Flight Price Prediction

Data Science With Python Lab Project Report

Bachelor

in

Computer Science

By

Team Name

S190767

S180995

Rajiv Gandhi University Of Knowledge And Technologies

S.M. Puram , Srikakulam -532410

Andhra Pradesh, India

Abstract

People who frequently travel through flight will hava better knowledge on best discount and right time to buy the ticket. For the business purpose many airline compines change price according to the seasons or time duration. They will increase the price when people travel more(like festival time or holidays and easy to trip).

Estimating the highest prices of the airline data for the route is collected with features such as Duration,Source,destination,arrival,department.features are taken from choosen dataset and this paper,we have used machine learning techniques and linear regression strategies for prediction of the price wherein the airline price ticket costs very overtime.Here we have different types of catagiores they are differ from their own price rates and facilities like food and confort.bussiness classes have more price sets and other classes have some what reasonable prices compared to bussiness classes.

So we have implented flight price prediction for users by using mechine learning model.The prime objective of our project"improvee flight price prediction system"is to make a prediction of the flight ticket fare for the future flights.

Contents

Chapter 1

Introduction

1.1 Introduction to Your Project

The aviation industry is highly dynamic, with airfares fluctuating rapidly due to various factors such as demand, fuel prices, competition, and seasonal variations. As a result, accurately predicting flight prices has become a challenging task for both travelers and airlines. However, advancements in data analysis and machine learning techniques have opened up new possibilities for making more precise flight price predictions.

Flight price prediction aims to forecast the future cost of air travel based on historical data and relevant factors influencing ticket prices. This predictive capability can empower travelers to make informed decisions by booking flights at the optimal time, ensuring cost savings and maximizing their travel budgets. For airlines, accurate price predictions can help with revenue management, demand forecasting, and optimizing ticket pricing strategies.

The outcome of this research will not only benefit individual travelers seeking affordable flights but also support airlines in optimizing their revenue management strategies and enhancing customer satisfaction. Moreover, this study contributes to the broader field of predictive analytics, showcasing the applicability of data-driven techniques in the aviation industry.

1.2 Application

1. Data Collection and Preprocessing: Gather relevant data such as historical flight prices, airline routes, departure/arrival times, travel durations, airline carrier information, customer reviews, and weather conditions. Preprocess the data by handling missing values, data normalization, feature engineering, and data cleaning.

2. Exploratory Data Analysis (EDA): Perform EDA to understand the dataset's characteristics, identify patterns, correlations, and outliers. Visualize the data using plots, histograms, scatter plots, and other graphical representations to gain insights into the relationships between variables and flight prices.

3. Feature Selection: Select the most relevant features that strongly influence flight prices. This can be done through statistical methods like correlation analysis or by employing feature selection techniques such as Recursive Feature Elimination (RFE), SelectKBest, or Principal Component Analysis (PCA).

4. Machine Learning Algorithms: Apply various machine learning algorithms to build predictive models. Some commonly used algorithms for flight price prediction include linear regression, decision trees, random forests, gradient boosting algorithms (such as XGBoost or LightGBM), support vector regression (SVR), and neural networks.

5. Model Training and Evaluation: Split the dataset into training and testing sets. Train the selected models on the training data and evaluate their performance using appropriate metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), or R-squared (R2) score. Perform cross-validation to assess the models' generalization capabilities.

1.3 Motivation Towards Your Project

1. Cost Savings for Travelers: Flight prices can vary significantly based on factors like booking timing, seasonality, and demand. Accurate price predictions empower travelers to make informed decisions and book flights at the most cost-effective times, potentially saving them a considerable amount of money.

2. Enhanced Planning and Budgeting: By having access to reliable flight price predictions, travelers can better plan their trips and allocate their budgets accordingly. They can

make informed decisions about the best time to travel, choose between different airlines and routes, and adjust their travel plans based on price fluctuations.

3. Revenue Management for Airlines: For airlines, accurate price prediction models are invaluable for revenue management. They allow airlines to optimize ticket pricing strategies, adjust fares based on demand patterns, and make data-driven decisions to maximize revenue.

4. Competitive Advantage: Airlines that can offer accurate price predictions gain a competitive edge in the market. By providing customers with transparency and reliable information about flight prices, airlines can build trust and loyalty, attracting more passengers.

