**ASSIGNMENT-1 BDM 2203**

**SUBMITTED BY-MS. GURLEEN KAUR**

**STUDENT ID-C0907014**

**SUBMITTED TO-MR. ISHANT GUPTA**

**FLIGHT PRICES DATASET:**

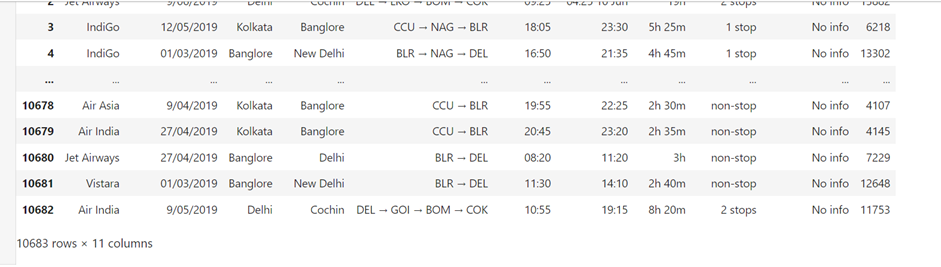
* Q1.) Loading the “flight\_price.xlsx” dataset and examining its dimensions:

1. Import all the necessary libraries
2. Use “df=pd.read\_excel” command to read the excel dataset for performing further analysis.



* Displaying the output:

#The “flight\_price” dataset contains **10683 rows and 11 colums.**



* Q2.) Plotting a Histogram to get insights of flight prices distribution in the dataset:

A green bar graph with white background

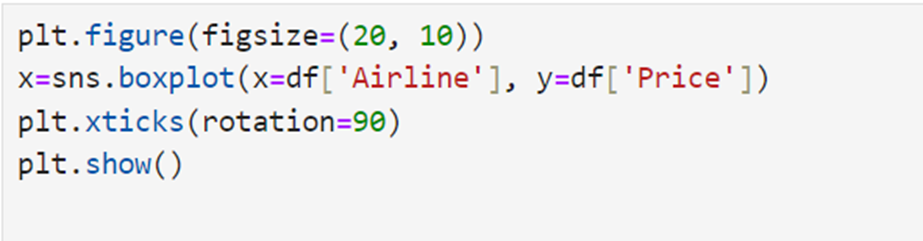
Description automatically generated

* Q3.) Let’s examine the range of prices along with the minimum and maximum price flight in the datset:

A white background with colorful text

Description automatically generated with medium confidence

* Q4.) Creating a Boxplot for contrasting the prices of different airlines:

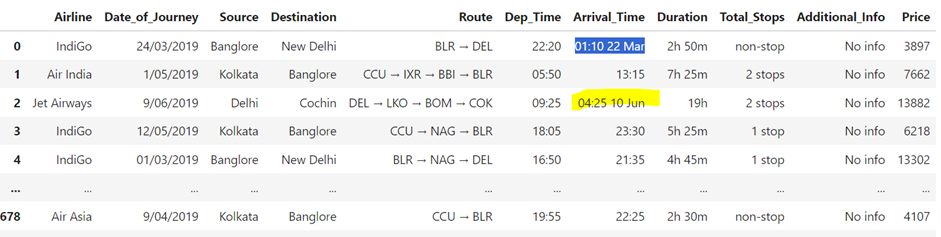


#Output:

A graph of blue boxes

Description automatically generated with medium confidence

* Q5.) Identifying the potential outliers and their impact in the dataset :
* The “flight\_price” dataset has some of the outliers for instance; the arrival time section contains most of the data in Time format but some of the entries are in unusual format such as 04:25 10 Jun.



These outliers can be identified with the help of boxplot as shown below:

A screenshot of a computer program

Description automatically generated

A blue and black graph

Description automatically generated with medium confidence

There are points which lie significantly below or above the **whiskers(**which refers to the lines in a box plot that extend to the highest and lowest data points within a certain range) these are the potential outliers present in the dataset which might skew the distribution while impacting the analysis.

Outliers may impact the Stastical measure,Model's performance and Interpretation and Visualization of the data.

