**Flight Price:**

**Q1. Load the flight price dataset and examine its dimensions. How many rows and columns does the dataset have?**

A close-up of a message

Description automatically generated

A screenshot of a computer

Description automatically generated

**Q2. What is the distribution of flight prices in the dataset? Create a histogram to visualize the distribution.**

A green bar graph with white background

Description automatically generated

**Q3. What is the range of prices in the dataset? What is the minimum and maximum price?**

A white background with colorful text

Description automatically generated with medium confidence

**Q4. How does the price of flights vary by airline? Create a boxplot to compare the prices of different airlines.**

A computer screen shot of a math equation

Description automatically generated

A graph of blue boxes

Description automatically generated with medium confidence

**Q5. Are there any outliers in the dataset? Identify any potential outliers using a boxplot and describe how they may impact your analysis.**

**Ans:-**  Yes there is some outliers in the data set that as for example in the arrival time every data is in time but some of the data have time like 01:10 22 Mar as shown below

A screenshot of a computer

Description automatically generated

It can be shown with the help of boxplot

A screenshot of a computer program

Description automatically generated

A blue and black graph

Description automatically generated with medium confidence

By examining the boxplot, you can identify any points that lie significantly above or below the whiskers. These points are potential outliers that may skew the distribution and impact your analysis.

Outliers can impact the Statistical Measure, Data visualization, Model performance and Data Interpretation.

**Q6. You are working for a travel agency, and your boss has asked you to analyze the Flight Price dataset to identify the peak travel season. What features would you analyze to identify the peak season, and how would you present your findings to your boss?**

**Ans:-**To identify the season of travel in the airline ticket dataset, I analyzed the following characteristics:

**Date or time**: Analyzing airline tickets from different months or seasons can help identify when prices tend to be higher or lower. This could mean grouping the data by month or time of year and calculating the average price for each period.

**Destination**: By understanding which destinations are more in demand at certain times of the year, you can get an idea of ​​the peak travel season. Some destinations may be more popular during holidays, festivals or certain weather conditions.

**Airline**: Different airlines may have different pricing strategies and seasonal differences. Analyzing average fares offered by different airlines over time can help identify peak travel periods.

**Special Events**: Identifying special events such as festivals, conferences or major holidays that coincide with times of increased travel demand can also help identify peak travel seasons..

**Q7. You are a data analyst for a flight booking website, and you have been asked to analyze the Flight Price dataset to identify any trends in flight prices. What features would you analyze to identify these trends, and what visualizations would you use to present your findings to your team?**

**Ans:-** To identify trends in flight prices in the price data, I analyzed the following characteristics: **Date or time, Route, Airline, Booking class**

To present my findings to the group, I use visualizations such as:

**Time Series Charts, Bar Charts, Boxplots, Heatmaps**

**Q8. You are a data scientist working for an airline company, and you have been asked to analyze the Flight Price dataset to identify the factors that affect flight prices. What features would you analyze to identify these factors, and how would you present your findings to the management team?**

**Ans:-** o identify factors affecting flight prices in the Flight Price dataset, I would analyze features such as:

**Date and Tim, Route, Airline, Booking Class, Advance Booking Period** and many more.

I would present my findings to the management team through a combination of:

**Visualizations:** Utilize graphs, charts, and heatmaps to visually represent relationships between flight prices and the identified factors.

**Summary Statistics:** Provide key statistical metrics such as mean, median, and standard deviation to summarize the data.

**Insights and Recommendations:** Offer actionable insights and strategic recommendations based on the analysis to optimize pricing strategies and enhance competitiveness.

**Below is the python file and excel which has been used in the assignment**

****

**Google Playstore:-**

**Q9. Load the Google Playstore dataset and examine its dimensions. How many rows and columns does the dataset have?**

A screenshot of a computer

Description automatically generated A screenshot of a computer

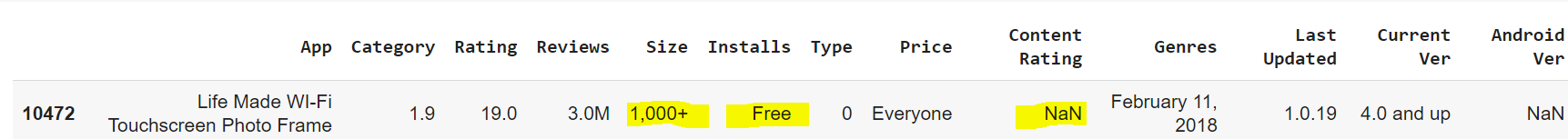
Description automatically generated

**Q10. How does the rating of apps vary by category? Create a boxplot to compare the ratings of different app categories.**

A screenshot of a graph

Description automatically generated

**Q11. Are there any missing values in the dataset? Identify any missing values and describe how they may impact your analysis.**



As in above figure the missing and error values are 1000+, installs like Free and content rating is NaN.