5. Demand Forecasting: Flight price prediction models also aid in demand forecasting. By understanding the relationship between ticket prices and demand, airlines can anticipate future passenger volumes, optimize resource allocation, and plan their operations effectively.

6. Data-driven Decision Making: Flight price prediction projects require the use of data analysis and machine learning techniques, which foster a data-driven approach to decision making. This allows stakeholders to rely on objective insights and make evidence-based choices when it comes to pricing, marketing, and operational strategies.

7. Advancements in Machine Learning: Flight price prediction projects provide an opportunity to apply and explore various machine learning algorithms, techniques, and methodologies. Through such projects, advancements can be made in the field of machine learning and predictive analytics, contributing to the broader scientific and technological community.

1.4 Problem Statement

1. Optimal Booking Timing: The problem is to predict flight prices accurately, enabling travelers to determine the optimal time to book their flights to obtain the best possible prices. This involves considering factors such as seasonality, demand patterns, and market fluctuations.

2. Cost-effective Travel Planning: The problem is to provide travelers with reliable flight price predictions to assist in cost-effective travel planning. By considering variables

such as departure/arrival locations, travel dates, and flight durations, the goal is to help travelers make informed decisions and optimize their travel budgets.

3. Revenue Management for Airlines: The problem is to develop accurate flight price prediction models that aid in revenue management for airlines. This involves forecasting ticket prices based on historical data, demand patterns, and market dynamics, allowing airlines to optimize their pricing strategies and maximize revenue.

4. Demand Forecasting: The problem is to predict flight prices based on historical data and external factors, enabling airlines to forecast future demand accurately. By considering variables such as seasonality, route popularity, and travel trends, airlines can effectively allocate resources and plan their operations accordingly.

5. Competitive Pricing Strategy: The problem is to analyze historical flight price data and competitor pricing information to develop a competitive pricing strategy for airlines. This involves predicting flight prices in relation to competitor fares, market demand, and other relevant factors to gain a competitive advantage.

6. Customer Price Transparency: The problem is to provide customers with transparent and accurate flight price predictions, fostering trust and loyalty. By offering reliable price information, airlines can enhance customer satisfaction, leading to increased bookings and improved customer relationships.

7. Real-time Price Updates: The problem is to develop a system that provides real-time updates on flight prices based on current market conditions, allowing travelers to make informed decisions on short notice. This involves integrating real-time data sources and employing efficient prediction algorithms to deliver up-to-date.

Chapter 2

Approach To Your Project

2.1 Explain About Your Project

The project "Flight Price Prediction" aims to develop a system or model that can accurately forecast the prices of airline tickets. The goal is to assist travelers in making informed decisions about when to book their flights, potentially saving them money by identifying the best time to purchase tickets.

The project likely involves the following key steps and components:

1. Data Preprocessing: Once the data is collected, it needs to be cleaned and preprocessed to ensure its quality and suitability for analysis. This process involves removing any inconsistencies, handling missing values, normalizing numerical data, and encoding categorical variables, among other necessary tasks.

2. Feature Engineering: In this step, relevant features or variables that may impact flight prices are extracted or generated from the available data. This can involve creating new features such as day of the week, month, or season of travel, flight duration, layovers, and other factors that may influence ticket prices.

3. Model Selection: Various machine learning algorithms or statistical models can be employed to predict flight prices. Commonly used models include regression models (e.g., linear regression, random forest regression, or gradient boosting regression), time series models (e.g., ARIMA or SARIMA), or more advanced techniques like neural networks.

4. Model Training: The selected model is trained using the preprocessed data. The

dataset is typically split into training and validation sets, and the model is fitted to the training data. The model's parameters are adjusted to minimize the difference between the predicted flight prices and the actual prices in the training set.

5. Prediction and Deployment: After the model has been trained and evaluated, it can be used to predict flight prices for future or unseen data. The model can be deployed as a standalone application, integrated into a website or mobile app, or used in an API (Application Programming Interface)

2.2 Data Set

1. Airline: So this column will have all the type of airlines like indiogo,jet Airways,Air India,and many more.

2. Data of journey: This column will let us know about the data on which the passenger's journey will start.

3. Source: This column holds the name of the place from where the passenger's journey will start.

4. Destination: This column holds the name of the place to where passenger wanted to travel.

5. Route: Here we can know about that what is the route through which passengers have opted to travel from his/her source to their destination.