* Q6.) To identify the flight prices during the peak season, the following characteristics can be analyzed:

1. Examining airline ticket data can offer valuable insights into pricing trends and peak travel seasons. By organizing the data based on months or seasons, one can detect price fluctuations. This allows for the identification of periods when prices are typically higher or lower, aiding travelers and industry experts in planning their trips effectively.
2. Understanding the popularity of destinations at different times of the year enables individuals to predict peak travel seasons. For instance, some destinations may attract more visitors during specific months due to holidays, festivals, or favorable weather conditions. Analyzing average prices during these periods helps travelers determine the most economical times to book flights.
3. Airlines employ various pricing strategies and offer diverse fare structures, making it essential to analyze average fares over time. By doing so, travelers can discern patterns in pricing and recognize peak travel periods. This information assists in making informed decisions about when to book flights and which airlines to choose.
4. Identifying special events such as festivals, conferences, or major holidays that coincide with heightened travel demand is crucial. These events can significantly influence flight availability and prices. Thus, travelers need to be aware of such events when planning their trips to avoid potential disruptions and optimize their travel experiences.

* Q7.) To identify trends in flight prices, I would analyze several key features in the Flight Price dataset. Here are the features I would focus on:

**Date/Time:** Analyzing flight prices over time can reveal seasonal trends, such as peak travel seasons and periods of lower demand. This could involve examining average prices by month, day of the week, or even time of day for flights.

**Route/Destination:** Investigating flight prices for different routes and destinations can uncover patterns in demand and pricing variations. Certain routes may experience higher demand and subsequently higher prices, especially during specific times of the year or for popular events.

**Airline:** Exploring flight prices by airline can highlight differences in pricing strategies and competitiveness. Some airlines may consistently offer lower prices or have more significant price fluctuations over time.

**Special Events:** Identifying any special events, such as holidays, festivals, conferences, or major sporting events, and analyzing their impact on flight prices can provide valuable insights into pricing trends during peak travel periods.

**For presenting my findings to the team**, I would use a combination of visualizations to effectively communicate the identified trends. Here are some visualization techniques I would consider:

**Time Series Plots:** Using line graphs to plot average flight prices over time (e.g., months or years) can illustrate seasonal trends and any long-term changes in pricing.

**Bar Charts:** Bar charts can be used to compare average flight prices across different routes, destinations, airlines, or special events. This visual representation makes it easy to identify variations and trends among different categories.

**Heatmaps:** Heatmaps can visualize flight prices across different dates and destinations, providing a comprehensive overview of pricing patterns. This can help identify peak travel periods and destinations with consistently high or low prices.

**Scatter Plots:** Scatter plots can be used to explore the relationship between flight prices and specific factors such as time of booking, distance traveled, or advance purchase period. This can help identify any correlations or trends that may not be apparent in other visualizations.

By analyzing these features and using appropriate visualizations, I would be able to provide valuable insights into flight price trends to the team, enabling them to make informed decisions about pricing strategies, marketing campaigns, and customer engagement.

* Q8.) To identify factors affecting flight prices, I'd analyze features like date/time, route/destination, airline, flight class, booking time, special events, duration, and competitor prices.

I'd present findings to management with visuals like time series plots for seasonal trends, bar charts for route comparisons, box plots for airline comparisons, scatter plots for booking patterns, and bar charts for special event impacts.

The presentation would include an executive summary, trend analysis, route/destination insights, airline comparisons, booking pattern analysis, special event impacts, and recommendations for pricing strategies.

**GOOGLE PLAYSTORE DATASET:**

* Q9.) Loading the “Google play store” dataset and examining its dimensions:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

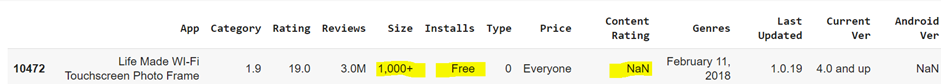
Thus, the above dataset has 10841 rows and 13 columns

* Q10.) Plotting a boxplot to contrast the ratings of different app categories:

A screenshot of a graph

Description automatically generated

* Q11.) Missing values in Dataset and their impact:



From the above fig., it is evident that there are 1000+ null values in the dataset which may have the following impacts on the dataset analysis:

1. Failure to appropriately address missing values in the data can result in biased analysis since the available data may not accurately represent the entire dataset.
2. Incomplete data due to missing values diminishes the sample size, potentially undermining the statistical power of the analysis and introducing greater uncertainty into the findings.
3. The presence of missing values can distort the relationships between variables, potentially misleading analysts about correlations or associations among the data.

