Raquel Ocasio
D213 – Advanced Data Analytics, Task 1: Time Series Modeling
March 11, 2024
Western Governors University

Part A1

Can daily revenue for the organization be predicted for the next year?

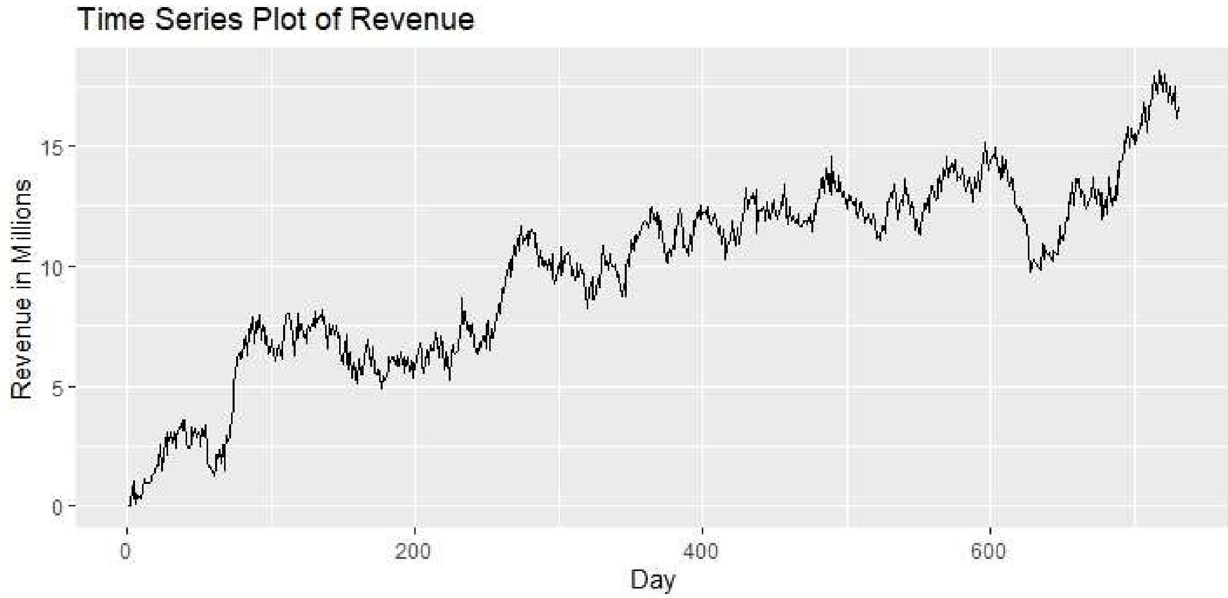
Part A2

The goal of the data analysis is to use time series modeling to predict the daily revenue of the organization for the next year.

Part B

Time series models rely on several assumptions to ensure their validity and reliability. Fundamental among these is stationarity, which asserts constancy in statistical properties over time, either strictly or with respect to mean. Autocorrelation, indicating correlation between a time series and its past values, is also crucial. Homoscedasticity assumes a consistent variance of residuals, while normality of residuals and independence between them facilitate effective statistical inference. Additionally, the absence of seasonality, or proper adjustment for it, is vital, as is addressing outliers that may disrupt model performance. Adequate sample size is required to accurately capture the complexity of the time series.

Part C1



Part C2

The time step formatting of the realization is a day. The length of the sequence is 731 days, and there are no gaps in measurement.

Part C3

The Augmented Dickey-Fuller test was applied to the data to determine stationarity of the time series. The test yielded a p-value of 0.02431, which is less than the p-value of 0.05. Therefore, the time series is likely stationary.