6. Arrival Time: Arriline time is when the passenger will reach his/her destination.

7. Duration: Duration is the whole period that a flight will take to complete its journey from source to destination.

8. Total stops: This will let us know in how many places flights will stop there for the flight in the whole journey.

9 .Price: Price of the flight for a complete journey including all the expenses before onboarding.

2.3 Prediction techniques

1. Regression Models: Regression techniques, such as linear regression, polynomial regression, or support vector regression, can be employed to predict flight prices. These models analyze historical flight data and try to find patterns and relationships between various factors (e.g., departure time, airline, route, etc.) and the corresponding prices.

2. Time Series Analysis: Flight prices often exhibit temporal patterns and trends. Time series analysis techniques, such as ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), or Prophet, can be used to capture and predict these patterns. These models consider factors like seasonality, trends, and historical price data to forecast future prices.

3. Machine Learning Algorithms: Various machine learning algorithms can be employed for flight price prediction, including decision trees, random forests, gradient boosting, or neural networks. These algorithms can handle complex relationships and nonlinearities in the data, allowing for more accurate predictions.

4. Ensemble Methods: Ensemble methods combine multiple prediction models to improve overall accuracy. Techniques like bagging (Bootstrap Aggregating) or boosting (e.g., AdaBoost or XGBoost) can be applied to flight price prediction models, where multiple models are trained on different subsets of the data or sequentially adjusted to make combined predictions.

5. Deep Learning Models: Deep learning models, such as recurrent neural networks (RNNs) or long short-term memory (LSTM) networks, can be used for flight price prediction. These models can capture sequential dependencies and handle large amounts of data, including text data from sources like airline reviews or news articles.

6. Hybrid Approaches: Combining multiple techniques, such as using regression models in conjunction with time series analysis or incorporating machine learning algorithms within a deep learning framework, can lead to more accurate predictions.

It's important to note that the choice of technique depends on the specific requirements of the project, the available data, and the desired level of accuracy. It's common to experiment with multiple techniques and compare their performance to select the most suitable model for flight price prediction.

2.4 Graphs

sns.barplot(x="days left",y="price",data=mydata)
plt.savefig('barplots.png')
plt.show()

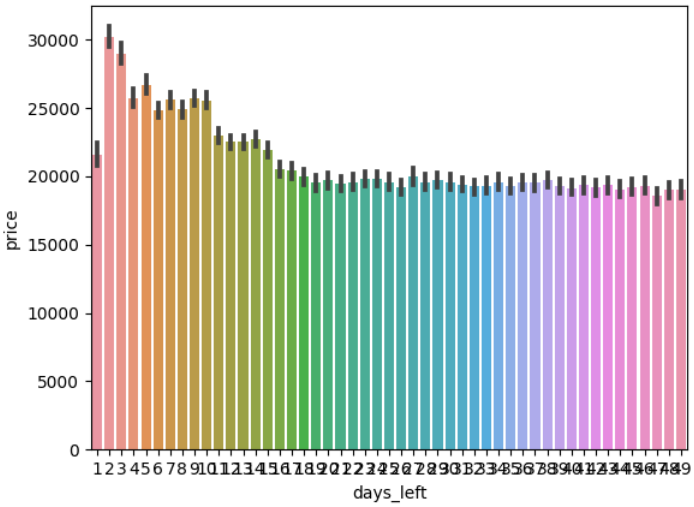

Figure 2.1: Boxplot

2.5 Visualization

1. Data visualization is the representaion of data through use of common graphs such as plots,histograms,scatters,infographics and even animations 2. TO visualization the data as graphs we have to import matplotlib.pyplot it is a visualization library in python is used to visualize the data and information 3. command:import matplotlib.pyplot as plt

```
sns.boxplot(x="days left",y="duration",data=mydata)
plt.savefig('boxplot4.png')
plt.show()
```

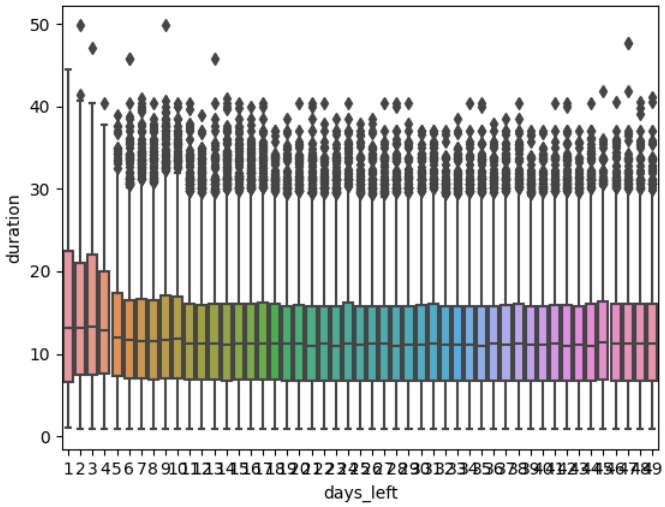

Figure 2.2: Boxplot

sns.set style(”ticks”)
sns.pairplot(mydata)
plt.savefig('pairplot.png')
plt.show()

