

Problem Statement: Forecasting Airline Performance Metrics

Background:

Airlines operate in a highly dynamic environment where historical trends, seasonality, and external shocks—such as the COVID-19 (Corona) pandemic—significantly impact operational and financial metrics. This problem involves a rich, monthly time series dataset spanning 2003 to 2023 (249 observations) that captures key performance indicators (KPIs) for the airline industry, including:

- Year & Month: Temporal markers for each observation.
- Passenger Counts:
 - Dom_Pax: Domestic passengers
 - Int_Pax: International passengers
 - Pax: Total passengers

Flight Operations:

Dom_Flt: Domestic flights

Int_Flt: International flights

Flt: Total flights

Revenue Metrics:

Dom_RPM: Domestic Revenue Passenger Miles

Operational Efficiency:

LF: Load Factor

Given the dataset includes the disruptive period of the COVID-19 pandemic, it presents an excellent opportunity to explore and compare advanced time series forecasting techniques to accurately model and predict these KPIs.

Objective:

Develop robust time series forecasting models using Python that can:

- Capture Seasonality and Trends: Identify and model the underlying seasonal patterns, long-term trends, and potential structural breaks, especially due to the COVID-19 period.
- Leverage Advanced Forecasting Techniques: Implement and compare different forecasting methods, including:
 - Exponential Smoothing Methods: Holt-Winters Seasonal Method
 - Bayesian Forecasting: Prophet module
 - Autoregressive Models: ARIMA
 - Multivariate Techniques: VAR (Vector Autoregression) to capture interdependencies among the KPIs
- Other Advanced/Hybrid Models: As applicable

Assess the Impact of COVID-19: Analyze how the pandemic period (around 2020 and its aftermath) influenced trends and seasonal patterns across different metrics and evaluate model performance before, during, and after this period.

Provide Actionable Insights: Generate forecasts that can inform strategic decisions regarding capacity planning, revenue management, and operational adjustments in a post-pandemic world.

Tasks to be Addressed:

Data Exploration and Preprocessing:

- Load the dataset and perform exploratory data analysis (EDA) to understand the distribution, trends, and seasonality of each variable.
- Identify any missing values or outliers, especially around the COVID-19 period, and decide on appropriate preprocessing techniques.
- Visualize the time series for each KPI, highlighting the period affected by the pandemic.

Univariate Time Series Analysis:

- Develop forecasting models for individual KPIs (e.g., Dom_Pax, Int_Pax, Flt) using:
 - Holt-Winters Exponential Smoothing
 - Prophet Module
 - ARIMA models
 - Neural Network Models

Evaluate the models using appropriate metrics (e.g., RMSE, MAPE) and validate the forecasts with a train-test split or cross-validation.

Multivariate Time Series Analysis:

- Explore potential interdependencies between variables (e.g., the relationship between flight counts and passenger numbers) using VAR models.
- Determine whether incorporating multiple KPIs into a joint forecasting model improves prediction accuracy.

Impact Analysis of the COVID-19 Period:

- Specifically analyze the structural changes and volatility during the COVID-19 pandemic.
- Compare model performance on data segments pre-, during, and post-pandemic to assess how each method handles disruptions.

Model Comparison and Selection:

- Systematically compare the forecasting performance of the various methods.
- Provide a rationale for selecting the most suitable forecasting approach(es) for each KPI and for integrated multivariate forecasting.

Reporting and Recommendations:

- Summarize key findings from the EDA and modeling phases.
- Provide actionable insights and recommendations for airline industry professionals to optimize planning and operations.
- Include visualizations, such as forecast plots and error metrics, to effectively communicate results.

Deliverables:

- A well-documented Python notebook/script that includes data preprocessing, model development, evaluation, and visualizations.
- A comprehensive report that discusses:
- Data insights and the impact of the COVID-19 period.
- Comparative performance of different time series models.
- Strategic recommendations for decision-makers in the airline industry.