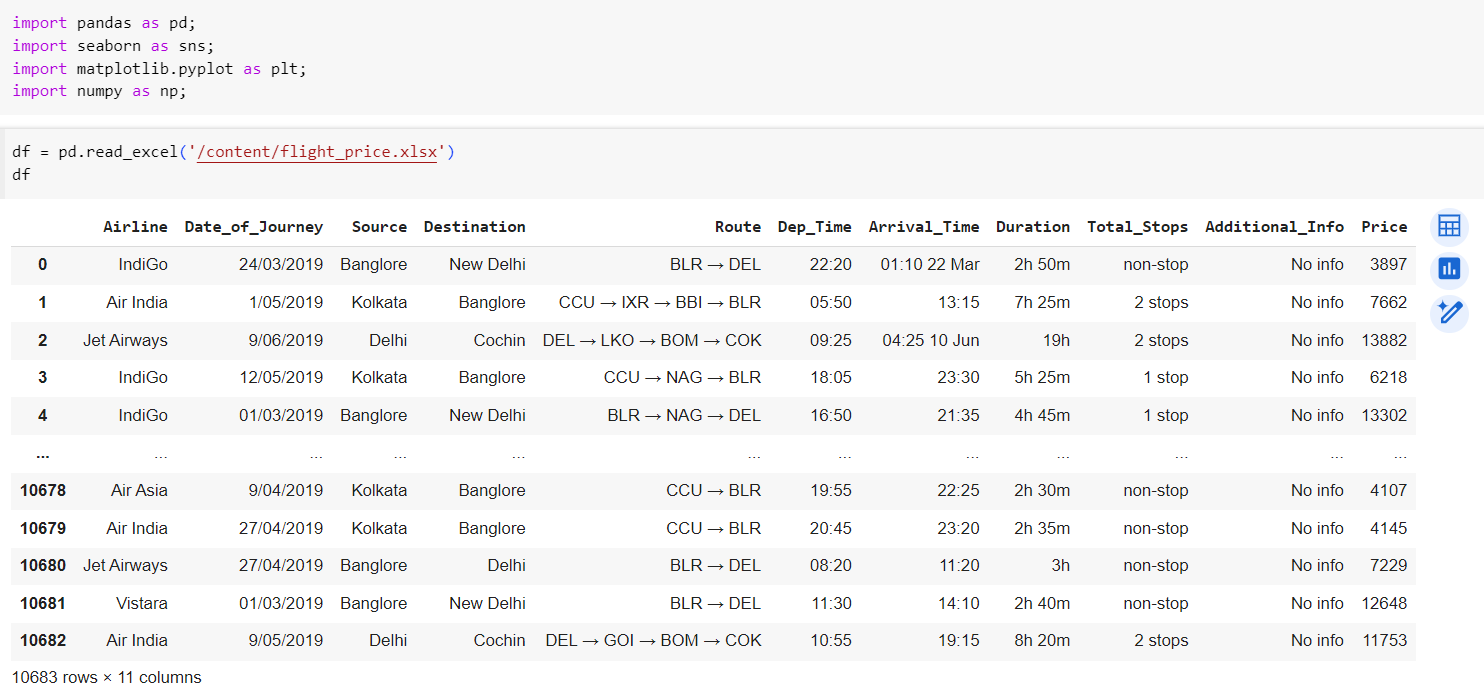
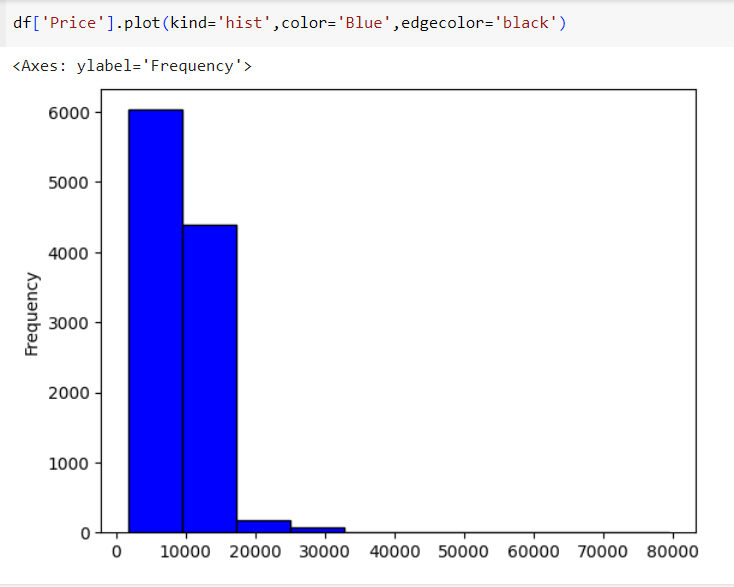
Question 1

