MA641 Final Project

Akanksha Wagh

Spring 2025

Project 1: Seasonal Dataset

Title: Forecasting Bus On-Time Performance

1. INTRODUCTION:

Public transportation systems rely heavily on timely operations to ensure commuter satisfaction and systemic efficiency. Bus On-Time Performance (OTP) is one of the most critical metrics in evaluating service reliability. This study focuses on time series modeling of NJ Transit OTP data collected monthly from 2009 to 2025, with the objective of producing actionable forecasts and identifying seasonal patterns [1]. This project focuses on analyzing monthly Bus On-Time Performance (OTP) data to uncover the forecast future service reliability. Using a combination of time series analysis techniques—including stationarity testing, ACF/PACF inspection, and Box-Jenkins modeling, we identified SARIMA as the best-fit model.

2. DATA DESCRIPTION:

Time Range: Starts from January 2009 to March 2025.

Columns:

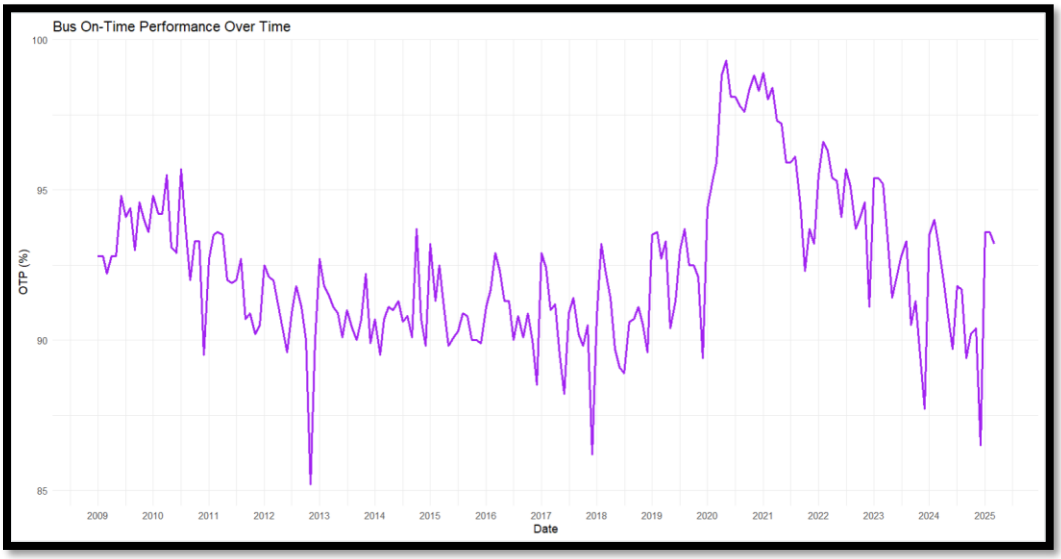
- OTP_YEAR: Year of data
- OTP_MONTH: Month of data
- OTP: On-Time Performance (%)
- TOTAL_TRIPS: Total number of bus trips
- TOTAL_LATES: Number of late trips

```
> head(data, 10)
   OTP_YEAR OTP_MONTH  OTP TOTAL_TRIPS TOTAL_LATES       DATE
2      2009         1 92.8        34119         2464 2009-01-01
3      2009         2 92.8        31932         2300 2009-02-01
4      2009         3 92.2        34065         2645 2009-03-01
5      2009         4 92.8        35134         2524 2009-04-01
6      2009         5 92.8        32451         2349 2009-05-01
7      2009         6 94.8        34731         1810 2009-06-01
8      2009         7 94.1        34596         2040 2009-07-01
9      2009         8 94.4        33953         1890 2009-08-01
10     2009         9 93.0        34198         2398 2009-09-01
11     2009        10 94.6        36034         1939 2009-10-01
```