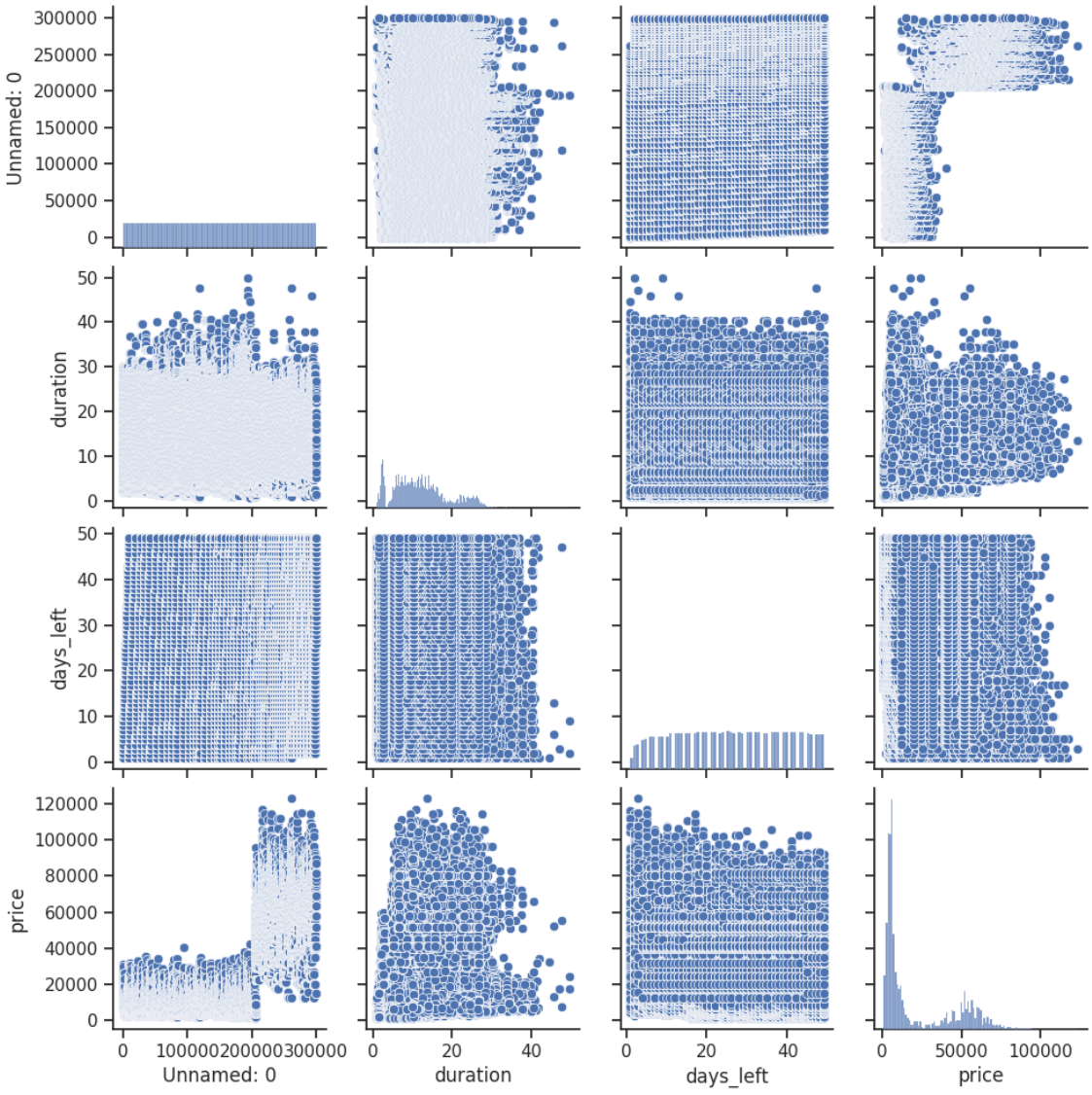

Figure 2.3: pairplot

```
sns.violinplot(x=mydata[”price”])
plt.savefig(’violinplot.png’)
plt.show()
```

