**7COM1079-0901-2024**

**Team Research and Development Project**

**Final report title: Analyzing the Correlation Between Flight Duration and Airfare for Domestic Flights in India**

**Group ID: A147**

**Dataset number: DS115 Flight\_data.csv**

**Prepared by: Harun Shaikh ()**

**Umar Shakeeb (23091095)**

**Mohammed Ashraf (23080299)**

**Aman (23076396)**

**Aashiq ()**

**University of Hertfordshire**

**Hatfield, 2024**

**Table of Contents**

[1. Introduction 3](#_Toc186831978)

[1.1. Problem Statement and Research Motivation 3](#_Toc186831979)

[1.2. The Dataset 3](#_Toc186831980)

[1.3. Research Question 3](#_Toc186831981)

[1.4. Null Hypothesis and Alternative Hypothesis 3](#_Toc186831982)

[2. Background Research 4](#_Toc186831983)

[2.1. Research Papers 4](#_Toc186831984)

[2.2. Research Gap and Interest 4](#_Toc186831985)

[3. Visualisation 4](#_Toc186831986)

[3.1. Scatter Plot Explanation 4](#_Toc186831987)

[3.2. Additional Insights 4](#_Toc186831988)

[3.3. Key Observations 4](#_Toc186831989)

[4. Analysis 4](#_Toc186831990)

[4.1. Statistical Test 4](#_Toc186831991)

[4.2. Interpretation of Results 4](#_Toc186831992)

[5. Evaluation – Group’s Experience 4](#_Toc186831993)

[5.1. What Went Well 4](#_Toc186831994)

[5.2. Points for Improvement 4](#_Toc186831995)

[5.3. Group’s Time Management 4](#_Toc186831996)

[5.4. Project’s Overall Judgement 4](#_Toc186831997)

[6. Conclusions 4](#_Toc186831998)

[6.1. Results Explained 4](#_Toc186831999)

[6.2. Interpretation of Results 4](#_Toc186832000)

[6.3. Future Work and Limitations 4](#_Toc186832001)

[7. Reference List 4](#_Toc186832002)

[8. Appendices 4](#_Toc186832003)

# 1. Introduction

## 1.1. Problem Statement and Research Motivation

The airline industry has seen fluctuating pricing designs, particularly in domestic markets where rivalry is savage. Figuring out factors that impact airfare pricing, like flight duration, can give important experiences to the two customers and organizations. Price transparency and predictability are basic for passengers, while airlines look to improve revenue management techniques. The relationship between flight duration and price has been generally bantered in literature, with some research areas of strength for demonstrating between these factors (Williams, 2022). The current study centers around the relationship between's flight duration and price for domestic flights in India.

## 1.2. The Dataset

The dataset utilized for this examination, "DS115 Flight\_data.csv," contains 389 rows and remembers information for domestic flights in India. Key variables in this dataset incorporate flight duration (independent variable) and price (dependent variable). The dataset likewise gives data on airlines and flight subtleties. It offers a reasonable sample size to look at possible relationships between flight duration and price, considering significant statistical examination of domestic airfare pricing patterns in India.

## 1.3. Research Question

The essential research question (RQ) of this study is: "Is there a correlation among price and duration for domestic flights in India?"

To answer this RQ, statistical methods, including Spearman's Rho correlation test, are used to assess the strength and significance of the relationship between flight duration and price

## 1.4. Null Hypothesis and Alternative Hypothesis

* Null Hypothesis (H0): There is no significant correlation between flight duration and price. In other words, flight duration does not influence the ticket price.
* Alternative Hypothesis (H1): There is a significant correlation between flight duration and price. This suggests that flight duration does impact ticket prices.

The hypothesis will be tested using Spearman's position correlation, which is a non-parametric test ideal for assessing monotonic relationships in the dataset. The invalid hypothesis will be dismissed if the p-value from the correlation test is less than the significance level (typically 0.05).