3. ANALYSIS OF INITIAL OTP DATASET:

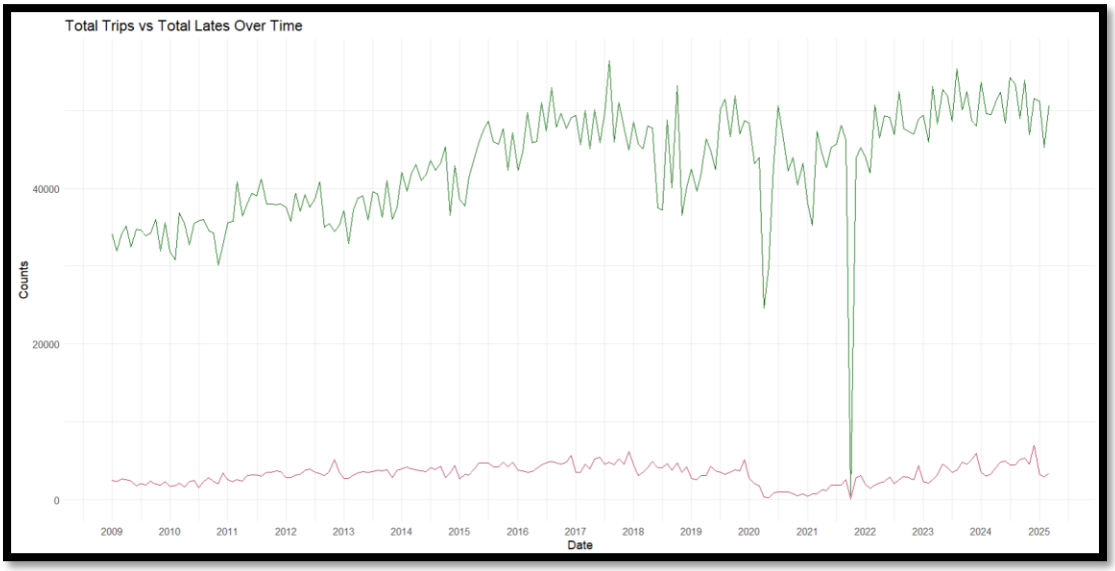
Bus On-Time Performance Over Time:

The OTP range is between 85% to 100% approximately, with noticeable seasonal dips. Sharp drops are observed around early 2013, mid-2017, early 2020, and 2024, indicating some uncontrollable changes like for example e.g., weather, labor issues, or COVID-19. Sudden dips in 2020-2021 are real, not errors.



Total Trips vs Total Lates

Total trips gradually increased from 2009 to 2019, peaking just before the pandemic. A sharp dip in early 2020–2021 indicates major service disruptions—likely due to COVID-19 lockdowns. Total lates mirror seasonal patterns and increase post-2022, suggesting strain on services despite recovery in trip volume.



4. STATIONARITY TEST:

Stationarity is an important factor when it comes to time series analysis because the stability in data makes it easier for the models to learn and extrapolate patterns. Models like ARIMA and SARIMA consider the data to be stationary and the estimates can be unreliable or biased. I used Augmented Dickey-Fuller (ADF Test) to check the stationarity of the series.

Ho: Series is non-stationary.

If the p-value < 0.05 then, we reject Ho which means our data is stationary and we can move ahead. But if the p-value >0.05, we fail to reject Ho which means the data is non-stationary.