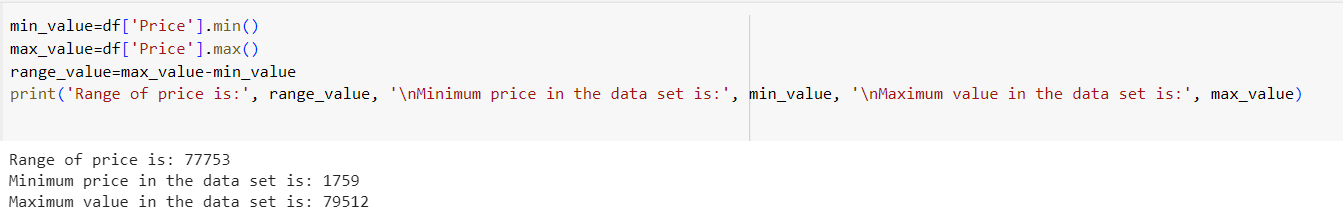


The sheet has 10683 rows and 11 columns

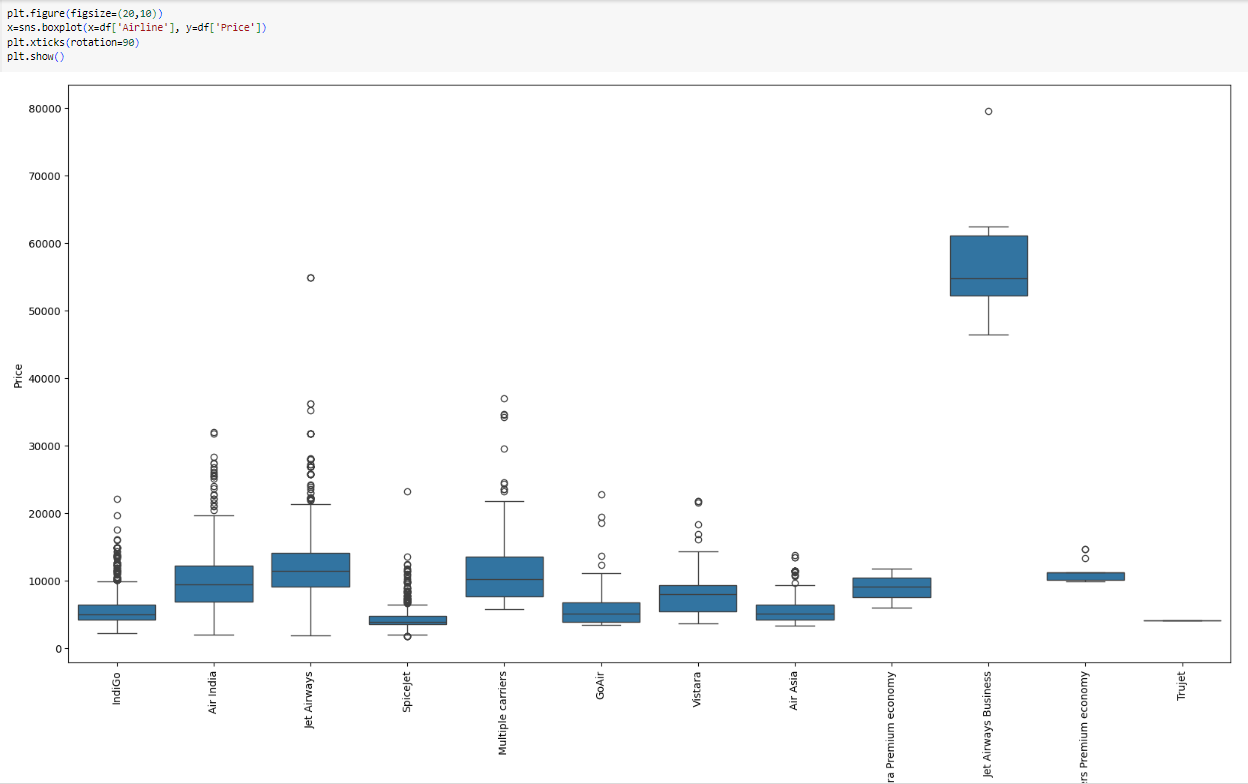
Question 2



Question 3



Question 4



Question 5

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 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.

