

CS5346 2023-24 OTOT Tasks C

Zhao Shanhe*

March 28th, 2024

1 Introduction

<i>Data Name</i>	<i>Description</i>	<i>Data Type</i>
Date_of_journey	The date on which the flight is scheduled to depart is stored in this column	Temporal
Journey_day	The day of the week on which the flight is scheduled to depart which is derived from the date of the journey.	Temporal
Airline	The name of the airline company is stored in the airline column.	Norminal
Flight_code	Flight stores information regarding the plane's flight code.	Norminal
Class	A categorical feature that contains the information of flight seat class	Discrete
Source	City from which the flight takes off	Norminal
Departure	This is a derived categorical feature obtained created by grouping time periods into bins	Temporal
Total_stops	The number of stops between the source and destination cities	Textual
Arrival	This is a derived categorical feature created by grouping time intervals into bins	Temporal

*Student ID: A0287084U

<i>Data Name</i>	<i>Description</i>	<i>Data Type</i>
Destination	The city where the flight will land	Continuous
Duration_in_hours	A continuous feature that displays the overall amount of time it takes to travel between cities in hours	Interval
Days_left	This is a derived characteristic that is calculated by subtracting the date of the journey from the booking date	Interval
Fare	This is a target variable that stores the price of the flight ticket in the (INR), including all taxes	Continuous

Data source: <https://www.kaggle.com/datasets/yashdharme36/airfare-ml-predicting-flight-fares>

This task is done through ***Python***.

The purpose of this visualization task is to find out the relationship between the price of a plane and the destination of the flight, the distance flown, the duration of the flight, etc.

Task on Github: <https://github.com/Edward-EH-Holmes/CS5346-2023-24-Own-Time-Own-Target-OTOT-Tasks/tree/main/Task%20C>

2 Visualization

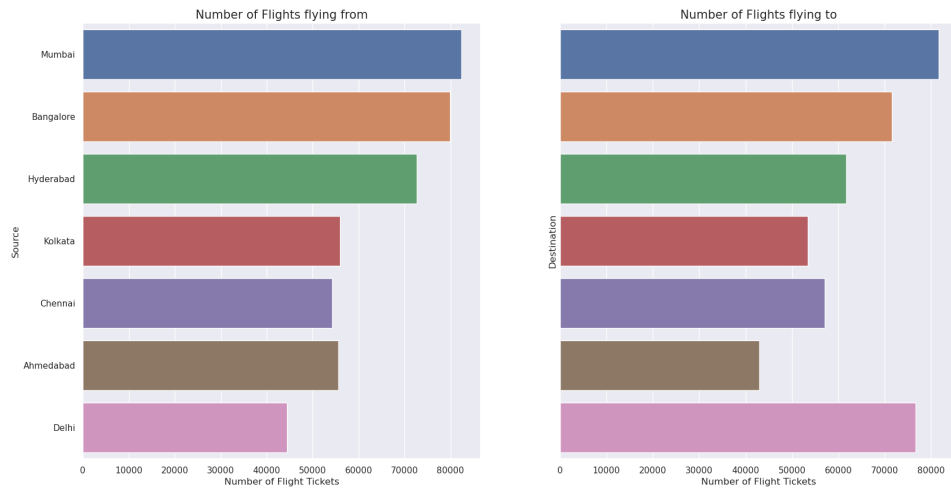


Figure 1: Number of tickets for different sources and different destinations (Bar chart)

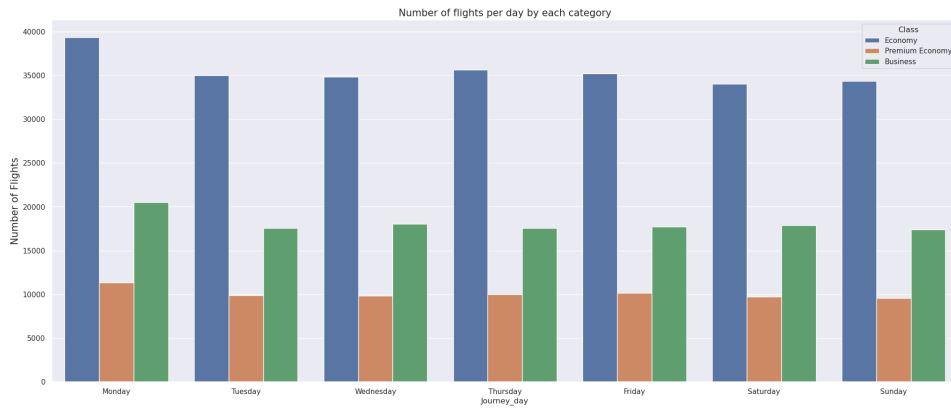


Figure 2: Number of flights per day by each category (Bar chart)

