**Project Title**

**Spring Airlines Flight Price Analysis**

**ADTA 5130 – Data Analytics I**

**Prof. Henrique Ewbank**

**12-02-2024**

**by Manasa Dontireddy**

**Context of Problem Statement:**

The airline business is so competitive; airlines need to comprehend price dynamics to maximize profits and raise customer satisfaction for an Agency. This dataset includes details about several flights, including their costs, which might vary depending on several criteria like:

1. The Airline dataset consist of various attributes of different airline service providers and their associated pricing strategies.
2. Travel characteristics: Airline prices are competitive and sensitive towards various characteristics such as on-spot booking versus advance booking, duration of light, class type of airline, holiday season versus non-holiday seasons etc.,
3. In-flight amenities: Customers are highly sensitive to airline prices in making their decision either to book it or not. In-general, amenities like meals, WIFI, checked bags, and in-flight entertainment plays a crucial role in booking status.
4. External events: External events such as weather, traffic have certain degree of variations on the flight prices.

**Objectives of Exploratory Data Analysis:**

The following are the main goals of the EDA on this dataset:

1. Recognize Pricing Variability: If helps to analyze how airline, class type, day of the week, time of day, and length of flight affect cost of flight.
2. Find Trends and Patterns: Keep an eye out for pricing trends related to peak travel times (holidays or weekends) or reservation practices.
3. Find Outliers: Check to see if any prices stand out that might point to incorrect pricing or premium services.

**Step-1: Data Dictionary:**

The **“Spring 2024 Airline Dataset”** includes different types of airline information related to prices and its metadata. The dataset covers 10000 records and has distinct features (22). Each feature consists of binaries, numerical and categorical values.

The below [**Table\_1**](#Table_1) data dictionary gives an overview of various attributes, types of data, description and sample values of the Airplane Prices

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Type of Data** | **Description** | **Sample Value** |
| Flight ID | Numerical | Unique identifier for each flight. | 1,2,3 |
| Airline | Categorical | Airline operating the flight. | American Airlines, Delta, United |
| Origin Airport | Categorical | Airport from where the flight departs. | HND, ORD, DFW |
| Destination Airport | Categorical | Destination airport. | LHR, DFW, JFK |
| Distance | Numerical | Distance of the flight in miles. | 339,322,431,245.60 |
| Date Of Flight | Date | Date of the flight. | 2/20/2021,3/6/2022 |
| Day Of Week | Categorical | Day of the week of the flight | Saturday, Monday |
| Time Of Day | Categorical | Time of day of the week | Morning, Afternoon, Evening |
| Flight Duration | Numerical | Duration of the flight in hours. | 1.87,2.24 |
| Aircraft Type | Categorical | Type of aircraft used for flight. | Airbus A380, Boeing 777 |
| Class Type | Categorical | Class of the ticket | Economy, Business, First Class |
| Seat Availability | Categorical | Seat availability in the flight | High, Medium, Low |
| Advance Booking Days | Numerical | Number of days the ticket was booked in advance. | 234, 56,78 |
| Holiday Period | Categorical | Whether the flight is in a holiday period | Yes, no |
| Luggage Allowance | Numerical | Luggage allowance in kilograms. | 18, 34 |
| Meal Included | Categorical | Whether a meal is included | Yes, no |
| WIFI Available | Categorical  Categorical | Whether WIFI is available | Yes, no |
| In Flight Entertainment | Categorical | Availability of in-flight entertainment | Yes, no |
| Fuel Surcharge | Numerical | Fuel surcharge (in USD). | 70,85, 23,45 |
| Airport Traffic | Categorical | Traffic level at the origin airport | High, Medium, Low |
| Weather Conditions | Categorical | Weather conditions during the flight | Clear, Rainy, Snowy |
| Flight Price | Numerical | Price of the flight ticket in USD (Dependent Variable). | 592.34, 123,45,678 |

**Table\_1. Data dictionary and its sample values**.

**Step-2: Data Inspection**

As a second step, need to check for any kind of missing values or NULLs in the dataset. Also, if any outliers are present, we can highlight those to find out the deviations. This helps to get accurate information. The list of outliers, null values and total filed count is displayed in [Table\_2](#Table_2_DI)

