

Description

The Air Passenger Forecasting project aims to predict the number of passengers traveling by air using time series data. This project is a typical example of Time Series Analysis, where historical data is leveraged to forecast future values. The analysis involves exploring and modeling the patterns in the data, such as trends, seasonality, and cyclic behavior, to make accurate predictions.

Dataset

The project utilizes the `airpassengers.csv` dataset, which consists of 144 monthly observations from January 1949 to December 1960. The dataset contains two columns:

- Month: Represents the month and year in the format `YYYY-MM`.
- #Passengers: Indicates the number of passengers who traveled by air during that particular month.

Compilation

This Jupyter Notebook is compiled using Python and requires the following libraries:

- pandas: For data manipulation and analysis.
- matplotlib and seaborn: For data visualization.
- statsmodels: For statistical modeling, particularly ARIMA and SARIMA models.
- numpy: For numerical operations.

The notebook is designed to be executed sequentially, ensuring that all dependencies are properly loaded and the analysis flows logically.

Format

The notebook is structured into the following sections:

1. Dataset Information: This section provides an overview of the dataset, including a summary of its structure and the first few entries.

2. Time Series Analysis: Here, the data is explored through visualizations, statistical tests, and decompositions to understand underlying patterns.
3. Modeling: In this section, ARIMA and SARIMA models are developed to forecast future passenger numbers. The notebook includes order selection methods and a comparison between different models.
4. Conclusion: This section summarizes the findings and the accuracy of the forecasting models.

Usage

To use this notebook:

1. Ensure that all required Python libraries are installed.
2. Download the `airpassengers.csv` dataset and place it in the working directory of the notebook.
3. Open the notebook in Jupyter and run the cells sequentially to reproduce the analysis and predictions.
4. The notebook allows customization of the model parameters and can be adapted for other time series datasets.

Statistics

Throughout the analysis, the notebook provides insights into:

- Statistical Tests for Time Series Analysis: Tests like ADF (Augmented Dickey-Fuller) are used to assess stationarity.
- Order Selection for ARIMA & SARIMA Models: AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are used to select the optimal order for the models.
- In-sample and Out-of-sample Forecasting: The notebook demonstrates both rolling and non-rolling forecasting techniques.
- ARIMA & SARIMA Functions: The distinction between the `forecast` and `predict` functions is explained, highlighting their usage in different scenarios.