

FINAL PROJECT REPORT

GROUP-14

Group Member / Student ID:

Vinay Mattapalli / 11636059

Keerthi Yeakambaram / 11805152

Nehemiah Ray / 11836540

Course: ADTA 5130 Section 001 – Data Analytics I

Henrique Ewbank Ph.D.

INTRODUCTION:

This project uses statistical techniques to examine trends in flight prices. The main goal is to find factors that affect prices such as distance airlines and booking behaviors. The analysis mainly focuses on practical insights for optimizing pricing strategy and handling real world situations. The analysis was done using the airline data set with 10,000 flight records containing different details such as airlines, origin and destination airports, distance, flight duration, class type, ticket price, and booking behavior. We used Excel to analyze and perform statistical calculations and generate visualizations which uncovered patterns in flight pricing. The goal is to provide actionable insights for optimizing pricing strategies and enhancing decision making in the airline industry from the analysis.

EXPLORATORY DATA ANALYSIS:

Here, we conducted an exploratory data analysis (EDA) on a dataset that includes flight data from multiple airlines. The main goal of the research is to comprehend how important features, including flight duration, distance, and fuel surcharge, are distributed and how they relate to flight costs. Below are the figures and tables for a more thorough examination.

DESCRIPTIVE STATISTICS:

The descriptive statistics for the key numerical variables in the data set are summarized below

Variable	Count	Mean	Std Dev	Min	25%	50% (Median)	75%	Max
Distance (miles)	10,000	4002.66	2290.67	100.52	1994.90	3977.83	5960.40	7999.64
Flight Duration (hours)	10,000	8.02	4.05	1.00	4.54	8.03	11.55	14.99
Advance Booking Days	10,000	182.10	105.72	0	90	181.50	274	364
Luggage allowance (KG)	10,000	21.95	4.30	15	18	22	26	29
Fuel Surcharge (USD)	10,000	54.50	25.78	10.01	32.07	54.50	76.82	99.99
Flight Price (USD)	10,000	1026.38	559.88	50.21	548.95	1030.80	1499.95	1999.88

KEY OBSERVATIONS:

Distance: The average flight distance is about 4003 miles, with a wide range from approximately 101 to 8000 miles.

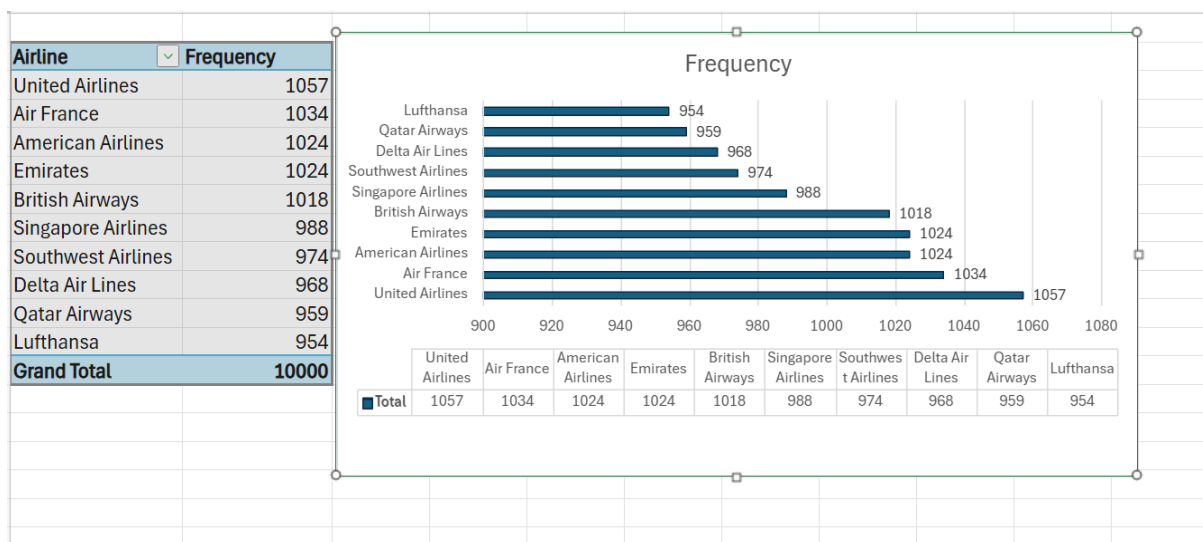
Flight duration: On average, flights last around 8 hours, with a maximum duration of nearly 15 hours.