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Name** | **Missing/Null Value** | **Outliers [Yes/No]** | **Total Field count** |
| Flight ID | 0 | No | 10000 |
| Airline | 0 | No | 10000 |
| Origin Airport | 0 | No | 10000 |
| Destination Airport | 0 | No | 10000 |
| Distance | 0 | No | 10000 |
| Date Of Flight | 0 | No | 10000 |
| Day Of Week | 0 | No | 10000 |
| Time Of Day | 0 | No | 10000 |
| Flight Duration | 0 | No | 10000 |
| Aircraft Type | 0 | No | 10000 |
| Class Type | 0 | No | 10000 |
| Seat Availability | 0 | No | 10000 |
| Advance Booking Days | 0 | No | 10000 |
| Holiday Period | 0 | No | 10000 |
| Luggage Allowance | 0 | No | 10000 |
| Meal Included | 0 | No | 10000 |
| WIFI Available | 0 | No | 10000 |
| In Flight Entertainment | 0 | No | 10000 |
| Fuel Surcharge | 0 | No | 10000 |
| Airport Traffic | 0 | No | 10000 |
| Weather Conditions | 0 | No | 10000 |
| Flight Price | 0 | No | 10000 |

**Table\_2: Data Inspection – Finding Nulls or Outliers**

**Step-3: Descriptive Statistics:**

As a third stage calculated the Descriptive Statistics (Mean, Mode, Median, Maximum Value, Minimum value, Standard Deviation, Range, Variance, Skewness, Total Count) for the numerical features and are displayed in [**Table\_3**](#Table_3).  
Descriptive statistics are used to more easily examine, summarize, and comprehend patterns and trends within the dataset.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptive**  **Metrics** | **Duration** | **Distance** | |  | | --- | | **Advance**  **Booking Days** | | |  | | --- | | **Luggage**  **Allowance** | | |  | | --- | | **Fuel**  **Surcharge** | | |  | | --- | | **Flight**  **Price** | |
| Average | 8.02245334 | 4002.655 | 182.0954 | 21.947 | 54.49919481 | 1026.381 |
| Median | 8.03019094 | 3977.8289 | 181.5 | 22 | 54.49717445 | 1030.798 |
| Mode | #N/A | #N/A | 123 | #N/A | #N/A | #N/A |
| Standard Deviation | 4.04626371 | 2290.671 | 105.724644 | 4.30342224 | 25.7812542 | 559.8764 |
| Variance | 16.37225 | 5247175 | 11177.7003 | 18.5194429 | 664.673066 | 313461.6 |
| Maximum | 14.9966236 | 7999.6440 | 364 | 29 | 99.99524101 | 1999.882 |
| Minimum | 1.00041090 | 100.51747 | 105.719357 | 15 | 25.77996507 | 559.8484 |
| Range | 13.9962127 | 7899.1266 | 14 | 364 | 89.9852 | 1949.674 |
| Total Count | 10000 | 10000 | 10000 | 10000 | 10000 | 10000 |
| 25th percentile | 4.5365858 | 1994.904 | 90 | 18 | 32.0699185 | 548.9523 |
| 50th percentile | 8.03019094 | 3977.828 | 181.5 | 22 | 54.4971744 | 1030.797 |
| 75th percentile | 11.5521856 | 5960.401 | 274 | 26 | 76.8186323 | 1499.946 |
| Skewness | 0.01205035 | 0.0203444 | 0.00504767 | 0.01555916 | 0.007789497 | -0.01342 |

**Table\_3. Descriptive Statistics for Numerical Overview**

**Step-4: Visualization of the plots**

**Visualization 1:**

From the [**Figure\_1**](#Figure_1)and[**Table\_4**](#Table_4) we can infer that the average flight price for each destination airports. The port ALT is highest over other ports and is not equal with any other ports. DFW stands as least.

|  |  |
| --- | --- |
| **Destination Airport** | **Average of Flight Price** |
| **ATL** | 1070.320859 |
| **HND** | 1044.161998 |
| **CDG** | 1037.096838 |
| **LAX** | 1033.3073 |
| **DXB** | 1026.105425 |
| **JFK** | 1023.510936 |
| **LHR** | 1021.188863 |
| **ORD** | 1017.817023 |
| **SIN** | 998.3137475 |
| **DFW** | 990.8125764 |
| **Grand Total** | **1026.381289** |

Table\_4: Average Flight Price vs Destination Port

Figure\_1. Bar plot representation of the average flight price.

