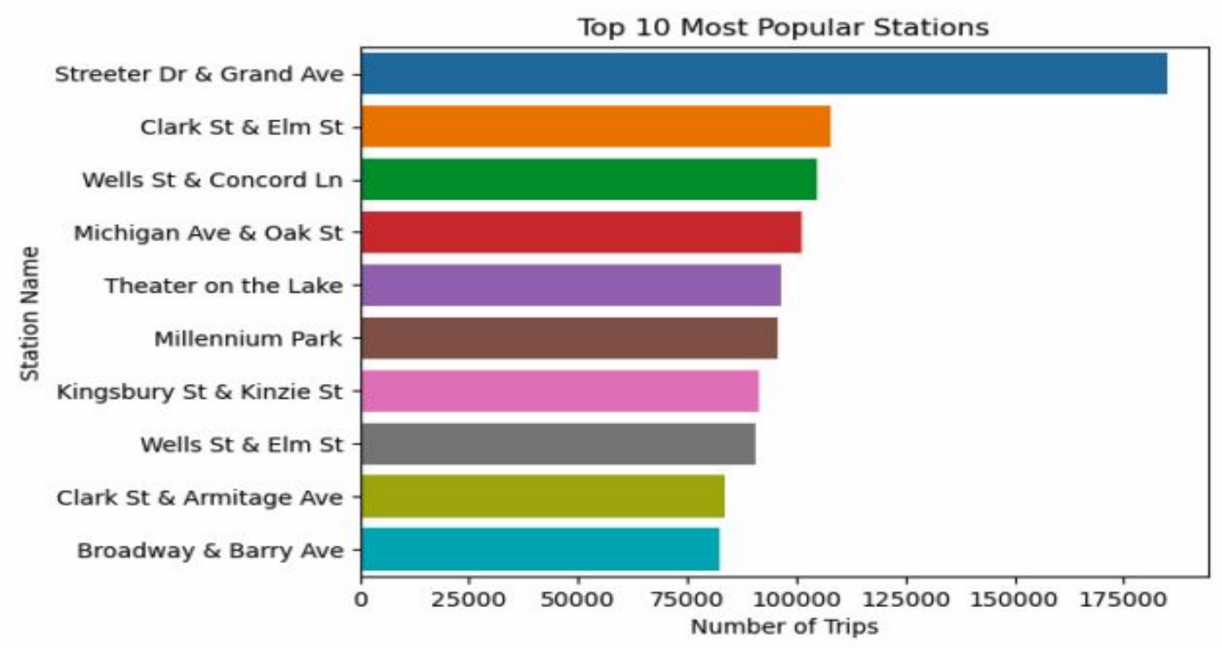
Trip Duration



The users tend to use divvy bikes only for short durations.

