**Airline Data Analysis**

Synopsis submitted.

to

**(college name) in the partial fulfilment of the requirement for**

**the degree of (stream)**

student name

college logo

**Supervisor:**

faculty name

(Associate Professor)

Department of Information Technology

**College name**

address

Batch ()

**College logo**

**College name**

**adresss**

**Certificate**

This is to certify that Mr./Ms Student name a student of Bachelor of Information Technology (BSc IT) under the Department of Information Technology at the college name will work on the topic entitled **“Airline data Analysis”** under my supervision and guidance at Department of Information Technology, college name for the partial fulfilment of requirement for the degree of BSc IT (Session).

**Date** **................ Supervisor**

**Declaration by Student**

I hereby declare that the synopsis of proposed project work entitled “**Airline data Analysis**” in fulfilment of the requirements for the award of degree of Bachelor of Information Technology to College name, is an authentic project work will be carried out by me under the expert guidance as well as direct supervision of faculty name (supervisor). I further declare that the material obtained from other sources has been duly acknowledged in the synopsis.

Signature..........................

Name:

Enrollment No.:

Department of IT

College name

**Title of the proposed project work:** Airline Data Analysis

**Name of the Student**

**Name and designation of guide:**

**Place of work**:

**Students Signature Guide Signature**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Topic** | **Page No.** |
| **1.** | Certificate | 2 |
| **2.** | Declaration | 3 |
| **3.** | Abstract | 6 |
| **4.** | Introduction | 6 |
| **5.** | Objective | 7 |
| **6.** | About the dataset | 7 |
| **7.** | Methodology | 8 |
| **8.** | Indian airline ticket price analysis | 10 |
| **9.** | Future Scope | 11 |
| **10.** | Conclusion | 11 |
| **11.** | Requirement Hardware & Software | 12 |
| **12.** | E-R Diagram | 13 |
| **13.** | Bibliography | 14 |

**Table Content**

**Abstract**

The airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers. Consequently, this has ushered in substantial economic growth, both nationally and internationally. The airline industry produces vast amounts of data, capturing a diverse set of information about their operations, including data related to passengers, freight, flights, and much more.

**Introduction**

The airline industry is complex, with multiple facets to manage and improve upon to satisfy passengers.

Data driven Decision-making has become increasingly essential in the airline industry to help airlines better understand their customers’ needs and expectations. The airline industry generates an enormous amount of data from various sources, such as customer data, social media, and customer feedback, which can be used to improve their overall experience.

Through the analysis of customer data, airlines can gain insight into customer behaviors, preferences, and patterns. This information can help airlines understand what customers need and expect, allowing them to tailor their services to better meet these expectations. For instance, airlines can use data to determine which routes are most popular among customers, which amenities are preferred, and which inflight entertainment options are most enjoyed.

Social media platforms have also become valuable sources of customer data for airlines. Airlines can monitor social media platforms to gain an understanding of how customers perceive their brand and services. Social media provides airlines with real-time feedback on customer experiences and opinions, allowing them to respond promptly to issues and improve customer satisfaction. In addition to providing real-time feedback, social media can also help airlines build better customer relationships.

**Objective**

As an aviation enthusiast, I always look forward to flying whenever I plan to travel. Since the Covid-19 pandemic ended, aviation market began to recover slowly from rock bottom ticket prices in peak of pandemic. With the war in Ukraine and rise in ATF prices, ticket prices have been through the roof reaching new high. As a result of this, I decided to do an exploratory data analysis for ticket prices to better understand the factors affecting them in India. I aim to answer questions like No. of flights available across India, Ticket availability across

I was able to work on my project.

**About the dataset**

Data is gathered from the Kaggle and considered as secondary data. The dataset includes details on the ticket booking alternatives available through the website "Easemytrip" for flights between India's top 6 metro areas. The cleaned dataset contains 11 characteristics and 300261 datapoints. Data was gathered in two stages: for business class tickets and for economy class tickets. The site provided a total of 300261 unique flight booking possibilities. 50 days of data were gathered, from February Kaggle.

The various features of the cleaned dataset are explained below:

1. **Airline:** The airline column contains the name of the airline firm. There are six different airlines, making it a category trait.
2. **Flight:** The flight code of the aircraft is stored in flight.
3. **Source City:** City where the flight departs from is a classification feature with 6 distinctive cities.
4. **Departure Time:** This is a categorical feature that was deduced from time periods being divided into bins. It has six different time labels and stores information about the departure time.
5. **Stops:** A category feature that holds the number of stops between the source and destination cities and has 3 different values.
6. **Arrival Time:** This derived categorical feature was developed by binning time intervals. It maintains information regarding the arrival time and has six different time labels.
7. **Destination City:** The location of the aircraft's landing. It is a classification feature with 6 distinctive cities.
8. **Class:** A permanent feature that shows the total number of hours needed to travel between cities.
9. **Duration:** A permanent feature that shows the total number of hours needed to travel between cities.
10. **Days Left:** The trip date is subtracted from the booking date to arrive at this derived feature.
11. **Price:** Information about the ticket price is stored in the target variable.

**Methodology**

The methodology to complete this project is as follows:

1. I explored jupyter notebook, concepts of python and PowerBI.
2. For further and a deeper understanding, I even referred to some articles, books and websites .

Below are the important concepts on which the work has been done and with the support of these I was able to work on my project.

