INT-353

EDA PROJECT

NAME: Nikhil Pathak

Reg. No.: 12102585

Section: K21UT

Roll No.: RK21UTB37



MY DATASET

INDIAN FLIGHT DATASET

Introduction:

Introduction to Aviation and Flight Data Analysis:

Aviation is a dynamic and complex industry that involves the design, development, production, operation, and maintenance of aircraft. Flight data analysis plays a crucial role in improving safety, efficiency, and overall performance in the aviation sector. Analyzing data generated during flight operations allows stakeholders, including airlines, manufacturers, and regulatory authorities, to gain valuable insights into various aspects of aviation.

Domain Knowledge:

1. Flight Data Monitoring (FDM): FDM involves the collection, analysis, and interpretation of data generated during flight operations. This includes parameters such as altitude, airspeed, heading, engine performance, and various other flight-related metrics. FDM is essential for identifying trends, anomalies, and potential safety issues.

2. Black Box Data Analysis: Aircraft are equipped with flight data recorders (FDR) and cockpit voice recorders (CVR), commonly known as "black boxes." These devices store critical data and communications from the aircraft during flight. Analyzing black box data is crucial for understanding the sequence of events in case of accidents or incidents.

3. Safety Management Systems (SMS): Safety is a top priority in aviation. SMS involves the systematic approach to managing safety, including data analysis to identify and mitigate risks. Analyzing safety-related data helps in developing proactive measures to enhance aviation safety.

4. Predictive Maintenance: Flight data analysis is instrumental in implementing predictive maintenance programs. By analyzing data from aircraft systems, engines, and components, operators can predict when maintenance is required, reducing unscheduled downtime and improving operational efficiency.

5. Fuel Efficiency Analysis: Airlines are constantly seeking ways to enhance fuel efficiency and reduce operational costs. Flight data analysis helps identify opportunities for optimizing routes, adjusting flight profiles, and implementing fuel-saving measures.

6. Regulatory Compliance: Aviation authorities impose strict regulations to ensure the safety and reliability of air travel. Flight data analysis is essential for compliance with these regulations, including performance monitoring, adherence to operational procedures, and validation of maintenance practices.

7. Aircraft Health Monitoring: Monitoring the health of aircraft systems in real-time through data analysis allows for early detection of potential issues. This proactive approach helps prevent in-flight failures and contributes to the overall reliability of the aircraft.

8. Operational Performance Optimization: Airlines use flight data analysis to optimize various aspects of their operations, including crew scheduling, route planning, and fleet management. This leads to improved efficiency and cost-effectiveness.

Data Understanding:

Data understanding in the context of aviation and flight involves acquiring, exploring, and comprehending the data collected from various sources within the aviation industry. Here are key aspects of data understanding in this domain:

1. Data Sources:

- Flight Data Recorders (FDR) and Cockpit Voice Recorders (CVR):\*\* These black box devices record parameters such as altitude, airspeed, heading, control inputs, and communication within the cockpit. Understanding the structure and format of data from these sources is crucial for accident investigation and performance monitoring.

- Aircraft Health Monitoring Systems: Modern aircraft are equipped with sensors that continuously monitor the health of various systems, including engines, avionics, and structural components. Understanding the data generated by these systems is essential for predictive maintenance and identifying potential issues.

- Operational Databases: Airlines maintain databases that store information on flight schedules, crew assignments, maintenance records, and other operational aspects. Analyzing this data provides insights into operational efficiency, crew performance, and overall airline management.

- Air Traffic Management Data: Data related to air traffic, including air traffic control communications, radar data, and airspace utilization, contributes to understanding the broader context of flight operations.

2. Data Preprocessing:

- Cleaning and Quality Assurance: Aviation data may have missing values, outliers, or inconsistencies. Preprocessing involves cleaning the data and ensuring its quality before analysis to avoid misleading insights.

- Normalization and Standardization: Different data sources may use varied units and scales. Normalizing and standardizing data help create a consistent framework for analysis.

3. Exploratory Data Analysis (EDA):

- Descriptive Statistics: Utilizing descriptive statistics helps in summarizing key features of the data, such as mean, median, standard deviation, and percentiles.

- Data Visualization: Creating visual representations, such as scatter plots, histograms, and time series plots, aids in understanding patterns, trends, and anomalies in the data.

5. Domain-Specific Analysis:

- Safety Metrics:\*\* Define and analyze safety-related metrics, such as the rate of descent, vertical acceleration, and proximity to safety thresholds, to assess and enhance flight safety.

- Fuel Efficiency Indicators: Develop metrics related to fuel consumption, flight efficiency, and emissions to optimize fuel usage.

- Operational Performance Metrics: Evaluate crew performance, on-time performance, and other operational metrics to enhance overall efficiency.

6. Data Integration:

- Integrating Diverse Data Sources: Combining data from different sources, such as flight operations, maintenance records, and crew schedules, enables a comprehensive understanding of the aviation ecosystem.

-Cross-Domain Analysis: Explore relationships between different aspects of aviation data to identify potential correlations and dependencies.

Understanding the nuances of aviation data is crucial for extracting meaningful insights that can inform decision-making, enhance safety, and optimize operational efficiency within the aviation industry.

Reasons:

One of the prominent reason was I wanted to choose some unique Industry and along with that my interest should lie, with the relevance in today’s world.

With the current relevance and growth in the Aviation Industry its new domains of knowledge, along with that the amount of the source available on the internet and also it is a multi-dimensional data set

This data set has a Real-World Impact, insights derived from flight data analytics can lead to tangible real-world improvements. Whether it's optimizing flight routes to reduce fuel consumption, predicting maintenance needs to enhance aircraft reliability, or improving on-time performance, our project holds the potential to positively influence the daily experiences of travellers and aviation personnel.

The realm of flight data analytics is closely intertwined with emerging technologies like predictive modelling, anomaly detection, and real- time data processing. Our project provides a platform to explore and leverage these cutting-edge tools, enhancing our proficiency in leveraging technology for impactful insights.

Lastly, I would say that potential for innovation, as technology and data analysis for innovation within the aviation industry emerge. By diving into flight data analytics, we position ourselves to identify uncharted opportunities for improvement and innovation, contributing to the continuous evolution of the aviation landscape.