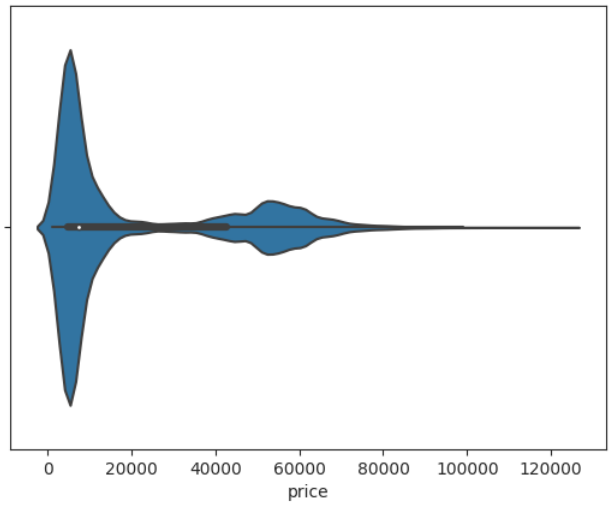

Figure 2.4: violinplot

```
sns.stripplot(x=’class’, y=’arrival time’ data=mydata)
plt.savefig(’stripplot4.png’)
plt.show()
```

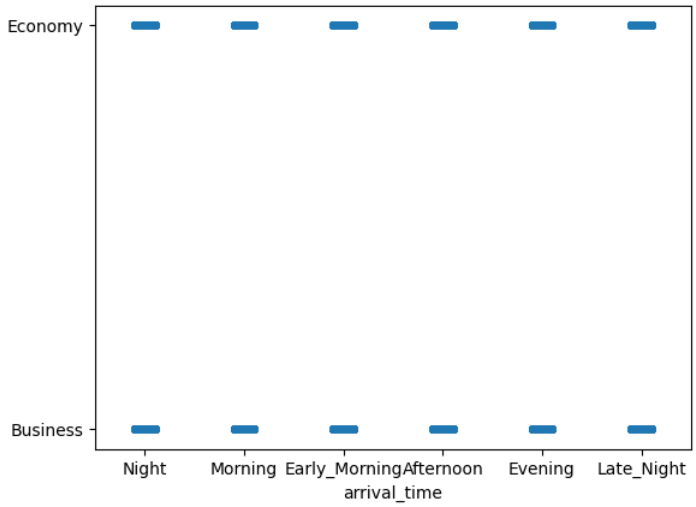

Figure 2.5: striplot1

```
sns.barplot(x='price' y='departure time' data=mydata)
plt.savefig('jv.png')
plt.show()
```

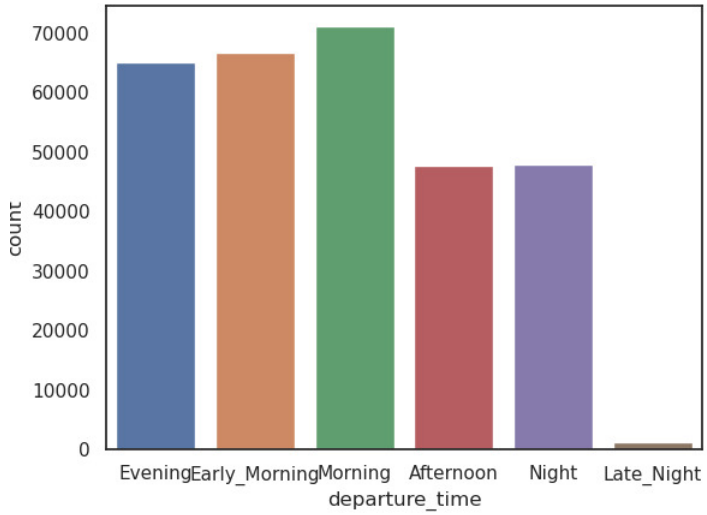

Figure 2.6: barplot

```
sns.boxplot(x="duration",y="class",data=mydata)
plt.savefig('boxplot3.png')
plt.show()
```