**Hypotheses Analysis: Flight Price vs Destination Ports**

**What is the relationship between Flight Price and Destination Ports?**

**Null Hypotheses (H0):** The average flight price does not vary with different departure airports.

# **μALT = μCDG = = μSIN**

where μALT = Mean flight price of ALT departure airport

**Alternative Hypotheses(H1):** The average flight price varies significantly across different departure airports.

**μj≠ μi where j ≠ i**

**Let’s say μALT ≠ μCDG**

where mean of one departure port ALT **≠** mean of another departure port CDG Not every μDeparture the same.

**ANOVA Single Factor Testing**: To compare whether the mean flight prices for all the airlines is equal or not, we use Anova single factor as there is only one categorical field (Destination Airport)

Segregated the Destination Airport of flight prices into different to make it easier for analysis

A screenshot of a computer

Description automatically generated

**Anova Results:**

A screenshot of a spreadsheet

Description automatically generated

# **Interpret the results**:

From ANOVA analysis, we can see that

Between-Groups Sum of Squares (SSE) is 4,191,370.65 and Mean Square is 465,707.85

Within-Groups Sum of Squares (SSE) is 3,073,648,792 and Mean Square is 312,871.42

F-value is 1.4885 and F-critical value is ​1.8808

P-value is 0.14565156

F-statistic value (1.4885) is less than F -critical value 1.8808

P-value (0.14565) is greater than alpha value (0.05 = 95% CI)

**Result: Fail to reject NULL hypothesis**

Since P-value (0.14565) is greater than alpha value (0.05 = 95% CI) and F-statistic value (1.4885) is less than F -critical value 1.8808, we fail to reject the NULL hypothesis which means average flight price does not vary with different departure ports 🡺 there is no significant difference. This indicates that the slight pricing variety of flights between these departure airports may just be the result of random chance instead of having a real underlying variation in mean prices.

**Visualization 2:**

The below [**Figure\_2**](#Figure_2)and[**Table\_5**](#Table_5) we can verify whether flight price varies with respect to distance or not. However, the below figure line plot talks about the relationship between the price and distance. The cost of the flight varies significantly, with highs and lows.

|  |  |
| --- | --- |
| Distance | Flight Price |
| 3396.67912 | 592.854383 |
| 7950.05833 | 1453.998701 |
| 6394.75801 | 75.56257961 |
| 7143.77437 | 1496.028578 |
| 5938.9955 | 431.8278464 |
| 5820.31725 | 1865.243659 |
| 3104.75463 | 1100.851949 |
| 1267.85872 | 1652.921382 |
| 2348.00808 | 746.9701499 |
| 7311.31821 | 1556.432874 |
| 6357.58319 | 151.3341593 |
| 2032.62382 | 172.564583 |

Table-5: flights Price vs Distance

Figure-2: Line Plot for Flight Price and Distance

**Hypotheses Analysis: Flight Price vs Distance**

**Is there any linear relationship between Flight Price and Distance?**

**Null Hypothesis (H0)**: – There is no linear relationship between distance and flight prices.

**H0​: β1⋅(Distance) = 0**

**Alternative Hypothesis (H1)**: - There is linear relationship between distance and flight prices.

**H1: β1⋅(Distance) ≠ 0**

**Performing Regression Analysis:** Comparing whether prices of flight are linearly corelated with the distance or not, hence we use regression analysis.

The equation for Regression model is **Y = β0​+β1⋅(Distance)+ϵ**

Where Y is predicted/dependent value = Flight Price

β0 = intercept

β1= Distance Coefficient 🡺 Independent variable

ϵ = error term

A screenshot of a computer

Description automatically generated

**Regression Results:**

A screenshot of a spreadsheet

Description automatically generated

Now the equation for Regression model is **Y = 970.5174​+ 0.01756⋅(Distance)**

From Regression Statistics and Anova Table values, we can say below

1. Multiple R (Correlation between the Distance and Flight prices) 🡺 0.072301 (It has weak positive relationship).
2. R square (coefficient of determination) 🡺0.005227 🡺 0.52% variation is observed in flight price (dependent variable) 🡺which means very less variation in flight price.
3. Adjusted R square (0.003214) is less than R square due to total number of predictors in model and has very minimal values.
4. Total observations used in model = 495
5. Regression(df) 🡺 total number of predictors which is 1 (Distance)
6. P-value is 0.10778 which is greater than alpha value. F value is same as P-value which means not significant.
7. Coefficient – Distance 🡺 0.0176 🡺 by Keeping all other factors identical, it says that the flight price(dependent) will rise by 0.0176 units for every unit increase in distance.
8. Sum of squares 🡺 Explained variation in flight price 🡺 832,564.56
9. Sum of squares (Residual) 🡺 Unexplained variation in flight price 🡺 158,437,514

**Result: Fail to reject NULL hypothesis**

Since P-value (0.10778) is greater than alpha value (0.05 = 95% CI), also the R square value is very low, we fail to reject the NULL hypothesis which means the flight prices does not vary with the distance 🡺 there is no linearity between these fields.