Question 6

Ans:-To identify trends in flight prices using the Flight Price dataset, I analyzed key features such as:

1.Date or Time: Examining flight prices over different months or seasons to identify price fluctuations throughout the year.

2.Route: Analyzing prices for different routes to understand variations in pricing based on destination and demand.

3.Airline: Investigating average fares offered by different airlines to identify pricing strategies and trends.

4.Booking Class: Understanding pricing differences between economy, business, and first-class bookings.

Question 7

1.Time Series Charts: Showing the trend of flight prices over time, potentially grouped by month or season, to visualize seasonal fluctuations.

2.Bar Charts: Comparing average flight prices across different routes, airlines, or booking classes to identify significant differences.

3.Boxplots: Displaying the distribution of flight prices for different routes or airlines, highlighting outliers and variability.

4.Heatmaps: Visualizing average flight prices across routes and time periods to identify peak travel seasons and popular destinations.

By analyzing these features and utilizing appropriate visualizations, we can effectively identify trends in flight prices and present actionable insights to the team for decision-making.

Question 8

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:

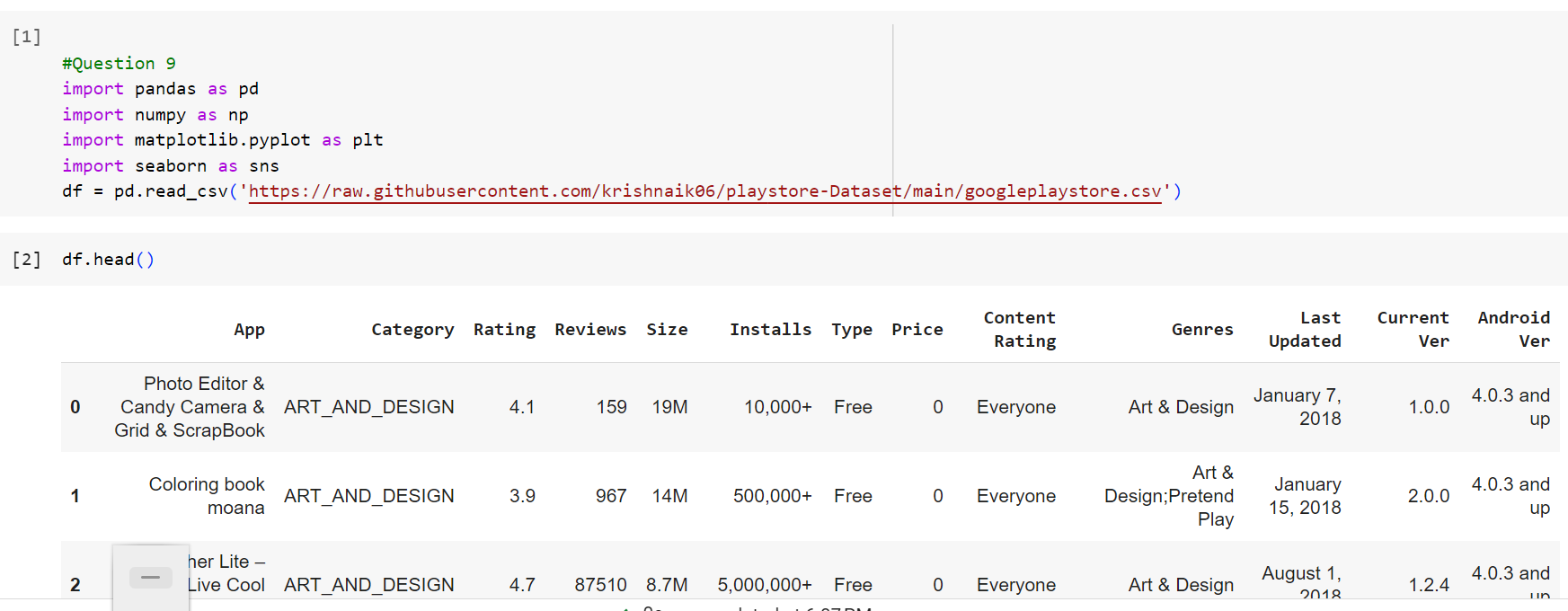
Visualizations: Utilize graphs, charts, and heatmaps to visually represent relationships between flight prices and the identified factors.