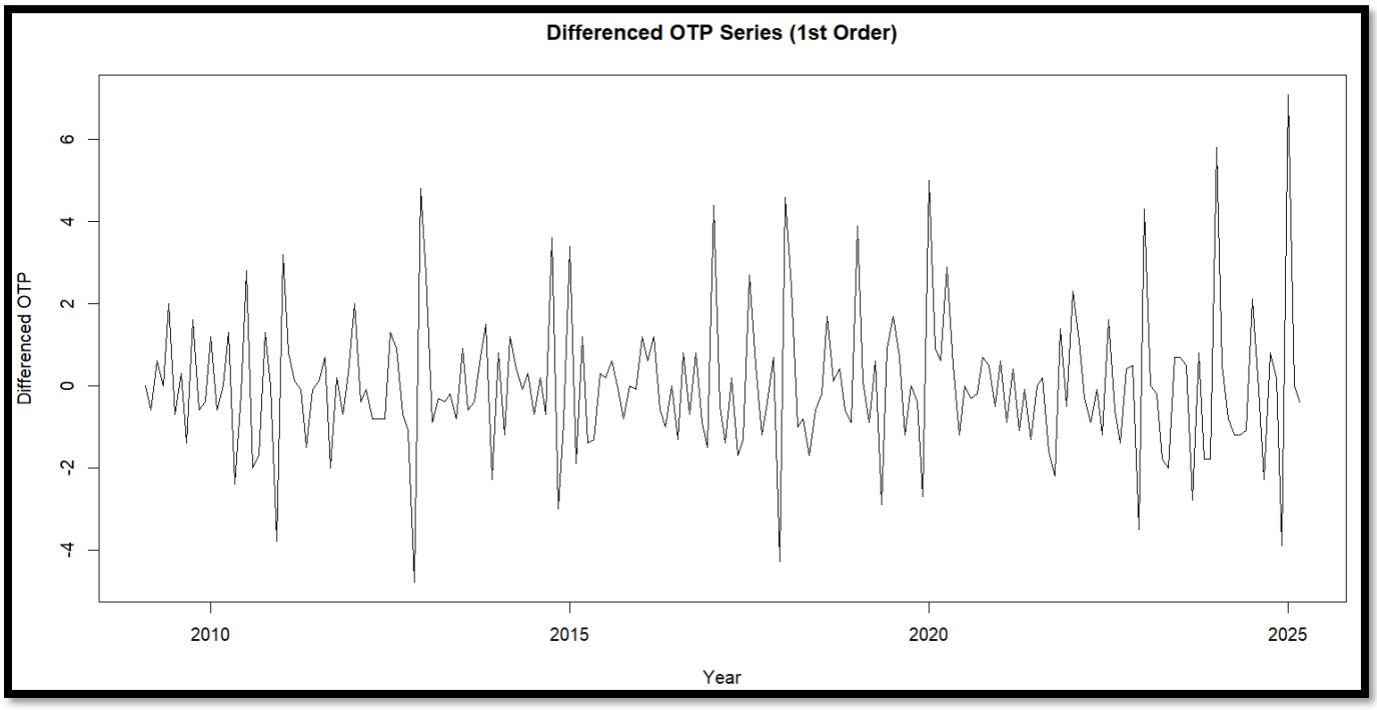
```
        Augmented Dickey-Fuller Test

data:  otp_ts
Dickey-Fuller = -1.8504, Lag order = 5, p-value = 0.6387
alternative hypothesis: stationary
```

As we can see the p-value is 0.6387. This means we fail to reject Ho, and our data is non-stationary.

5. DIFFERENCING:

Differencing is used for non-stationary data because such data often shows trends or fluctuations in mean over time. Differencing helps to remove these trends, effectively flattening any upward or downward movement in the series and making the mean constant over time.



This transformation is important for maintaining the stationarity condition required by ARIMA, which assumes the data has stable statistical properties. Differencing prepares the time series for more accurate and reliable modeling.

After differencing, we do the ADF Test again to check the stationarity.

Ho: Series is non-stationary.

If the p-value < 0.05 then, we reject Ho which means our data is stationary and we can move ahead. But if the p-value >0.05, we fail to reject Ho which means the data is non-stationary.