Questions for Analysis:

Question 1: Which part of the year is the busiest?

**Insights:**

With the bar graph we conclude that June is the busiest part of the year is May, probably because of the vacation period

**Question 2: Which part of the week is the busiest?**

**Insights:**

With the bar graph we conclude that Weekends are the busiest part of the week

**Question 3 : Calculating average flight time for a journey**

**Insights:**

367 minutes

**Questions 4: at what time of the day the travelers, love to travel?**

**Insights:**

# Extracting whether its a morning,evening,night or afternoon flight from departure time & arrival time of the flight

# (Evening)

**Question 5: how many times a flight stops in a journey?**

**Insights:**

Air India has more stops than any other company

**Question 6: In Vacation months(March-June), when are the price highest?**

**Insights:**

May is the highest

**Question 8: In Vacation months (March-June), when are the price at lowest point?**

**Insights:**

April is lowerst

# Question 10 : How are the price trends changing?

# Insights: We see that the average fare price is highest the month of March , this can be because people usually book the fights 2-3 months prior to their date of journey which leads to higher demand and hence higher fare prices.(since the count of tickets booked for travelling in May is the most, prior bookings are done in March)

**Question 11 :** In which city, price are maximum

Insights : The airfare price range in Delhi & New Delhi is the maximum, this can be due to: Jet fuel prices in Delhi had increased in the year of 2018 by 26.4%, it is also the National Capital, political seat of power and a highly visited place for vacations(same for Bangalore & cochin) The same reasoning can be given for higher price range in Delhi as the source of the flight

**Question 12 : Which Airline company is preferred the most?**

**Insights** : Jet airways is the most preferred one

**Question 13: Which company has the Highest price?**

**Insights** : Jet Airways

Question 16: Which part of the year is the busiest?

**Insights:**

With the bar graph we conclude that June is the busiest part of the year is May, probably because of the vacation period

**Question 17: Which part of the week is the busiest?**

**Insights:**

With the bar graph we conclude that Weekends are the busiest part of the week

**Question 18 : Calculating average flight time for a journey**

**Insights:**

367 minutes

**Question 19:** When people prefer to travel in day or night?

**Insights:** People prefer both evening and morning, but evening a bit more

**Question 20:** which day of the week is costliest to travel?

# Insights: Analysis : We see that prices are highest on a friday, that may be because this is when airlines start raisin g prices to make up for the lower prices before and also because weekdays are off for most people so they are mostly travelling to or travelling back

**Question 21:** Which day of the month is the most travelled day ?

**Insights:**

Analysis : It looks like that there's a trend in the air fare when compared to the day of respective months, prices are higher in the start of month but this is not a trend if you see from the broader perspective as this might be due to various reasons. For eg. the date of Journey is 10th March and people are booking towards 5th March or so, this will lead to higher flight prices.(Prices increase as near you date of booking is to the date of journey). So flight prices don't follow any particular pattern towards any time of the month

Libraries used and approaches:

Libraries Used:

Pandas: Pandas is a fundamental library for data manipulation and analysis. It provides powerful data structures like DataFrames and Series, making it easy to work with structured data. Pandas allows us to load, clean, transform, and analyze the dataset efficiently.

Matplotlib: Matplotlib is a widely-used library for creating static, animated, and interactive visualizations in Python. It offers various types of plots, including line plots, bar charts, scatter plots, histograms, and more. Matplotlib is crucial for presenting the results of our analysis visually.

Seaborn: Seaborn is built on top of Matplotlib and provides a higher-level interface for creating aesthetically pleasing statistical graphics. It offers specialized functions for creating complex visualizations like heatmaps, violin plots, and pair plots. Seaborn simplifies the process of creating informative and visually appealing plots.

NumPy: NumPy is a fundamental library for numerical computing in Python. It provides support for working with arrays and matrices, which are essential for mathematical operations, statistical analysis, and data manipulation. NumPy is often used in conjunction with Pandas to perform numerical calculations.

Jupyter Notebook: Jupyter Notebook is an interactive coding environment that allows us to write, execute, and document our Python code in a notebook-style format. It's particularly useful for exploratory data analysis (EDA), as it allows us to combine code, visualizations, and explanations in a single document. Jupyter Notebook facilitates the creation of data science reports and presentations.

Approach to Solve Problems:

1) Data Loading and Inspection:

* Loading the dataset containing MBA program rankings and related attributes using Pandas.
* Inspecting the dataset's structure, including columns, data types, and initial data samples.

2) Data Cleaning and Preprocessing:

* Handling missing values by imputing or removing them based on the context.
* Converting data types to ensure they are suitable for analysis (e.g., converting string representations of numbers to numeric data types).
* Addressing data inconsistencies and outliers that could affect analysis results.

3) Exploratory Data Analysis (EDA):

* Performing EDA to understand the dataset's characteristics and uncover initial insights.
* Generating summary statistics, including measures of central tendency and dispersion.
* Creating visualizations (e.g., histograms, box plots, scatter plots) to visualize data distributions and relationships between variables.

4) Data Visualization:

* Utilizing Matplotlib and Seaborn to create a wide range of visualizations tailored to the specific questions being addressed.
* Visualizing trends, patterns, and correlations in the data.
* Customizing plots with labels, titles, legends, and other elements for clarity.

5) Answering Questions and Problem-Solving:

* Addressing each question or problem statement one by one, writing code to perform the necessary calculations and visualizations.
* Applying domain knowledge and analytical techniques to gain insights.
* Interpreting the results of analyses and visualizations to draw meaningful conclusions.
* Presenting the findings in a clear and organized manner, often using Jupyter Notebook to create comprehensive reports.
* Providing explanations and context for the observed trends and relationships.

6) Iterative Process:

* Data analysis is often iterative. It may involve revisiting previous steps, refining analysis, or exploring additional questions as new insights are uncovered.

This is the outlined approach to address each question or problem systematically and leverage the capabilities of the selected Python libraries to conduct a comprehensive analysis of the MBA program rankings dataset.