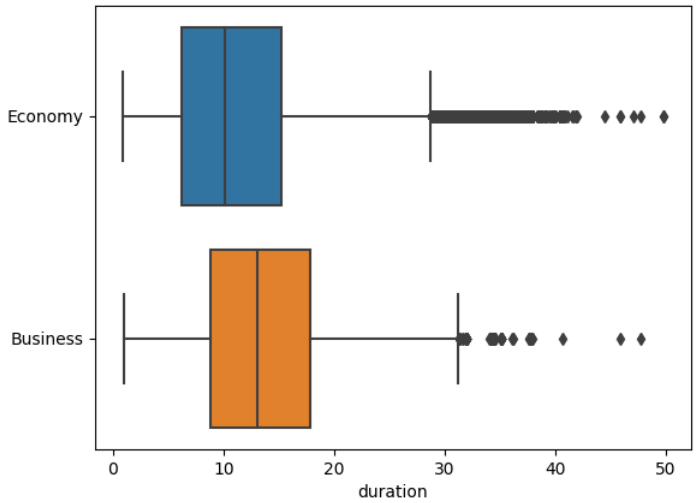

Figure 2.7: Visualization Example

```
sns.relplot(x='price',y='airline' data=mydata)
plt.savefig('grp1.png')
plt.show()
```

Figure 2.8: graph

```
sns.histogram(x='price',y='airline',data=mydata)
plt.savefig('Histogram.png')
plt.show()
```

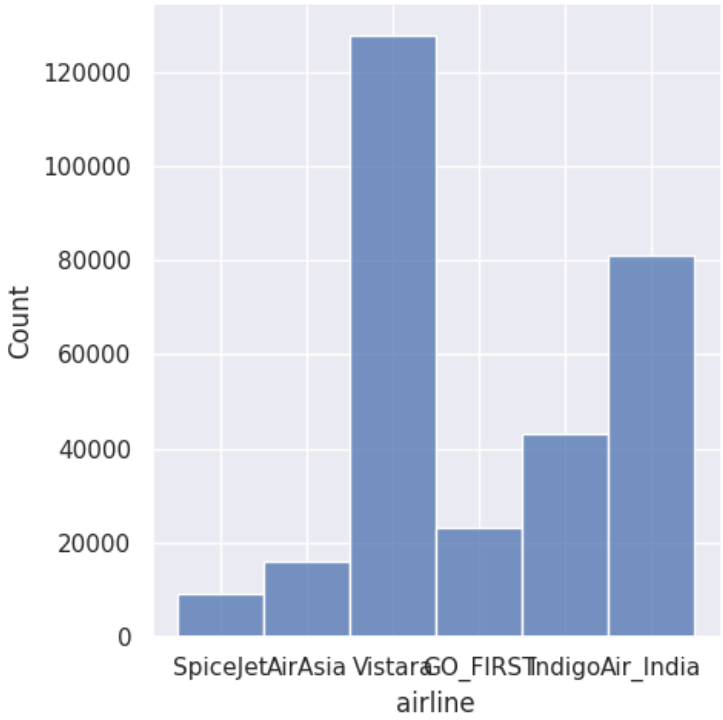

Figure 2.9: Histogram

```
sns.relplot(x='price',y='duration',data=df, hue='days left')
plt.savefig('grp.png')
plt.show()
```

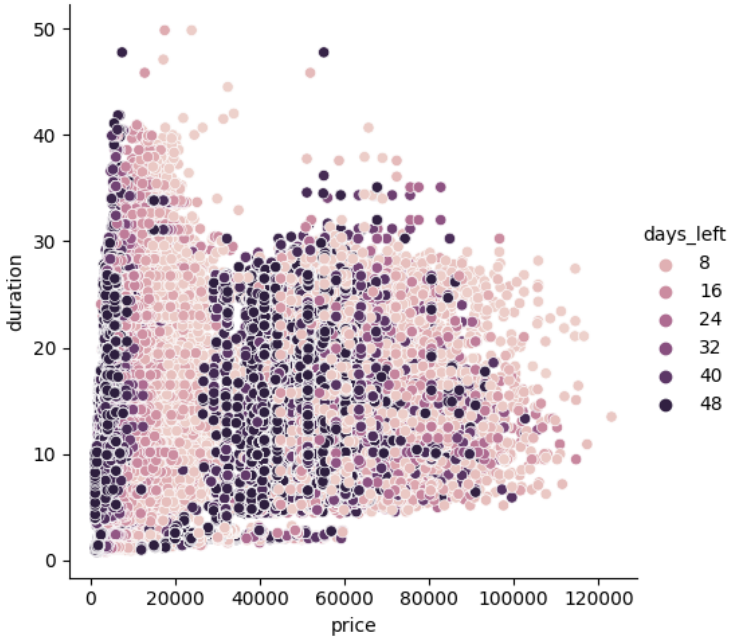

Figure 2.10: drpplot

```
sns.histogram(x="price",y="airline",data=mydata)
plt.savefig('Histogram.png')
plt.show()
```