**Visualization 3:**

The below [**Figure\_3**](#Figure_3)and[**Table\_6**](#Table_6) helps in understanding the flight prices over airport traffic and holiday period whether prices are changing or have significant effect on these factors.

|  |  |  |
| --- | --- | --- |
| Holiday Period | Airport Traffic | Flight Price |
| No | Low | 592.854383 |
| Yes | Medium | 1453.9987 |
| No | Low | 75.5625796 |
| Yes | Low | 1496.02858 |
| No | Medium | 431.827846 |
| No | High | 1865.24366 |
| Yes | Low | 1100.85195 |
| No | High | 1652.92138 |
| Yes | Low | 746.97015 |

Table\_6: Prices of airlines vs Airport Traffic vs Holiday period

A graph of different colored lines

Description automatically generated

Figure-3: Line plots for the prices vs Traffic Period vs Holiday

**Hypotheses Analysis: Flight Price vs Airport Traffic vs Holiday Period**

**Do Airport traffic and holiday season have a significant impact on flight costs?**

**Null Hypothesis (H0)**: – Flight prices does not vary based on Airport Traffic and Holiday Period

**H0​: β⋅ (Airport Traffic) = β⋅ (Holiday Period) = 0**

**Alternative Hypothesis (H1)**: - At least one of the components has a significant impact on flight prices.

**H1​: At least one βj ≠ 0 for J ∈ {Airport Traffic and Holiday Period}**

**Performing Regression Analysis:** Comparing whether prices of flight are linearly corelated with the airport traffic and Holiday Period or not, hence we use regression analysis.

The equation for Regression model is **Y = β0​+β1⋅ (Airport Traffic) + β2⋅ (Holiday Period)+E**

Where Y is predicted/dependent value = Flight Price

β0 = intercept

β1 = Airport Traffic Coefficient 🡺 Independent variable

β2 = Holiday Period Coefficient 🡺 Independent variable

ϵ = error term

A screenshot of a computer

Description automatically generated

**Regression Results:**

A screenshot of a spreadsheet

Description automatically generated

Now the equation for Regression model is **Y = -51.45371​+526.23182⋅ (Airport Traffic) + 578.62314⋅ (Holiday Period)**

From Regression Statistics and Anova Table values, we can say below

1. Multiple R (Correlation between the actual and expected prices) 🡺 0.6493195 (moderate positive relationship).
2. R square (coefficient of determination) 🡺0.4216158🡺 42% variation is observed in flight price (dependent variable) which is fair percent, still unexplained variation is seen in flight price.
3. Adjusted R square (0.3894834) is less than R square due to total number of predictors in model, not every predictor is significant to model.
4. Total observations used in model = 40
5. Regression(df) 🡺 total number of predictors which is 2 (Airport Traffic and Holiday Period)
6. P-value is 0.10778 which is greater than alpha value. F value is same as P-value which means not significant.
7. Coefficient – Distance 🡺 0.0176 🡺 by Keeping all other factors identical, it says that the flight price(dependent) will rise by 0.0176 units for every unit increase in distance.
8. Sum of squares (Regression) 🡺 Explained variation in flight price 🡺 5046943.3
9. Sum of squares (Residual) 🡺 Unexplained variation in flight price 🡺 6923535.9
10. Total sum of squares 🡺 overall variation in flight price 🡺 11970479

**Results**: **Reject NULL Hypothesis**

**Airport Traffic:** Airport Traffic has a p-value of 0.00006973 and a t-statistic of 4.4938215. This suggests that the flight price - dependent variable is significantly impacted by airport traffic.  
**Holiday Period:** The p-value for Holiday Period is 0.00055416 and the t-statistic is 3.7897455. This further suggests that the flight price - dependent variable is significantly impacted by the holiday period.

Since both P-values (0.00006973 and 0.00055416) are less than alpha value (0.05 = 95% CI), we reject the NULL hypothesis which means the flight prices vary with the airport traffic and holiday period 🡺 there is linearity between these fields.

