**Part I:  Research Question**

A.  Describe the purpose of this data analysis by doing the following:

1.  Summarize **one** research question that is relevant to a real-world organizational situation captured in the selected data set and that you will answer using time series modelling techniques.

A. What is telecom’s expected revenue forecast for next year?

2.  Define the objectives or goals of the data analysis. Ensure that your objectives or goals are reasonable within the scope of the scenario and are represented in the available data.

A. Create a predictive model to forecast next year’s revenue and identify trends and seasonality.

**Part II:  Method Justification**

B.  Summarize the assumptions of a time series model including stationarity and autocorrelated data.

Time Series Assumptions:

1. The Time Series data must not have outliers as they can cause inaccurate results (DataCamp, 2021).

2. The Residuals are not autocorrelated.

Stationarity and Autocorrelated Assumptions:

1. The data set series must not have any shirking or growing trends.

2. The variance of the series is constant.

3. The autocorrelation is also constant. The relationship between the values in the series must be the same (DataCamp, 2021).

**Part III:  Data Preparation**

C.  Summarize the data cleaning process by doing the following:

1.  Provide a line graph visualizing the realization of the time series.

A. The data is non-stationary which means it has trends. I will later use adfuller to validate and then use the means to make it stationary.

Chart, scatter chart

Description automatically generated

2.  Describe the time step formatting of the realization, including any gaps in measurement and the sequence length.

1. There are two series in the data frame. The Day series contains 731 days (2 years) worth of data. Each row equals a unique day with no gaps. The Revenue series has the same count of rows with no null values. Both series do not have any outliers.

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Table

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

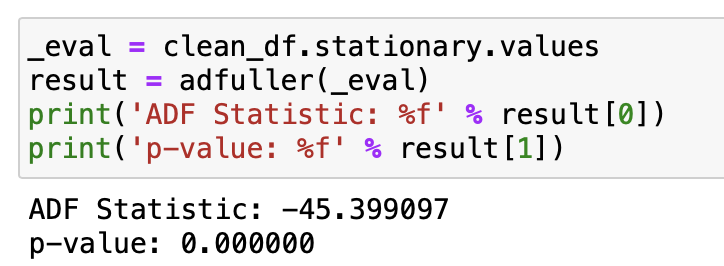
3.  Evaluate the stationarity of the time series.

1. The p-value is not <.05; hence the data is not stationary and rejects the H0 hypothesis.

Text

Description automatically generated

Adjusted the p-value using the panda's diff function



4.  Explain the steps used to prepare the data for analysis, including the training and test set split.

A. The steps to prepare the data for analysis are:

1. Import the ‘teleco\_time\_series .csv’ into Pandas dataframe using Python in Jupyter.
2. Perform EDA then visualize the data using histogram and boxplots.
3. Add a new datetime column based on the day’s series.
4. Identify null values and remove them (if applicable)
5. Check if there any outliers using the stats.zscore module and remove anything below and above -3 and +3 zscore.
6. Review the Stationarity using ‘adfuller’ module either reject or expect the null hypothesis.
   1. Coerce the Stationarity (if needed) by taking the difference from the mean.
   2. Drop the null values from the new ‘stationarity’ series.
   3. Re-evaluate the Stationarity using ‘adfuller’ by ensuring the p-value is < .05.
7. Text

   Description automatically generatedSplit the data using ‘train\_test\_split’ module and apply 80/20 rule.

5.  Provide a copy of the cleaned dataset.

A. ‘D213\_timeSeriesData.csv' attached.

A picture containing text

Description automatically generated

**Part IV:  Model Identification and Analysis**

D.  Analyze the time series dataset by doing the following:

1.  Report the annotated findings with visualizations of your data analysis, including the following elements:

•   The seasonality still exists even after taking the difference.

Chart, bar chart

Description automatically generated

* The line graph does not have any up or down trends. The data is horizontal and/or stationary.

Chart, line chart

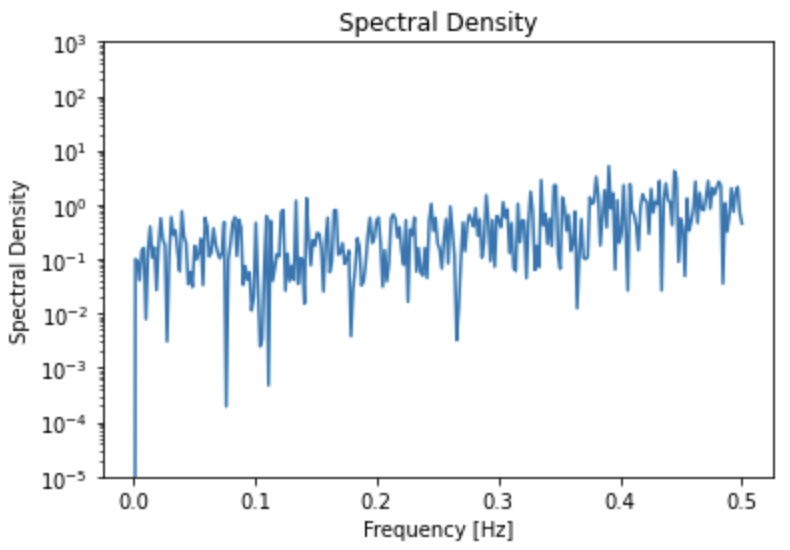
Description automatically generated

•    The plot shows that the values are within the blue shade and are not statistically significant. It also validates the stationarity, as the ACF drops quickly to zero.

Chart

Description automatically generated

•   The below graph shows the Spectral density of the stationary series.



•   Decomposed time series

Graphical user interface, timeline

Description automatically generated

•   This horizontal graph shows that lack of trend in the residuals

Chart, line chart

Description automatically generated

2.  Identify an autoregressive integrated moving average (ARIMA) model that takes into account the observed trend and seasonality of the time series data.

A. Lowest AIC score: 876.596

Best model: ARIMA(1,0,0)(2,1,0)[12]

3.  Perform a forecast using the derived ARIMA model.

4.  Provide the output and calculations of the analysis you performed.

5.  Provide the code used to support the implementation of the time series model.

**Part V:  Data Summary and Implications**

E.  Summarize your findings and assumptions, including the following points:

1.  Discuss the results of your data analysis, including the following:

•   the selection of an ARIMA model

•   the prediction interval of the forecast

•   a justification of the forecast length

•   the model evaluation procedure and error metric

2.  Provide an annotated visualization of the forecast of the final model compared to the test set.

3.  Recommend a course of action based on your results.

**Part VI:  Reporting**

F.  Create your report from part E using an industry-relevant interactive development environment (e.g., a Jupyter Notebook). Include a PDF or HTML document of your executed notebook presentation.

G.  List the web sources used to acquire data or segments of third-party code to support the application.

Correlation and Autocorrelation - Introduction to Course. (2021). DataCamp. https://campus.datacamp.com/courses/time-series-analysis-in-python/correlation-and-autocorrelation?ex=1

H.  Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

I.  Demonstrate professional communication in the content and presentation of your submission.