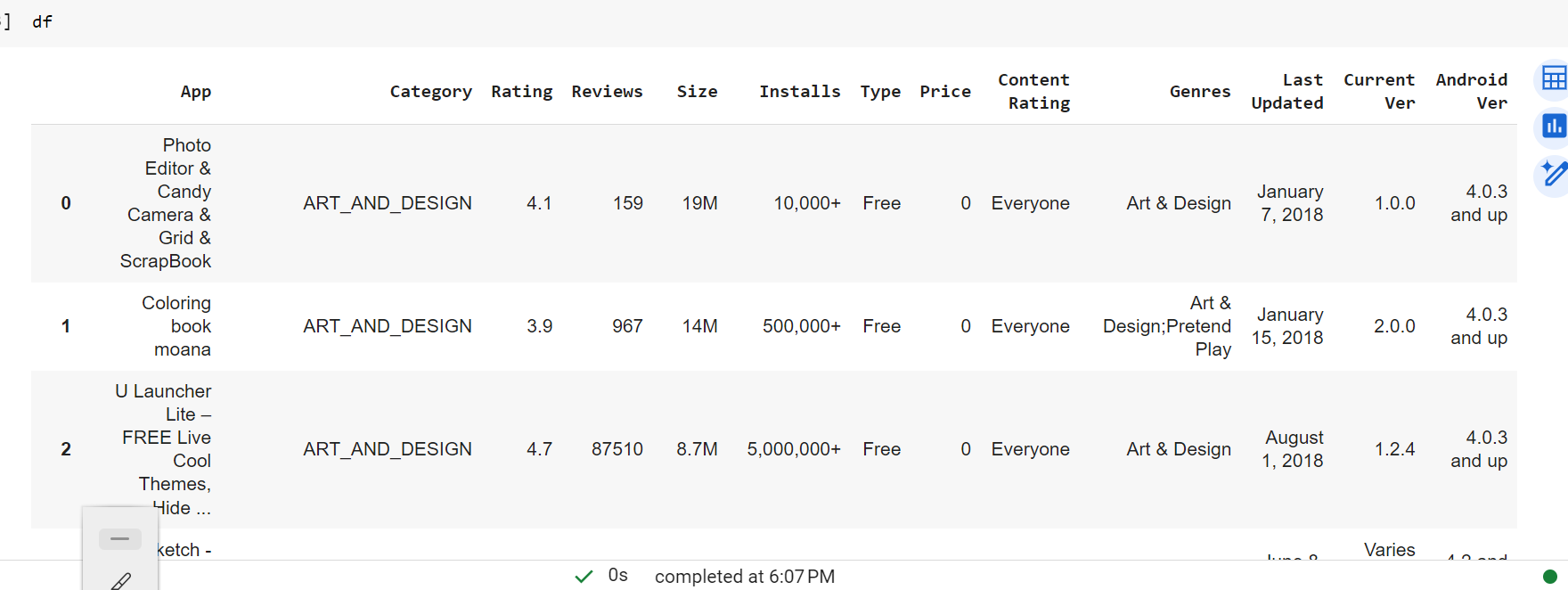
Summary Statistics: Provide key statistical metrics such as mean, median, and standard deviation to summarize the data.