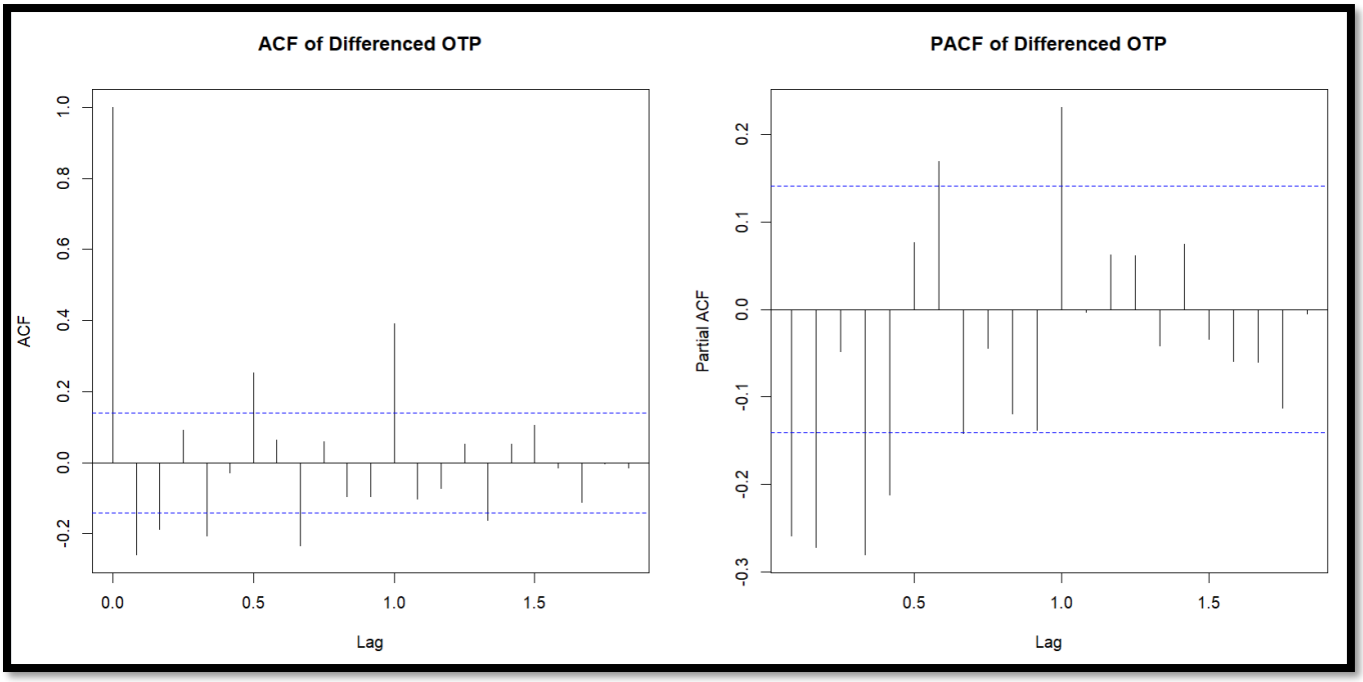
```
        Augmented Dickey-Fuller Test

data:  otp_diff1
Dickey-Fuller = -7.4947, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

As we can see the p-value is 0.01. This means we reject Ho and our data is stationary.

6. ACF & PACF PLOTS:

Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots are crucial diagnostic tools for time series forecasting because these plots help use understand the structure and pattern of our data and directs us to choose the best model – especially in the case of ARIMA & SARIMA.



Let's suppose we are selecting parameter for the model ARIMA(p, d, q); **ACF** helps you choose **q (MA order)**, **PACF** helps you choose **p (AR order)** AND **d** is decided by differencing

to make the series stationary.

If we observe our ACF Plot, there is a sudden drop after lag 1 which indicates MA(1) component. In PACF Plot, there are significant spikes at lag 1 and 2 which indicate AR(2).

From the plots, we will consider the following ARIMA model candidates:

ARIMA (2,1,1) - based on AR (2) and MA (1) structure

ARIMA (1,1,1) - simpler alternative

ARIMA (2,1,0) - if MA term isn't needed

7. MODEL SELECTION & COMPARISON:

Since we have 3 ARIMA models, we will compare them using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). We use AIC and BIC to select the best models for prediction, simplicity and interpretability. The model which has the lowest AIC and BIC is the best for the dataset we are working on.

