***Time Series Modeling***

*WGU*

*Course Number: 603*

*Trayvonious Pendleton*

*011205284*

**B1: Research Question**

“What are the historical trends and patterns in daily revenue over a telecommunications company’s first two years of operation, and how can time series modeling be used to forecast revenue over the next 30 days?”

This question is relevant because understanding revenue trends over time helps the organization evaluate the impact of customer churn and plan strategies for retention and profitability. A 30-day forecast enables short-term planning for operational decisions, resource allocation, and customer engagement initiatives.

**B2: Objectives**

**Objectives:**

1. To identify and visualize historical revenue trends, seasonality, and anomalies in the first two years of operation.
2. To assess the stationarity and autocorrelation of the daily revenue data.
3. To build and evaluate a time series forecasting model for predicting future revenue trends.
4. To provide actionable insights on the expected revenue trajectory, which can help evaluate the impact of churn and guide decision-making around customer retention efforts.

**C. Assumptions of a Time Series Model**

When working with time series models such as ARIMA, several core assumptions must be understood and validated:

**Stationarity:**

* **Definition:** A time series is stationary if its statistical properties remain constant over time.
* **Why it matters:** Most time series models, especially ARIMA, assume that the series is stationary, so forecasts do not depend on time-based trends or volatility shifts.

**How to check:**

* **Visual inspection:** Look for stable mean and variance on a time series plot.
* **Statistical tests:** Use the Augmented Dickey-Fuller Test or the Kwiatkowski–Phillips–Schmidt–Shin test.
* **If not stationary:** Use differencing transformation or detrending methods to make the series stationary.

**Autocorrelation**

* **Definition:** Autocorrelation refers to the correlation of the time series with its past values
* **Why it matters:** Time series models like ARIMA rely on past data to make future predictions. Significant autocorrelation indicates that past values help in predicting future values.

**How to check:**

* **Autocorrelation Function:** Measures the correlation between the time series and its lags.
* **Partial Autocorrelation Function:** Measures the correlation between the series and its lags after removing the effects of intermediate lags.

**Linearity and Additivity:**

* Time series models like ARIMA assume a linear relationship between past values and future outcomes.
* It also assumes that the effects of components like trend, seasonality, and noise are additive.

**No or Low Multicollinearity:**

* In multivariate time series models, variables should not be highly collinear with each other.

**Residuals Should Be White Noise:**

After fitting a model, the residuals should be:

* Normally distributed
* Uncorrelated
* Have constant variance

This indicates that the model has captured all meaningful information from the data.

To use a time series model effectively, especially ARIMA, the data must be stationary, exhibit significant autocorrelation, and produce white noise residuals post-modeling. These assumptions help ensure reliable forecasts and accurate insights.

**D1: Provide a line graph visualizing the realization of the time series**

A graph showing the growth of the stock market

AI-generated content may be incorrect.

**D2: Describe the time-step formatting of the realization**

The time series dataset contains 731 observations, representing daily revenue values over two years. Each entry in the dataset corresponds to one day, with the Day variable serving as the time index. The time steps are evenly spaced at one-day intervals. A check for continuity confirmed that there are no missing days or gaps in the time sequence.

The sequence is complete and covers two years, making it suitable for time series analysis without imputation or date gap handling. Since the Day column was treated as a simple sequential index rather than a date format, the x-axis of the time series graph uses numeric day labels, which is sufficient for analysis in this context.

**D3: Stationarity Evaluation**

The Augmented Dickey-Fuller test was applied to evaluate whether the daily revenue time series is stationary. The results of the test are shown below:

* **ADF Statistic:** -1.9246
* **p-value:** 0.3206

**Critical Values:**

* 1%: -3.4394
* 5%: -2.8655
* 10%: -2.5689

**Interpretation:**

The ADF test’s p-value is greater than 0.05, so we fail to reject the null hypothesis that a unit root is present in the data. Therefore, the revenue time series is not stationary. Since stationarity is a core requirement for time series modeling, the data must be transformed using differencing or other techniques before applying models like ARIMA.

**D4: Explain the steps used to prepare the data for analysis**

To prepare the dataset for time series modeling, first-order differencing was applied to the daily revenue series to address non-stationarity, as confirmed by the Augmented Dickey-Fuller test. The differencing removed trends, allowing for stationarity.

The dataset was then split into a training set (80%) and a test set (20%) based on index position. This split supports the validation of forecasting models in later steps. The cleaned dataset, with missing values removed after differencing, was saved to a separate file for future analysis.

**E1: Visualizations**

**Trends:**

**A graph showing a blue line

AI-generated content may be incorrect.**

This plot shows the different daily revenue values across 731 days. After differencing, the data no longer shows a clear upward trend, confirming that the transformation removed the non-stationary behavior. The mean and variance appear relatively stable, a key assumption for time series modeling.

**The Autocorrelation Function:**

**A graph with a line

AI-generated content may be incorrect.**

The ACF plot shows no significant autocorrelation at any lag, suggesting that the differenced revenue series is mainly random and stationary. This further supports the conclusion that the time series is well-suited for ARIMA modeling without additional transformations.

**The Spectral Density:**

**A graph with purple lines

AI-generated content may be incorrect.**

The spectral density plot displays the frequency components in the differenced revenue data. No intense seasonal cycles are detected, as there are no sharp peaks at any frequency. This indicates the absence of dominant seasonal patterns in the data.