Steps of EDA:

Exploratory Data Analysis (EDA) is a critical phase in data analysis that helps us understand the dataset, identify patterns, and gain insights. In the project involving MBA program rankings, the following steps of EDA are used in detail:

1) Importing libraries:

The very first step of EDA was importing Python libraries that are used in MBA program rankings analysis project:

* Data handling: Pandas, NumPy
* Visualization: Matplotlib, Seaborn

2) Data Loading:

* Loading the dataset into a Pandas DataFrame.
* Ensuring that the data is read correctly by checking the first few rows and the column names.
* Understanding the size of the dataset (number of rows and columns).

3) Data Cleaning:

Checking the data types of each column using (“df.dtypes”)

Displaying summary statistics of the DataFrame using (“df.describe()”)

Identifying and handling missing data:

Check for missing values in each column.

Deciding whether to impute missing values, remove rows with missing values, or handle them differently based on context.

* I used Mode fillna to fill missing values of “Rank of 2022” and “Audit year \*”
* I used Median fillna to fill missing vales of “Rank of 2021” and “Three-year average rank”

Checking for duplicate rows and removing them if necessary.

In my dataset there were no duplicate rows to remove them. So, this step doesn’t apply to my dataset

Ensure data consistency and uniformity:

Converting data types to appropriate formats (e.g., converting strings to numeric types).

             Standardizing data (e.g., removing unnecessary characters, whitespace).

* To convert salary strings to numeric data we use

# Ensuring that 'Salary today (US$)' columns are numeric

             df['Salary today (US$)'] = pd.to\_numeric(df['Salary today (US$) \*\*'].str.replace(',', '',

             regex=True), errors='coerce')

4) Data Exploration:

             Generating summary statistics:

* Calculating descriptive statistics such as mean, median, mode, count etc. to understand the data's central tendencies and spread.

Visualizing data distributions:

* Creating histograms,  pie charts to visualize the distribution of numerical variables.
* Using bar plots or count plots for categorical variables.

Exploring relationships:

* Constructing scatter plots, line charts or correlation matrices to identify associations between variables.
* Computing correlation coefficients to quantify relationships.

Handling categorical data:

* Examining the frequency distribution of categorical variables using bar plots.
* Exploring relationships between categorical variables using Heatmaps.

5) Data Visualization:

         Creating a variety of plots and charts to visualize data:

* Line plots, scatter plots, and bar plots to show trends and relationships.
* Box plots and violin plots to depict data distributions and identify outliers.
* Heatmaps to visualize correlations between variables.

         Customize visualizations with titles, labels, legends, and color palettes to enhance

          clarity.

6) Summary and Insights:

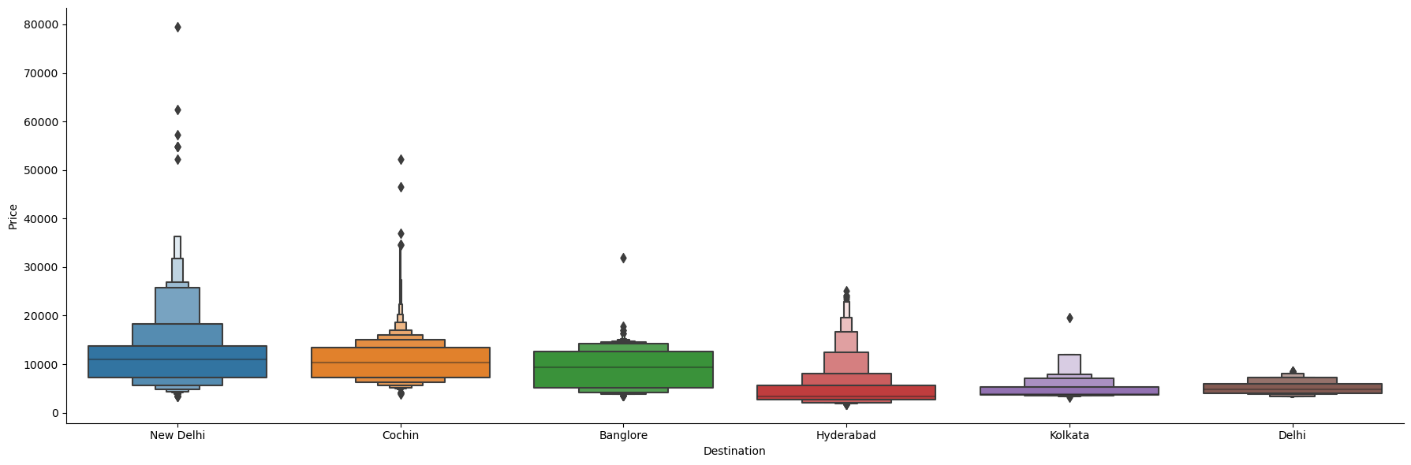
* Summarizing key findings, trends, and insights from the EDA process.
* Providing context and explanations for observed patterns.
* Formulating preliminary conclusions for further analysis.

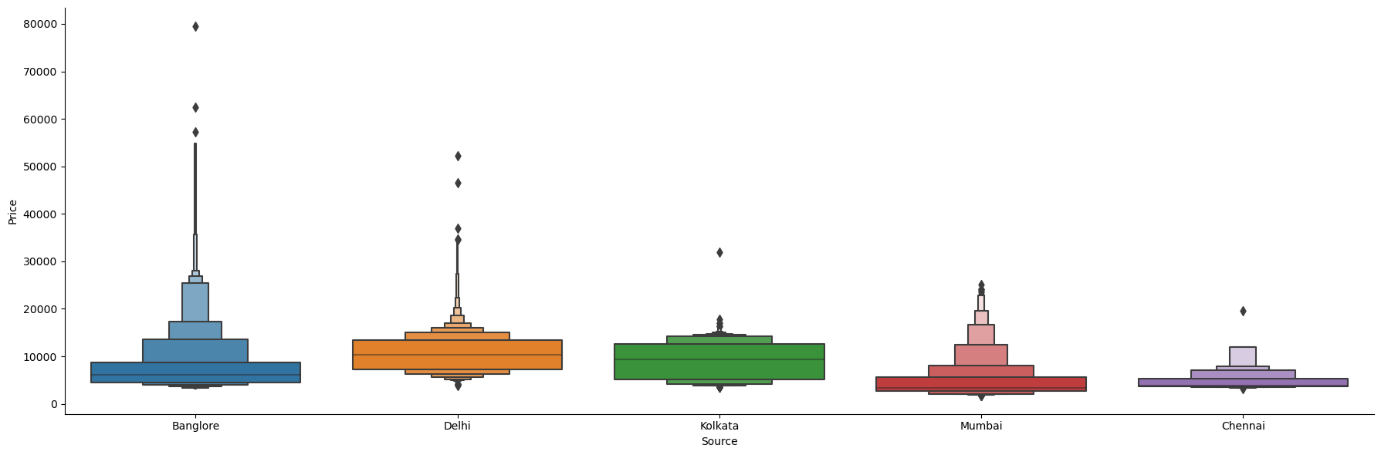
7) Documentation:

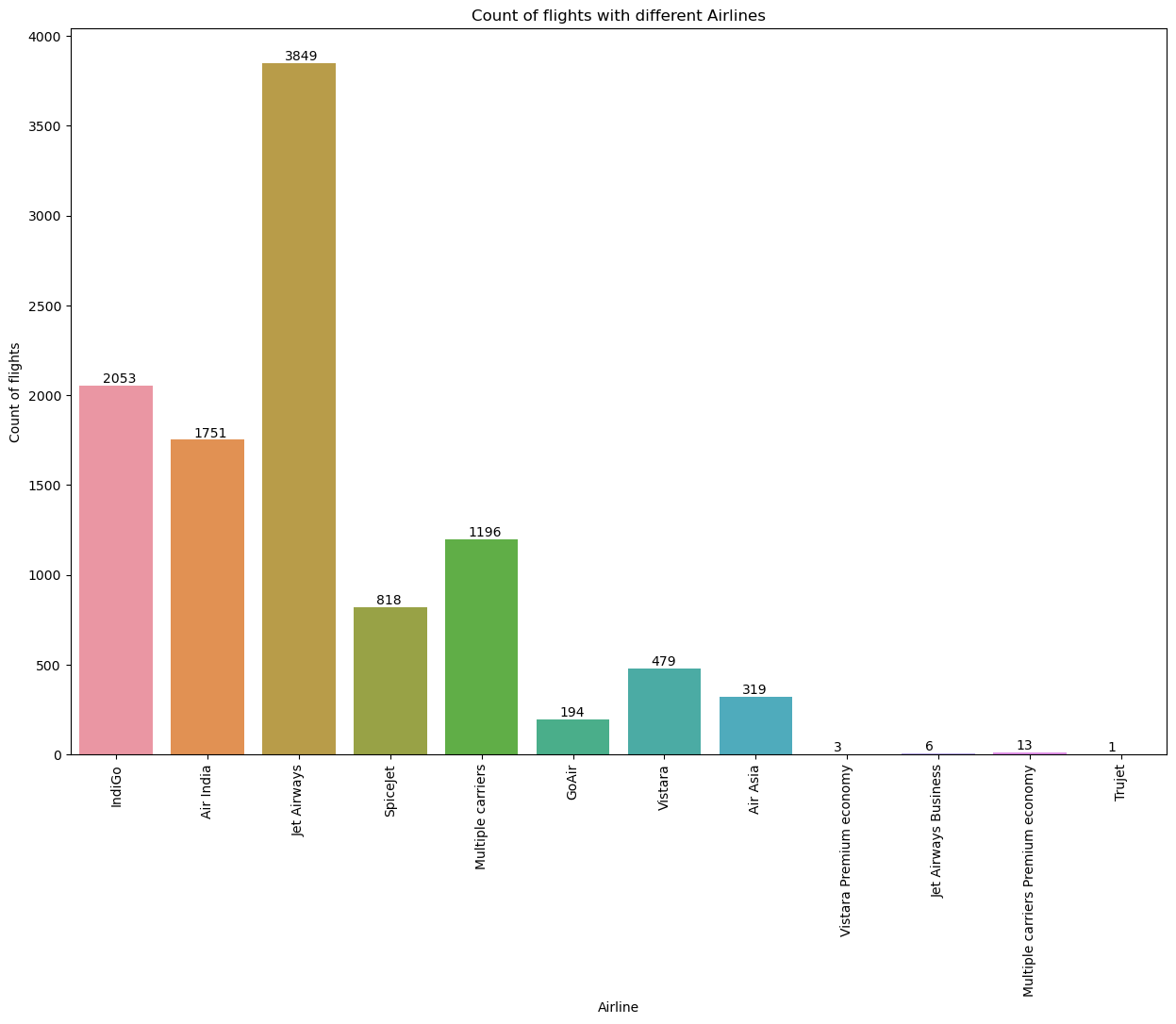
* Documenting the EDA process, including code, visualizations, and insights, in a clear and organized manner.
* Using Jupyter Notebook to create a comprehensive report that can be easily shared and referenced.

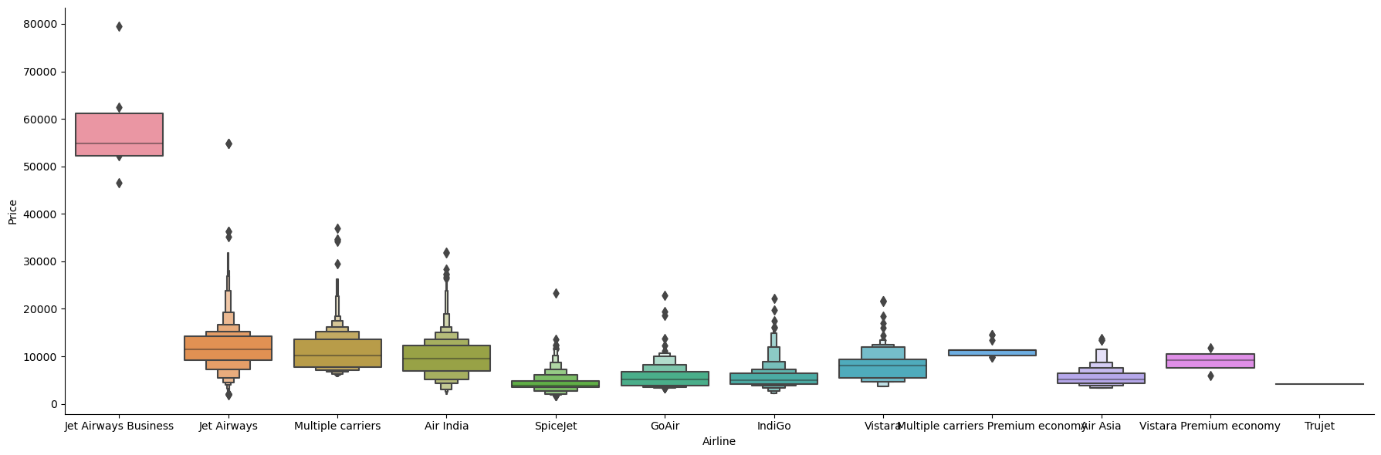
The EDA process in this project is systematic and thorough, involving a combination of statistical analysis and data visualization techniques to uncover meaningful insights from the MBA program rankings dataset.