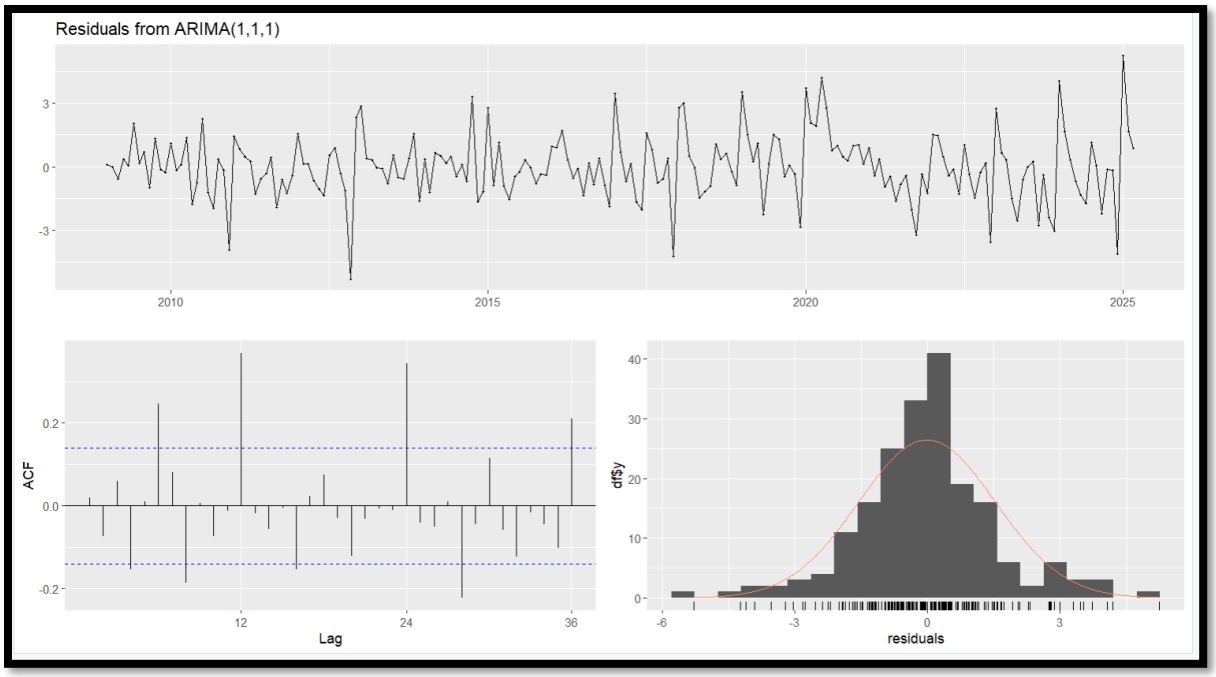
	Model	AIC	BIC
1	ARIMA(2,1,1)	727.5601	740.6315
2	ARIMA(1,1,1)	726.8072	736.6108
3	ARIMA(2,1,0)	736.1124	745.9160

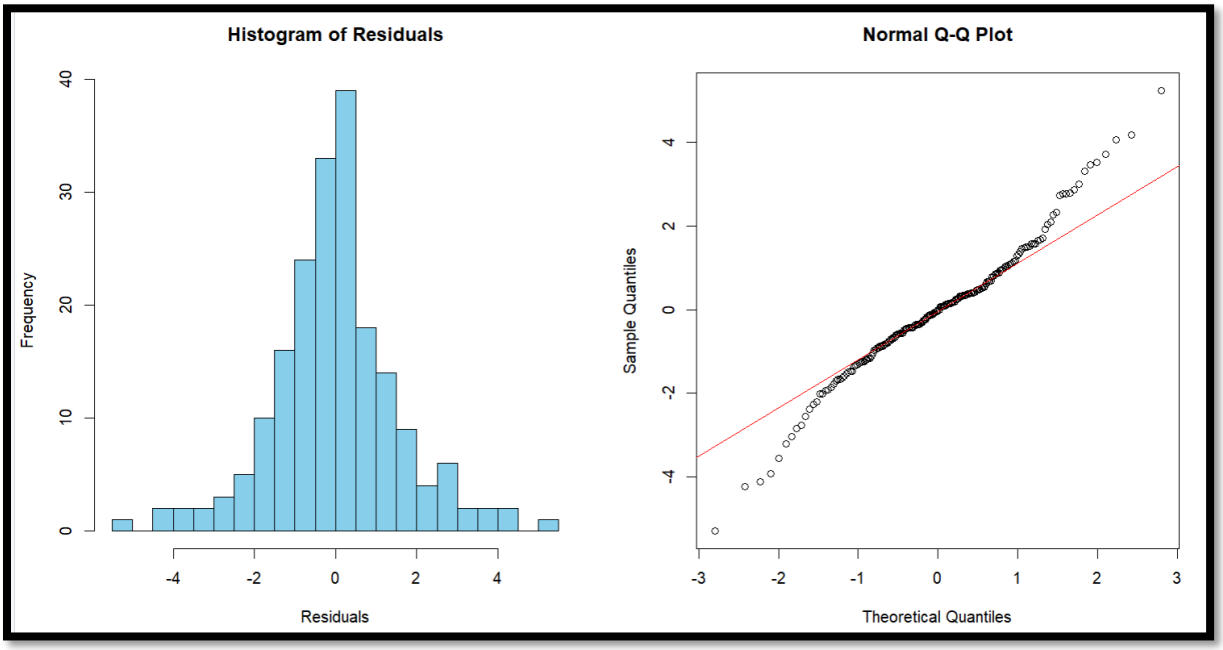
ARIMA (1,1,1) has the lowest values of AIC & BIC i.e., 726.80 & 736.61 respectively. It suggests balance in model complexity, and it is the best fit.

8. RESIDUAL DIAGNOSTICS & MODEL FITTING:

Even if the model is selected, we will perform Residual Diagnostic. It is important because it checks whether the selected model is model is appropriate, reliable, and avoids overfitting or underfitting. What are residuals? Residual is the error between actual and predicted values which should be random, uncorrelated and normally distributed. If any one condition is not satisfied that means, there were patterns which were not captured. We will check the autocorrelation by plotting ACF of residuals, normality by QQ-Plot and Histogram, white noise by Ljung-Box Test.

In our case, the residuals plot fluctuates around zero which is good and there's no visible trend. The ACF of Residuals spikes within the confidence line. Although there are spikes at few lag 12, 24, 36, suggesting possibly missed seasonality. The histogram is roughly bell-shaped, indicating the normal distribution. QQ-plot has minor deviation from the line but all together it is a sign of normality. Moving ahead with the Ljung-Box Test, the p-value is 3.9e-07 which means the p-value is much smaller than 0.05, indicating no significant autocorrelation in residuals.