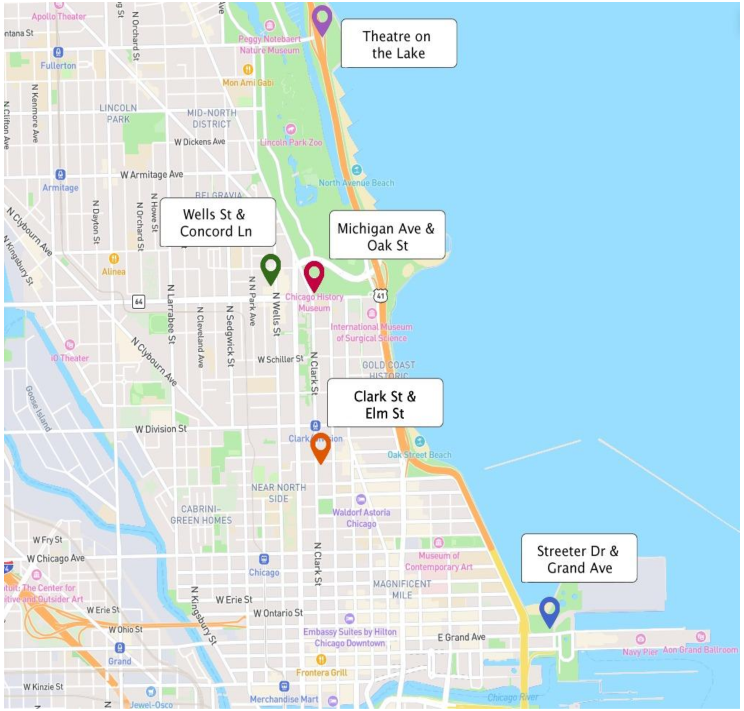
EDA

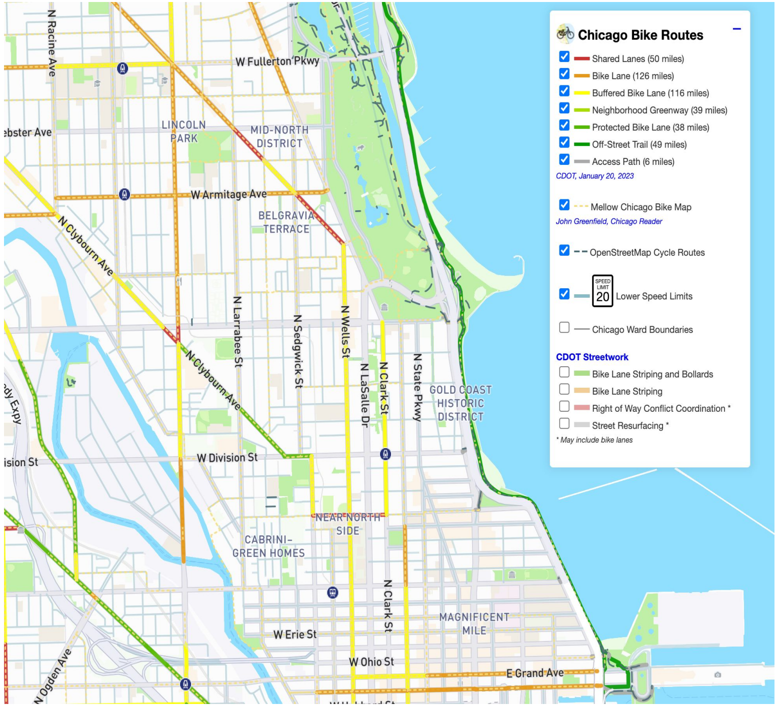
Most Popular Stations



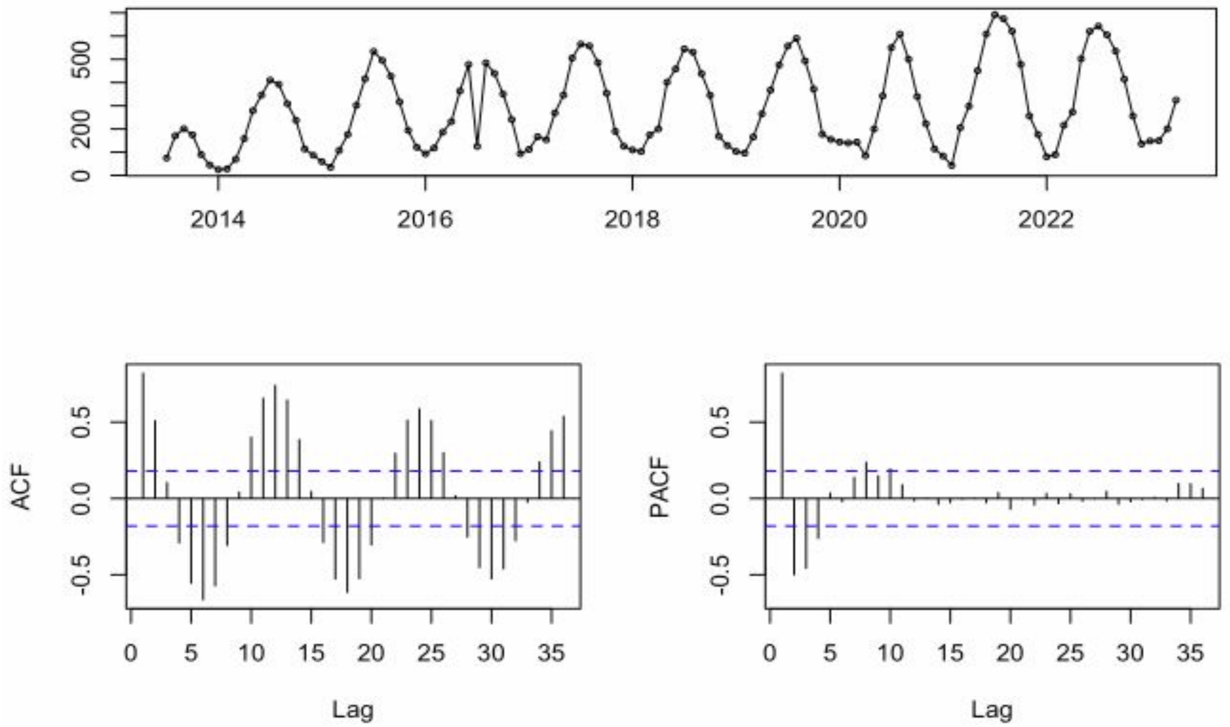
EDA

Most Popular Stations





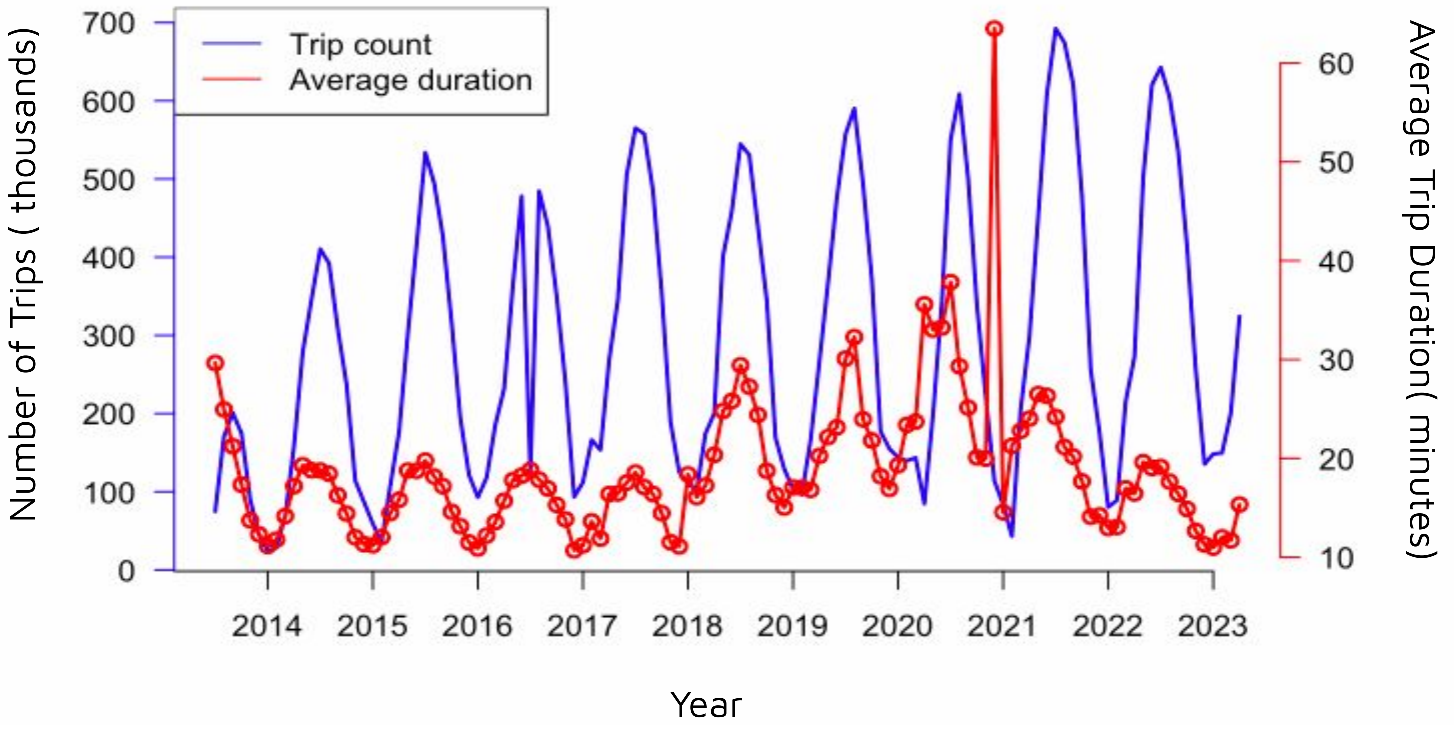
Yearly Divvy Rides in Chicago



- **Trend**: Slightly Increasing
- **Seasonal** patterns occurring on an annual basis
- **KPSS Test** : Non-stationary time series