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

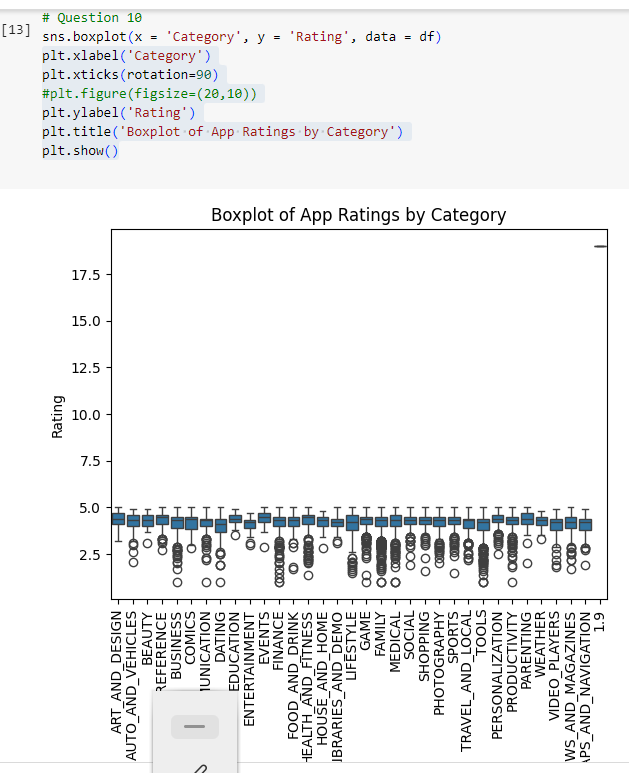


Question 9

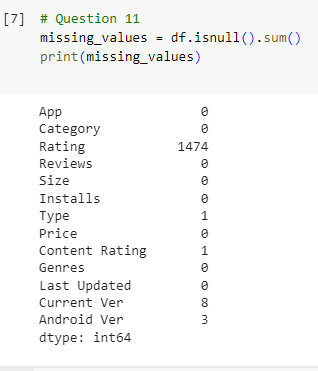




Question 10



Question 11



We can see a lot of null values.

**Reduced Sample Size**: Missing values reduce the number of observations available for analysis. This reduction in sample size can lead to less reliable statistical estimates and potentially bias your results.

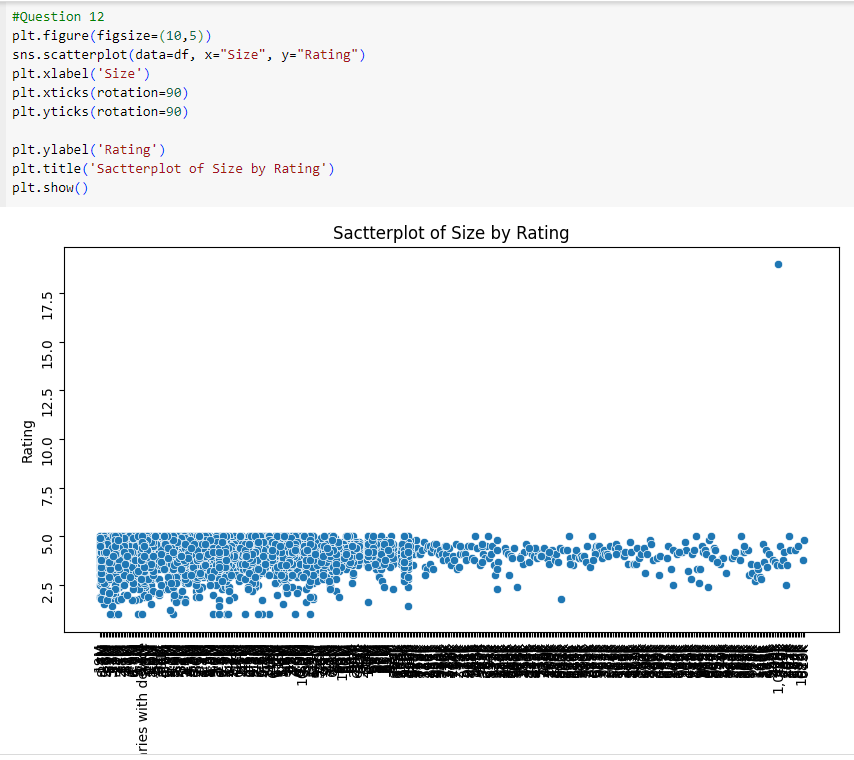
**Biased Estimates**: If missing values are not randomly distributed across the dataset, they may introduce bias into your analysis. For example, if certain demographic groups are more likely to have missing data, it could skew your findings related to those groups.

Impaired Statistical Power: Missing data can reduce the statistical power of your analysis, making it more difficult to detect significant effects or relationships between variables.

**Misleading Results**: Ignoring missing values or improperly handling them can lead to misleading results. For instance, if missing values are systematically related to certain variables, omitting them from the analysis could distort the relationship between variables.