Advance booking days: Tickets are typically booked around 182 days in advance, with a maximum of 364 days.

Flight price: The average flight price is approximately \$1026.38, with prices ranging from about \$50.00 to nearly \$2000.

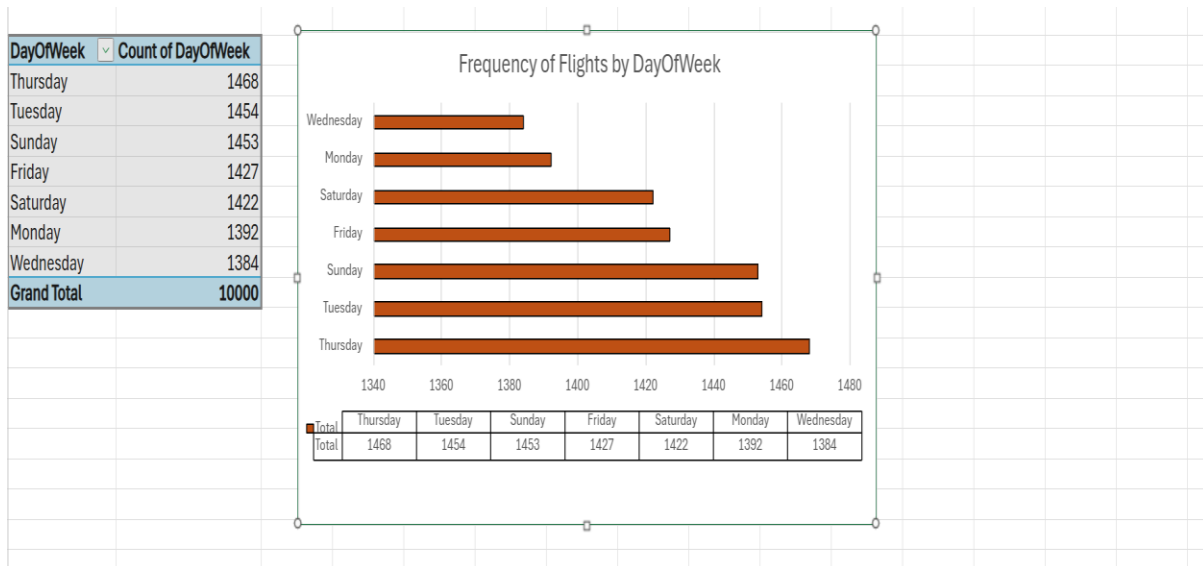
FREQUENCY ANALYSIS:

Frequency of Flights by Airlines



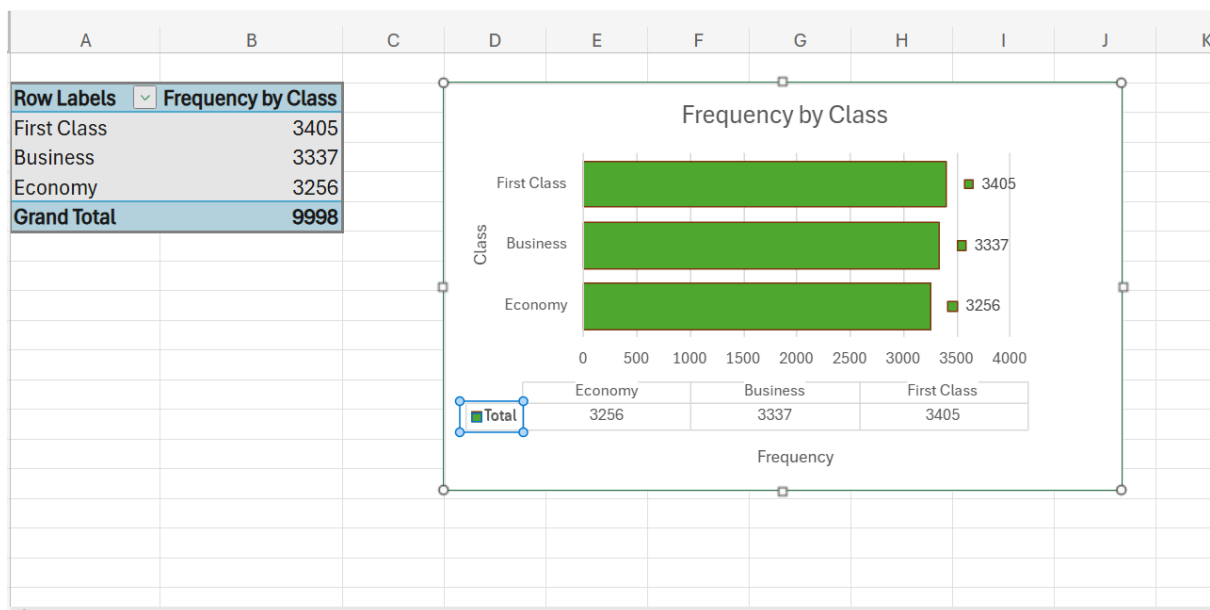
Analysis- Some airlines operated significantly more flights which shows varying market shares.

Frequency of Flights by DayOfWeek



Analysis- We observed Thursdays, Tuesdays, and Sundays with higher frequency of flights compared to other days in a week.

