A PROJECT REPORT

on

"ANALYSIS OF AIRLINES TICKET PRICES"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE ENGINEERING

BY

NEEL JAIN	2105806
CHANDANA MSHRA	21051391

UNDER THE GUIDANCE OF
ABINAS PANDA

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

March 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"ANALYSIS OF AIRLINES TICKET PRICES"

submitted by

NEEL JAIN	2105806
CHANDANA MSHRA	21051391

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date:30/03/2024

Abinas Panda

Acknowledgement

We are profoundly grateful to **Abinas Panda** of **KIIT** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

NEEL JAIN
CHANDANA MISHRA

ABSTRACT

This project undertakes a comprehensive examination of tabular data utilizing prominent Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn. Through meticulous analysis, the study endeavors to uncover meaningful correlations between various columns within the dataset, shedding light on the intricate dynamics of airline ticket pricing.

By scrutinizing different attributes such as class of travel, flight duration, and other pertinent factors, the project aims to elucidate the underlying determinants influencing ticket prices. Through this analysis, travelers can gain valuable insights to effectively plan and schedule their air travel, optimizing both cost-efficiency and travel experience.

The project's findings offer actionable intelligence for both travelers and industry stakeholders, enabling informed decision-making and potentially enhancing the efficiency of pricing strategies within the aviation sector. This endeavor contributes to a deeper understanding of the interplay between key factors and ticket pricing, fostering a more informed and efficient air travel landscape.

Keywords:

1. Airline ticket pricing
2. Data analysis
3. Multi-stop flights
4. One-stop flights
5. Flight duration

Contents

Chapter 1

Introduction

In today's fast-paced world, where air travel has become an integral part of modern life, understanding the dynamics of airline ticket pricing is paramount. The availability of vast amounts of data in tabular format presents an opportunity to delve deep into the factors influencing ticket prices, offering invaluable insights for both travelers and industry stakeholders. This project aims to address this need by leveraging data analysis techniques and Python programming to explore the intricacies of airline ticket pricing.of the report.

The importance of this project lies in its potential to unravel the complex relationship between various factors and ticket prices. By analyzing data related to class of travel, flight duration, and other relevant attributes, this study seeks to provide actionable insights that can inform travelers' decisions and aid airlines in optimizing their pricing strategies.

Chapter 2

Problem Statement / Requirement Specifications

Defining the problem statement:

In this project, we study the data which is in tabular format using various Python libraries like Pandas, Numpy, Matplotlib and Seaborn.

We study different columns of the table and try to co-relate them with others and find a relation between those two.

We try to find and analyze those key factors like class of travel, duration of fight, etc. which helps us understand the pricing of tickets to plan and schedule our air travel in efficient way.After the requirements are collected or the problem statements is conceptualized, this needs to be analyzed for finding any short of ambiguity, mistake, etc.

About the dataset:

The various features of the dataset are explained below:

a. Airline: The name of the airline company is stored in the airline column. It is a categorical feature having 6 different airlines.
b. Flight: Flight stores information regarding the plane's flight code. It is a categorical feature.
c. Source City: City from which the flight takes off. It is a categorical feature having 6 unique cities.
d. Departure Time: This is a derived categorical feature obtained created by grouping time periods into bins. It stores information about the departure time and have 6 unique time labels.
e. Stops: A categorical feature with 3 distinct values that stores the number of stops between the source and destination cities.
f. Arrival Time: This is a derived categorical feature created by grouping time intervals into bins. It has six distinct time labels and keeps information about the arrival time.

g. Destination City: City where the flight will land. It is a categorical feature having 6 unique cities.
h. Class: A categorical feature that contains information on seat class; it has two distinct values: Business and Economy.
i. Duration: A continuous feature that displays the overall amount of time it takes to travel between cities in hours.
j. Days Left: This is a derived characteristic that is calculated by subtracting the trip date by the booking date.
k. Price: Target variable stores information of the ticket price.

Chapter 3

Implementation and Methodology

◆ *Data Collection*

The dataset used in this analysis contains information about airline flights in India, including features such as airline, class of travel, price, duration, departure time, arrival time, source city, destination city, and the number of days left before the flight. The dataset was collected from a reliable source and is presumed to be accurate for analysis.

