**Question No. 4**  
**Observations:**

The observation can vary depending on the correlation coefficient values and the scatter plot visualization. Here are possible interpretations based on different scenarios:

1. **Positive Correlation (Correlation coefficient close to 1)**:
   * If both active users and users who have uninstalled the app show a positive correlation between spending capacity and screen time, it suggests that as the amount spent by users increases, their screen time also tends to increase. This may indicate that users who spend more money on the app tend to use it more frequently or for longer durations.
   * The scatter plot would show a general upward trend, indicating that users with higher spending capacity tend to have higher screen time.
2. **Negative Correlation (Correlation coefficient close to -1)**:
   * If both active users and users who have uninstalled the app show a negative correlation between spending capacity and screen time, it suggests that as the amount spent by users increases, their screen time decreases. This could imply that users who spend more money may be more efficient in their app usage or have other priorities, leading to less screen time.
   * The scatter plot would show a general downward trend, indicating that users with higher spending capacity tend to have lower screen time.
3. **No Correlation (Correlation coefficient close to 0)**:
   * If there is no significant correlation between spending capacity and screen time for either group, it suggests that there is no clear relationship between these two variables. This could mean that spending capacity and screen time are independent of each other, and other factors might influence users' behavior.
   * The scatter plot would show a scattered distribution of points with no clear trend, indicating that spending capacity does not have a consistent effect on screen time

**Question No. 5**

**Observations:**

Based on the correlation coefficient and the scatter plot:

**Positive Correlation:** If the correlation coefficient is close to 1, it suggests a positive correlation between ratings and average screen time. Users who give higher ratings tend to have higher average screen time.

**Negative Correlation:** If the correlation coefficient is close to -1, it suggests a negative correlation between ratings and average screen time. Users who give higher ratings tend to have lower average screen time.

**No Correlation**: If the correlation coefficient is close to 0, it suggests no significant correlation between ratings and average screen time. Ratings and average screen time are independent of each other.

**SUMMARY**

Here's a summary of the steps performed:

## **Data Import and Preliminary Analysis:**

Loaded the dataset "userbehaviour.csv".

Checked for null values, column information, and descriptive statistics of the data to understand its structure and characteristics.

## **Exploratory Data Analysis (EDA):**

Explored the distribution of features such as "Average Screen Time", "Average Spent on App (INR)", and "Ratings".

Visualized relationships between different features using scatter plots to gain insights into user behavior.

## **App User Segmentation with K-means Clustering:**

Selected relevant features for clustering, including "Average Screen Time", "Average Spent on App (INR)", and "Ratings".

Scaled the features using StandardScaler to ensure that all features have the same scale.

Determined the optimal number of clusters (K) using the Elbow method.

Applied K-means clustering with the optimal number of clusters to segment users into distinct groups based on similarities in their behavior.

Identified the number of segments obtained from the clustering.

## **Visualization of Segments:**

Visualized the segments using scatter plots for pairs of features, such as "Average Screen Time" vs "Average Spent on App (INR)" and "Average Screen Time" vs "Ratings".

Added the "Last Visited Minutes" feature to the scatter plots to further analyze the segments.

## **Explanation of Observations:**

Observed patterns and relationships between different features to understand user behavior.

Analyzed the characteristics of each segment to identify users that the app retained and lost forever.

Explored the relationship between user engagement metrics (such as screen time and spending) and user satisfaction (ratings) to gain insights into factors influencing user retention.

Overall, the working involved data exploration, clustering-based segmentation, and visualization to understand user behavior and identify strategies for user retention. The insights obtained from this analysis can help in making informed decisions to improve app engagement and user satisfaction.