**Jupyter notebook-** *The*Jupyter Notebook is a server-client application that allows editing and running notebook document via a web browser. The Jupyter Notebook App can be executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet.

In addition to displaying/editing/running notebook documents, the Jupyter Notebook App has a “Dashboard” (Notebook Dashboard), a “control panel” showing local files and allowing to open notebook documents or shutting down their Kernels.

**PowerBi-** Power BI is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and BI reports. Power BI suite provides multiple software, connector, and services - Power BI desktop, Power BI service based on Saas, and mobile Power BI apps available for different platforms. These set of services are used by business users to consume data and build BI reports.

Power BI desktop app is used to create reports, while Power BI Services (Software as a Service - SaaS) is used to publish the reports, and Power BI mobile app is used to view the reports and dashboards.

Power BI Desktop is available in both 32-bit and 64-bit versions.

**Python-**Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is available on a wide variety of platforms including Linux and Mac OS X.

1. **Numpy-**NumPy stands for numeric python which is a python package for the computation and processing of the multidimensional and single dimensional array elements.

With the revolution of data science, data analysis libraries like NumPy, SciPy, Pandas, etc. have seen a lot of growth. With a much easier syntax than other programming languages, python is the first choice language for the data scientist.

NumPy provides a convenient and efficient way to handle the vast amount of data. NumPy is also very convenient with Matrix multiplication and data reshaping. NumPy is fast which makes it reasonable to work with a large set of data.

There are the following advantages of using NumPy for data analysis.

1. NumPy performs array-oriented computing.
2. It efficiently implements the multidimensional arrays.
3. It performs scientific computations.
4. It is capable of performing Fourier Transform and reshaping the data stored in multidimensional arrays.

**Data visualization**

1. **Pandas-** Python Pandas is defined as an open-source library that provides high-performance data manipulation in Python.

Data analysis requires lots of processing, such as **restructuring, cleaning** or **merging**, etc. There are different tools are available for fast data. processing, such as **Numpy, Scipy, Cython**, and **Panda**. But we prefer Pandas because working with Pandas is fast, simple and more expressive than other tools.

1. **Seaborn-** Seaborn is one of an amazing library for visualization of the graphical statistical plotting in Python. Seaborn provides many color palettes and defaults beautiful styles to make the creation of many statistical plots in Python more attractive.

Seaborn library aims to make a more attractive visualization of the central part of understanding and exploring data. It is built on the core of the matplotlib library and also provides dataset-oriented APIs.

Seaborn is also closely integrated with the Panda's data structures, and with this, we can easily jump between the various different visual representations for a given variable to better understand the provided dataset.

1. **Matplotlib -** **Matplotlib**is a Python library which is defined as a multi-platform data visualization library built on Numpy array. It can be used in python scripts, shell, web application, and other graphical user interface toolkit.

Matplotlib is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc. It supports a very wide variety of graphs and plots namely - histogram, bar charts, power spectra, error charts etc. It is used along with NumPy to provide an environment that is an effective open source alternative for MatLab. It can also be used with graphics toolkits like PyQt and wxPython.

**Indian Airline Ticket Price Analysis**

* Defining the problem statement

**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

**Future Scope**

Data Science technologies such as machine learning and AI with help airlines gain valuable insights on fuel-burn, weather, navigation, and operations data to deliver valuable thereby helping optimize fuel utilization & reduce operational costs.

One of the key issues airlines must resolve to thrive is how to price flights and determine traveller demand for particular city pairs. Carries must analyse data while considering thousands of parameters to do this.

**Conclusion**

1. 'Air Asia' offers the cheapest flight tickets while flying Economy class while 'Air India' is cheapest while flying Business class.

2. Booking tickets 3-7 weeks before travel will be cheaper than buying them within 3 weeks of travel as prices rise rapidly in 2-20 days period. Tickets can be cheap when bought just 1 day before however they might not be as cheap as when bought more than 3 weeks before

3. Ticket price grow linearly with duration of flight peaking when duration of flight reaches 20 hours. However due to some outliers they again fall for flights with duration of more than 20 hours. Relation can be approximated by 2nd degree curve

4. Flight departing late at night and arriving early morning or late at night are cheapest.

5. Flight prices increase with increase in number of stops.

6. Delhi offers the cheapest flights while Hyderabad is most expensive city to fly to

**Hardware And Software Requirement**

**The hardware required for the development of the project is**

* Keyboard- Any Standard
* Monitor- Any Color Monitor
* Operating system- Windows
* RAM required – 2GB (minimum)
* Processor- Single core or Dual Core .
* Disk space – 20 GB

**Local Area Network Preferable Software Requirements of the software required for the development of the project is:**

* Jupyter notebook or Pycharm
* PowerBI
* Python libraries

**ER DIAGRAM**

**Business Class**

**Indian Airline**

economy

**BIBLIOGRAPHY**

**Google for problem solving**

* [**https://www.geeksforgeeks.org/data-visualization-with-python/**](https://www.geeksforgeeks.org/data-visualization-with-python/)
* [**https://gilberttanner.com/blog/introduction-to-data-visualization-inpython/**](https://gilberttanner.com/blog/introduction-to-data-visualization-inpython/)
* [**https://www.javatpoint.com/packages-of-data-visualization-in-python**](https://www.javatpoint.com/packages-of-data-visualization-in-python)
* [**https://www.datacamp.com/tutorial/power-bi-dashboard-tutorial**](https://www.datacamp.com/tutorial/power-bi-dashboard-tutorial)