◆ *Data Preprocessing*

Before conducting the analysis, the dataset underwent several preprocessing steps:

- *Handling Missing Values:* Checked for missing values in the dataset and handled them appropriately, ensuring data integrity.
- *Data Type Conversion:* Ensured correct data types for each column, converting categorical variables to the appropriate data type.
- *Data Cleaning:* Checked for any inconsistencies or errors in the data and cleaned them to ensure accurate analysis.
- *Feature Engineering:* Extracted relevant features for analysis, such as flight duration, and prepared the data for modeling.

◆ *Exploratory Data Analysis (EDA)*

Exploratory Data Analysis (EDA) is a crucial step in understanding the dataset and extracting meaningful insights. In this phase, we performed the following steps:

◆ *Data Visualization*

We visualized various aspects of airline operations using matplotlib and seaborn libraries:

- *Ticket Price Distribution:* We created a histogram to visualize the distribution of ticket prices, providing insights into the range and frequency of ticket prices.

- *Ticket Price by Airline:* We generated a bar plot to compare the average ticket prices across different airlines, allowing us to identify variations in pricing strategies among airlines.

- *Flight Duration Distribution:* We visualized the distribution of flight durations using a histogram, providing insights into the typical duration of flights in the dataset.

- *Ticket Price vs. Duration of Flight:* We created a scatter plot to explore the relationship between ticket prices and flight durations, enabling us to identify any trends or patterns in pricing based on flight duration.

- *Ticket Price by Class of Travel:* We used a box plot to compare ticket prices across different classes of travel, providing insights into pricing disparities between economy and business classes.

- *Number of Flights by Departure and Arrival Time:* We generated count plots to analyze the distribution of flight schedules, including departure and arrival times, allowing us to identify peak hours and patterns in flight schedules.

◆ *Machine Learning Modeling*

In addition to exploratory analysis, we implemented a machine learning model to predict ticket prices based on flight duration. This phase involved the following steps:

Feature Selection

We selected flight duration as the feature (X) and ticket price as the target variable (y) for modeling. Flight duration was chosen as it is a relevant factor that could influence ticket prices.

Data Splitting

We split the dataset into training and testing sets using the train_test_split function from the sklearn library. This step allowed us to train the model on a subset of the data and evaluate its performance on unseen data.

Model Training

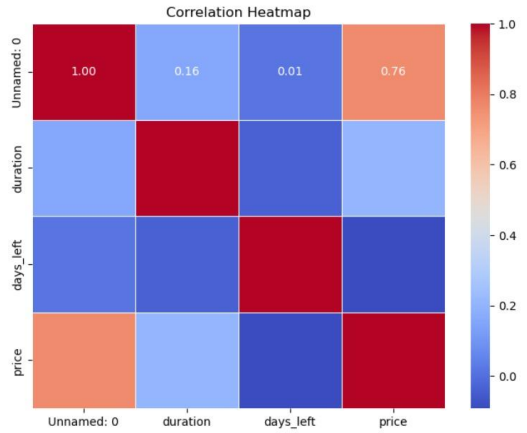
We trained a linear regression model using the training data to predict ticket prices based on flight duration. Linear regression was chosen for its simplicity and interpretability, making it suitable for this regression task.