Regression

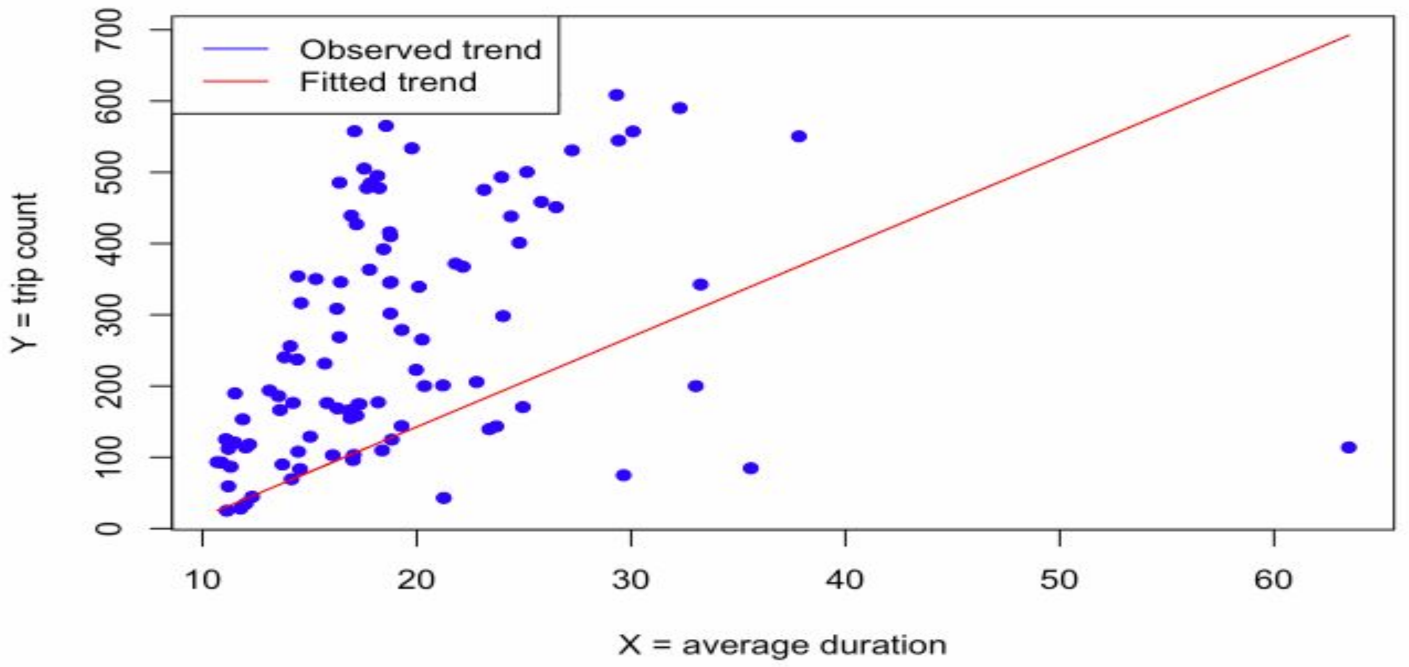
r = 0.342



DIVVY

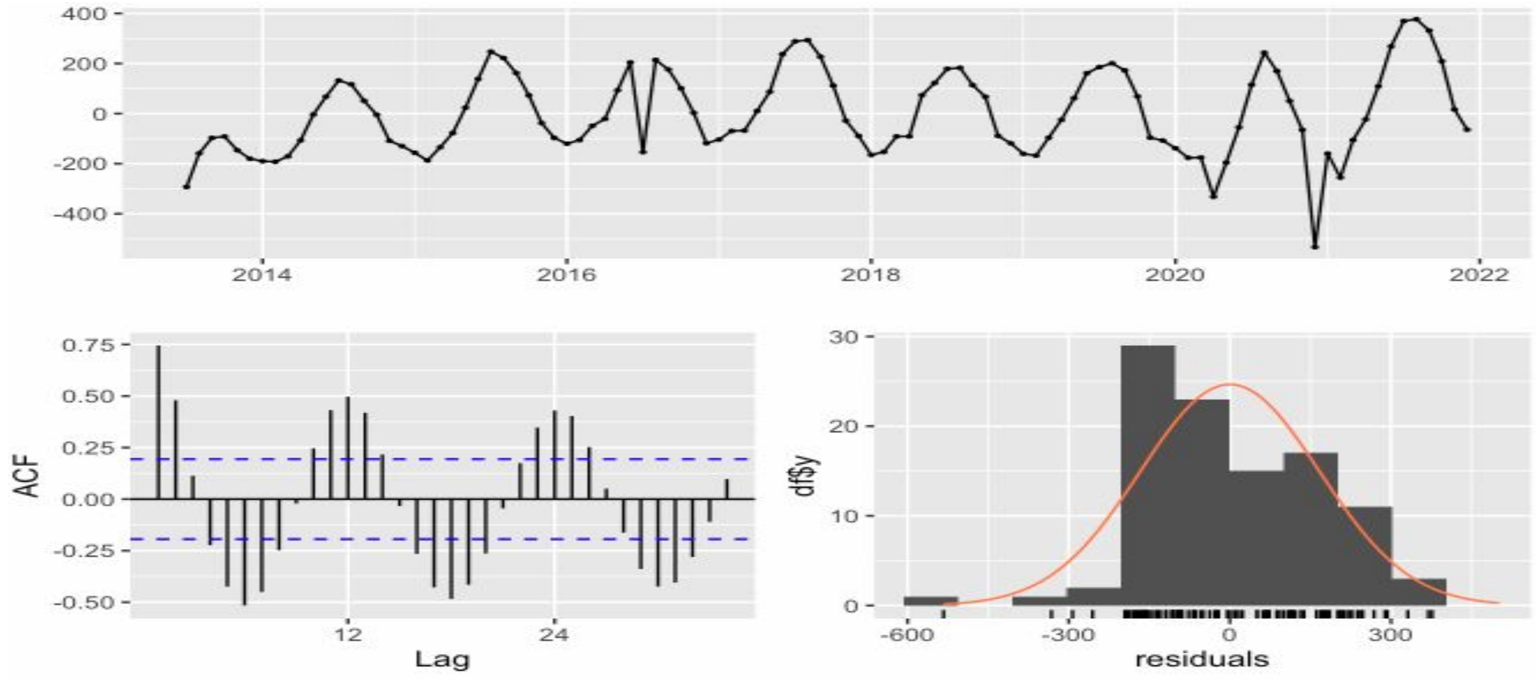
Regression

Observed V/s Fitted Trend



Setting up for Regression with ARIMA error

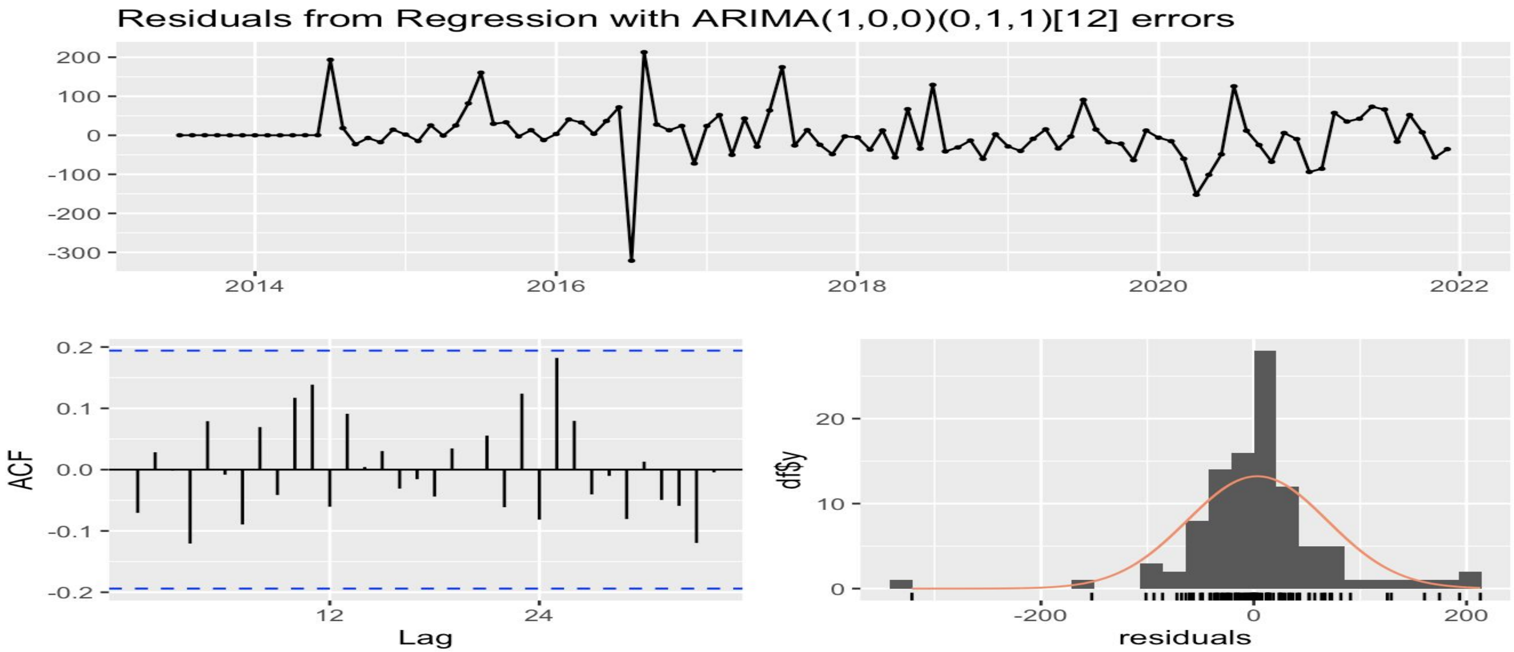