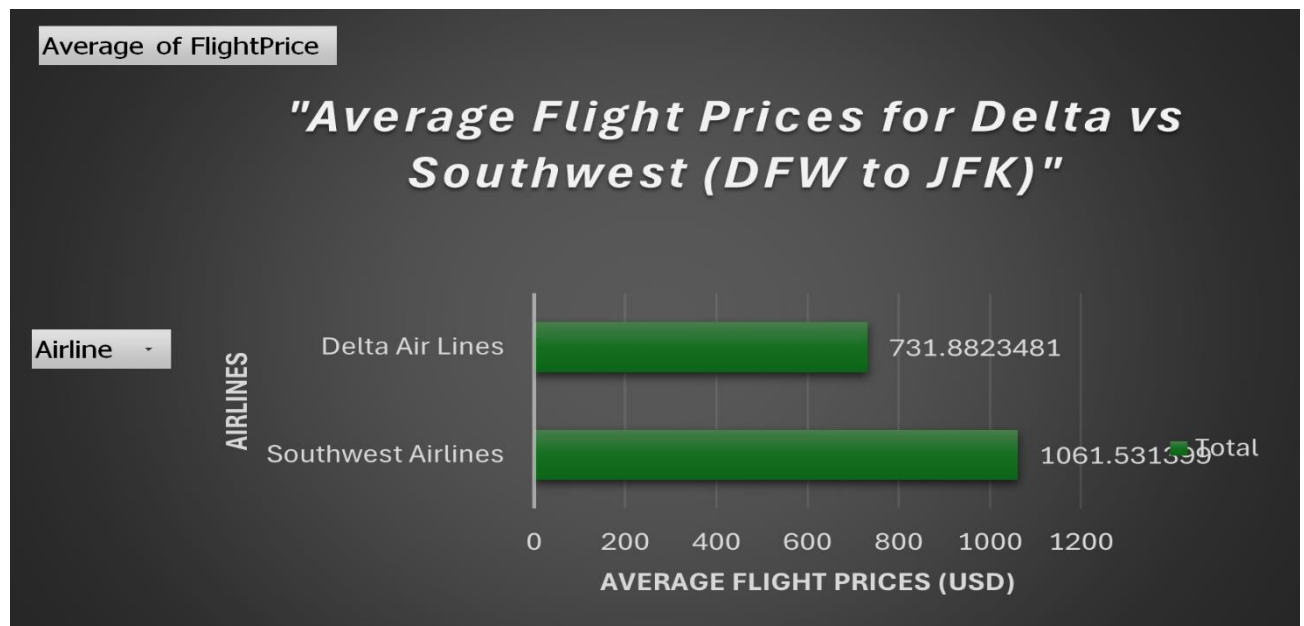
Frequency by Class



Analysis- Usually in general, Economy is expected to dominate other booking class, but First class has taken the highest frequency giving a markable representation, might be due to promotions.

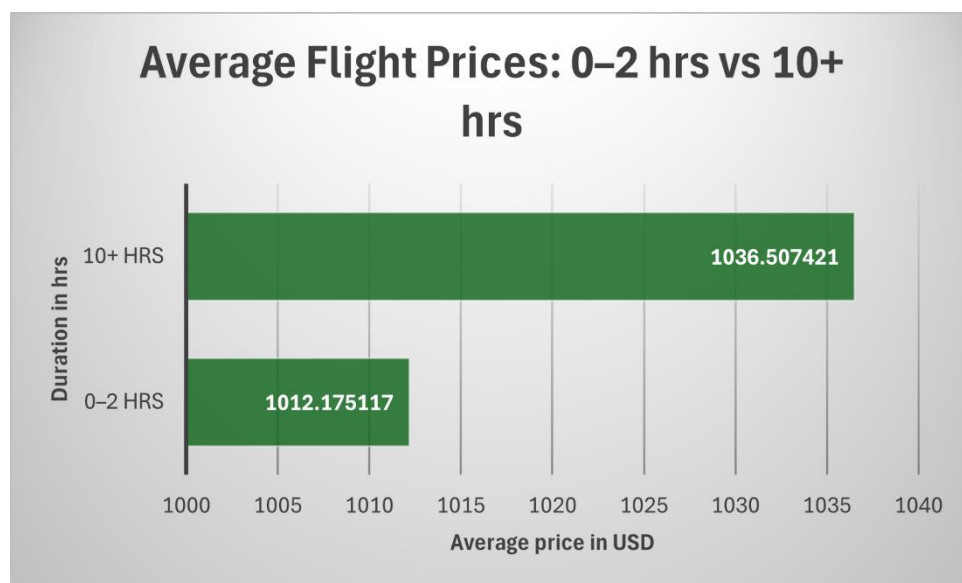
HYPOTHESES:

- 1) **STATEMENT:** Southwest Airlines from Dallas-Fort Worth (DFW) to New York (JFK) have higher average prices compared to Delta airlines for the same route.
Average prices from DFW to JFK – Delta – \$731.88,
Southwest- \$1061.53



- 2) **STATEMENT:** flights with durations of more than 10 hours(10+hrs) are priced more than short -haul flights (0-2hrs).