# 2. Background Research

## 2.1. Research Papers

Several studies have analyzed factors affecting airfare pricing. Williams, (2022) examined domestic airfare trends and tracked down a powerless yet significant correlation among duration and price. Similarly, Agrawal, (2020) investigated the determinants of airline pricing in India, presuming that price sensitivity is impacted by both flight duration and rivalry among airlines. The relationship between flight duration and pricing has been investigated in various studies. In this ongoing study analyzed how airline competition and request impact pricing dynamics, while this study focused on the streamlining of ticket pricing comparable to advertise segmentation. This investigated pricing trends in minimal expense carriers, emphasizing the role of price sensitivity among consumers. Be that as it may, these studies didn't straightforwardly address flight duration as an essential variable impacting prices. This research aims to fill that hole by focusing specifically on the correlation between flight duration and ticket prices. One more study by Phillips, (2021) assessed airline revenue management strategies and emphasized that flight duration, course contest, and seasonal interest are basic drivers of admission variety. These studies underscore the significance of investigating duration as a determinant of price. While the legitimate dataset has not been specifically referred to in previous studies, its structure aligns with the information used in these research papers, making it fitting for this analysis. Each study emphasized the necessity for additional investigation of domestic flights in quickly developing markets like India, where pricing strategies are still advancing.

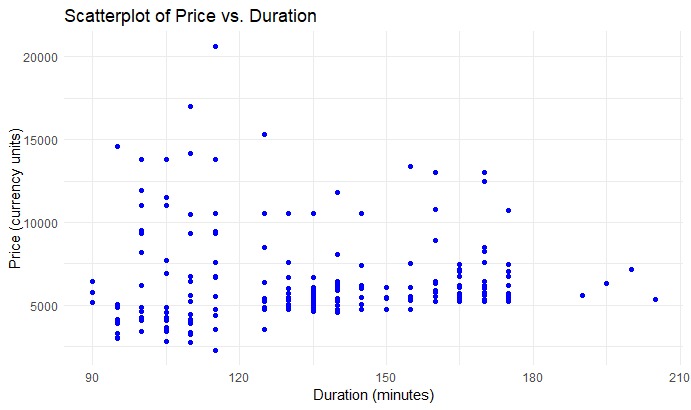
## 2.2. Research Gap and Interest

Despite extensive research on airline pricing, some studies have focused on India's domestic flights as of late. The quickly extending avionics industry and developing working class have increased competition, yet the effect of flight duration on price remains underexplored. This research seeks to fill that hole by looking at the correlation among price and duration for domestic flights. The findings might give insights to future studies, possibly impacting charge forecast models and dynamic pricing strategies.

# 3. Visualisation

## 3.1. Scatter Plot Explanation

The scatter plot of price vs. duration was chosen to visually represent the relationship between flight duration (independent variable) and airfare price (dependent variable). The x-axis represents flight duration in minutes, while the y-axis represents price in cash units. This plot allows for easy observation of any patterns or trends.

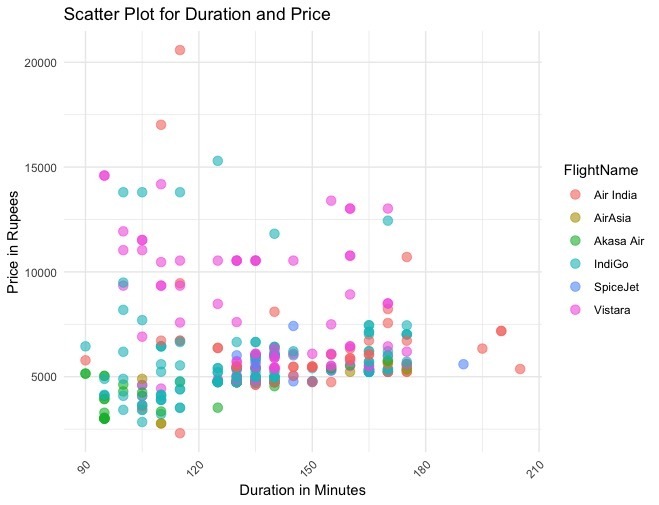


**Figure: Scatterplot of Price vs. Duration**

(source: self-created)

## 3.2. Additional Insights

The scatter plot indicates that, despite the fact that there is some variety in price at various durations, more costly flights will quite often be clustered at specific duration ranges. The changeability in prices for flights of similar durations suggests different factors may also add to price, such as interest, airline, or season of booking.



**Figure: Scatterplot for Duration and Price**

(source: self-created)

## 3.3. Key Observations

The scatter plot suggests a positive, but powerless, correlation between flight duration and price. Shorter flights regularly have lower prices, while longer flights show greater cost fluctuation. This observation aligns with the overall assumption that more extended flights will generally have higher functional costs, thus justifying the higher fares.