Residuals



DIVVY

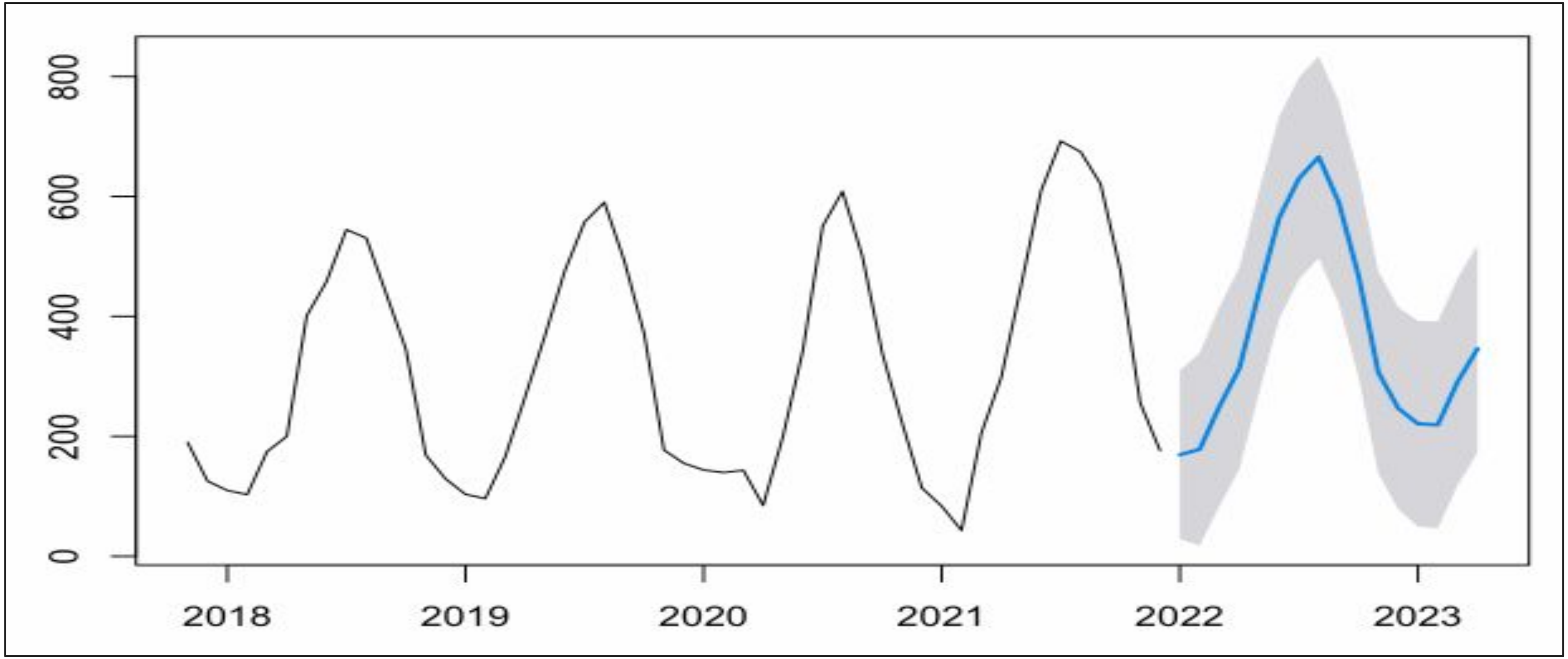
Regression with ARIMA error

ARIMA(1,0,0)(0,1,1)[12] residuals



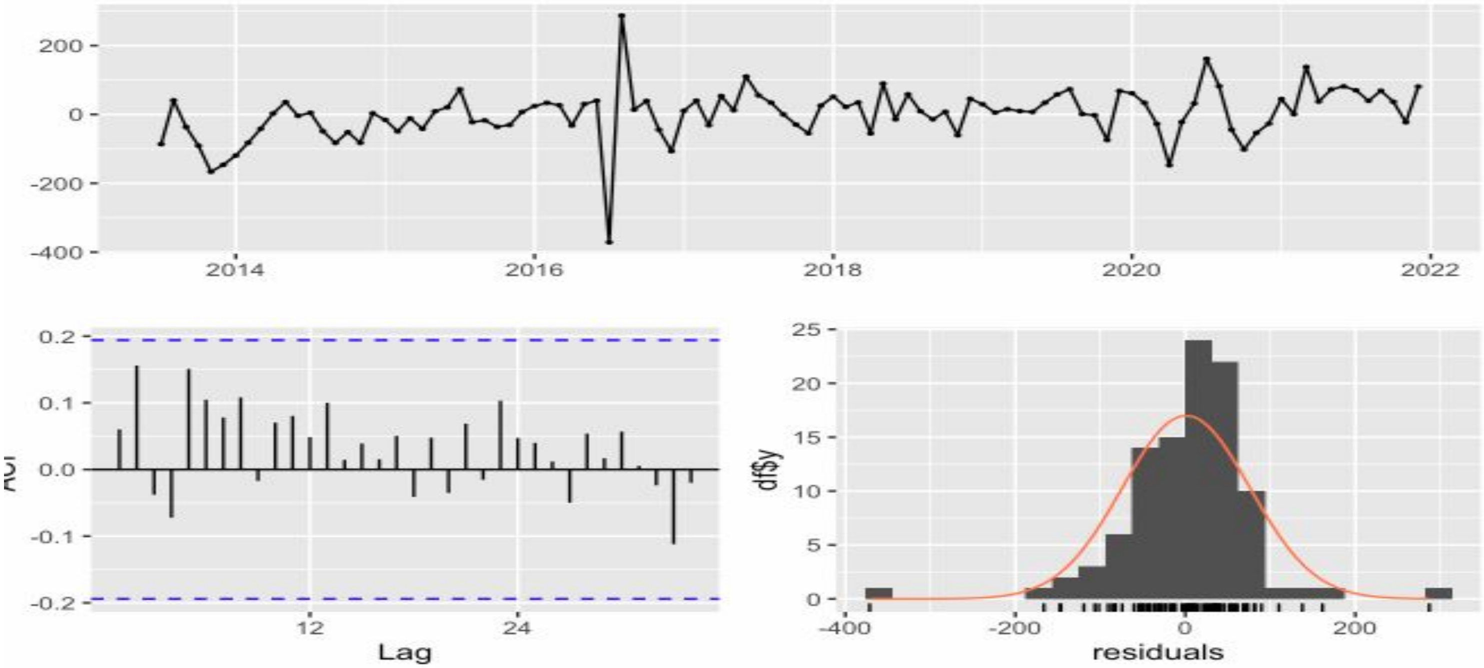
Regression & Regression with ARIMA Error