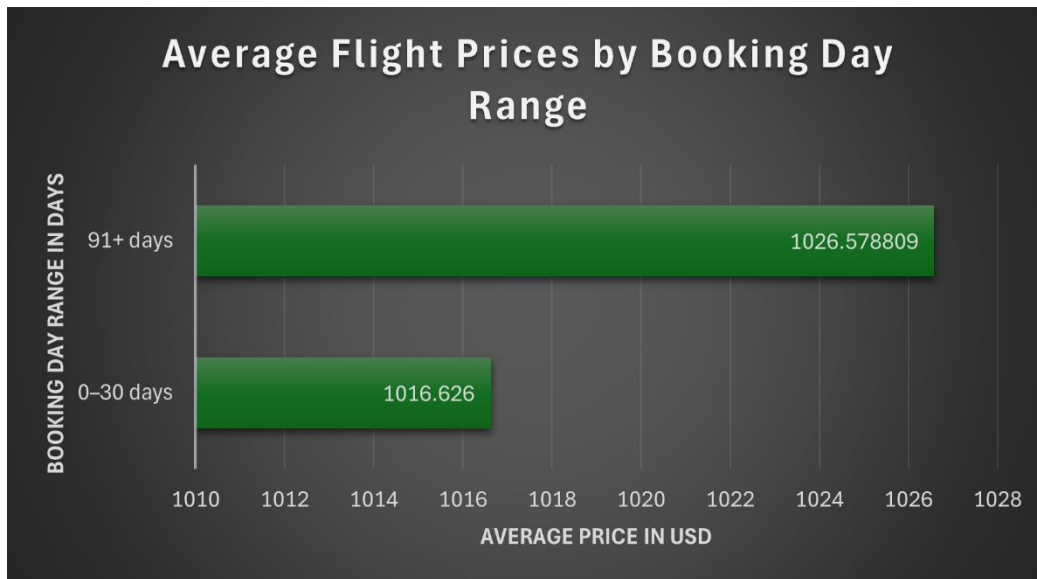
Average flight prices 0-2hrs- \$1012.17 whereas 10+hrs- \$1036.50.



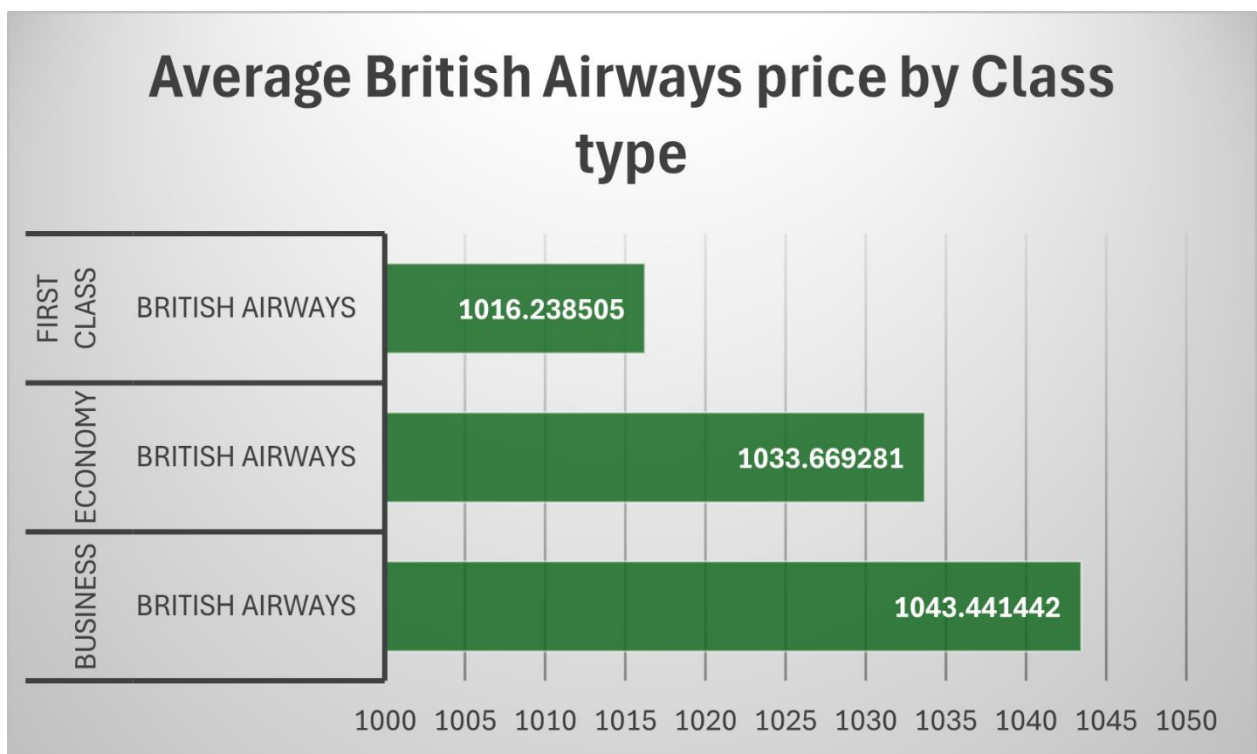
- 3) **STATEMENT:** The flights booked further in advance (91+ days) are slightly more expensive than those booked closer to the departure date (0 -30 days).