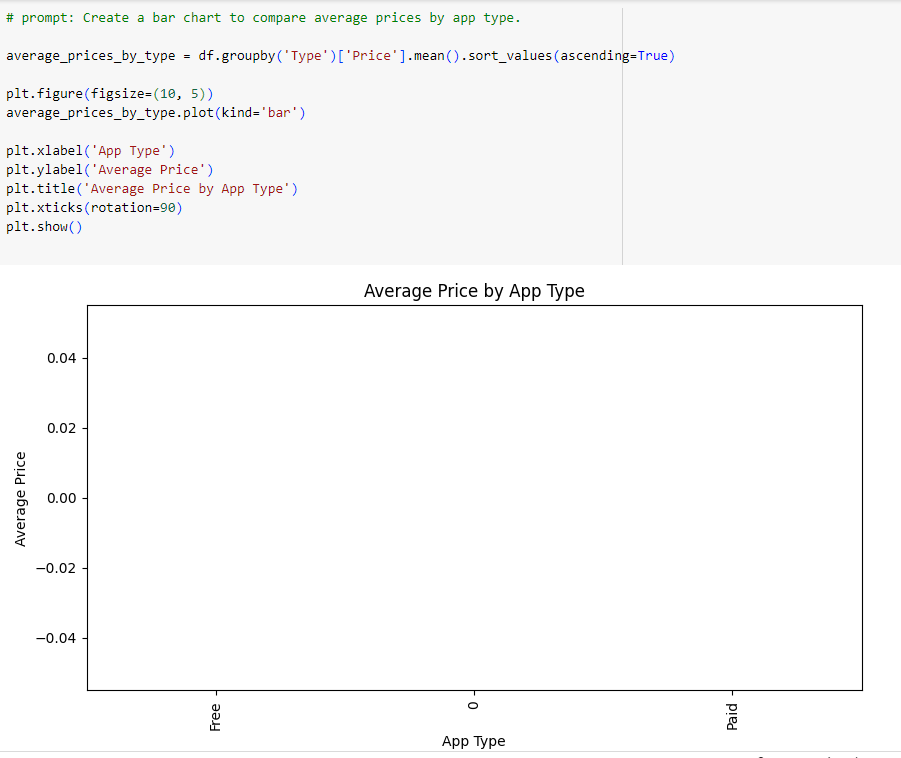
**Imputation Challenges**: Imputing missing values (replacing them with estimated values) can be a solution, but it comes with its own set of challenges. The method used for imputation can affect the results, and imputed values may introduce uncertainty into the analysis.

Question 12

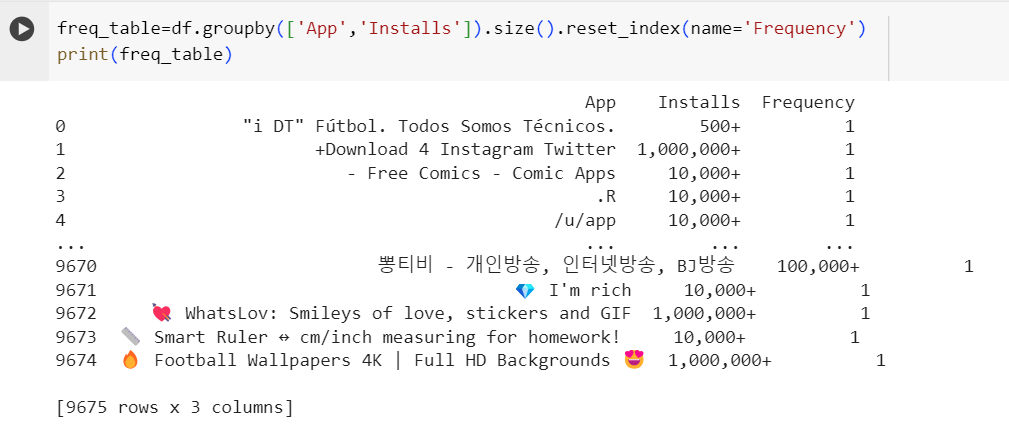


As the size of the app increases the ratings have decreased.

Question 13



Question 14



Question 15

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:

Question 16

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.

Question 17

To determine the best time to launch a new app in the Google Play Store:

* Analyze trend patterns in app installations, ratings, and reviews over time.
* Assess competition and category performance to identify potential niches.
* Utilize line charts for trend analysis, bar charts for comparison, and heatmaps to visualize distribution.
* Evaluate user engagement metrics such as ratings and reviews.
* Employ box plots to understand variability within app categories and make informed launch decisions.