**4. Analysis**

**4.1. Statistical Test**

Considering that the dataset does not follow a typical distribution, as proven by the skewed histogram, Spearman's Rho test was selected as the proper non-parametric test to break down the relationship among price and duration. Spearman's Rho evaluates the position correlation between variables, giving insight into the monotonic relationship between them, without assuming ordinariness (Kapoor, 2020).

A graph showing a blue line

Description automatically generated

**Figure: Histogram for Price**

(source: self-created)

**4.2. Interpretation of Results**

The results of Spearman's Rho test returned a p-value of 0.00000000004161, which is far beneath the threshold of 0.05. This significant result leads to the dismissal of the invalid hypothesis, supporting the elective hypothesis that there is a correlation among price and flight duration. While the correlation is statistically significant, the strength of the correlation is frail, showing that in spite of the fact that duration influences price, different factors may also assume a part in deciding airfare for domestic flights in India.

**5. Evaluation – Group’s Experience**

**5.1. What Went Well**

The group successfully teamed up in breaking down the dataset and carrying out the statistical tests using R. Task appointment was clear, and the group successfully used web-based resources, including GitHub, for version control and code sharing. The task objectives were met, and the visualizations gave significant insights into the dataset. Besides, the selection of the proper statistical test (Spearman's Rho) was a key accomplishment, improving the robustness of the analysis.

**5.2. Points for Improvement**

One field for development is the underlying time spent on cleaning the dataset. A more coordinated way to deal with information preprocessing might possess saved energy for the analysis phase. Also, while Spearman's Rho was a proper decision, investigating elective statistical tests or models, such as direct regression for comparison, might have given extra insights. Further developing correspondence during information preprocessing would also streamline future group collaborations.

**5.3. Group’s Time Management**

Generally speaking, the group oversaw time well, designating sufficient opportunity to the two information analysis and report composing. Week after week meetings ensured steady progress, and deadlines were consistently met. An unmistakable division of tasks among group members considered productive cooperation, adding to a convenient submission.

**5.4. Project’s Overall Judgement**

The project successfully accomplished its objective of dissecting the correlation between flight duration and price using a very much structured dataset and fitting statistical methods. The insights acquired from the analysis have viable significance for understanding pricing patterns in India's domestic airline market, and the project's findings are all around supported by the information.

**6. Conclusions**

**6.1. Results Explained**

The results of the Spearman's Rho test show a statistically significant correlation between flight duration and price for domestic flights in India. This correlation, however powerless, suggests that more drawn out flights will generally have greater costs, supporting the hypothesis that flight duration influences pricing.

**6.2. Interpretation of Results**

The observed correlation implies that flight duration is one of the factors contributing to price variability in the domestic airline market. While the relationship is statistically significant, it may not be the dominant factor, as prices are likely affected by additional factors such as airline reputation, demand, and seasonal fluctuations.

**6.3. Future Work and Limitations**

Future research should incorporate additional variables such as demand, competition, and seasonal trends to provide a more comprehensive understanding of airfare pricing. The current study is limited by its focus solely on flight duration and price, which may not capture the full complexity of pricing strategies.

# 7. Reference List

Agrawal, A., 2020. Sustainability of airlines in India with Covid-19: Challenges ahead and possible way-outs. *Journal of Revenue and Pricing Management*, *20*(4), p.457.<https://pmc.ncbi.nlm.nih.gov/articles/PMC7481765/pdf/41272_2020_Article_257.pdf>

Kapoor, M., 2020. Spearman’s rank correlation and factorial analysis: socio-demographic characteristics of diabetic hypertensive patients presenting to a tertiary care hospital. *Indian Journal of Pharmacy Practice*, *13*(2). <https://ijopp.org/files/IndJPharmPract-13-2-125.pdf>

Phillips, R.L., 2021. *Pricing and revenue optimization*. Stanford university press. <https://business.columbia.edu/sites/default/files-efs/pubfiles/4169/pricing%20and%20revenue%20optimization.pdf>

Williams, K.R., 2022. The welfare effects of dynamic pricing: Evidence from airline markets. *Econometrica*, *90*(2), pp.831-858. <https://www.nber.org/system/files/working_papers/w28989/w28989.pdf> .