Part C4

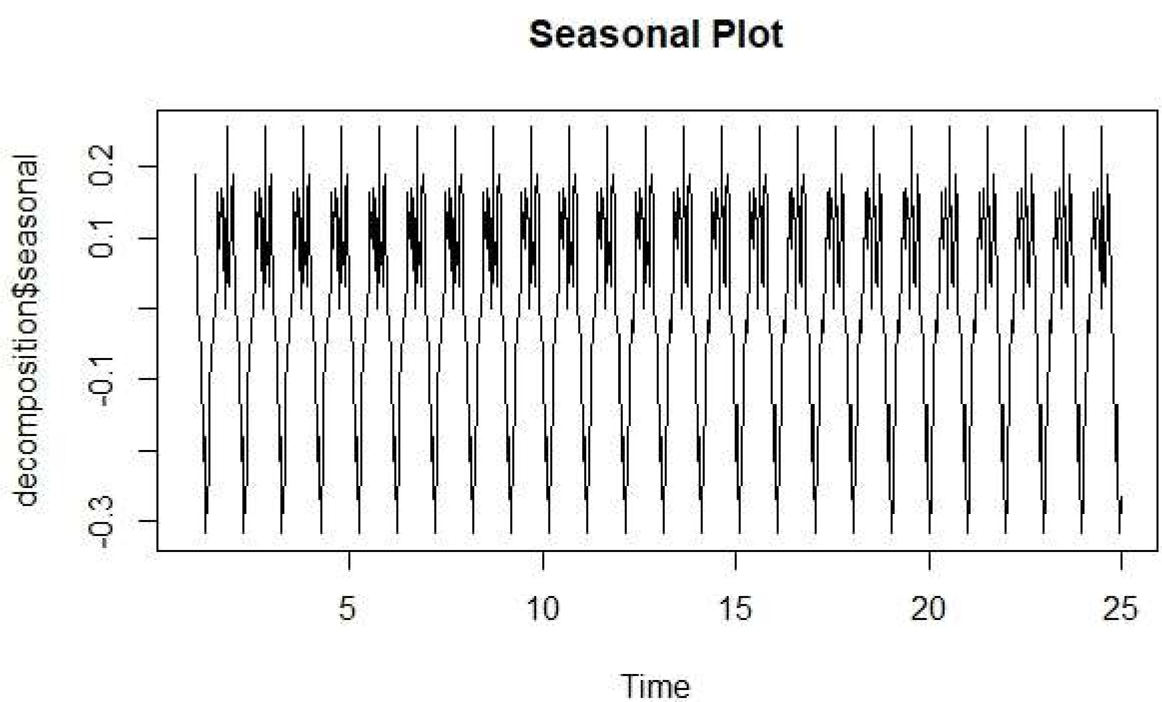
The data was prepared for analysis by checking for: NA values, null values, duplicates, stationarity, and splitting the data into test and training sets.

Part C5

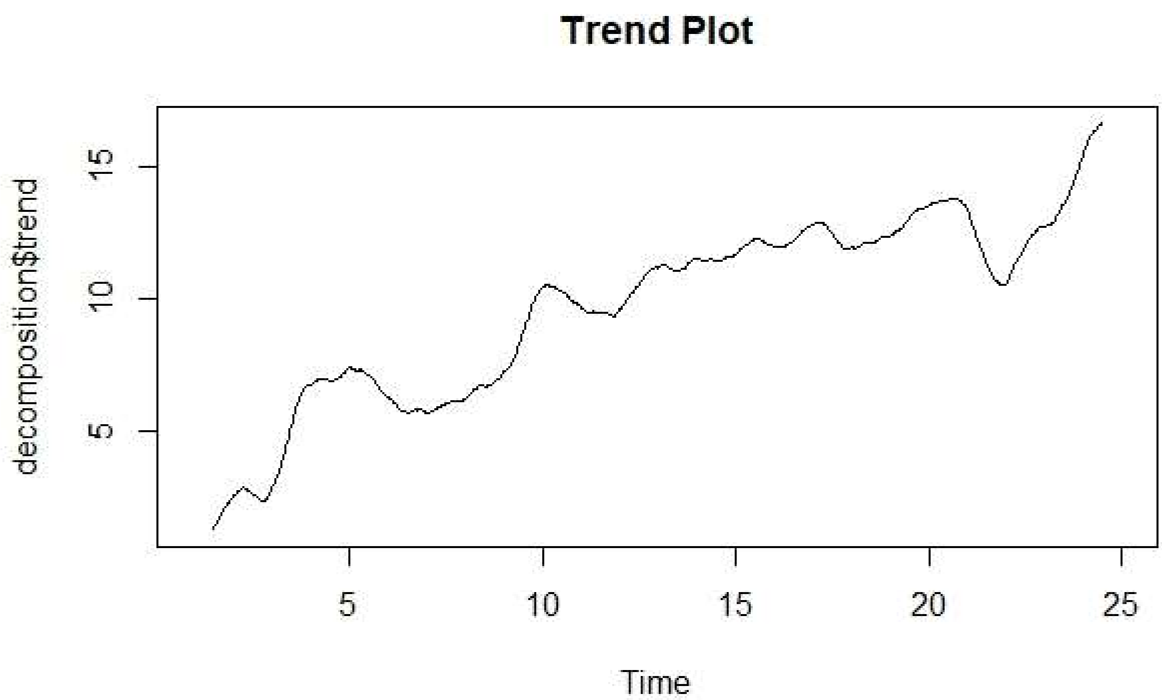
See attached files.