Forecast



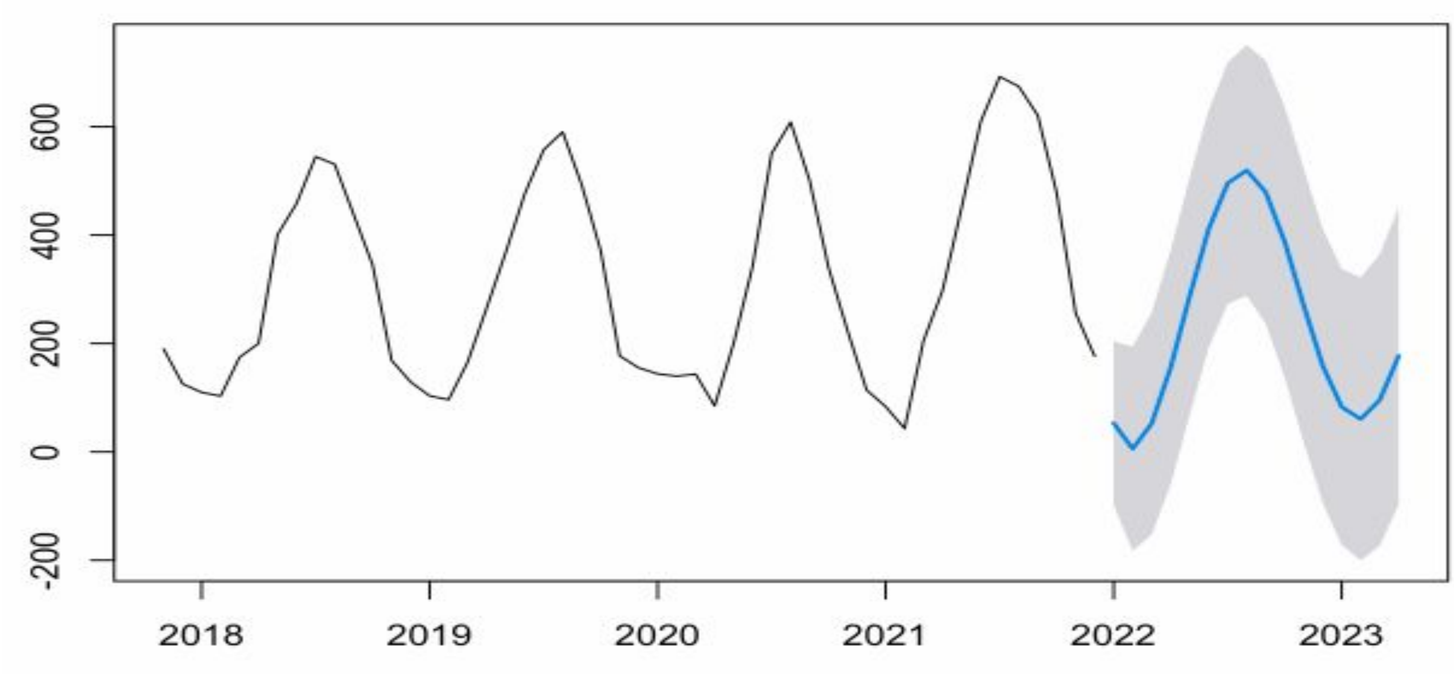
ARIMA Model

ARIMA(2,0,3) with non-zero mean residuals



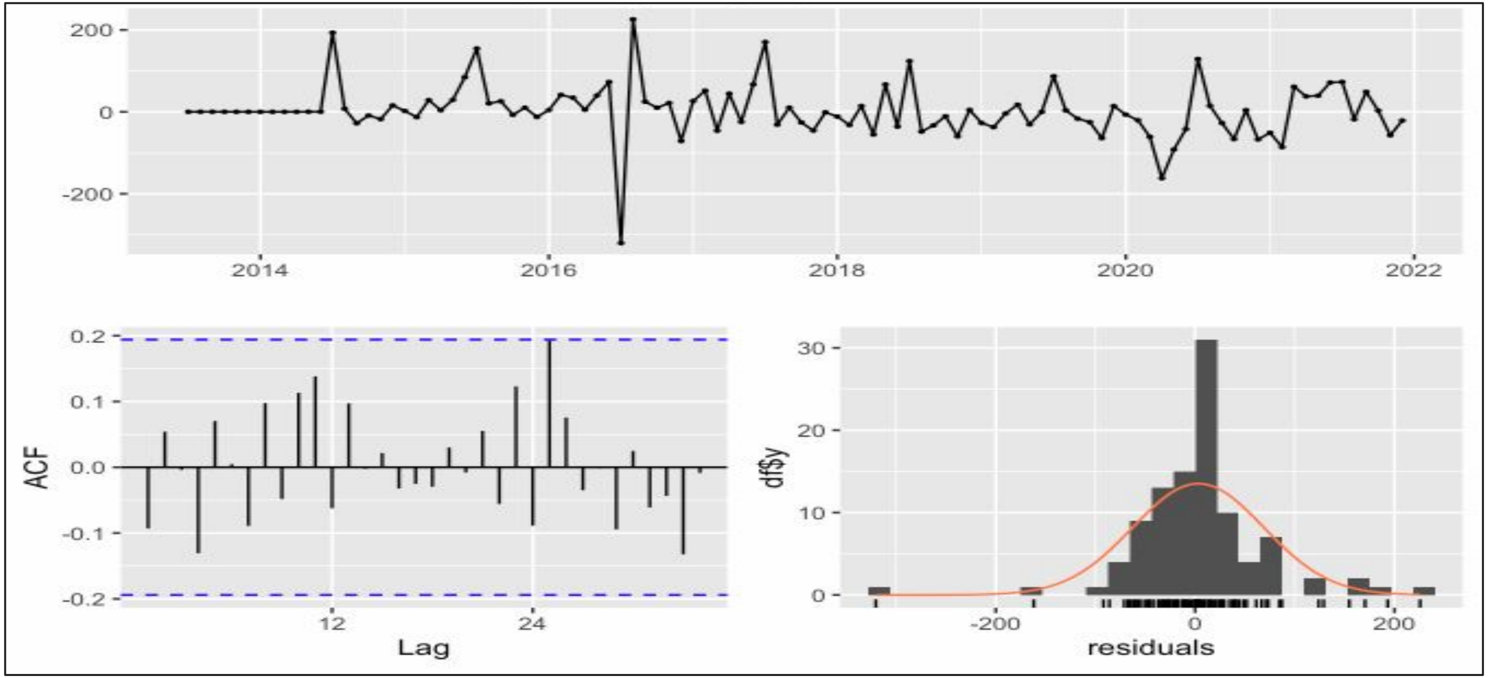
ARIMA (2,0,3) Model

Forecast



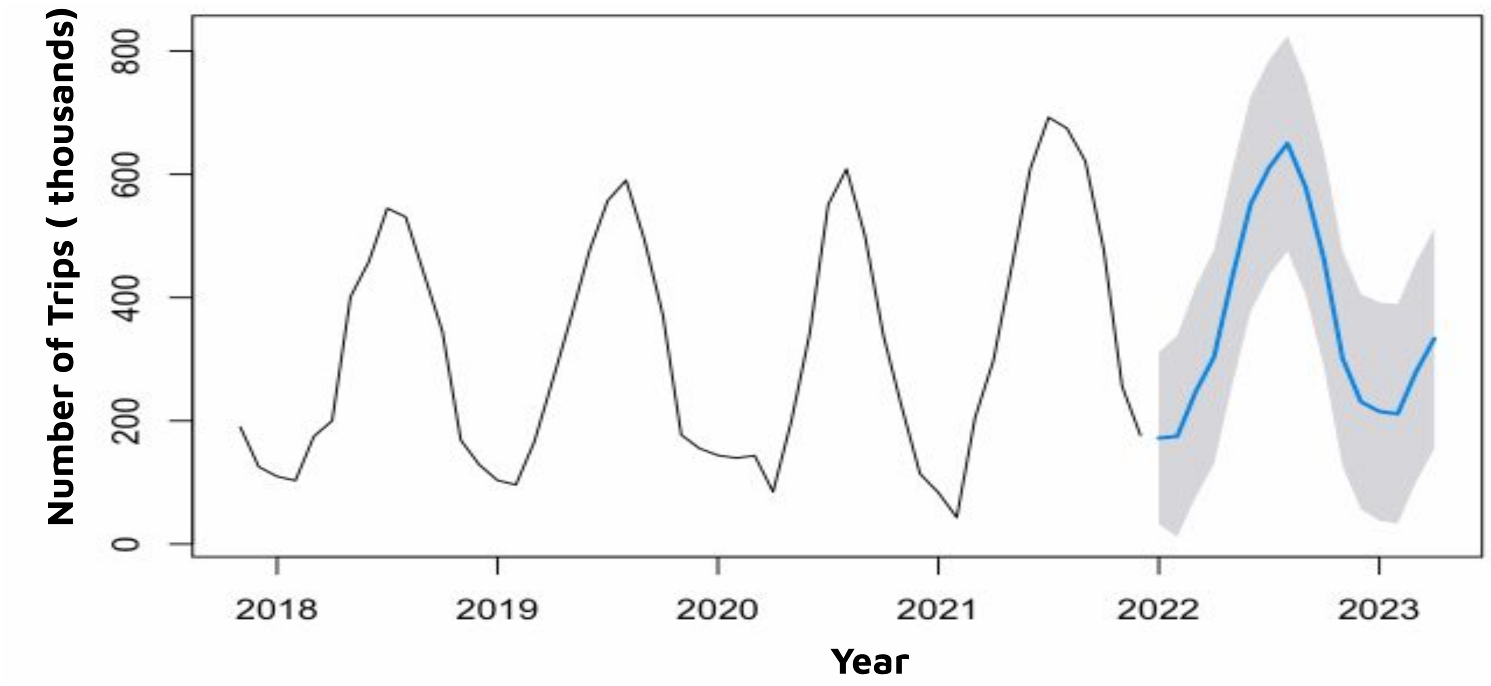
SARIMA Model

- **ARIMA (1,0,0)(0,1,1)[12] with drift**



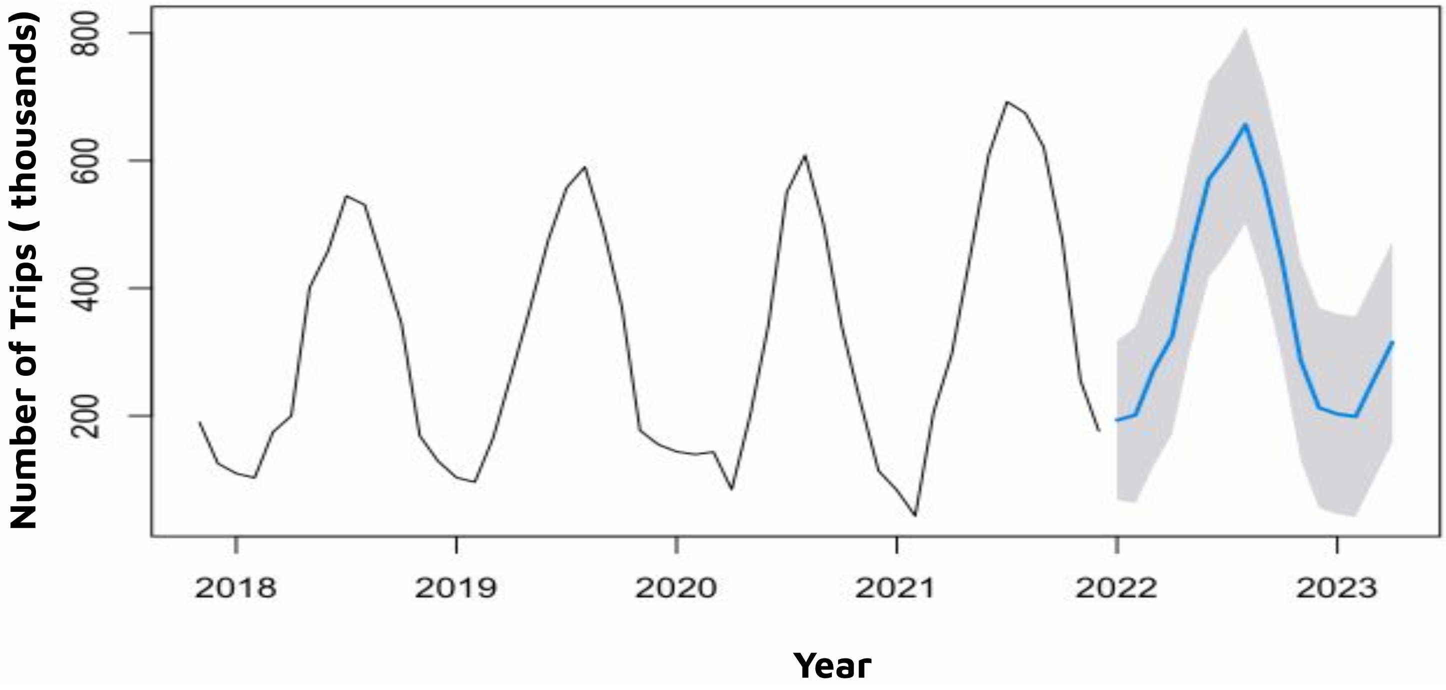
SARIMA (1,0,0)(0,1,1)[12] Model

Forecast



VAR (10) Model

