# **Dynamic Pricing Strategies for US Airlines**

# **Team 2:**

Abhishek Tikam Ramchandani

Devansh Manoj Thard

## **Introduction:**

**Background:**

In the competitive landscape of the U.S. airline industry, dynamic pricing has emerged as a key strategy for maximizing revenue, leveraging advanced machine learning algorithms to adjust ticket prices in real time based on demand, competitive actions, and market trends. This approach utilizes a sophisticated analytical framework, integrating data from diverse sources including historical sales, search frequency, passenger booking patterns, and external economic factors. By applying predictive models such as regression analysis, time series forecasting, and machine learning techniques like random forests and gradient boosting, airlines can now more accurately forecast demand and set prices that optimize both load factors and profitability. This shift towards a data-driven pricing strategy reflects the industry's adaptation to the digital age, where the ability to quickly analyze and act on large volumes of data becomes a critical competitive advantage.

**Motivation:**

The driving force behind our endeavor is to harness advanced data analytics to revolutionize airline pricing structures, thereby enhancing revenue management and customer satisfaction. Our project seeks to delve into the complexities of pricing in the airline industry by applying sophisticated machine learning models to identify the most effective pricing strategies. The insights gained from this project aim to empower airlines with the agility to respond to market dynamics in real time, offering optimized pricing that benefits both the airline's bottom line and the consumer's value for money. Ultimately, this initiative is poised to contribute substantially to the efficiency and economic success of the U.S. airline industry.

**Goal:**

The primary objective of our project is to develop a dynamic pricing model for the U.S. airline industry by scrutinizing and learning from our expansive dataset. Our goal is to uncover the multifaceted factors that influence flight pricing and to construct a predictive model that can accurately forecast the optimal prices for airline tickets. With a focus on applying machine learning algorithms, we aim to provide airlines with actionable insights that allow for real-time pricing adjustments, ultimately boosting profitability and market responsiveness. Through our analysis, we aspire to craft a model that not only benefits airlines by driving revenue growth but also enhances the consumer experience through fair and responsive pricing.

## Methodology

1. **Data preprocessing and cleaning:** Our initial step involves rigorously cleaning and preprocessing the airline dataset to ensure data integrity and usefulness. This includes handling missing values, normalizing fare amounts, encoding categorical variables such as airport codes and fare basis codes, and parsing date-time columns for subsequent temporal analyses. Outlier detection and removal will be performed to prevent skewing of our model outcomes.
2. **Exploratory data analysis:** Our exploratory data analysis (EDA) will leverage statistical and visualization techniques to discern underlying structures within the data and to surface insights that could inform our pricing model. We will utilize histograms, box plots, and time series plots to understand the distribution of key variables like base fare and total fare, and we will create heat maps and correlation matrices to reveal relationships between different pricing factors and flight characteristics. Our EDA will also include the examination of seat occupancy rates, the impact of layovers on pricing, and a deep dive into the fare elasticity evident from historical data.
3. **Feature engineering:** We will select features informed by industry knowledge and statistical tests that will help us pinpoint crucial predictors for our pricing model. We anticipate key features will include time factors (like days to departure and time of booking), route specifics (such as departure and destination airports), and flight details (including stopovers and available seats). Beyond selection, we'll also craft new variables from the current dataset, such as anticipated demand based on seasonal trends, which could significantly bolster the precision of our price forecasts.
4. **Model Selection and Evaluation:** We wish to employ a range of regression models and machine learning techniques tailored for time series forecasting and price optimization, such as ARIMA, Random Forest Regression, Gradient Boosting Machines, and Neural Networks. These models will be rigorously trained and tuned on our historical airline pricing data to identify the most accurate and robust approach for predicting ticket prices. Evaluation metrics such as RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) will be utilized to assess each model's performance, ensuring we select the best-performing model that aligns with the dynamic nature of airline pricing.
5. **Time series analysis:** For our dynamic pricing model, we will harness time series analysis to forecast pricing trends, considering the temporal dependencies inherent in the data. We'll utilize methods suited to handle seasonality and trend components in our dataset, such as ARIMA for its adaptability to non-stationary data, and LSTM networks for capturing long-term dependencies. This phase will be pivotal in understanding how prices change over time and in predicting future fare prices with an eye toward maximizing revenue while maintaining market competitiveness.

## **Description of the Dataset**

Our dataset comprises a hefty 31.09 GB of data on one-way flights sourced from Expedia, spanning from April 16, 2022, to October 5, 2022, to/from the following airports: ATL, DFW, DEN, ORD, LAX, CLT, MIA, JFK, EWR, SFO, DTW, BOS, PHL, LGA, IAD, OAK. It details a spectrum of variables including flight and booking dates, departure and arrival information, pricing, seat availability, and more. This rich dataset is poised to fuel our analysis, offering a detailed snapshot of the U.S. airline industry's pricing patterns for a dynamic pricing model.

These are some relevant columns from the dataset:

* legId: Essential for uniquely identifying each flight leg.
* searchDate and flightDate: Crucial for understanding booking patterns and forecasting demand.
* startingAirport and destinationAirport: Important for gauging route popularity and pricing strategies.
* isBasicEconomy, isRefundable, and isNonStop: Significant for segmenting the market and price differentiation.
* baseFare and totalFare: Key to analyzing pricing trends and determining price elasticity.
* seatsRemaining: Indicative of demand and potential for price adjustments.
* totalTravelDistance: Can be correlated with pricing, as longer distances may have different pricing strategies.
* segmentsDepartureTimeEpochSeconds and segmentsArrivalTimeEpochSeconds: Timing can affect pricing due to varying peak and off-peak periods.
* segmentsAirlineCode: Airline identity may influence pricing due to brand positioning.
* segmentsDurationInSeconds: Flight duration could impact the pricing, especially for different market segments.
* segmentsDistance: Like total travel distance, this can influence fare calculations.

## Data Source:

## <https://www.kaggle.com/datasets/dilwong/flightprices>

## <https://www.dropbox.com/scl/fo/mybc5v9s800orsu78b6ao/h?rlkey=1an4ndcscd5uw9yi7oxx8ypfn&e=1&dl=0>