**Introduction**

Different airline industries in different countries have different rules and regulations due to which airline flight pricing strategies have developed into complex structures with sophisticated rules and mathematical models that drive the price strategies of flight tickets if all the factors and characteristics are taken into consideration. Traditional factors or variables such as distance, duration, and ticket class (economy and business) play a significant role but are no longer the only factors that dominate the pricing strategy. Economic, marketing, and sociological factors have been increasingly influential in determining flight ticket rates. The majority of studies on forecasting flight prices have focused on the market. The airport pair between the flight origin and destination is defined as a market. Airlines must be able to forecast airfare trends in order to alter their strategies and resources for the market along a particular route. However, research that already exists on flight prediction uses conventional statistical models such as linear regression and is based on the assumption that the relationship between dependent variables and independent variables is a linear relationship, which is not true in many cases. The advancement of statistical models and machine learning (ML) makes it possible to let us infer rules and simulate fluctuations in flight costs based on a huge variety of factors, frequently automatically finding hidden links between them. The difficulties in acquiring access to the data in the former case make replicating the results and expanding the study practically impossible. The problem with the latter is that each online booking site's transaction records represent a small percentage of total ticket sales in the entire market because there are several sites and people also book tickets from agents and directly from the airport. For different reasons, data collection from the airport and thousands of agents and sites makes it time-consuming and impossible. This causes the data to be skewed, as most of the data is not normally distributed and thus does not reflect the true nature of the entire market. The purpose of this research paper is to analyze and investigate the relationship between several factors that might be playing an important role in the pricing strategy of airlines and to develop a statistical modeling framework to predict the flight price with the help of publicly available datasets from Kaggle. This research paper will include background information about the dataset, preprocessing, and cleaning of the dataset to make it the best fit to use in our research. Later, we will do some exploratory data analysis and build models and interpret and analyze them based on the knowledge we have gained in the data science course this semester, and finally, we will conclude and end the research paper with a list of references that we considered and used.

**Why did we choose this topic?**

When we are traveling by flight, we are always confused by shifting airplane ticket prices. Why are flight prices so exorbitant one minute and reasonable the next? What is the optimal time to book a flight to obtain the best price? When is the best time to get those great airplane ticket offers and discounts? As a result, we have compiled an interesting list of points to consider when purchasing an airline ticket. As a result, we are inclined to conduct research on predicting flight prices because it is relevant to our daily lives.

**An overview of the dataset**