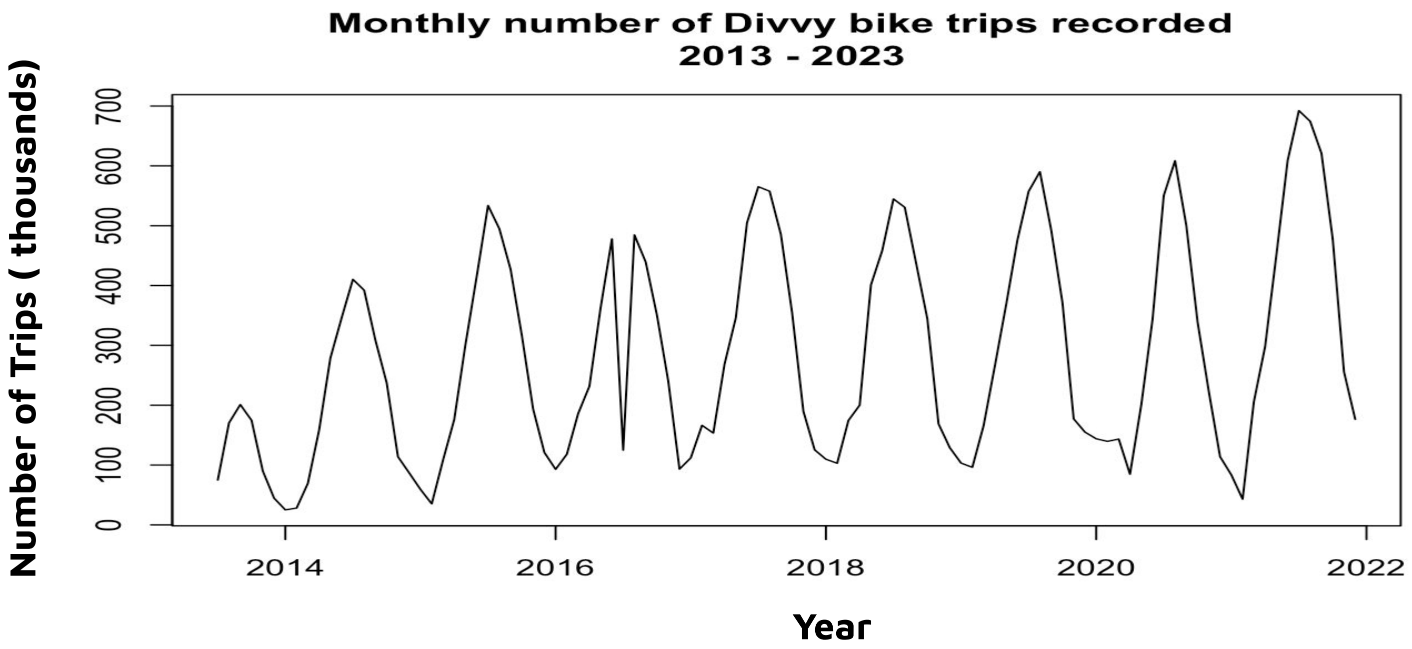
Forecast



DIVVY

Seasonal Holt-Winters Model

Multiplicative v/s Additive



Seasonal Holt-Winters Model

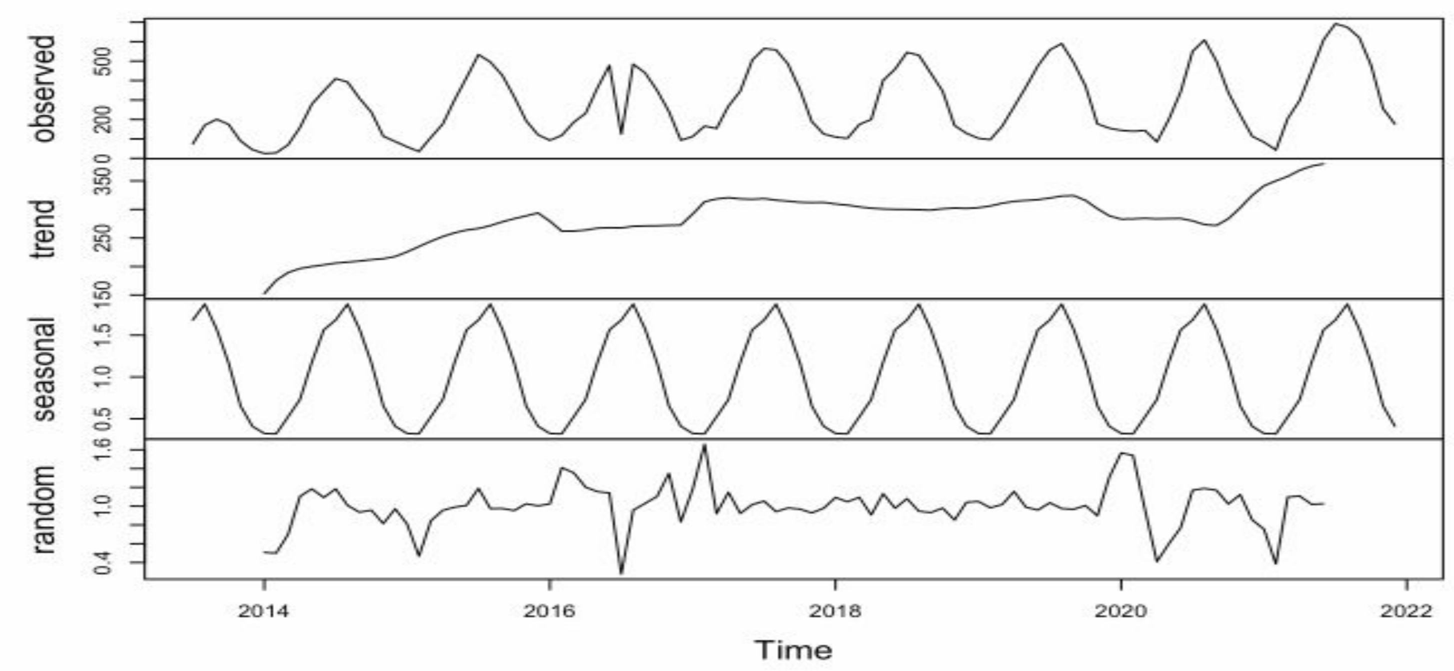
Accuracy Comparison

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Additive without damping	-1.368893	65.85313	43.22621	-5.772363	26.12261	0.624781	0.0763598