Part D1

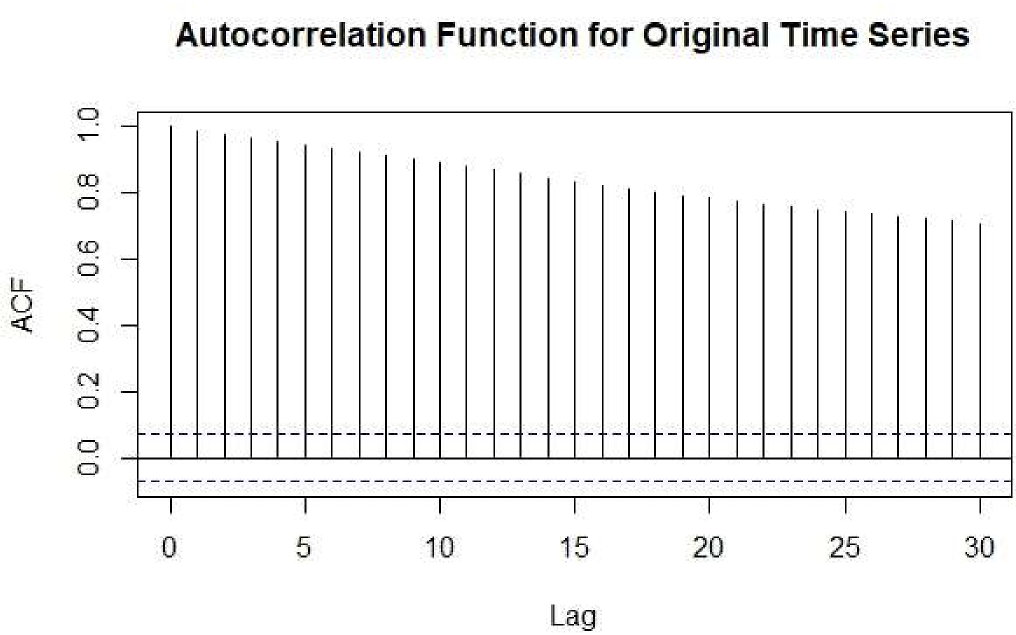
Plot for seasonal component



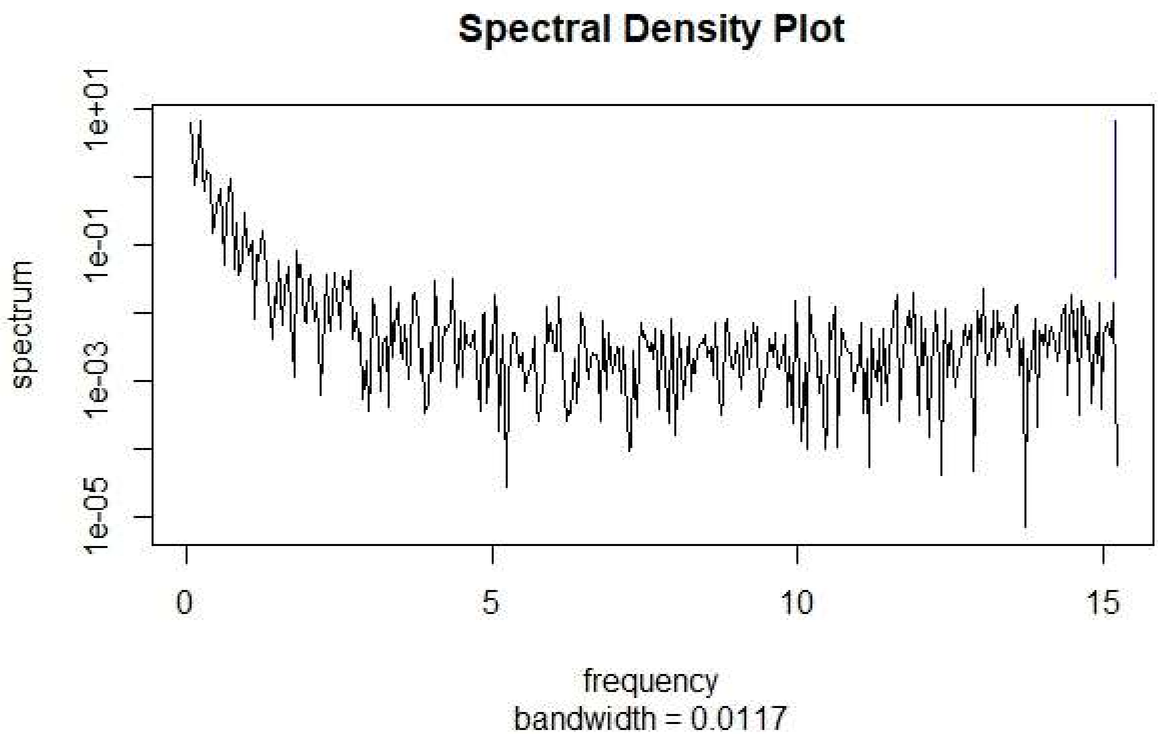
Plot for trends



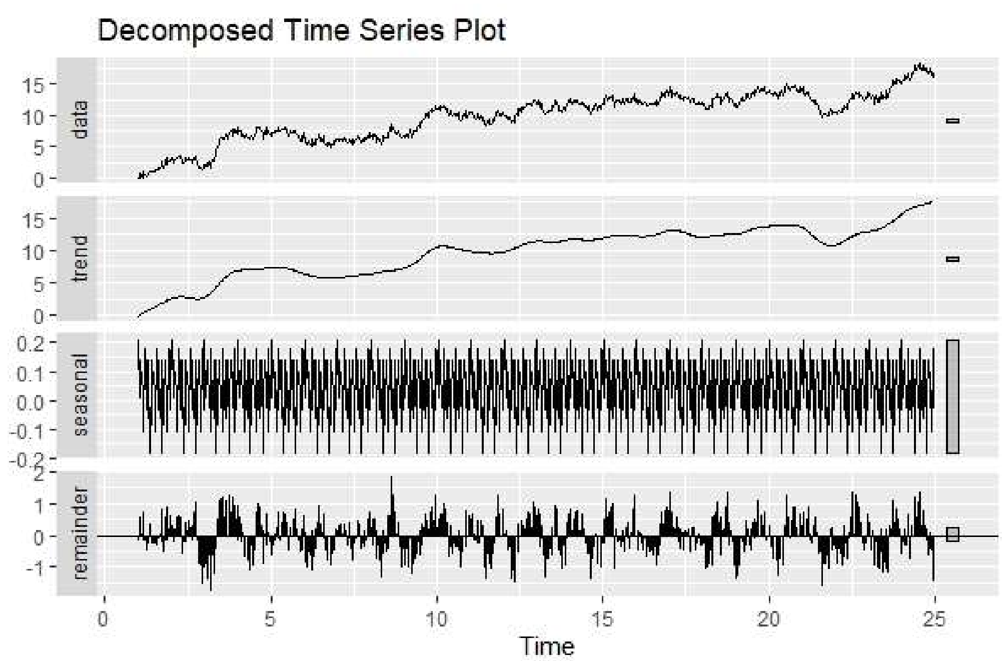
Plot for autocorrelation function



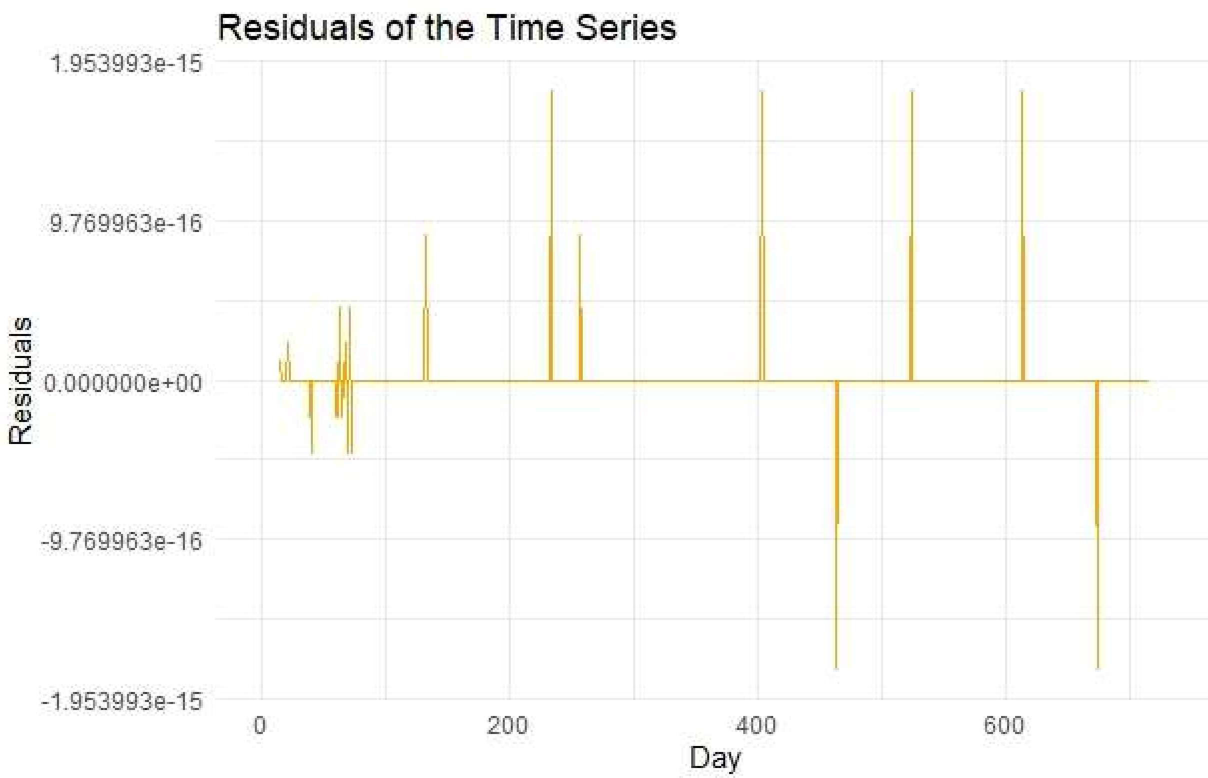
Plot for spectral density



Plot for decomposed time series



