Analysis of Airline Passenger Data Using Fourier Transform and Statistical Techniques

Name of the student: Shivani Guduru

Student ID: 23081838

Data Description

The dataset includes two key attributes: Date, that corresponds to the observation date in yyyy-mm-dd format, and Number, corresponding to the number of passengers that were flown on that particular day in thousands. In order to determine the periodicity, the Month was derived from the Date column to group and find the monthly average of the number of passengers. Furthermore, the daily passenger numbers dataset was pre-processed for Fourier analysis to identify patterns and contribution frequencies. The data gives a picture of short-run and long-run periodic fluctuations.

Methods

Data Preprocessing

The daily airline passenger numbers were normalized and underwent other pre-processing before the analysis. The Date column was converted to datetime format to allow time-based analysis and new columns, Year and Month, added to allow for seasonal trend analysis. As part of quantitative data, the Number, containing daily passenger counts, were obtained as numerical data in the form of arrays for mathematical analysis.

Fourier Transform

Applying Fourier Transform on the daily passengers enabled the breakdown of time series data into its frequency content. This mathematical operation defines cycles or frequencies in data such as annual or weekly patterns in the data. When these components have been identified, the patterns that govern the rates of passenger change become apparent. For the sake of analysis, only the first eight terms of the Fourier series were applied as an approximation of changes in the passenger number. This approach is thus a good compromise between complexity and simplicity while capturing the main features of passenger Utilization.

Monthly Average Calculation

The examination of monthly average passengers demonstrates strong monthly fluctuations, which basically represent the seasonal changes in people's mobility. July and August represent the month of vacation and holiday hence the average daily passenger for July was 27.30k and August 23.50k. On the other hand, the off-peak period of January and February revealed considerably lower means of 15.66k and 14.47k as expected due to cold weather that discourages travel. The moderate demand also falls under the shoulder seasons, with April having 19.47 thousand visitors and September having 19.23 thousand visitors due to travel during spring and after summer.

Power Spectrum

The power spectrum was then determined from the Fourier coefficients to estimate how much of the overall passenger variation is accounted for by each periodicity. Weekly periods and monthly periods, and periods between one week and one year were examined and the annual cycle was found to be an essential factor. This establishes beyond doubt that daily passenger variation is heavily influenced by seasonality.

Results

Monthly Passenger Distribution with Fourier Series Approximation

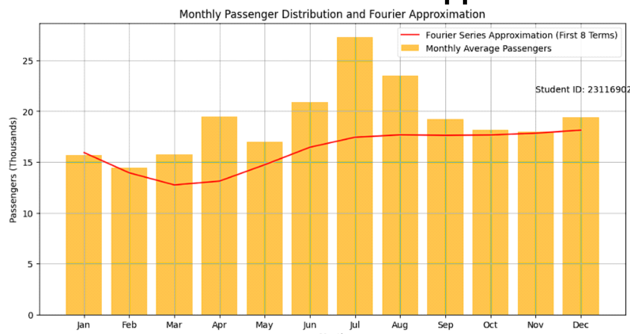

Figure 1: Monthly Passenger Distribution with Fourier Series Approximation
(Developed in python)

The graph below shows the monthly passenger distribution; the red line represents the Fourier series approximation. Monthly average number of passengers has been depicted by a bar chart and the trend shows higher values in summer and lower values in winter. The red line corresponds to the Fourier series fit taken to the first eight terms which gives the general pattern of fluctuations in passengers during a year. The approximation used here is able to replicate the cyclical behaviour in the data and reduce variability. These two visualizations combined demonstrate the impact of seasonality on travel, where the Fourier series gives a smooth continuous version of the passenger trend.

Power Spectrum Analysis

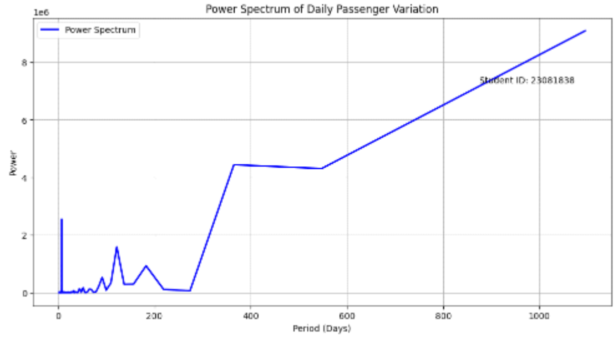

Figure 2: Power Spectrum Analysis

(Source: Developed in Python)

The graph illustrates the decomposition of the daily fluctuations of passengers and how many periodic cycles contributed to the result. On the x- axis is the period in days while on the y-axis is a power, which reveals the importance of each period. There is a clear upward spike at approximately 365 days, underlining that the annual cycle plays the most important role in the changes of passenger numbers due to seasonal fluctuations. Smaller one characterizes shorter-term periodic contributions such as daily modulations while the larger ones characterize long term modulations such as weekly modulations.

Key Values X and Y

The power spectrum yields two major values, X = 1095.00 days, a measure of the dominant annual rhythm of the time series and Y = 9087685.18, an index of the maximum power or variability of the series. From these results it is discernible that the annual periodic variation exerts the largest impact on passenger movements, pointing to the high seasonality of demand for travel throughout the year.

Formula

Identify X (Period of Max Contribution): X=T[argmax(P(f))]

argmax(P(f)) is the index of the maximum value in the power spectrum.

T is the array of periods.

Identify Y (Max Power): Y=max(P(f))

Discussion

The analysis used in the coursework provides some focus on important aspects of airline passengers. Daily averages further show a clear seasonality whereby the summer months of July and August receive lots of traffic while the winter months of January and February have low traffic. The Fourier series, which require only eight terms, provide a practical way to model daily fluctuations, and more with surprising levels of accuracy and computational ease.