* Q12.) Plotting a Scatter plot in order to visualize the relationship between the size of an app and its rating:

A screenshot of a computer program

Description automatically generated

A graph with blue dots

Description automatically generated

* Q13.) Creating a Bar chart to compare the average prices by app type:

A screen shot of a computer code

Description automatically generated

A ruler with text on it

Description automatically generated

* Q14.) Top 10 apps in the dataset:

A screenshot of a computer

Description automatically generated

* Q15.) Identifying the most popular app categories:

To identify the most popular app categories on the Google Play Store, I would approach the task by analyzing several key features in the dataset. Here's how I would do it:

**Category Analysis:** First, I would analyze the distribution of apps across different categories. This involves counting the number of apps in each category to understand which categories are the most prevalent on the Play Store.

**Rating Analysis:** Next, I would examine the average user ratings for apps in each category. Higher ratings generally indicate greater user satisfaction and popularity. Analyzing the distribution of ratings within each category can provide insights into user preferences.

**Installs Analysis**: I would also analyze the number of installs for apps in each category. This metric gives a direct measure of app popularity in terms of user adoption. Categories with a higher number of installs are likely to be more popular among users

**Review Analysis:** Analyzing the number of user reviews for apps in each category can also provide insights into popularity. Categories with a higher volume of user reviews may indicate greater user engagement and interest.

**Revenue Analysis (if available):** If revenue data is available in the dataset, I would analyze the revenue generated by apps in each category. This can provide additional insights into the popularity of different app categories in terms of monetization potential.

Based on the analysis of these features, I would make recommendations to the company regarding the most popular app categories on the Google Play Store. For example, if the analysis shows that the "Games" category has the highest number of installs, ratings, and user reviews, I recommend considering the development of gaming apps. Similarly, if the "Social" or "Communication" categories have high user engagement metrics, I would recommend exploring opportunities in those categories.

**Overall,** by analyzing features such as category distribution, user ratings, installs, reviews, and revenue, I can provide valuable insights to the company to inform their decision-making process regarding the launch of a new app on the Google Play Store.

* Q16.) Identifying the most successful app developers:

To identify the most successful app developers in the Google Play Store dataset, I would analyze several key features and use appropriate data visualizations to present my findings. Here's how I would approach it:

**Number of Downloads/Installs:** Analyzing the total number of downloads or installs for each app developer can provide a measure of their success in reaching and engaging users. I would sum up the installs for all apps associated with each developer.

**Average Rating:** Calculating the average user rating for each app developer's portfolio can indicate the overall satisfaction level of users with their apps. Higher average ratings suggest greater success in delivering quality apps.

**Number of Apps:** Examining the total number of apps published by each developer can provide insights into their productivity and level of activity in the app market.

**Revenue (if available):** If revenue data is available in the dataset, analyzing the total revenue generated by apps from each developer can offer insights into their financial success and profitability.

**Data Visualizations:**

**Bar Chart:** Use a bar chart to visualize the total number of downloads or installs for each app developer. This will allow for easy comparison of the success metrics across different developers.

**Horizontal Bar Chart:** Present a horizontal bar chart to display the average rating of apps by each developer. This visualization will highlight developers with the highest average ratings.

**Pie Chart:** Create a pie chart to show the distribution of the total number of apps among different developers. This visualization will help identify developers with the largest app portfolios.

**Scatter Plot:** Use a scatter plot to visualize the relationship between the number of downloads/installs and the average rating for each developer. This can help identify developers who achieve high success metrics in both user engagement and satisfaction.

**Stacked Bar Chart (if revenue data is available):** If revenue data is included, use a stacked bar chart to visualize the total revenue generated by each developer, broken down by app categories or individual apps. This will provide insights into the financial success of developers in different market segments.

By analyzing these features and presenting the findings using appropriate data visualizations, the mobile app development company can identify the most successful app developers in the Google Play Store dataset and make informed decisions about potential partnerships or collaborations.

* Q17.) Identifying the best time to launch the new app:

To identify the best time to launch a new app using the Google Play Store dataset, I'd analyze **app release dates, category trends, user engagement metrics, and seasonal patterns.**

For recommendations, I'd use visualizations like a **time series line chart for app releases, a stacked area chart for category trends, a bar chart for user engagement metrics, and a heatmap for seasonal patterns.**

This approach would help the marketing research firm pinpoint optimal launch times based on trends in app releases, category popularity, user engagement, and seasonal factors.