**Visualization 4:** The below [**Figure\_4**](#Figure_4) which is bar chart and [**Table\_7**](#Table_7)which is a pivot table gives us the understanding of variation of flight prices on different days of a week.

|  |  |
| --- | --- |
| **Day** | **Average Flight prices** |
| Sunday | 1040.631247 |
| Monday | 1027.370883 |
| Tuesday | 996.9797809 |
| Wednesday | 1041.167978 |
| Thursday | 1042.34833 |
| Friday | 1016.035706 |
| Saturday | 1020.722854 |
| **Grand Total** | **1026.424646** |

Table\_7: Average Flight Prices on different days of a week

Figure\_4: Bar Chart for Average Flight Price on different days of a week

**Hypotheses Analysis: Weekend Flight Price vs Weekday Flight Price**

**Does the average flight price differ between weekend and weekday flights?**

**Null Hypothesis (H0)**: There is no significant difference in the average flight prices between weekend and weekday flights **(β1=0)**

**Alternative Hypothesis (H1)**: There is a significant difference in the average flight prices between weekend and weekday flights **(β1≠0)**

**Performing Regression Analysis:**

Segregated the data into two columns of weekday (0) and weekend (1) to make it easier for analysis

A screenshot of a computer

Description automatically generated

**Regression Results:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.00482704 |  |  |  |  |  |  |  |
| R Square | 2.33E-05 |  |  |  |  |  |  |  |
| Adjusted R Square | -7.672E-05 |  |  |  |  |  |  |  |
| Standard Error | 559.897893 |  |  |  |  |  |  |  |
| Observations | 10000 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 73030.31 | 73030.31 | 0.23296221 | 0.62934661 |  |  |  |
| Residual | 9998 | 3134229536 | 313485.651 |  |  |  |  |  |
| Total | 9999 | 3134302566 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 1024.66465 | 6.63309777 | 154.477544 | 0 | 1011.66245 | 1037.66686 | 1011.66245 | 1037.66686 |
| Weekend | 5.97090571 | 12.3707911 | 0.48266159 | 0.62934661 | -18.278335 | 30.2201464 | -18.278335 | 30.2201464 |
|  |  |  |  |  |  |  |  |  |

From the above table,

* **Multiple R (0.0048):** A very low correlation between weekend flights and flight prices.
* **R Square (2.33E-05):** Indicates that only 0.0023% of the variability in flight prices can be explained by whether the flight occurs on a weekend or weekday. This is extremely low.
* **Adjusted R Square (-7.67E-05):** A negative adjusted R-square reflects no meaningful relationship after adjusting for the number of predictors.
* **Standard Error (559.89):** This is the standard deviation of the residuals
* **F-Statistic (0.23296):** Very low F-statistic indicates the model does not explain much variance.
* **Significance F (0.6293):** A p-value higher than 0.05 indicates the predictor variable (Weekend) is not statistically significant.
* **Intercept (1024.66):** The average flight price when the weekend variable is 0 (weekday).
* **Weekend Coefficient (5.97):** The average difference in flight prices between weekend and weekday flights.
* **t-Stat (0.482):** Indicates the predictor's influence is weak.
* **P-value (0.629):** Not statistically significant (p > 0.05).
* **Confidence Interval (-18.28 to 30.22):** The interval crosses zero, further supporting the insignificance of the effect.

**Results: Fail to reject the NULL Hypotheses.**

The regression results suggest no statistically significant relationship between flight prices and whether a flight occurs on a weekend or weekday.

**Visualization 5:** From the [Table\_8](#Table_8) which is pivot table and [Figure\_5](#Figure_5), which is Bar chart, we can get the information of variation of flight prices during different seasons.

|  |  |
| --- | --- |
| **Season** | **Average flight prices** |
| Fall | 1011.029447 |
| Spring | 1036.563325 |
| Summer | 1015.882418 |
| Winter | 1041.25251 |
| **Grand Total** | **1026.424646** |

Table\_8: Average Flight Prices in different seasons

Figure\_5: Bar Chart for Average Flight Price in different seasons

**Hypotheses Analysis: Flight prices vs Prime Holiday Seasons (Summer and Winter)**

**Do flight prices vary significantly between two main holiday seasons (Summer and Winter)?**

**Null Hypothesis (H0)**: Seasonal differences do not significantly affect flight prices **(β1​=β2​=0).**

**Alternative Hypothesis (H1​)**: Seasonal differences significantly affect flight prices **(at least one βi≠0).**

**Performing Regression Analysis:**  
Segregated data into separate columns of Winter and Summer to make it easier to perform analysis.