Multiplicative without damping	-4.53528	62.63694	40.20102	-12.49822	24.29647	0.5810557	0.1876421
Additive with damping	2.804036	62.78255	39.91608	-8.187681	23.32441	0.5769372	0.1316823
Multiplicative with damping	2.804036	62.78255	39.91608	-8.187681	23.32441	0.5769372	0.1316823

DIVVY

Seasonal Holt-Winters Model

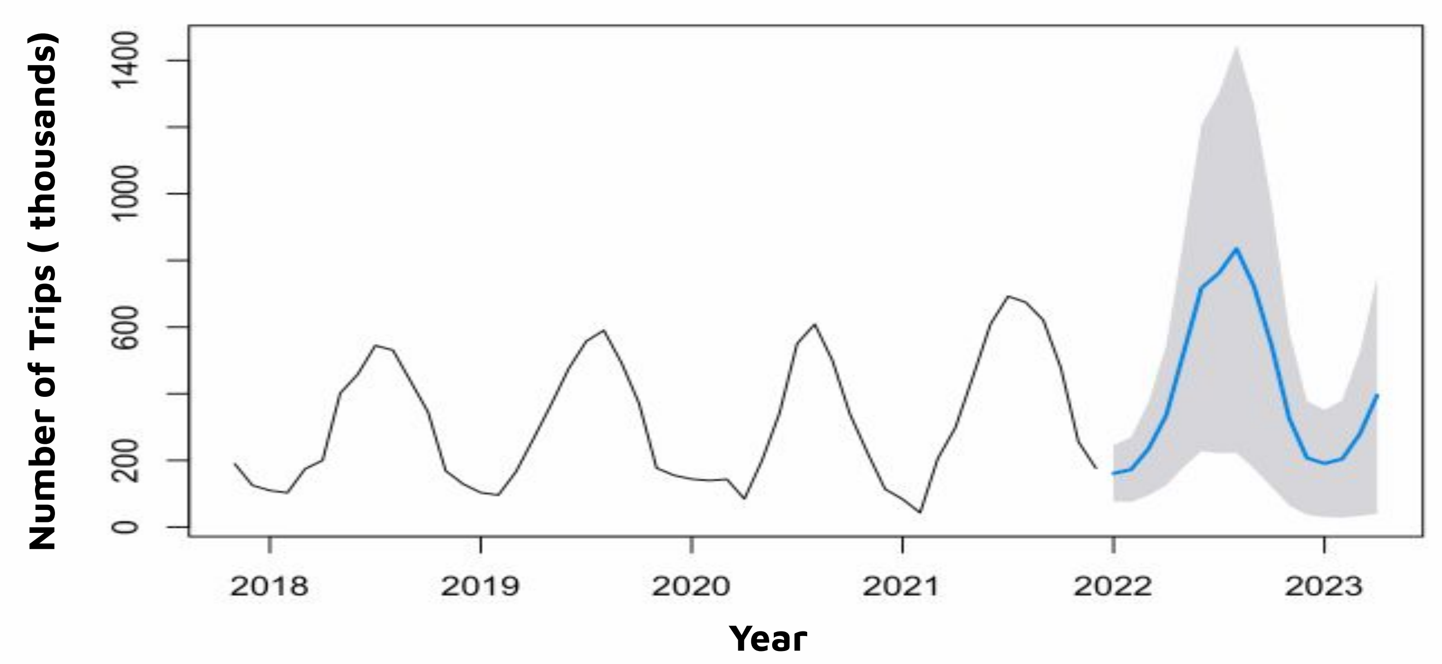
Decomposition of Multiplicative Time Series