```
        Box-Ljung test

data:  residuals_111
X-squared = 67.965, df = 20, p-value = 3.9e-07
```

SARIMA:

ARIMA (1,1,1) fits the data well but doesn't capture **seasonality.** So, let's use SARIMA and incorporate seasonality in the ARIMA model. We will be trying SARIMA (1,1,1) (1,0,1) [12]. We used (1,1,1) in SARIMA from non-seasonal ARIMA and (0,1,1) [12] because, differencing 1 is added for stability, MA with seasonality of 12 months and AR is removed to avoid non-stationarity.

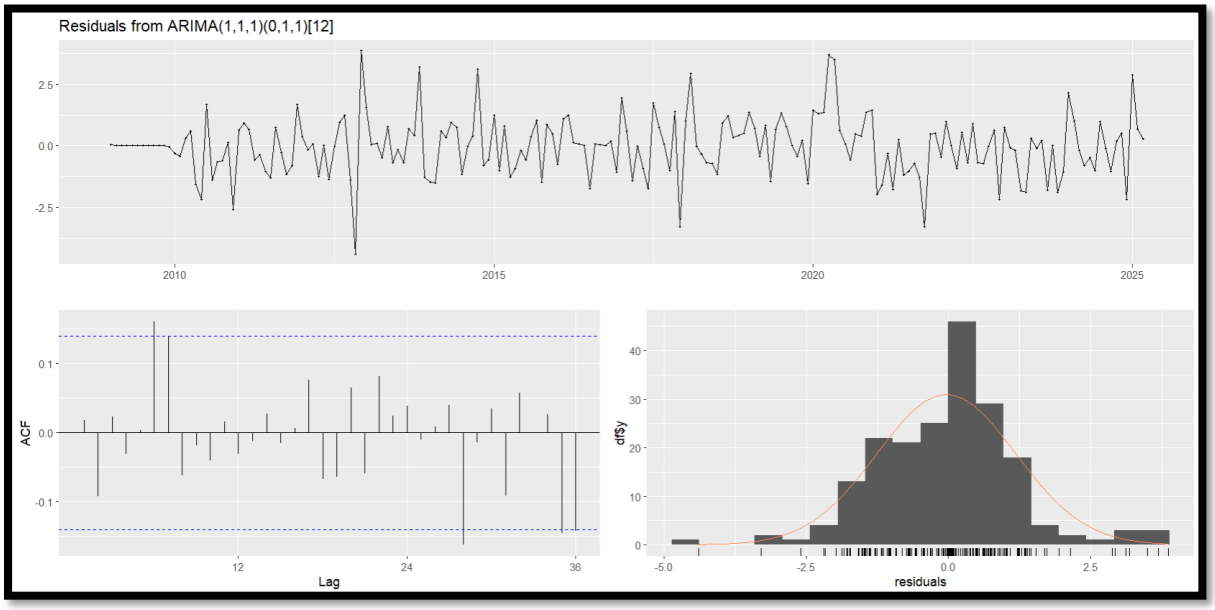
```
> summary(sarima_model)
Series: otp_ts
ARIMA(1,1,1)(0,1,1)[12]

Coefficients:
          ar1      ma1     sma1
      0.2847  -0.6627  -0.8525
s.e.  0.1266   0.0932   0.0669