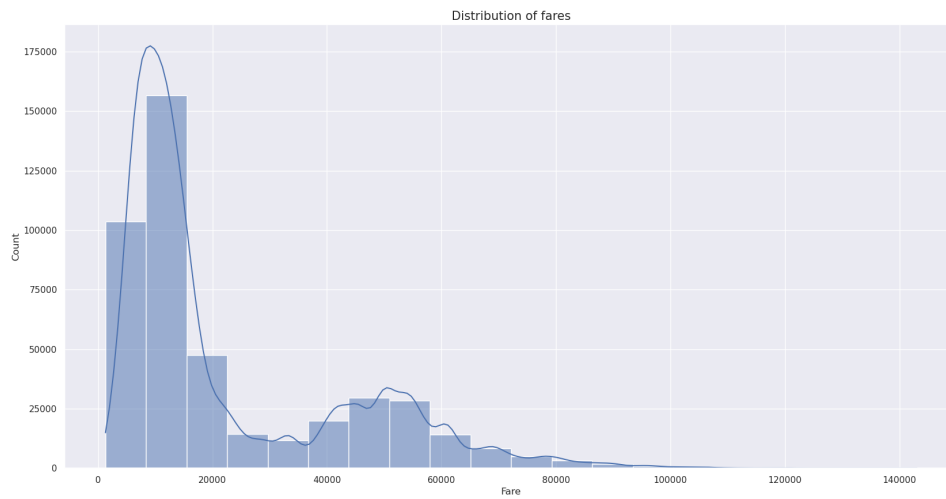


Figure 3: Distribution of fares (Line-Bar chart)

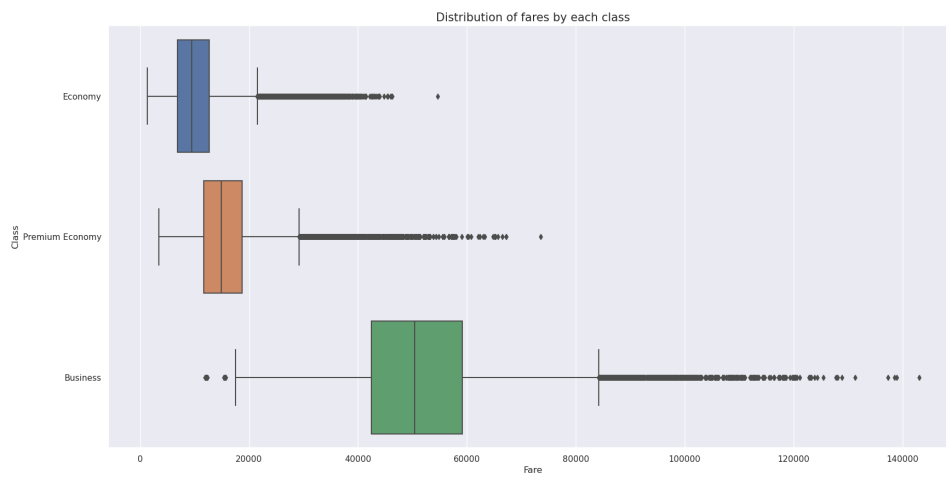


Figure 4: Distribution of fares by each class (Box plot)

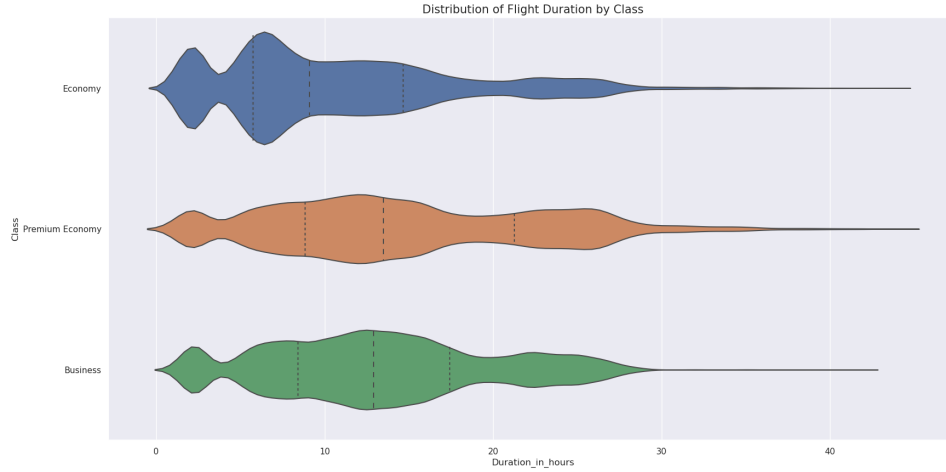


Figure 5: Distribution of Flight Duration by Class (Violin plot)

3 Conclusion

From this, we can draw the following conclusions:

- From Figure 1, the dataset includes 7 cities of India. Mumbai has the highest number of flight tickets, both to and from the city. It makes sense that Mumbai, India's most populous city and one of the world's top 10 business centers, is the most populous. Delhi, on the other hand, is the main destination for flight tickets. As the ancient capital of India, tourism is probably strong here.
- From Figure 2, number of Flights and the ticket prices are not effected by the day. In other words, they remain same on weekdays as on weekends.
- From Figure 3, ticket prices are mostly concentrated in the range of 0 20,000 Indian rupees. This may be related to sources, destinations, the distance flown, the airline, the month, etc.
- From Figure 4, generally speaking, $Business\ Tickets\ Price > Premium\ Economy\ Price > Economy\ Price$
- From Figure 5, generally speaking, the longer the Flight Duration, the more expensive the ticket. The Economy Class fare fluctuated less, followed by the Business Class.