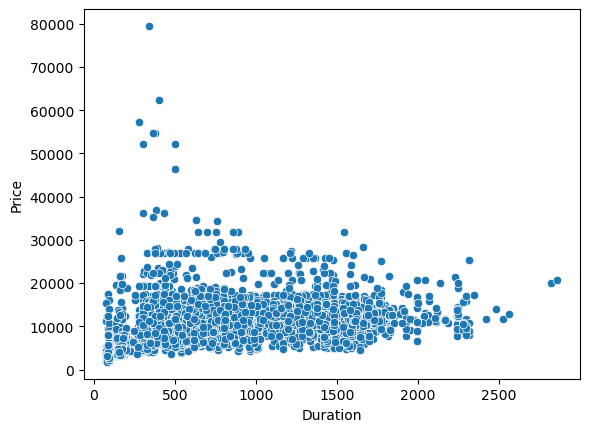
Visualization:

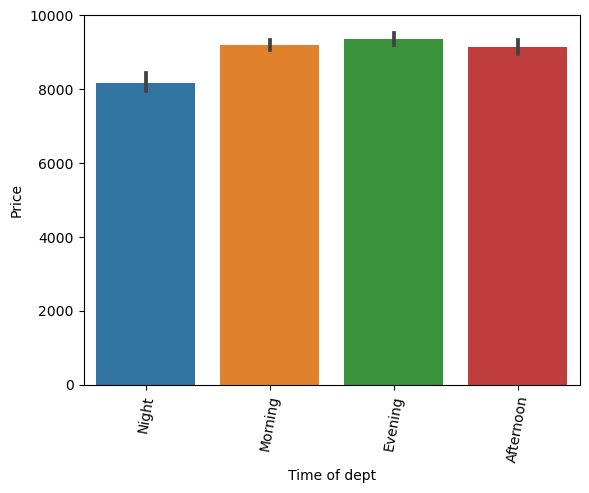


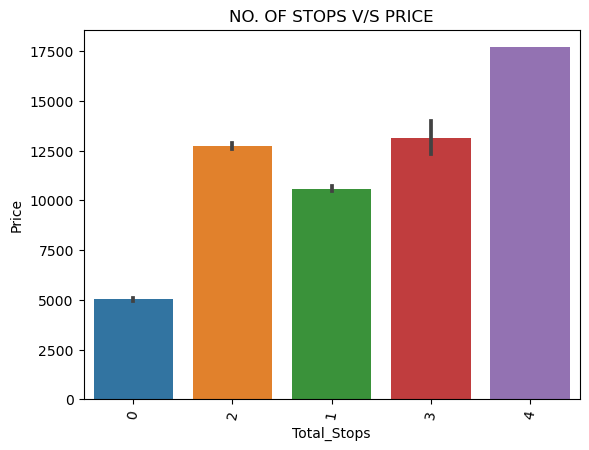


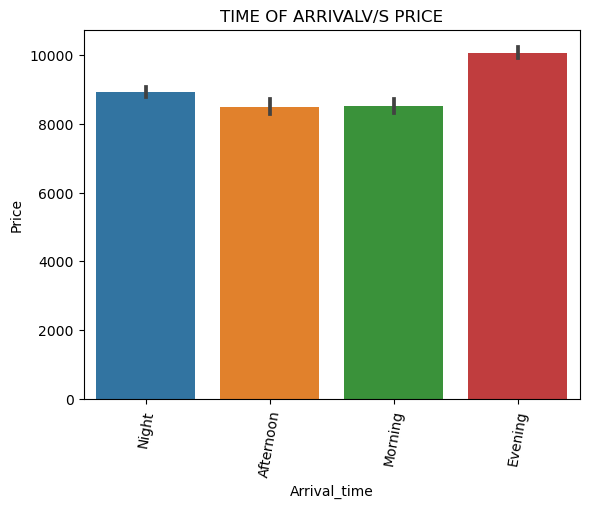


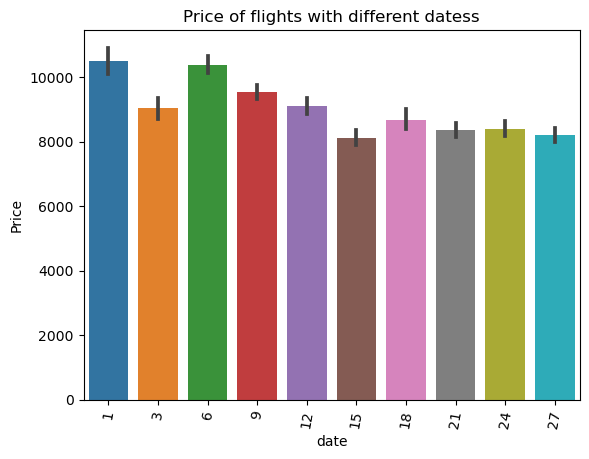


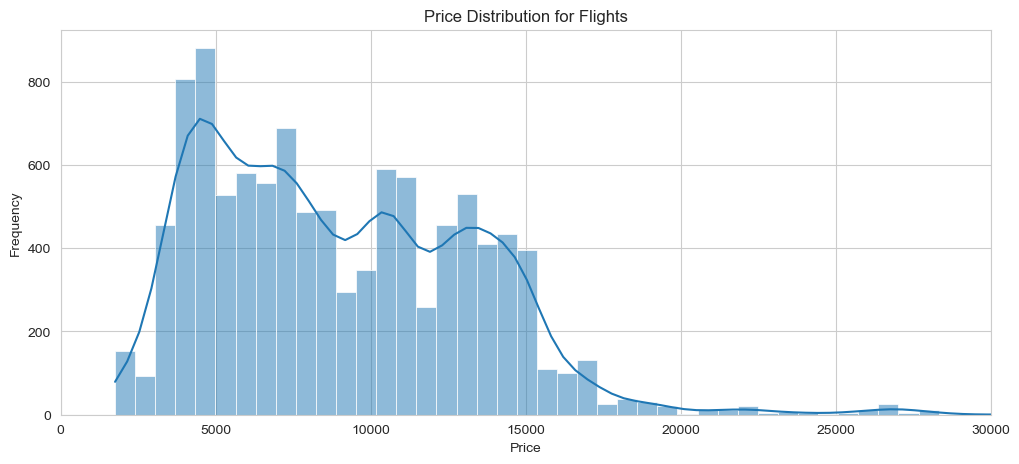


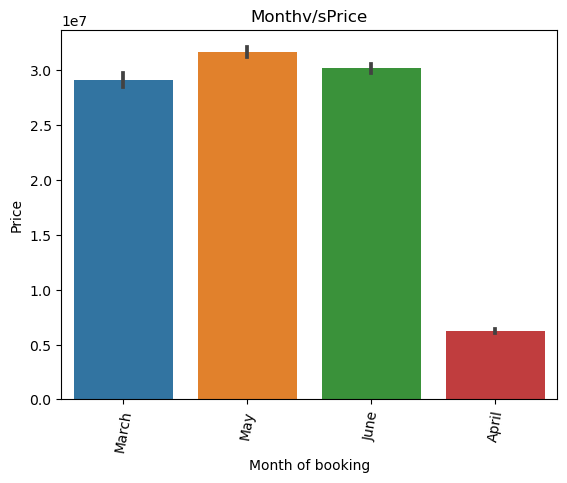


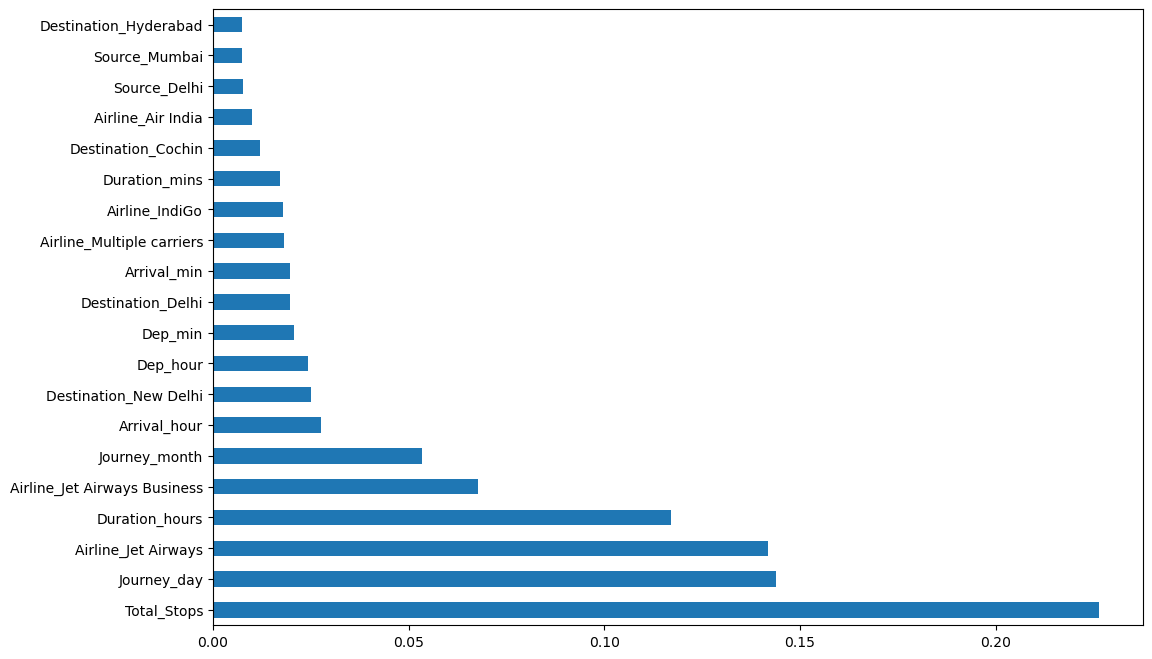