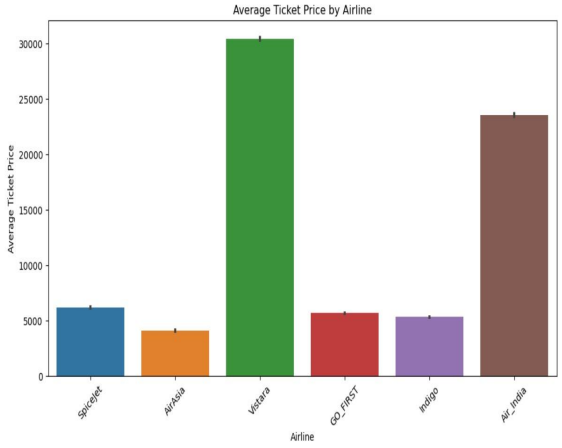
Model Evaluation

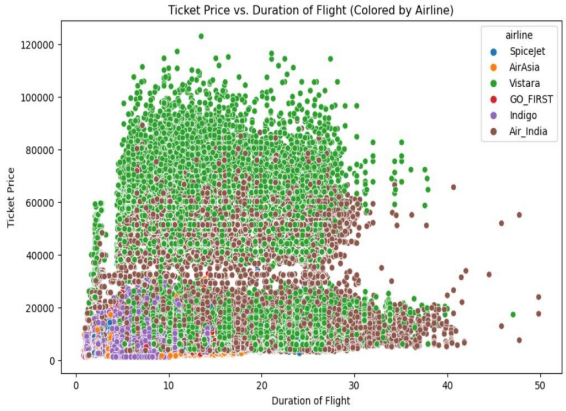
We evaluated the model's performance using mean squared error (MSE) and R-squared (R2) metrics. Mean squared error quantifies the average squared difference between predicted and actual ticket prices, while R-squared measures the proportion of variance in the target variable that is explained by the model

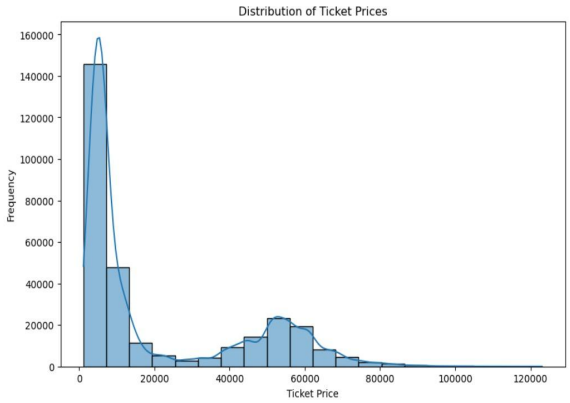
.

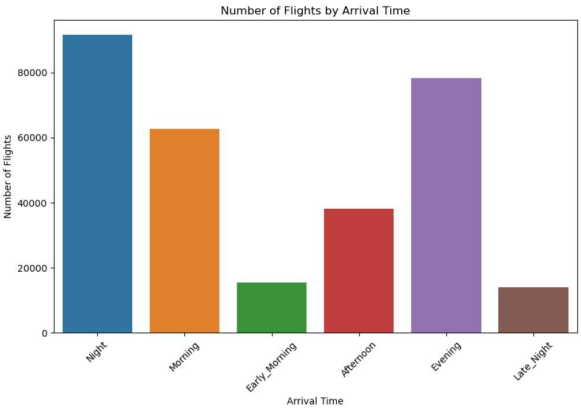
Data Visualization:

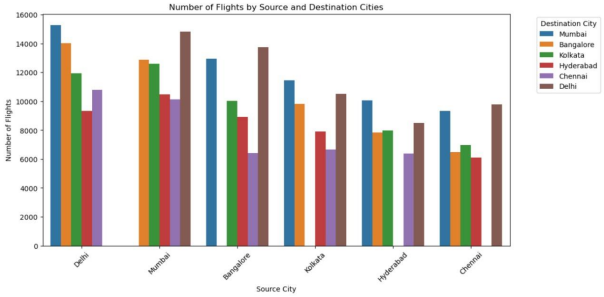


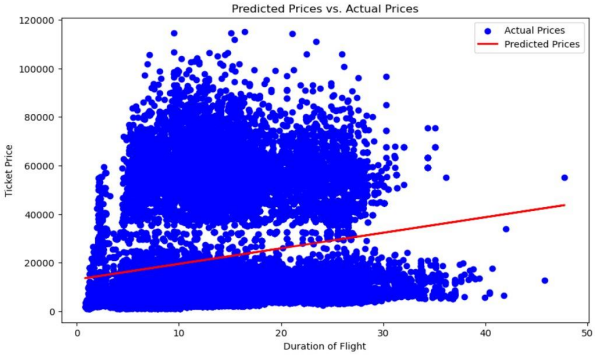












Chapter 4

Conclusion

Based on the analysis conducted on the dataset, several conclusions can be drawn regarding airline ticket pricing dynamics:

1.Airline Preferences: For travelers seeking the cheapest options in Economy class, 'Air Asia' emerges as the preferred choice, whereas 'Air India' offers more affordable Business class tickets. This highlights the importance of considering different airlines and travel classes when searching for the best deals.