Average Flight Price booked before 30 days – \$1016.63

Booked after 91+ days - \$1026.58



- 4) **STATEMENT:** Economy Class has highest average price than First Class in British Airways. Average price in British Airways for Economy Class - \$1033.67
First Class – \$1016.24



ANOVA (Analysis of Variance)

1. Flight price or Advanced Booking days

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	FLIGHTPRICE VS ADVANCED BOOKING DAYS						
4							
5	SUMMARY						
6	Groups	Count	Sum	Average	Variance		
7	FlightPrice	10000	10263812.89	1026.381289	313461.6028		
8	AdvanceBookingD	10000	1820954	182.0954	11177.70027		
9							
10							
11	ANOVA						
12	Source of Variation	SS	df	MS	F	P-value	F crit
13	Between Groups	3564093308	1	3564093308	21957.25086	0	2.705794134
14	Within Groups	3246068391	19998	162319.6515			
15							
16	Total	6810161699	19999				
17							
18							
19							

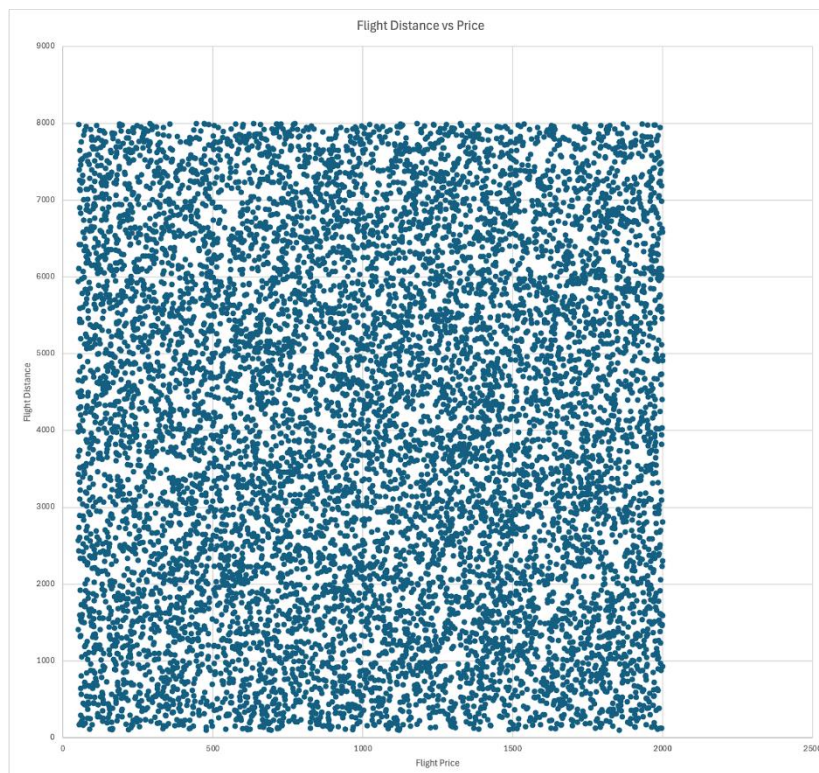
The above analysis contains the single-factor analysis performed between the Flight price and Advanced Booking Days. The F-statistics (21957.25086) exceeds the F-critical value (2.70579), and the P value is 0, indicating strong proof against the null hypothesis. This means that the variation in flight prices is shifting efficiently influenced by the number of days booked in advance, which supports the idea that earlier bookings often result in lower flight prices.

