Flight Data Analysis Project

Group: Meri Asatryan, Lili Kostanyan, Jeni Balaban

Project Overview:

Welcome to the Flight Data Analysis project, a comprehensive exploration of meticulously curated flight data from Zvartnots International Airport. This project aims to extract meaningful insights from the dataset through thorough data cleaning, basic analysis, time series analysis, and advanced ARIMA modeling.

Project Goal:

The overarching goal of this Flight Data Analysis project is to provide actionable insights and foresight into the patterns, trends, and performance metrics of flights departing from Zvartnots International Airport.

Data Cleaning:

The project begins with a meticulous data cleaning process, ensuring the dataset's reliability and consistency. Operations include merging datasets, handling missing values, and creating new columns for enhanced analytics. The result is a refined dataset ready for in-depth exploration.

Merging Datasets:

The merging process involves the seamless integration of disparate data sources, fostering a holistic view without compromising data integrity. This strategic operation sets the stage for a comprehensive understanding of the intricacies within the dataset.

Handling Missing Values:

Addressing missing values is approached judiciously, employing methods tailored to the dataset's nature. The goal is completeness without introducing bias, ensuring that subsequent analysis are built upon a robust foundation.

Creating New Columns:

Strategic column creation goes beyond procedural formalities; it introduces variables that significantly amplify the depth of subsequent analysis. These new columns serve as valuable dimensions, enriching the dataset with nuanced perspectives for detailed exploration.

Basic Analysis:

The project conducts fundamental analysis to unravel key aspects of the dataset. This includes identifying popular destinations, evaluating the frequency of flights to specific countries, and visualizing the distribution of flights and airline performance. Bar charts provide intuitive visualizations for effective interpretation.

Identifying Popular Destinations:

This involves a nuanced examination of travel preferences, considering factors such as seasonality, passenger demographics, and economic indicators. The objective is to identify trends that influence travel choices and, subsequently, the airline industry's strategic decisions.

Flight Frequency Evaluation:

Understanding the frequency of flights over time provides strategic insights for both airlines and airport management. Trends in flight frequency reveal operational dynamics, guiding decisions related to capacity planning, resource allocation, and overall operational efficiency.

Visualizing Distribution and Airline Performance:

Visualizations, including bar charts, present a narrative of flight distribution and airline performance. These visuals serve as a compass, guiding stakeholders through the operational landscape, providing insights that transcend numerical summaries.

Time Series Analysis:

Delving into time series analysis, the project unveils temporal patterns and trends in flight data. Techniques like moving averages, exponentially weighted moving averages, and differencing are applied to detrend the data. Line plots, time series plots and autocorrelation plots facilitate a comprehensive understanding of the dataset's temporal dynamics. Integral to this process is

the meticulous evaluation of stationarity, with methods like the Dickey-Fuller test ensuring the time series' readiness for advanced modeling. This collective effort offers a profound understanding of the temporal evolution of flight data, setting the stage for the subsequent application of ARIMA modeling techniques.

ARIMA Modeling:

The project employs advanced time series analysis through ARIMA modeling. Leveraging differencing, autocorrelation, and partial autocorrelation, the project fits ARIMA models to make insightful predictions. Visualizations showcase the effectiveness of ARIMA modeling in forecasting future flight counts, providing stakeholders with valuable foresight. The pivotal role of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) becomes apparent in this stage. ACF helps unveil the correlation between a time series and its lagged values, while PACF highlights the direct relationship, enabling the project to discern and quantify the temporal dependencies within the flight data. These critical analysis pave the way for insightful predictions, as visualized through effective and illustrative representations, offering stakeholders a valuable tool for forecasting future flight counts with enhanced precision and foresight.

Tableau Visualizations:

The project strategically incorporated Tableau visualizations to enrich the depth and clarity of our analysis. Specifically, Tableau was employed to present the distribution of flight destinations in a visually compelling manner. By leveraging the dynamic and interactive features of Tableau, stakeholders were provided with a powerful tool to explore intricate flight patterns with ease. The ability to dynamically filter, zoom, and explore the data in real-time enhanced the interpretability of the information. Through intuitive visualizations, Tableau facilitated a more profound understanding of the distribution of flight destinations, enabling stakeholders to derive actionable insights.

Code Highlights:

The Python code exemplifies proficiency in data science and analytics. Utilizing libraries such as Pandas, NumPy, Statsmodels, and Matplotlib, the code showcases expertise in data manipulation, cleaning, and advanced time series analysis. It serves as a valuable resource for enthusiasts seeking to expand their skills in data science.

Conclusion:

This Flight Data Analysis project serves as a comprehensive resource for extracting actionable insights from flight data. Whether exploring popular destinations, evaluating airline performance, or forecasting future flight counts, the project offers a robust foundation for further analysis. The detailed code and analysis encourage exploration and adaptation based on specific analytical needs.