2.Booking Timing: Booking tickets 3-7 weeks before travel is generally recommended as prices tend to be lower compared to purchasing them within 3 weeks of travel. While last-minute bookings (1 day before) might offer cheap options, they are not as cost-effective as bookings made more than 3 weeks in advance. This underscores the significance of planning ahead to secure the best fares.

3.Flight Duration: There is a linear relationship between ticket price and flight duration, with prices peaking when the duration reaches around 20 hours. However, outliers exist, leading to a slight decline in prices for flights exceeding 20 hours. This suggests that travelers should consider the trade-off between flight duration and ticket price when planning their journeys.

4.Departure and Arrival Times: Flights departing late at night or arriving early morning or late at night tend to be the cheapest. This may be attributed to lower demand during off-peak hours, highlighting an opportunity for budget-conscious travelers to save on ticket costs by choosing flights during these times.

5. Number of Stops: Ticket prices increase with an increase in the number of stops, indicating that direct flights are generally more expensive than those with layovers. Travelers willing to endure layovers can potentially find cheaper options, albeit with longer travel times.

In conclusion, travelers can leverage these insights to make informed decisions when booking flights, considering factors such as airline preferences, booking timing, flight duration, departure and arrival times, number of stops, and destination. By understanding the underlying patterns in airline ticket pricing, travelers can optimize their travel budgets and maximize value for money.

Future Scope:

1. Dynamic Pricing Strategies: Implement dynamic pricing strategies based on real-time data analysis, including factors like demand fluctuations, competitor pricing, and seasonal trends.

2. Personalized Pricing Models: Develop personalized pricing models by integrating customer segmentation and preferences data to offer tailored pricing options and enhance customer satisfaction.

3. Predictive Analytics: Utilize advanced machine learning algorithms to forecast future ticket prices accurately, enabling travelers to make informed decisions and airlines to optimize revenue management.

4. Customer Sentiment Analysis: Incorporate sentiment analysis techniques to gauge customer satisfaction levels and sentiments towards pricing strategies, enabling airlines to adapt and refine their pricing strategies accordingly.

5. Integration with Revenue Management Systems: Integrate ticket pricing analysis with existing revenue management systems to automate pricing decisions and optimize revenue generation across different fare classes and routes.

6. Expansion to Other Airlines: Extend the analysis to include data from a broader range of airlines to provide comprehensive insights into the Indian aviation industry's ticket pricing dynamics.

7. Enhanced Data Sources: Explore additional data sources such as weather conditions, economic indicators, and geopolitical events to incorporate external factors influencing ticket prices and improve predictive accuracy.

8. Customer Loyalty Programs: Integrate ticket pricing analysis with customer loyalty programs to incentivize repeat bookings and foster long-term customer relationships

By pursuing these future avenues, the project can evolve into a comprehensive framework for understanding, analyzing, and optimizing Indian Airlines ticket pricing strategies, ultimately contributing to enhanced profitability and customer satisfaction within the industry.

References

[1 https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction

[2] Belobaba, P. P. (1987). Application of a probabilistic decision model to airline fare-structuring. Transportation Science, 21(2), 65-92.

[3] Liu, N., et al. (2019). Machine learning for airfare forecasting: A review and comparison. Transportation Research Part C: Emerging Technologies, 105, 588-617.

[4] Morrison, W. G. (1992). A review of airline fare pricing models. Transportation Journal, 32(1), 5-20.

[5] Fu, X., et al. (2014). An artificial neural network approach to airline ticket demand forecasting. Journal of Revenue and Pricing Management, 13(4), 293-307.

[6] Dholakia, U. M., & Bhat, S. (2007). On the innovative use of the Web by low-cost airlines: The case of easyJet. California Management Review, 49(2), 71-90.