2. Flight Pries Vs Flight Duration

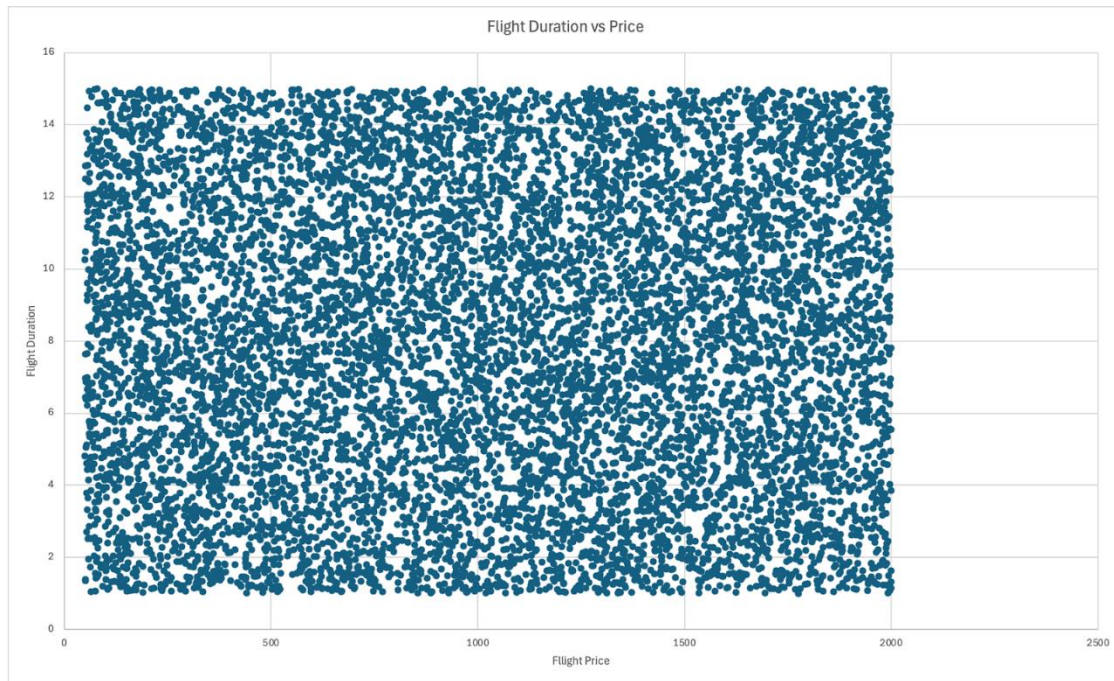
	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3		FLIGHTPRICE VS FLIGHT DURATION					
4							
5	SUMMARY						
6	Groups	Count	Sum	Average	Variance		
7	FlightPrice	10000	10263812.89	1026.381289	313461.6028		
8	FlightDuration	10000	80224.5334	8.02245334	16.37224998		
9							
10							
11	ANOVA						
12	Source of Variation	SS	df	MS	F	P-value	F crit
13	Between Groups	5185273586	1	5185273586	33082.21949	0	2.705794134
14	Within Groups	3134466272	19998	156738.9875			
15	Total	8319739858	19999				
16							
17							
18							

The analysis of flight price against flight duration also indicates a significant relationship. The F-statistic (33082.21949) greatly surpasses the critical value (2.70579), with a P value of 0, showing that the null hypothesis is strongly rejected this gives flight duration significantly affects flight prices, reflecting how longer flights tend to be priced higher than short haul flights.