# 

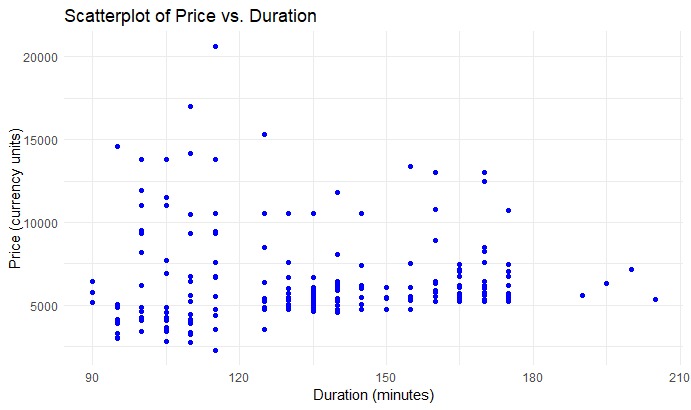
# 8. Appendices

***Appendix A: Analysis R code-***

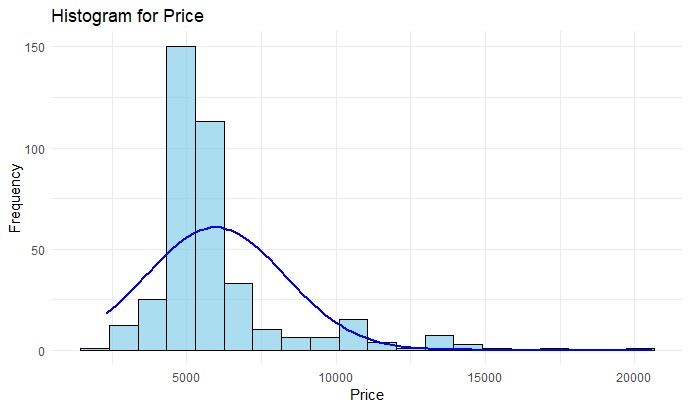
|  |
| --- |
| # Load necessary libraries  library(ggplot2)  # Read the dataset  data <- read.csv("A:/TR&D/Group---A147/flight\_data.csv")  # Combine all datasets into one  combined\_data <- bind\_rows(data\_list)  # Save the combined dataset to a new Excel file  write\_csv(combined\_data, "A:/TR&D/Group---A147/flight\_data.csv")  # View the combined data  print(combined\_data)  # Remove rows with N/A values  clean\_data <- na.omit(data)  # Save the cleaned dataset  write.csv(clean\_data, "A:/TR&D/Group---A147/flight\_data.csv", row.names = FALSE)  # Create a scatterplot  ggplot(data, aes(x = Duration, y = Price)) + geom\_point(color = "blue") + labs(title = "Scatterplot of Price vs. Duration", x = "Duration (minutes)", y = "Price (currency units)") + theme\_minimal()  # Histogram  ggplot(data, aes(x = Price)) + geom\_histogram(aes(y = ..count..), bins = 20, fill = "skyblue", color = "black", alpha  = 0.7) + stat\_function(fun = function(x) {dnorm(x, mean = mean(data$Price, na.rm = TRUE), sd = sd(data$Price, na.rm = TRUE)) \*nrow(data) \* diff(range(data$Price)) / 20  }, color = "blue", size = 1) + labs(title = "Histogram for Price", x = "Price", y = "Frequency") + theme\_minimal()  # Spearman Rho  spearman\_rho <- cor.test(data$Duration, data$Price, method = "spearman")  print(spearman\_rho) |

***Appendix B: visualisation-***

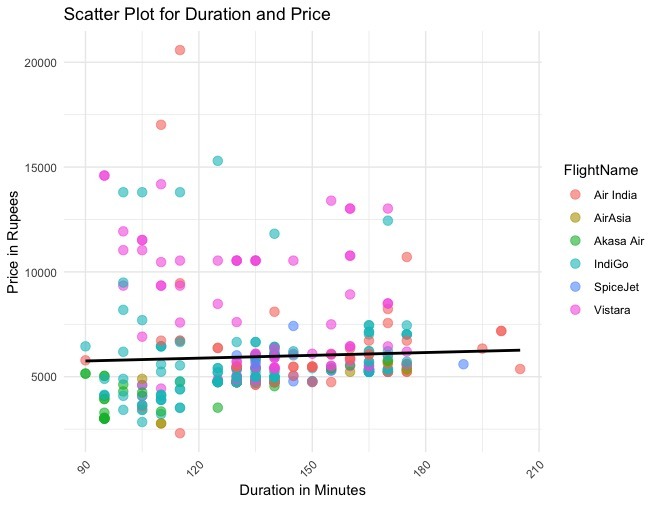
***Scatterplot of Price vs. Duration:***

******

***Histogram for Price:***

******

***Scatterplot for Duration and Price:***

******