CS3 - Air Travel Case Study Rubric

DS 4002 - Fall 2024 - Emily Fredenburgh
Submission Format: Presentation and Github Repository

Purpose: Completing this assignment will let you practice time series analysis techniques with a real world application. Your goal is to produce a forecasting model that predicts airline passenger trends for the next 3 years based on the data available using the SARIMA technique. By the end of this case study, you will have honed your data cleaning, exploratory visualization, and predictive modeling generation skills with a practical application in airline traffic trends.

Task: You will navigate the materials provided in the Github repository to follow the process of analyzing the data and forecasting future trends in the data using Python scripts and the provided dataset. Upon the completion of this project, you will be able to deliver a report targeted to the airline industry that is well supported by your visualizations and provides an accurate forecast based on the data.

Tips For Success: Make sure to be concise but clear in explanations to avoid any confusion. Generate plenty of visualizations to support your conclusions. Document the reasoning behind decisions made in your process.

How Will I Know I've Succeeded? (Deliverables): You will provide a nice, professional report that summarizes all of your relevant findings and possible recommendations for future implications. You will have plenty of well titled visualizations, including exploratory plots, data visualizations, and forecasts. Good luck!

Spec Category	Spec Details

Repository & Submission	You will submit a complete, well-structured Github repository that contains several elements:
	● All of the scripts (likely in Python)
	  ○ Thoroughly commented and have headings describing the contents of the script and necessary libraries
	● All data (pre-cleaned, cleaned -any modifications and versions of the data included)
	● Analysis Visualizations
	  ○ Graphs, results of statistical analyses, forecasts
	● Exploratory Visualizations
	  ○ Initial plots, histograms, summary statistics - whatever applies

	- Properly named and documented files and folders - A map of documentation explaining where each element is and what each folder's purpose is (READme) - References I provided, and additional references you may have needed to find (READme) - Outline the steps you followed to reproduce my results (READme)
Data Preparation & Cleaning	Script execution and cleaning with 1-InitialDataCleaning.py - Successfully execute the provided script and clean the dataset (remove unnecessary variables) - cleaned_air_traffic_data.csv - Clear and consistent documentation of the cleaning steps
Exploratory Data Analysis	Visualizations outputted with 2-InitialDataPlots.py - Create clean, quality plots by executing the proper script - Consider many situations, such as airline vs. passenger counts, as well as monthly and yearly trends - Discuss any initial trends found in the data
Time Series Preparation	Transform the data and perform a stationarity analysis - Successful execution of 3-DataPreparationAggregation.py to generate a well structured dataset - Includes verification that activity_period was converted to the datetime format - monthly_passenger_data.csv - Successful execution of 4-StationarityDifferencingAnalysis.py that includes statistical and visualized plot checks that ensure stationarity - Explanation of the purpose of stationarity is not required, but appreciated
Modeling/Forecasting/Evaluation	Successful SARIMA model fitting and trend forecasting - The script 5-SARIMAModelFittingForecasting.py was executed and the parameters were well chosen for the context - Successfully forecasted a feasible 3 year forecast with confidence intervals included - Evaluate residuals with ADF and PACF plots - Discuss the model's overall performance, as well as confidence in the model