Comparative Analysis of ARIMA and Neural Network Models for forecasting International Flight Departures

Anisha Dsouza, Prathamesh Joshi, Saideepti Seelam Veera Naga

Introduction

Project Focus: Comparative analysis of ARIMA and Neural Network models for forecasting international flight departures.

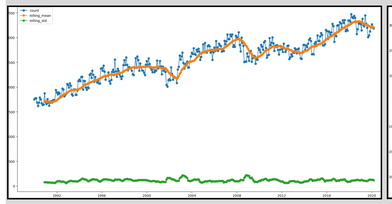
Goal: Evaluate and compare the predictive accuracy, efficiency, and adaptability of each model for flight departure patterns.

Motivation: Gain valuable insights for aviation forecasting strategies and operational decisions.

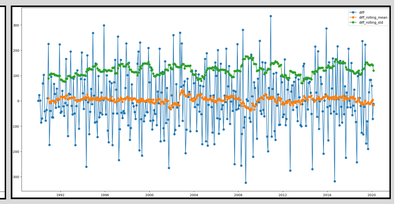
Primary Analysis

Stationarity of Time Series:

- Stationary series provide a stable foundation, reducing the impact of trends and improving the reliability of predictions.
- Utilized three approaches to assess stationarity: original series, differencing, and log transformation.
- Calculated rolling mean and rolling standard deviation for each transformed series.



Time series and rolling Mean and rolling Standard deviation for Original Series



Time series and rolling Mean and rolling Standard deviation for Differenced Series, where difference is First Order

Augmented Dickey–Fuller (ADF) test:

- Common method to assess stationarity in time series datasets.
- The p-value in the ADF test helps to determine the significance of the test results.
- A smaller p-value strengthens the evidence for stationarity, while a larger p-value suggests non-stationarity.
- For our data, the ADF test outputs gave:

Original Time Series
ADF Statistic: -1.829860
p-value: 0.365741

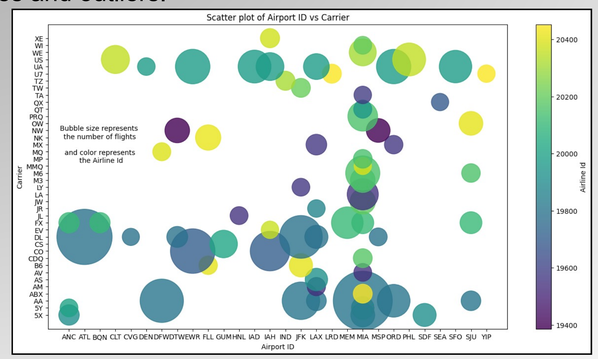
Differenced Time series
ADF Statistic: -4.895917
p-value: 0.000036

Log Time series
ADF Statistic: -2.313881
p-value: 0.167530

As a result, it can be asserted that First Ordered Differenced Time Series is a stationary time series.

Exploratory Data Analysis

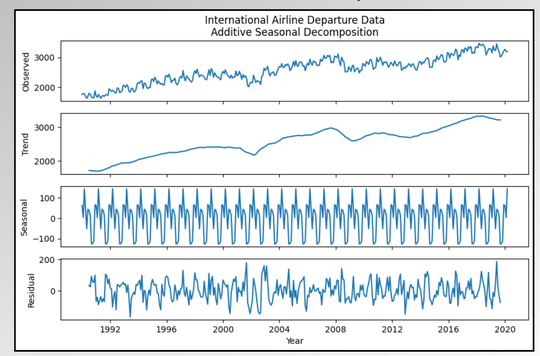
Transform the dataset into a time series format by leveraging flight timestamps, considering date, month, or a combined representation. We further handles the missing values and outliers.



Scatter plot of Top 70 International Departure Airports and Carriers

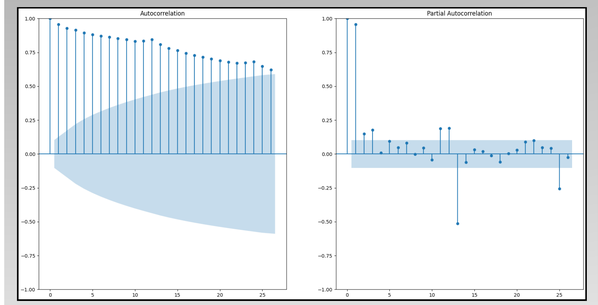
Time Series Analysis

- Time series modelling entails the analysis and prediction of sequentially ordered data points collected over time.
- Leverage decomposition insights to improve forecasting accuracy, identifying trends, seasonal, residual and observed patterns crucial for predictive models.

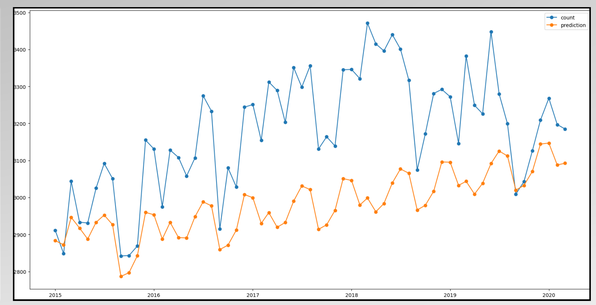


ARIMA

- ARIMA models are widely used in time series analysis and have shown promising results in forecasting international flight departures.
- Explore the predictive capabilities of different ARIMA configurations, including variations in auto-regression (AR), differencing (I), and moving average (MA) components.
- Examine the relationship between stationarity assumptions and ARIMA's performance, evaluating how well the model handles volatile aspects of the flight departure dataset.



ACF and PACF components of Original time series

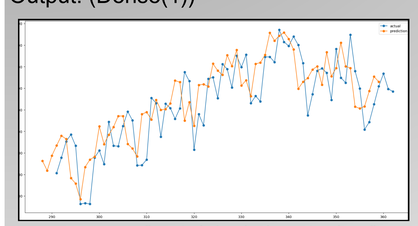


Test output for Original Time Series for ARIMA (12, 3, 12)

Test output for Differenced Time Series for ARIMA (12, 0, 12)

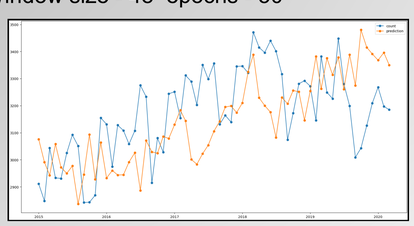
Neural Network

1. Dense Neural Network:
Layer1: (Dense(64, activation='relu'))
Layer2: (Dense(32, activation='relu'))
Output: (Dense(1))



Test output for Original Time Series for Dense Neural Network

2. LSTM Neural Network: Well-suited for effective sequence handling, they learn from historical patterns in time series data to make predictions.
Window size - 48 epochs - 50



Test output for Original Time Series for LSTM Neural Network

Layer (type)	Output Shape	Param #
lstm_9 (LSTM)	(None, 64)	16896
dense_10 (Dense)	(None, 1)	65

Results

Comparative Analysis of Accuracy and Speed for various models.

Model	RMSE	Execution Time
ARIMA (12, 2, 12) Original TS	232.99	2.65 s
ARIMA (12, 3, 12) Original TS	227.76	2.48 s
ARIMA (12, 0, 12) Differenced TS	159.96	2.14 s
Dense Neural Nets	121.15	35 s
LSTM Neural Nets	174.27	109 s

- Neural networks reign in accuracy, especially for complex data, but ARIMA wins on speed for simpler cases.
- Neural networks are also more adaptable to diverse data types and noise.