Missing values in a dataset can have various impacts on the analysis:

**Biased Analysis**: If missing values are not handled properly, it can lead to biased analysis because the available data may not be representative of the entire dataset.

**Reduced Sample Size**: Missing values reduce the effective sample size, which can affect the statistical power of the analysis and increase uncertainty in the results.

**Distorted Relationships**: Missing values can distort relationships between variables, leading to inaccurate conclusions about correlations or associations.

**Q12. What is the relationship between the size of an app and its rating? Create a scatter plot to visualize the relationship.**

A screenshot of a computer program

Description automatically generated A graph with blue dots

Description automatically generated

**Q13. How does the type of app affect its price? Create a bar chart to compare 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. What are the top 10 most popular apps in the dataset? Create a frequency table to identify the apps with the highest number of installs.**

A screenshot of a computer

Description automatically generated

Q15. **A company wants to launch a new app on the Google Playstore and has asked you to analyze the Google Playstore dataset to identify the most popular app categories. How would you approach this task, and what features would you analyze to make recommendations to the company?**

Ans;-

1.**Data Exploration**:

Start by loading the dataset and exploring its structure. Understand the columns and the type of data available. Check for any missing values and handle them appropriately, either by removing or imputing them.

2.**Data Cleaning:**

Clean the data by addressing any inconsistencies, such as misspellings, duplicates, or outliers.

Ensure that relevant columns are of the correct data type for analysis.

3.**Feature Selection**:

Identify the features that are relevant to determining the popularity of an app category. These may include:

App category (e.g., Education, Games, Social)

4. **Data Aggregation:**

Aggregate the data by app category to calculate metrics such as the average number of installs, average ratings, and total number of reviews for each category.

5.**Analysis:**

Analyze the aggregated data to identify the most popular app categories based on various metrics. For example:

Determine the categories with the highest average number of installs or ratings.

Identify categories with a large number of apps but relatively low average ratings or installs, which may indicate a less saturated market.

Consider trends over time by analyzing changes in popularity across different time periods.

6.**Visualization:**

Visualize the findings using appropriate charts and graphs, such as bar plots, scatter plots, or pie charts. This will help stakeholders easily understand the insights derived from the data.

Recommendations:

**Q16. A mobile app development company wants to analyze the Google Playstore dataset to identify the most successful app developers. What features would you analyze to make recommendations to the company, and what data visualizations would you use to present your findings?**

Ans:

To identify the most successful app developers in the Google Play Store dataset, I would analyze the following features:

**Number of Downloads/Installs**:

Apps with a high number of downloads or installs indicate popularity and success.

**Ratings and Reviews:**

Higher ratings and a large number of positive reviews suggest user satisfaction and engagement with the app.

**App Updates:**

Regular updates indicate developer commitment to improving the app and addressing user feedback.

**App Size:**

Smaller app sizes are generally preferred by users, potentially leading to higher adoption rates.

**Price:**

Free or low-cost apps may attract more users, while premium apps could indicate higher quality or specialized offerings.

**Retention Rate:**

Apps with high user retention rates are likely to be successful in the long term.

For data visualizations, I would use the following:

**Bar Charts:**

To compare the number of downloads/installs, ratings, and reviews for different developers.

**Scatter Plots:**

To visualize the relationship between app size, ratings, and number of installs.

**Pie Charts:**

To show the distribution of free vs. paid apps by each developer.

**Line Charts:**

To track the frequency of app updates over time for each developer.

**Q17. A marketing research firm wants to analyze the Google Playstore dataset to identify the best time to launch a new app. What features would you analyze to make recommendations to the company, and what data visualizations would you use to present your findings?**

Ans:

To identify the best time to launch a new app in the Google Play Store dataset, I would analyze the following features:

**Trend Analysis, Number of Installs, Ratings and Reviews, Competition, Category Analysis**

For data visualizations, I would use the following:

**Line Charts, Bar Charts, Heatmaps, Box Plots**

****