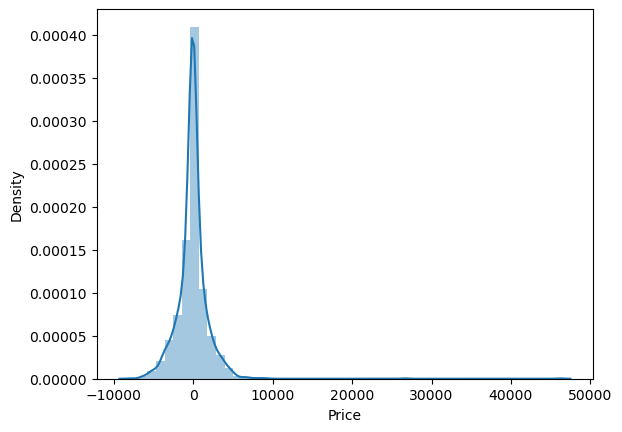












Univariate Analysis:

Univariate analysis explores each variable in the dataset, separately.

In this we use appropriate visualizations (histograms, box plots, etc.) to explore their distributions.

Bivariate Analysis:

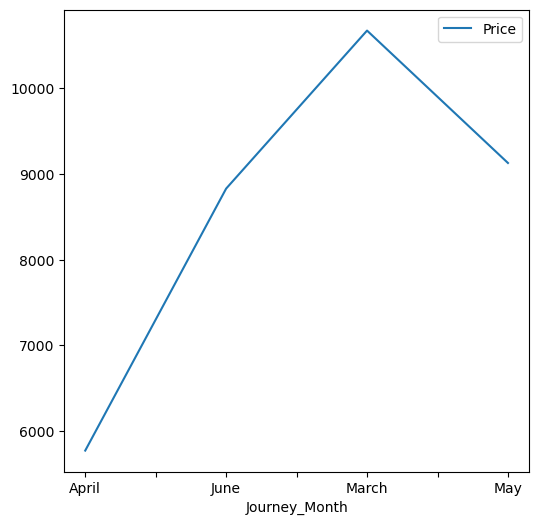
Bivariate analysis explores two different variables in the dataset, at the same time.

