Project Proposal

1. **Group 3 Information**
   Group Name: TrendTakeoff
   Group Members: Andi Zhang, Serena Liu, Victoria Fu, Charlotte Liao
2. **Project Overview**
   The United States airline industry sees swings in passenger traffic, flight volume, and revenue passenger-miles as a result of economic conditions, seasonal demand, and worldwide events. Learning these trends is critical for improving operations and projecting demand. This study examines past airline traffic to identify key patterns and then uses time series models to build forecasts for the next 12 months, offering valuable insights for aviation decision-making.
3. **Data Collection and Sources**
   a. **Airline Traffic Data:** The dataset is downloaded from Kaggle.
   The dataset contains monthly U.S. airline traffic data from January 2003 to December 2023, covering domestic and international passenger volume, flight counts, revenue passenger-miles (RPM), available seat miles (ASM), and load factors (LF). It provides insights into seasonal trends and long-term changes in air travel.
   b. **Jet Fuel Price:** The dataset is downloaded from the U.S. Energy Information Administration (EIA).
   The dataset provides monthly U.S. Gulf Coast jet fuel prices from EIA's historical petroleum data. It tracks fuel costs per gallon, serving as a key indicator of airline operational expenses. This data will help analyze fuel price fluctuations and their impact on airline passenger volume, flight counts, and revenue passenger-miles over time.
   c. **Real GDP:** The dataset is downloaded from S&P Global.
   The dataset provides monthly U.S. GDP Index which is an indicator of real aggregate output that is conceptually consistent with real Gross Domestic Product (GDP) in the National Income and Product Accounts (NIPA). This data will help analyze the influence the macroeconomic environment can have on airline passenger volume, flight counts and revenue passenger-miles over time.
   - **Rights to Use and Share the Data:** The datasets from Kaggle, S&P Global Market Intelligence, and the U.S. Energy Information Administration (EIA)are publicly available for non-commercial research and educational purposes, with usage subject to each source's data-sharing policies.
4. **Proposed Models and Methodology**
   a. **Exploratory Data Analysis:** EDA will assess data integrity, identify missing values, visualize variable distribution and reveal key patterns in the dataset.
   b. **Time Series Models and Machine Learning Models**
   - **Seasonal ARIMA (SARIMA):** Extends ARIMA by incorporating seasonality, making it well-suited for airline traffic, which follows periodic fluctuations driven by holidays, summer travel, and other recurring factors.
   - **Bayesian Structural Time Series (BSTS):** Offers a flexible model for forecasting air travel, capturing trends, seasonality, and external factors like economic conditions and policy changes.
   - **Extreme Gradient Boosting (XGBoost)**: A powerful gradient boosting algorithm that excels at time series forecasting by handling complex relationships, non-linear patterns, and large datasets efficiently, making it ideal for predicting air travel demand with multiple influencing factors.
   - **Recurrent Neural Network (RNN):** RNNs are ideal for time series forecasting as they capture temporal patterns in sequential data. We will use advanced RNN architectures like LSTM and GRU to model complex, non-linear trends in airline traffic, addressing long-term dependencies and impacts from events like COVID-19.
5. **Research Questions**
   a. **Main Research Question:** How have U.S. airline passenger traffic, flight volume, and revenue passenger-miles evolved over time, and how can time series forecasting models predict their values for the next 12 months by incorporating macroeconomic indicators and fuel prices to help airlines optimize flight scheduling, adjust capacity, and respond to market demand efficiently?
   b. **Three Specific Smaller Questions:**
   - Analyzing the seasonal and cyclical patterns in domestic and international airline travel, with a focus on how COVID has influenced these trends;
   - Examining the long-term trends in U.S. airline travel, with a focus on identifying differences between the pre-COVID and post-COVID periods;
   - Assessing the accuracy of time series forecasting models in predicting airline passenger traffic, flight volume, and revenue passenger-miles for the next 12 months.