sigma^2 = 1.621:  log likelihood = -308.55
AIC=625.1   AICc=625.33   BIC=637.92

Training set error measures:
                      ME     RMSE       MAE          MPE       MAPE
Training set -0.03499008 1.219936 0.8979458 -0.05021043 0.9747167
                    MASE        ACF1
Training set 0.5074864 0.01853477
```

The model is better than ARIMA (1,1,1) as the value of AIC and BIC is smaller than the previous model. The residuals plot fluctuates around zero, which is good and there's no visible trend. The ACF of Residuals spikes within the confidence line. As you can see, the spikes at few lags like 12, 24, 36 are not there anymore. The histogram is bell-shaped, indicating the normal distribution.

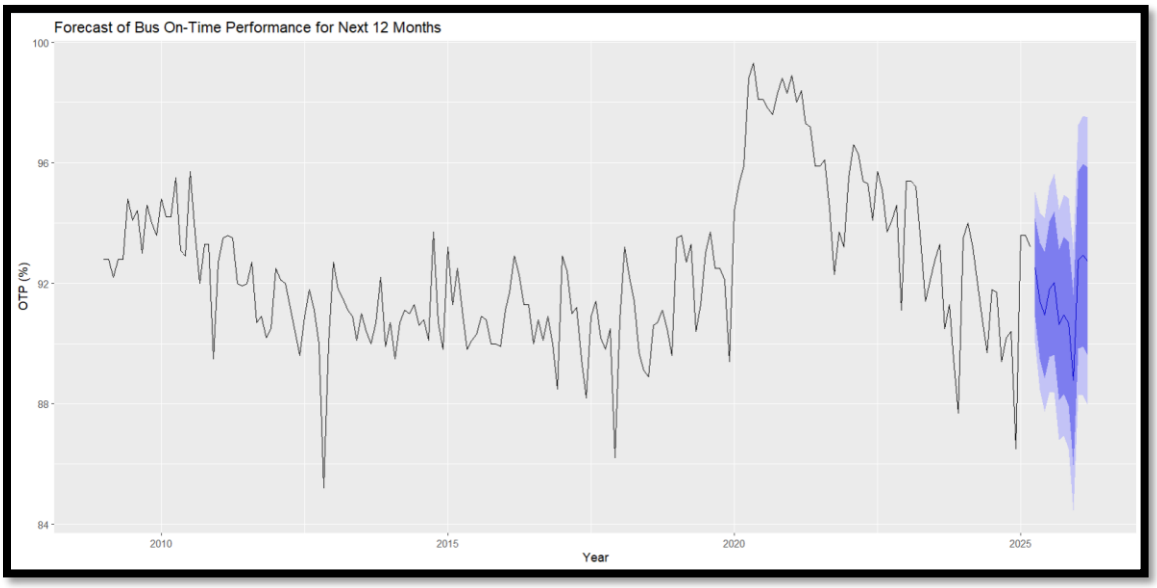



```
        Box-Ljung test

data:  residuals(sarima_model)
X-squared = 17.146, df = 20, p-value = 0.6434
```

Moving ahead with the Ljung-Box Test, the p-value is 0.64 which means the p-value is greater than 0.05, indicating significant autocorrelation in residuals. SARIMA (1,1,1) (0,1,1) [12] is a better and statistically valid model compared to ARIMA (1,1,1). It captures seasonality and gives white-noise residuals.

9. FORECASTING FOR THE NEXT 12 MONTHS:



The blue shaded part of the graph is Bus On-Time Performance forecasting for the next 12 months, which is up till March 2026. The range of the forecast is between 88% to 93% approximately. The dark blue part of the forecast is 80% confidence interval whereas the light blue part of the forecast is 95% confidence interval.

10. CONCLUSION:

Seasonal trends suggest early months are typically riskier for delay which is likely due to weather or operational issues. I read an article which mentioned that NJ transit was almost on the verge of bankruptcy during COVID-19 pandemic and that may have caused difficulties in functionality of the services. At the end of the forecast, you can see there is a significant rise backing up the facts mentioned in the article that it is going to be a profitable year for NJ transit in the later part of 2025.

11.REFERENCE:

[1] NJ Transit Source of Dataset

APPENDIX [Code]

```
#Seasonal Dataset of NJ Transit On-Time Bus Performance
#Forecasting the On-Time Performance for net 12 months

library(tidyverse)
library(lubridate)
library(ggplot2)
library(scales)
library(tseries)
library(urca)
library(forecast)

setwd("C:/Users/akank/Downloads")
data <- read.csv("BUS_OTP_DATA.csv", stringsAsFactors = FALSE)

#Cleaning the data and converting the data type
names(data) <- toupper(trimws(names(data)))
data <- data[data$OTP_YEAR != "----------", ]
data <- data %>%
  mutate(
    OTP_YEAR = as.integer(OTP_YEAR),
    OTP_MONTH = as.integer(OTP_MONTH),
    OTP = as.numeric(OTP),
    TOTAL_TRIPS = as.numeric(TOTAL_TRIPS),
    TOTAL_LATES = as.numeric(TOTAL_LATES),
    DATE = make_date(year = OTP_YEAR, month = OTP_MONTH)
  )

head(data, 10)