DIVVY

Seasonal Holt-Winters Model

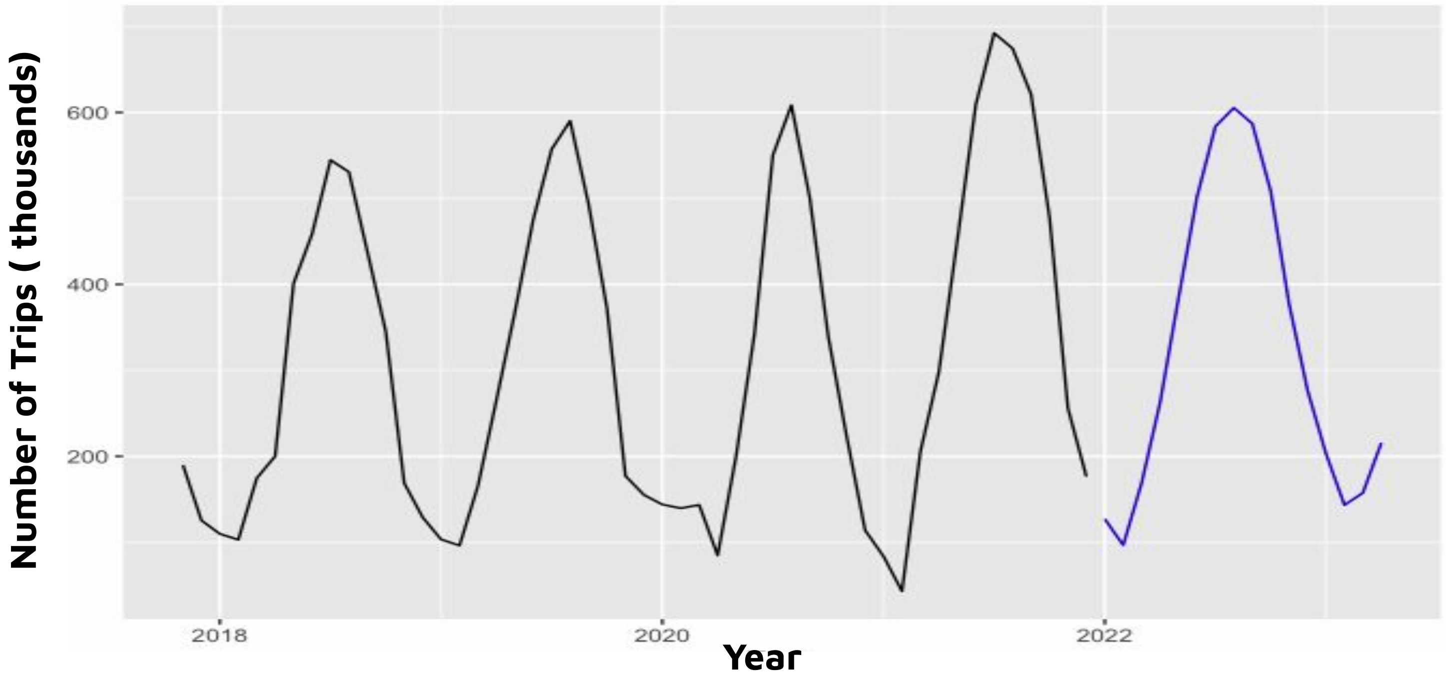
Forecast



DIVVY

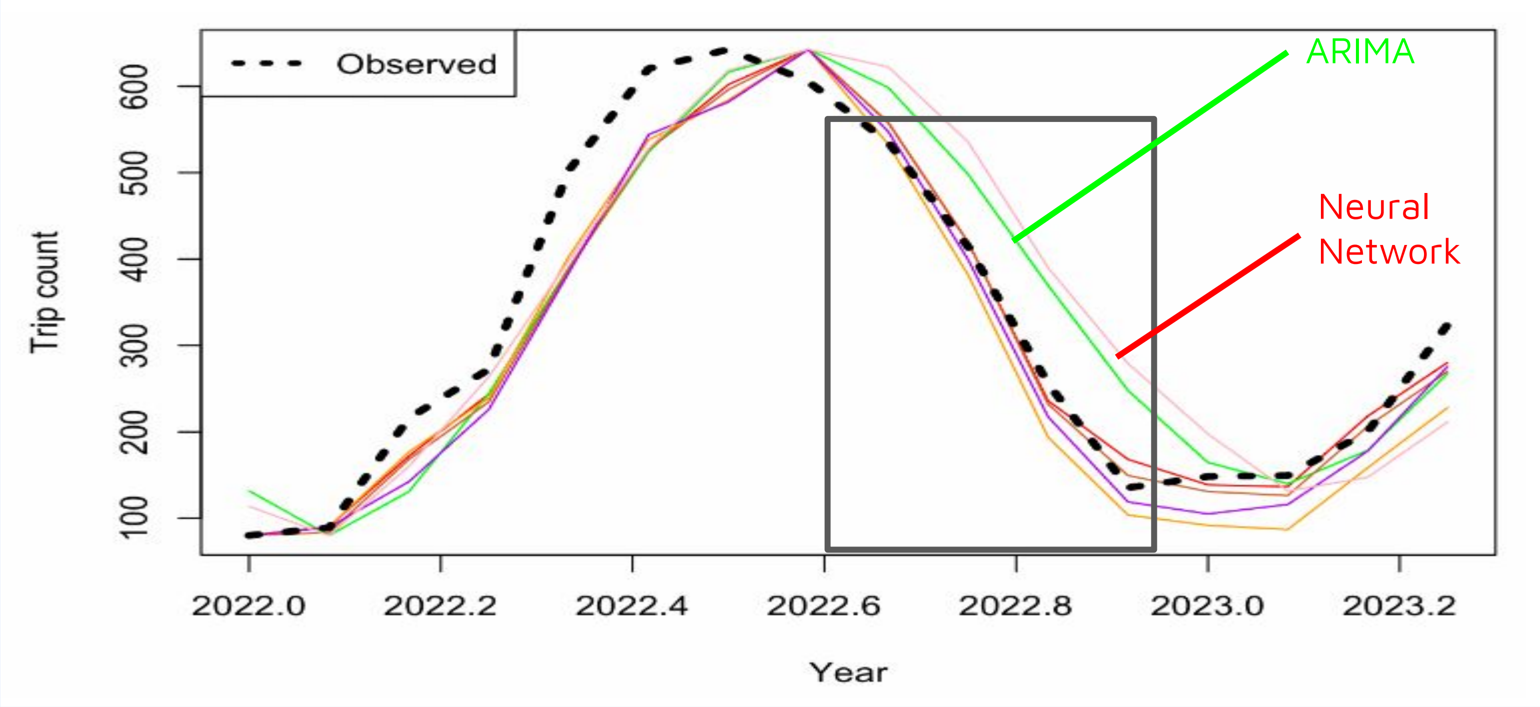
Neural Network Autoregression(1,1,2) [12] Model

Forecast



Model Selection

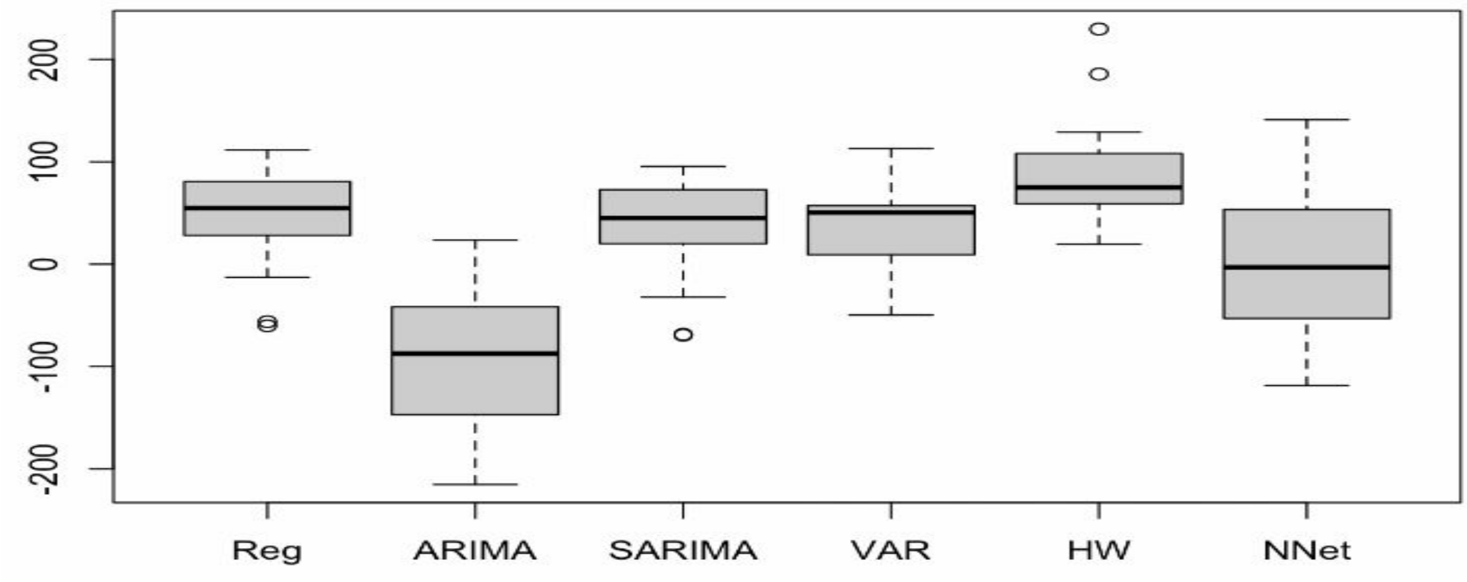
Comparing Forecast Residual in a plot



All the models , except for ARIMA and Neural Network seem to be give very similar results anc closer to the actual values.

Model Selection

Observed Vs Predicted trip count values for test period Jan 2022 - April 2023



RMSE :	65.76	116.40	61.19	65.76	103.65	78.66

SARIMA model seems to be the best model as it has the least RMSE value

DIVVY

Future Work

Model

- Can apply expanding and sliding window to validate prediction (currently using only one time period to test prediction)
- Extend testing period to more than one year to see if the models can factor in seasonality for longer forecast period

Business Indication

- Fine Grained Temporal Analysis : Explore granular time intervals such as weekly, daily, or even hourly, helping Divvy optimize resources on a smaller time scale.
- Incorporating External Factors: Expand the predictive model by incorporating external factors that may influence trip counts
- User Behavior Analysis: Analyze user behavior and preferences from trip count data to inform marketing, infrastructure, and service improvements

Thank You