In this we use scatterplots, correlation matrices, and other visualizations to uncover associations.

* Firstly, let’s see the correlations between the career progress rank and alumni network

Distributions:

Let’s look at, probability density function (PDF) of Female Faculty (%):



Hypothesis Testing:

T-Test: The t-test is used to compare the mean. In this we perform T-test on “Salary today” section. And the results of the test are as follows:-

Mean salary: 220376.6

t-statistic: 1.6248497153861678

p-value: 0.1795194563593551

Fail to reject null hypothesis

Insights and Findings:

Price

• The price of a flight is influenced by a variety of factors, including the airline, the time of day, the day of the week, the time of year, and the route.

• In general, non-stop flights are more expensive than connecting flights.

• Flights that depart during peak hours (e.g., in the morning and evening) are more expensive than flights that depart during off-peak hours (e.g., overnight).

• Flights that depart during the holiday season are more expensive than flights that depart during the off-season.

• Flights on popular routes are more expensive than flights on less popular routes.

Time

• The flight time is influenced by the route and the type of aircraft.

• Non-stop flights are generally shorter than connecting flights.

• Flights that depart during peak hours (e.g., in the morning and evening) may take longer due to congestion.

• Flights that depart during the holiday season may take longer due to increased air traffic.

• Flights on popular routes may take longer due to congestion.

Days

• The day of the week that you fly can also affect the price and flight time.

• Flights that depart on weekends and holidays are generally more expensive than flights that depart on weekdays.

• Flights that depart on weekends and holidays may also take longer due to increased air traffic.

Route

• The route that you fly can affect the price, flight time, and number of airlines that offer flights.

• Popular routes have more flights and are more competitive, which can lead to lower prices.

• Shorter routes are generally cheaper and shorter in flight time than longer routes.

• Less popular routes may have fewer flights and higher prices.

How to use this information to your advantage

• If you are flexible with your travel dates and times, you can often find cheaper flights.

• If you are able to fly on a less popular route, you may also be able to find cheaper flights.

• If you are looking for the shortest flight time, you may want to consider a non-stop flight.

• If you are flying during peak hours or during the holiday season, you may want to book your flight well in advance to avoid higher prices and longer flight times.

Limitations:

Here are some of the limitations I faced during performing EDA:

Limited or missing data: The dataset might have missing or incomplete information for some schools, which can limit the scope of my analysis.

Data quality issues: Inaccuracies, inconsistencies, or errors in the dataset can affect the reliability of my analysis.

Data bias: The dataset might be biased toward specific regions, types of business schools, or criteria used for ranking, which could skew my analysis.

Privacy concerns: The dataset may contain sensitive or private information about individuals associated with the business schools. Ensuring compliance with data privacy regulations is essential.

Ethical considerations: Be mindful of potential ethical issues when working with data, especially if the dataset contains information that could harm the reputation of schools or individuals.

Rankings can be subjective: Rankings of business schools are often based on criteria that are subjective or proprietary. Understanding the ranking methodology and its limitations is crucial for a meaningful analysis.

Limited Variables:

The dataset may not include all the variables of interest. For a comprehensive analysis, need to supplement it with external data sources.

Time Sensitivity:

The dataset may not be up-to-date, and rankings can change over time. It's essential to consider the relevance of the data to the present day.

Recommendations:

here are some general recommendations that I might consider:

In-Depth Analysis of High-Impact Factors: Identify the factors that strongly correlate with the ranking or performance of business schools. Explore these factors in more detail to understand their impact. This may involve conducting regression analyses, hypothesis testing, or using machine learning models to predict rankings.

Benchmarking and Peer Comparisons: While comparing the top 100 business schools with their peers or with schools that are close to the ranking threshold. This can help identify areas where schools can improve and potentially move up in the rankings.

Identify Outliers: While examining schools that deviate significantly from the expected patterns. These outliers may provide insights into what distinguishes high-performing schools or schools that have room for improvement.

Qualitative Research: Supplement quantitative analysis with qualitative research, such as interviews or surveys with school administrators, students, or employers who hire graduates from these schools. This can provide insights into the qualitative aspects that contribute to a   school's reputation and ranking.

Conclusion:

Flight data analysis is a valuable tool for understanding and improving air travel safety, efficiency, and profitability. By analyzing data from flight recorders, airlines can identify areas where they can improve their operations, such as reducing delays, fuel consumption, and maintenance costs. They can also use this data to develop new products and services that better meet the needs of their customers.

Here are some specific conclusions that can be drawn from flight data analysis:

* Safety: Flight data analysis can be used to identify and correct potential safety hazards. For example, airlines can use flight data to track the performance of aircraft components and systems, and to identify pilots who may need additional training.
* Efficiency: Flight data analysis can be used to improve the efficiency of flight operations. For example, airlines can use flight data to optimize flight routes, reduce taxi times, and improve fuel efficiency.
* Profitability: Flight data analysis can be used to improve the profitability of airlines. For example, airlines can use flight data to price tickets more accurately and to identify areas where they can reduce costs.

Overall, flight data analysis is a powerful tool that can be used to improve air travel in a variety of ways. By understanding and analyzing flight data, airlines can make their operations safer, more efficient, and more profitable.

In addition to the above, flight data analysis can also be used to:

* Develop new air traffic control procedures
* Improve aircraft maintenance programs
* Train pilots more effectively
* Design new aircraft and avionics systems
* Investigate accidents and incidents

Flight data analysis is an essential part of the aviation industry, and it plays a vital role in ensuring the safety and efficiency of air travel.

Reference:

* Kaggle
* Python Libraries: Common libraries for data analysis in Python include Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn.
* Jupyter Notebooks: Jupyter notebooks are widely used for interactive data analysis and reporting.