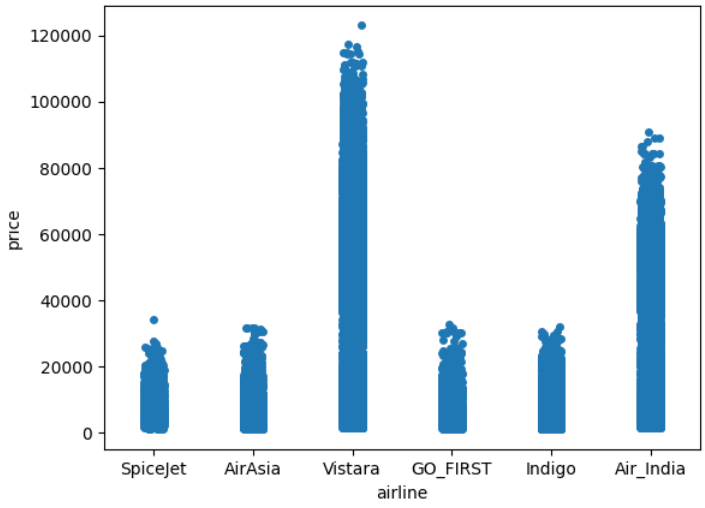

Figure 2.11: Histogram

Chapter 3

Code

3.1 Explain Your Code With Outputs

PANDAS • Pandas is free open source library written for python programming language for data manipulation and data analysis.

• Pandas provides functions to read data from various file formats, including CSV,Excel, etc...

• pandas has two data structures i.e series and dataframe.

• A Series is a one-dimensional labeled data structure that can hold any data type. It is similar to a column in a DataFrame and can be used to perform various operations, such as indexing, slicing, and applying mathematical operations.

• A DataFrame is a data structure that organizes data into two dimensional table of rows and columns.

• Pandas offers a wide range of functions to handle missing values, duplicate values and other data cleaning tasks. It provides methods like dropna(), fillna(), drop du- plicates(),etc.

• command for installation of pandas: pip install pandas.

• Syntax for pandas:import pandas.

• Here the operations of a pandas are defined as follows:

1.Import a csv file to jupyter notebook using pandas we are importing pandas for data manuplating numpy for calculations , pytplot for graphs , seaborn for effective plots

3.2 READING DATA SAMPLE

Now I want to see some rows and columns To see the first 5 rows, we use data.head(5) function

1 df.head()

2 df.info

3 df.discribe()


/home/rgukt/Downloads/Screenshot 2023-08-04 at 12-41-38 graps - Jupyter Notebook.p

Chapter 4

Conclusion and Future Work

"In this flight price prediction project, we aimed to develop a model that can accurately predict flight prices. We began by performing exploratory data analysis (EDA) on the 'Flight Price Data EDA' dataset. The dataset provided information on various variables such as departure city, arrival city, airline, flight duration, and others. We analyzed the data, calculated summary statistics, and visualized the distribution of flight prices.

In conclusion, this project successfully developed a flight price prediction model based on the 'Flight Price Data EDA' dataset. The model's predictions can assist airlines, travel agencies, and travelers in estimating flight prices and planning their travel budgets effectively. However, ongoing monitoring and updating of the model's performance are recommended to ensure its continued accuracy in a rapidly evolving market."

Throughout the project, data science techniques, including data cleaning, feature engineering, exploratory data analysis, machine learning modeling, and evaluation, are applied using Python libraries such as pandas, scikit-learn, TensorFlow, or PyTorch. Additionally, visualization libraries like Matplotlib or Seaborn can be used to visualize the data and model results.

By successfully completing the flight price prediction project, one can provide valuable insights to travelers, enable airlines to optimize their pricing strategies, and contribute to the field of data science by demonstrating the application of techniques in the aviation industry.