Plot for residuals

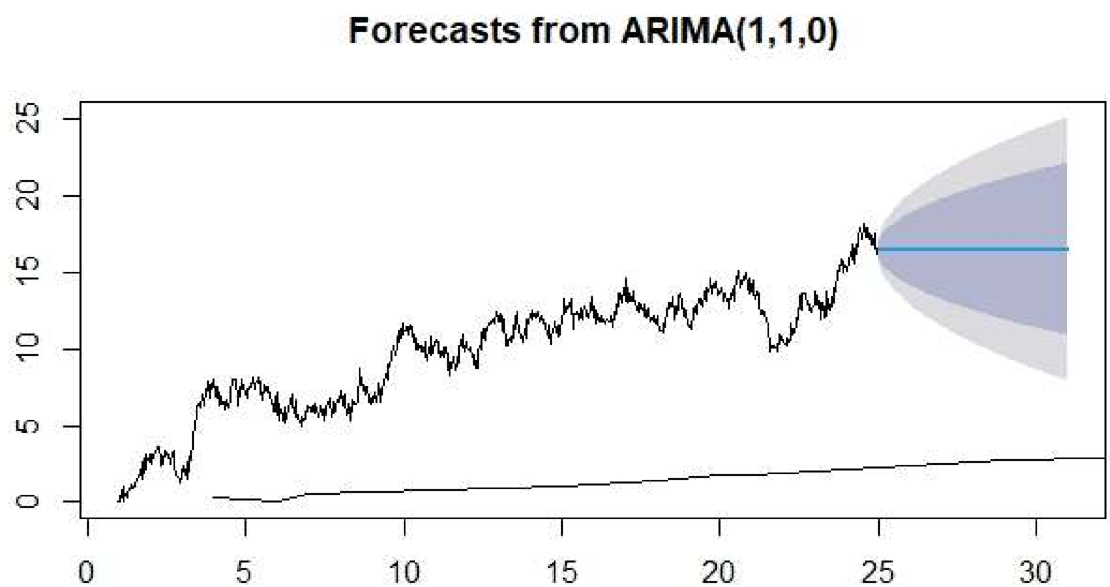


Part D2

```
Series: telecodata_ts
ARIMA(1,1,0)

Coefficients:
          ar1
      -0.4667
s.e.   0.0327

sigma^2 = 0.2246:  log likelihood = -490.36
AIC=984.71   AICc=984.73   BIC=993.9

Training set error measures:
                     ME      RMSE       MAE        MPE     MAPE      MASE
Training set 0.03307548 0.4732782 0.3805007 -0.1994301 6.246503 0.2591873
                      ACF1
Training set -0.0009638057
```

Part D3

```
        Box-Ljung test

data:  forecast
X-squared = 2.9038, df = 1, p-value = 0.08837
```



Part D4

See attached code file.