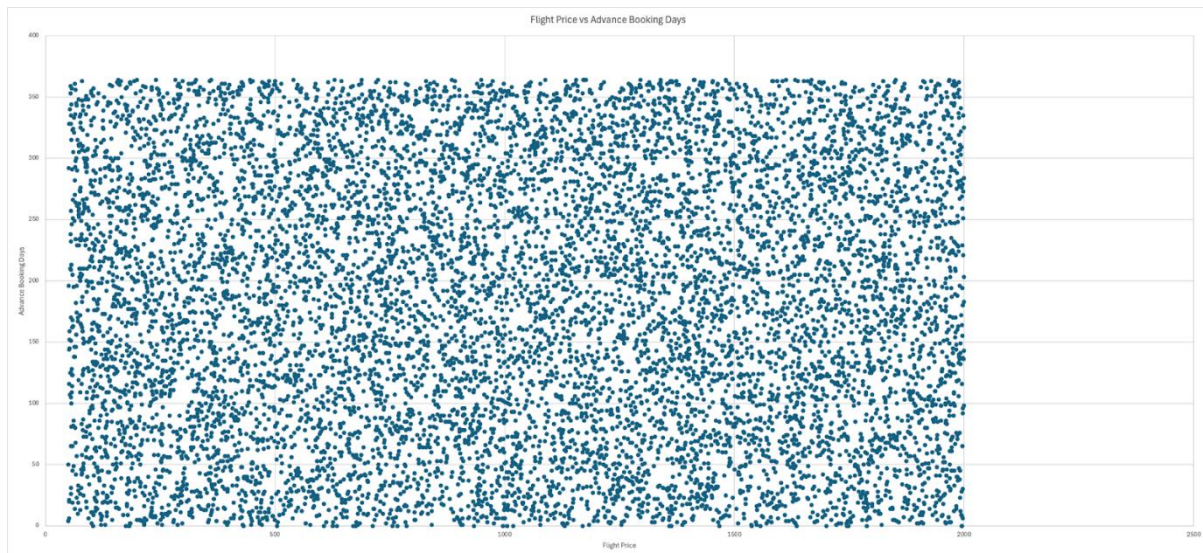
REGRESSION ANALYSIS:



Regression Analysis 1-The distance predicting flight price showed an R squared of 0.000. Meaning that contrast in flight prices is not caused by changes in distance. -0.0003 was the coefficient for distance which is not significant. The scatterplot confirmed there was no obvious trend, which suggests that distance does not significantly influence flight price.



Regression Analysis 2 - The regression analysis of the flight duration as a predictor of the flight price resulted in a R squared of 0.000. This suggests that flight duration also has little to no influence for the variations in the flight price. 1.48 was the coefficient which statistically is not significant ($p=0.285$). The regression line of the scatterplot confirmed that there is not a meaningful relationship between the flight price and duration.



Regression Analysis 3 - The correlation between flight prices and advance booking days seems to be insignificant. The regression analysis produced 0.000 for the R squared value. This shows that the number of days the flight was booked in advance did not influence the various flight prices. At a 95% confidence level the p value of 0.884 shows that the variable is not of importance. The results show that the timing, no matter how far in advance, does not alter or have any effect on the flight prices in this dataset.

CONCLUSION

The analysis on the airline dataset has given important variables affecting flight prices. Significant correlations between flight pricing and factors such as advanced booking days and flight duration were found by the ANOVA study. Cheaper rates are linked to earlier reservations, highlighting the need to encourage early reservations as part of pricing strategies, in a similar manner lengthier flights often cost more.

Interestingly, regression analysis has revealed that neither flight distance nor advanced booking days showed a strong predictive capability for price variations, as found by their R-squared values being near zero. The difference implies that, even if there are some links, pricing fluctuations are probably greatly influenced by other unmeasured factors.

These results indicate the necessity for airlines to improve their pricing strategies and consider other factors, including consumer demographics or industry trends. It is important to recognize the datasets limitations, which include possible biases and exclusions of outside variables like seasonality. These factors may be included in future studies to create a more thorough comprehension of the dynamics of flight prices. Considering all these things, this project offered useful information to airlines that are looking to enhance revenue management and pricing tactics.