Project Proposal (C): Time Series Analysis of Hotel Data Using SARIMA Models

Plan of Analysis

For this project, we will analyze the hotel_data time series dataset, which includes daily information on hotel operations with five variables: Date, Weekday, Occupancy, ADR (Average Daily Rate), and Total_Revenue. Among these, we will focus my SARIMA modeling on the Total_Revenue variable, as it exhibits clear seasonal behavior and non-stationarity—making it suitable for seasonal ARIMA modeling.

1. Preliminary Analysis

- Visualization: We will begin by plotting the time series of Total_Revenue to identify visible patterns such as trend, seasonality, and variance instability.

- Stationarity Testing: We will formally assess stationarity using both the Augmented Dickey-Fuller (ADF) and KPSS tests. Prior assessments suggest the series is neither stationary nor trend-stationary.

- Seasonality Check: Based on ACF plots provided, there is strong weekly seasonality (peaks at lags 7, 14, 21, etc.), indicating that the data may benefit from a SARIMA model with weekly seasonality (s = 7).

- Distribution Shape: Since Total_Revenue is skewed right, we may also consider log-transforming the series to stabilize variance if necessary.

2. Transformation & Differencing

- Apply a log transformation (if needed) to reduce heteroscedasticity.

- Perform seasonal and non-seasonal differencing to achieve stationarity, informed by ACF and PACF plots:

    - Seasonal differencing (lag-7) to address weekly seasonality.

    - First differencing to remove overall trend.

- ACF/PACF plots of the transformed series will help determine candidate SARIMA model orders.

3. Model Selection

Using the ACF and PACF plots of the transformed series, we will identify at least two SARIMA model candidates. For instance:

- Model 1: SARIMA(p=1,d=1,q=1)(P=1,D=1,Q=0)[7]
- Model 2: SARIMA(2,1,2)(1,1,1)[7]

These models will be fit to the data, and we will evaluate them using:

- AIC/BIC values
- Residual diagnostics (white noise checks using Ljung-Box test and residual ACF plots)
- Forecast accuracy measures (e.g., RMSE on a holdout set)

4. Final Model Selection and Forecasting

We will choose the best-fitting model based on goodness-of-fit metrics and diagnostic plots. The selected model will then be used to **forecast the next five days** of total revenue. Confidence intervals will be included to assess forecast uncertainty.

Project Proposal (GR): Advanced Time Series Analysis Using GARCH Models

Plan of Advanced Analysis (Analysis C)

To extend the SARIMA-based modeling of Total_Revenue in Part B, we will incorporate an advanced time series method that addresses time-varying volatility—specifically, the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. As observed in our preliminary analysis, Total_Revenue exhibits heteroscedasticity, with higher variance during certain periods of the year (e.g., April–May and September–November). These characteristics make the GARCH framework an appropriate extension.

---

1. Motivation for ARCH/GARCH Modeling

- Visual inspection of the residuals from the SARIMA model in Part B may reveal volatility clustering, a key signal of conditional heteroscedasticity.
- Preliminary tests such as the ARCH LM test will be used to formally check for ARCH effects in the residuals of the SARIMA model.
- Since the variance appears to change over time with periods of increased fluctuation, a GARCH-type model will help capture this dynamic.

2. Modeling Approach

- We will use the residuals from the SARIMA model fit in Part B as input to the GARCH modeling step.
- Two GARCH-family models will be considered:
    - ARCH(1): A simple model that captures first-order autoregressive volatility.
    - GARCH(1,1): A more flexible model that incorporates both lagged squared residuals and lagged conditional variances.

3. Estimation and Diagnostics

- We will estimate model parameters using maximum likelihood.
- Model selection will be based on:
    - AIC/BIC comparison between ARCH and GARCH models.
    - Diagnostic checks on standardized residuals (e.g., ACF, Ljung-Box test) to ensure white noise behavior.
    - QQ plots and normality tests for residual distribution evaluation.

4. Interpretation and Forecasting

- We will use the final model to forecast conditional variance for the next five days.
- We will compare this to the variance implied by the SARIMA model alone, and evaluate whether GARCH improves our understanding or prediction of volatility in hotel revenue.
- If time permits, we will consider combining the SARIMA and GARCH components into a hybrid model (e.g., SARIMA-GARCH), depending on residual behavior.

5. Literature Review

- We will include a brief review of GARCH applications in financial and hospitality demand forecasting to contextualize this modeling decision.
- This includes relevant studies on volatility modeling in daily revenue or occupancy data within the travel and tourism industries.