Part D5

See attached code file.

Part E1

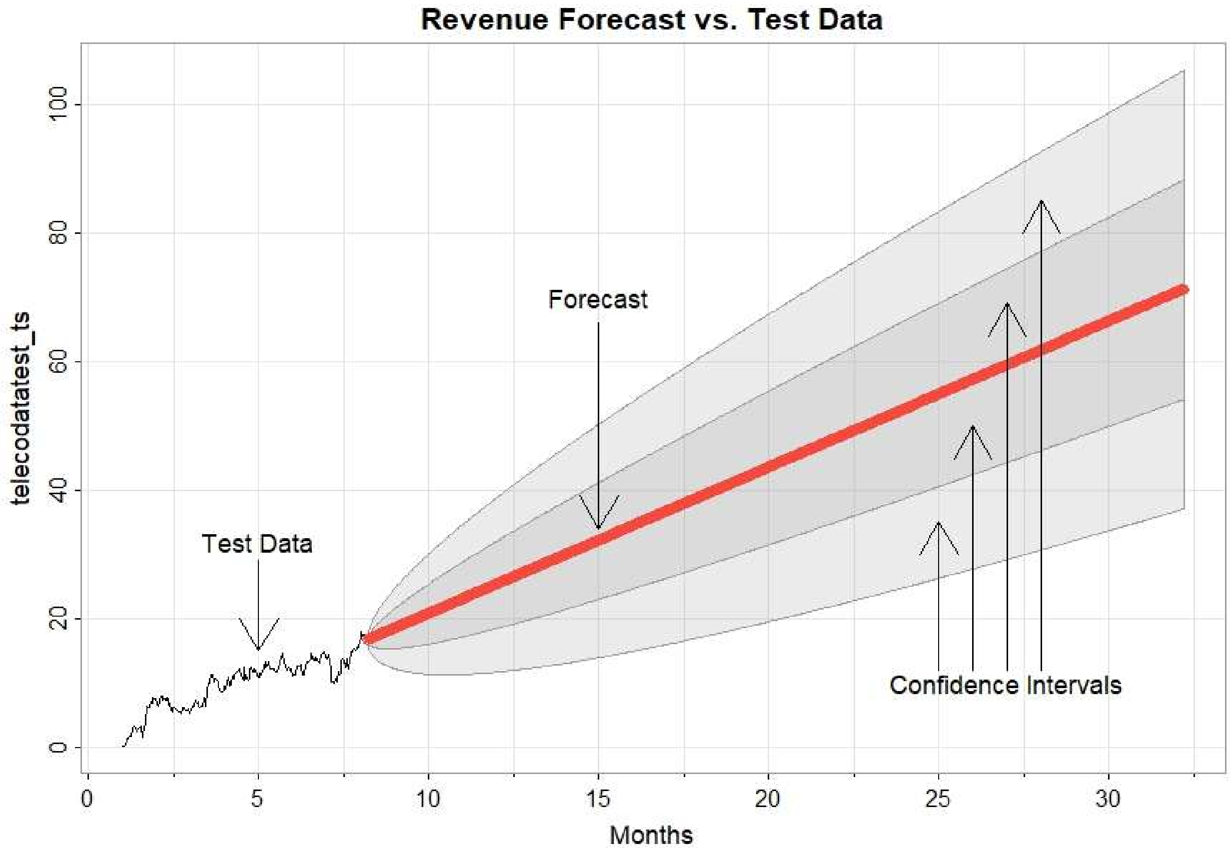
The ARIMA model was selected based on the output of the auto.ARIMA() function.

The forecast's prediction interval is set at 1 day, and our dataset consists of 2 years of daily revenue information. Consequently, the ARIMA model is utilized to uncover correlations and seasonality within the data, enabling the prediction of revenue at a daily interval.

Given a dataset spanning 2 years of daily revenue, TM has the capability to forecast future revenue for up to 1 year. Predictions made within the first year are expected to be more accurate. For more reliable long-term predictions, a larger historical dataset is necessary.

The model was evaluated automatically by the auto.ARIMA() function, which produced the model with the best AIC. The standard error (s.e.) was used as the error metric. The model produced by the auto.ARIMA() function produced a low s.e. value of 0.0327.

Part E2



Part E3

Based on the forecasted growth in revenue, stakeholders should continue to maintain current business operating procedures, and consider expansion.

Part F

See attached HTML document.

Part G

Part H

Part I