# GG Plot for OTP over time
ggplot(data, aes(x = DATE, y = OTP)) +
  geom_line(color = "purple", linewidth = 1) +
  labs(title = "Bus On-Time Performance Over Time", x = "Date", y = "OTP (%)") +
  scale_x_date(date_breaks = "1 year", labels = date_format("%Y")) + theme_minimal()

# GG Plot for total trips and total lates
ggplot(data, aes(x = DATE)) +
  geom_line(aes(y = TOTAL_TRIPS), color = "darkgreen") +
  geom_line(aes(y = TOTAL_LATES), color = "maroon") +
  labs(title = "Total Trips vs Total Lates Over Time", y = "Counts", x = "Date") +
  scale_x_date(date_breaks = "1 year", labels = date_format("%Y")) + theme_minimal()

# A monthly time series object for OTP
otp_ts <- ts(data$OTP, start = c(min(data$OTP_YEAR), min(data$OTP_MONTH)), frequency = 12)
otp_ts <- ts(data$OTP, start = c(2009, 1), frequency = 12)
summary(otp_ts)
plot(otp_ts, main = "Monthly Bus On-Time Performance", ylab = "OTP (%)", xlab = "Year")

# Check Stationarity

#ADF Test (Augmented Dickey-Fuller)
#H0: Series is non-stationary.(If p-value<0.05 -> Reject H0)
adf_result <- adf.test(otp_ts)
print(adf_result)

# In ADF test our OTP Time series is non-stationary.
# First-order differencing
otp_diff1 <- diff(otp_ts, differences = 1)
plot(otp_diff1, main = "Differenced OTP Series (1st Order)",
     ylab = "Differenced OTP", xlab = "Year")

# Check stationarity tests
#H0: Series is non-stationary.(If p-value<0.05 -> Reject H0)
adf.test(otp_diff1)
```

```
# In ADF test our OTP Time series is now stationary after differencing.

# ACF and PACF plots of differenced OTP series
# AR (p): At PACF - cut-off at lag p, gradual decay in ACF.
# MA (q): At ACF - cut-off at lag q, gradual decay in PACF.
par(mfrow = c(1, 2))  # Side-by-side plots
acf(otp_diff1, main = "ACF of Differenced OTP")
pacf(otp_diff1, main = "PACF of Differenced OTP")

#MA(1) AND AR(2)
#ARIMA Models: ARIMA(2,1,1) , ARIMA(1,1,1), ARIMA(2,1,0)
model_211 <- Arima(otp_ts, order = c(2,1,1))
model_111 <- Arima(otp_ts, order = c(1,1,1))
model_210 <- Arima(otp_ts, order = c(2,1,0))

# Compare models using AIC and BIC
model_comparison <- data.frame(
  Model = c("ARIMA(2,1,1)", "ARIMA(1,1,1)", "ARIMA(2,1,0)"),
  AIC = c(AIC(model_211), AIC(model_111), AIC(model_210)),
  BIC = c(BIC(model_211), BIC(model_111), BIC(model_210))
)
print(model_comparison)

#Model ARIMA(1,1,1)

# Residual diagnostics
checkresiduals(model_111)
residuals_111 <- residuals(model_111)

# Histogram for Normality
hist(residuals_111, main = "Histogram of Residuals", xlab = "Residuals",
     col = "skyblue", breaks = 20)

# QQ plot for Normality
qqnorm(residuals_111)
qqline(residuals_111, col = "red")

# Ljung-Box Test for White noise
Box.test(residuals_111, lag = 20, type = "Ljung-Box")

# Few Seasonal lags at 12, 24. and p-value is < 0.05 which indicates no significant
autocorrelation due to seasonal effects.

# Spikes repeat at lags 12, 24, 36; indicating SARIMA may be needed.

# Fit SARIMA with seasonal differencing
sarima_model <- Arima(otp_ts, order = c(1,1,1),
                      seasonal = list(order = c(0,1,1), period = 12))
summary(sarima_model)
checkresiduals(sarima_model)

# Ljung-Box test for SARIMA
Box.test(residuals(sarima_model), lag = 20, type = "Ljung-Box")

# Forecast for next 12 months
otp_forecast <- forecast::forecast(sarima_model, h = 12)
autoplot(otp_forecast) +
  ggtitle("Forecast of Bus On-Time Performance for Next 12 Months") +
  ylab("OTP (%)") +
  xlab("Year")
```