Project Code:

import warnings

warnings.filterwarnings('ignore')

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

data = pd.read\_excel('Data\_Train.xlsx')

data.head()

data.describe()

data.info()

data["Duration"].value\_counts()

print(data.isnull().sum())

data=data.dropna()

print(data.isnull().sum())

data['Journey\_Day'] = pd.to\_datetime(data.Date\_of\_Journey, format='%d/%m/%Y').dt.day

data['Journey\_Month'] = pd.to\_datetime(data.Date\_of\_Journey, format='%d/%m/%Y').dt.month

data['weekday']= pd.to\_datetime(data.Date\_of\_Journey, format='%d/%m/%Y').dt.weekday

data.drop(labels = 'Date\_of\_Journey', axis = 1, inplace = True)

data.head()

def duration(test):

test = test.strip()

total=test.split(' ')

to=total[0]

hrs=(int)(to[:-1])\*60

if((len(total))==2):

mint=(int)(total[1][:-1])

hrs=hrs+mint

test=str(hrs)

return test

data['Duration']=data['Duration'].apply(duration)

data['Duration'].nunique()

def deparrtime(x):

x=x.strip()

tt=(int)(x.split(':')[0])

if(tt>=16 and tt<21):

x='Evening'

elif(tt>=21 or tt<5):

x='Night'

elif(tt>=5 and tt<11):

x='Morning'

elif(tt>=11 and tt<16):

x='Afternoon'

return x

data['Dep\_Time']=data['Dep\_Time'].apply(deparrtime)

data['Arrival\_Time']=data['Arrival\_Time'].apply(deparrtime)

def stops(x):

if(x=='non-stop'):

x=str(0)

else:

x.strip()

stps=x.split(' ')[0]

x=stps

return x

data['Total\_Stops']=data['Total\_Stops'].apply(stops)

data=data.drop(['Route'], axis=1)

data["Duration"] = data["Duration"].astype(int)

data["Journey\_Day"] = data["Journey\_Day"].astype(object)

data["Journey\_Month"] = data["Journey\_Month"].astype(object)

data["weekday"] = data["weekday"].astype(object)

df1["Journey\_Month"]=df1["Journey\_Month"].replace({1: "Jan", 3:"March",4:"April",5:"May",6:"June"})

df1["Journey\_Month"]=df1["Journey\_Month"].astype(object)

#Journey month v/s total fare

v1=sns.barplot(x='Journey\_Month', y='Price', data=df1,estimator=sum)

v1.set\_title('Monthv/sPrice')

v1.set\_ylabel('Price')

v1.set\_xlabel('Month of booking')

v1.set\_xticklabels(v1.get\_xticklabels(), rotation=80)

#Journey month v/s Averagefare

monthly\_avg.plot(x='Journey\_Month',y='Price',figsize=(6,6))

# Destination vs AveragePrice

sns.catplot(y='Price',x='Destination',data= df1.sort\_values('Price',ascending=False),kind="boxen",height=6, aspect=3)

plt.show

# Source vs AveragePrice

sns.catplot(y='Price',x='Source',data= data.sort\_values('Price',ascending=False),kind="boxen",height=6, aspect=3)

plt.show

#Count of flights v/s Airline

plt.figure(figsize = (15, 10))

plt.title('Count of flights with different Airlines')

ax=sns.countplot(x = 'Airline', data =data)

plt.xlabel('Airline')

plt.ylabel('Count of flights')

plt.xticks(rotation = 90)

for p in ax.patches:

ax.annotate(int(p.get\_height()), (p.get\_x()+0.25, p.get\_height()+1), va='bottom',

color= 'black')

# Airline vs AveragePrice

sns.catplot(y='Price',x='Airline',data= data.sort\_values('Price',ascending=False),kind="boxen",height=6, aspect=3)

plt.show

#duration v/s AveragePrice

sns.scatterplot(data=data, x='Duration', y='Price')

#Deptarure time v/s AveragePrice

v2=sns.barplot(x='Dep\_Time', y='Price', data=data)

v2.set\_ylabel('Price')

v2.set\_xlabel('Time of dept')

v2.set\_xticklabels(v2.get\_xticklabels(), rotation=80)

#TIME OF ARRIVAL V/S average price

v3=sns.barplot(x='Arrival\_Time', y='Price', data=data)

v3.set\_title('TIME OF ARRIVALV/S PRICE')

v3.set\_ylabel('Price')

v3.set\_xlabel('Arrival\_time')

v3.set\_xticklabels(v3.get\_xticklabels(), rotation=80)

#total stops v/s average price

v4=sns.barplot(x='Total\_Stops', y='Price', data=data)

v4.set\_title('NO. OF STOPS V/S PRICE')

v4.set\_ylabel('Price')

v4.set\_xlabel('Total\_Stops')

v4.set\_xticklabels(v4.get\_xticklabels(), rotation=80)

#WEEKDAY V/S average price

v4=sns.barplot(x='weekday', y='Price', data=data)

v4.set\_title('WEEKDAY V/S PRICE')

v4.set\_ylabel('Price')

v4.set\_xlabel('WEEKDAY')

v4.set\_xticklabels(v4.get\_xticklabels(), rotation=80)

#Journey\_Day v/s Average price

v5=sns.barplot(x='Journey\_Day', y='Price', data=data)

v5.set\_title('Price of flights with different datess')

v5.set\_ylabel('Price')

v5.set\_xlabel('date')

v5.set\_xticklabels(v5.get\_xticklabels(), rotation=80)

plt.figure(figsize=(12, 5))

sns.set\_style("whitegrid")

sns.histplot(price\_data, kde=True)

plt.xlim(0, 30000)

plt.xlabel('Price')

plt.ylabel('Frequency')

plt.title('Price Distribution for Flights')

import scipy.stats as stats

\_ = sns.jointplot(x='Duration',y='Price',data=data,kind = 'reg')

\_.annotate(stats.pearsonr)

plt.show()

Dataset link:

<https://www.kaggle.com/datasets/salikhussaini49/flight-data>

Presentation link:

<https://docs.google.com/presentation/d/1-K1aEoU5pZ1FFG408iYtSP-9sM_gzwZT/edit?usp=sharing&ouid=116749743125678079226&rtpof=true&sd=true>

Project link:

<https://colab.research.google.com/drive/1YgQNTLZXcqKwtZ3G9J2O1cLOmAJN7cax>