**The Decomposed Time Series:**

**A graph of blue and white lines

AI-generated content may be incorrect.**

The original (non-differenced) revenue series decomposition reveals a strong trend and regular seasonal fluctuations. The trend line shows steady growth in revenue over time, while the seasonal component shows repeating cycles approximately every 30 days, likely reflecting monthly customer behavior. The residuals appear random and mostly centered around zero.

**Residual:**

**A graph showing the fall of a number of fall

AI-generated content may be incorrect.**

This plot of the residuals from the seasonal decomposition shows a random scatter with no visible trends or patterns. This is expected if the model has successfully captured both trend and seasonality. This confirms that the residual component behaves like white noise, supporting the assumption of stationarity for the differenced series.

**E2: Auto ARIMA**

The auto\_arima function was used to identify the optimal ARIMA model for the differenced revenue series. It evaluated multiple combinations of parameters and selected ARIMA(1,0,0) with an intercept based on the lowest AIC (983.122). This model includes one autoregressive term and no differencing or moving average components, indicating that the series is already stationary and that short-term dependencies exist.

The AR (1) coefficient was statistically significant (p < 0.001), validating the presence of short-term temporal correlation. The residuals showed no significant autocorrelation (Ljung-Box p = 0.96), confirming a good model fit.

**E3: Forecast**

To meet the out-of-sample forecasting requirement, the ARIMA(1,1,0) model was fit to the whole historical revenue dataset (non-differenced). The model forecasted revenue for 30 days, providing projected values and a 95% confidence interval.

The resulting plot shows the historical daily revenue trend followed by a forward-looking forecast (orange line) with a shaded confidence cone. This forecast enables the organization to anticipate short-term revenue performance and make informed decisions regarding customer engagement and churn mitigation strategies.

A graph showing the growth of the stock market

AI-generated content may be incorrect.

**E4: Output and Calculations of the Analysis**

**A screenshot of a computer code

AI-generated content may be incorrect.**

The ARIMA(1,0,0) model was trained using 80% of the differenced revenue data and evaluated on the remaining 20% (test set). The model’s predictive accuracy was assessed using the following error metrics:

• **Mean Squared Error (MSE):** 0.3246

• **Root Mean Squared Error (RMSE):** 0.5698

• **Mean Absolute Error (MAE):** 0.4670

These values indicate a low average error, supporting the model’s effectiveness in capturing short-term changes in the stationary revenue series. This confirms the ARIMA(1,0,0) model’s suitability for near-term forecasting of differenced revenue data.

**F1: Discuss the Results of Your Data Analysis**

The ARIMA(1,0,0) model was selected as the best-fitting time series model using the auto\_arima function, which evaluates multiple combinations of autoregressive (AR), differencing (I), and moving average (MA) terms on the differenced revenue series, which achieved stationarity after first-order differencing. This model was chosen based on the lowest AIC value (983.122), indicating a good balance between model complexity and fit. The ARIMA(1,0,0) structure includes one autoregressive term and no differencing or MA components, which is appropriate given the stationary nature of the transformed series.

The forecast was generated over a horizon equal to 20% of the dataset, which is standard practice in time series analysis to preserve a meaningful test set for validation. While the initial forecast on the test set did not include a confidence interval, a later out-of-sample forecast for the next 30 days included a 95% prediction interval to quantify uncertainty and support more informed decision-making.

Model evaluation was performed by comparing the forecasted values to actual test set values using three error metrics:

* **Mean Squared Error (MSE):** 0.3246
* **Root Mean Squared Error (RMSE):** 0.5698
* **Mean Absolute Error (MAE):** 0.4670

These results suggest that the AR(1) model performed well for short-term forecasting in a near-white-noise revenue series. The model effectively captures general volatility while avoiding overfitting and provides a reliable baseline for monitoring revenue stability over time.

**F2: Annotated Forecast Visualization**

A graph with blue lines

AI-generated content may be incorrect.

The above visualization shows the ARIMA(1,0,0) model’s forecast compared to actual values in the test set. The blue line represents actual differenced revenue, while the orange line shows the forecasted values. The shaded region represents the 95% confidence interval (prediction cone), indicating the expected range of future values.

This annotated plot provides a clear visual comparison between the model’s predictions and observed data. The close tracking of the forecast line with actual values and a well-bounded confidence interval supports the model’s reliability in short-term forecasting.

**F3: Recommendation**

Based on the ARIMA(1,0,0) analysis of the differenced revenue data, it is recommended that the organization focus on maintaining revenue stability through customer retention strategies. The forecast results suggest that the series behaves in a relatively stationary and random pattern, meaning large swings in revenue are unlikely unless there are structural business changes.

Revenue appears consistent since the model does not detect strong trends or seasonality. Therefore, rather than investing in aggressive forecasting strategies, the company should prioritize monitoring for sudden changes that could indicate rising churn rates or operational disruptions.

As a forward-looking strategy, implementing real-time monitoring of daily revenue and updating the model regularly will ensure responsiveness to unexpected shifts. Additionally, supplementing the revenue model with churn-specific features may allow for more targeted interventions to reduce churn and protect long-term profitability.