A screenshot of a computer

Description automatically generated

**Regression Results:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.01629929 |  |  |  |  |  |  |  |
| R Square | 0.00026567 |  |  |  |  |  |  |  |
| Adjusted R Square | 6.566E-05 |  |  |  |  |  |  |  |
| Standard Error | 559.858036 |  |  |  |  |  |  |  |
| Observations | 10000 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 2 | 832680.719 | 416340.359 | 1.32828932 | 0.26497684 |  |  |  |
| Residual | 9997 | 3133469885 | 313441.021 |  |  |  |  |  |
| Total | 9999 | 3134302566 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 1024.26004 | 7.88688916 | 129.8687 | 0 | 1008.80015 | 1039.71993 | 1008.80015 | 1039.71993 |
| Summer | -8.3776215 | 13.7533129 | -0.6091348 | 0.54244894 | -35.336883 | 18.5816406 | -35.336883 | 18.5816406 |
| Winter | 16.8125352 | 13.7106499 | 1.22623911 | 0.22013759 | -10.063099 | 43.6881692 | -10.063099 | 43.6881692 |

From the above table,

* **R-Squared (0.000266)**: Only 0.0266% of the variation in flight prices is explained by seasonality. This indicates the model has very poor explanatory power.
* **Adjusted R-Squared (0.000066)**: After adjusting for the number of predictors, the explained variance remains negligible.
* **Significance F (0.26498)**: The p-value for the overall model is 0.265 (> 0.05), indicating that the regression model is not statistically significant. This suggests that seasonality does not significantly impact flight prices.
* **Intercept (1024.26)**: The average flight price during the baseline season (e.g., Fall) is approximately $1,024.26.
* **Summer Coefficient (-8.38)**: Summer flight prices are $8.38 lower than the baseline season, but this difference is not statistically significant (p=0.542p=0.542).
* **Winter Coefficient (16.81)**: Winter flight prices are $16.81 higher than the baseline season, but this difference is also not statistically significant (p=0.220p=0.220).

**Results: Fail to reject the NULL Hypotheses**

Both p-values are greater than 0.05, which means the coefficients for these variables are **not statistically significant**. This indicates that neither Summer nor Winter has a significant impact on flight prices.

**Visualization 6:** From the [Table\_9](#Table_9) which is pivot chart and [Figure\_6](#Figure_6), which is Bar chart, we can understand the variation of flight prices with the inclusion of meal and without the inclusion of meal.

|  |  |
| --- | --- |
| **Meal Included** | **Average flight prices** |
| No | 1031.852232 |
| Yes | 1021.116241 |
| **Grand Total** | **1026.424646** |

Table\_9: Average Flight Prices with and without meal inclusion

Figure\_6- Bar Chart for Average Flight Price with or without meal inclusion

**Hypotheses Analysis:** **Flight Price vs Meal Inclusion**

**Does including a meal with the flight significantly affect the flight price?**

**Null Hypothesis (H0​)**: There is no significant difference in flight prices between flights with and without meal inclusion (**μFlight Price​ = μMeal Inclusion​).**

**Alternative Hypothesis (Ha)**: There is a significant difference in flight prices between flights with and without meal inclusion **(μFlight Price ≠ μMeal Inclusion​).**

**Performing Anova- Single factor testing:**  
  
A screenshot of a computer

Description automatically generated

Significance level = 0.05

**Anova Results:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Flight price | 10000 | 10263812.9 | 1026.38129 | 313461.603 |  |  |
| Meal Inclusion | 10000 | 5056 | 0.5056 | 0.24999364 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 5262104642 | 1 | 5262104642 | 33574.1309 | 0 | 3.84192387 |
| Within Groups | 3134305066 | 19998 | 156730.926 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 8396409708 | 19999 |  |  |  |  |
|  |  |  |  |  |  |  |

From the above table,

* The **F-value (33,574.1309)** is much larger than the **F crit (3.8419)**, and the **P-value (0.000)** is less than the significance level (α=0.05).

**Results: Reject the NULL Hypotheses**

This indicates that there is a significant difference in flight prices between flights with and without meal inclusion.