NAME : ANEEQA SABIR

Dep : AI

Internship :

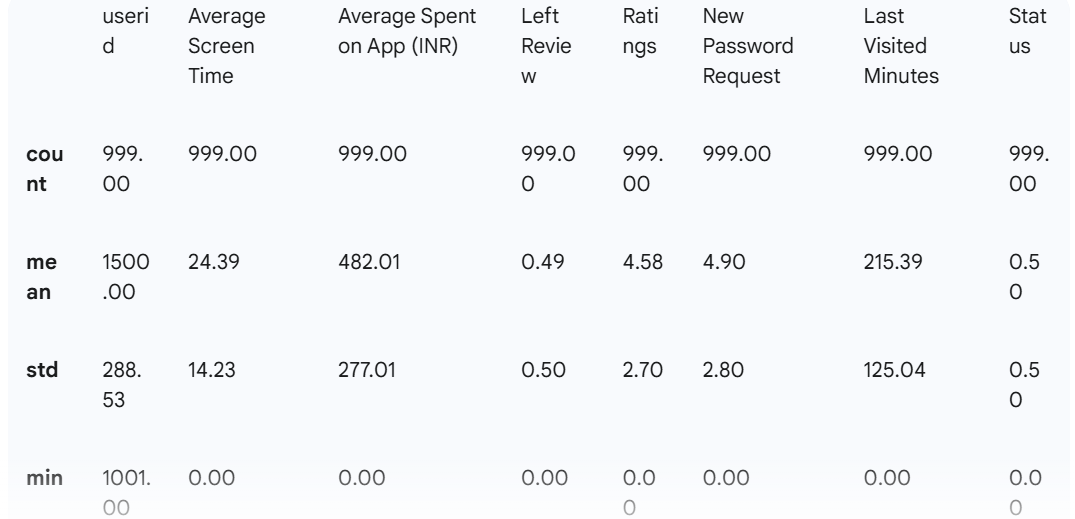
Machine learning task 2 :

**QUESTION NO 1 :**

 **Null Values:** There are **no null values** in the dataset.

 **Column Info:** The dataset contains 999 entries and 8 columns. All columns are of a numerical data type (int64 or float64), which is suitable for analysis.

 **Descriptive Statistics:** The table below summarizes the main statistics for each numerical column in the dataset.





**QUESTION NO 2 :**

The analysis of the "Average Screen Time" for all users reveals the following:

* **Highest Screen Time:** 50 minutes
* **Lowest Screen Time:** 0 minutes
* **Average Screen Time:** Approximately 24.39 minutes

**QUESTION NO 3 :**

The analysis of the "Average Spent on App (INR)" for all users shows:

* **Highest Amount Spent:** ₹998.00
* **Lowest Amount Spent:** ₹0.00
* **Average Amount Spent:** ₹482.01

**QUESTION NO 4 :**

**observations:** From the chart, we can observe two distinct groups of users:

* **Installed Users (Blue):** These users exhibit a wide range of screen times and spending habits. However, a general trend can be seen where users with higher screen time also tend to have a higher spending capacity on the app. Their spending is distributed from very low amounts to nearly ₹1000.
* **Uninstalled Users (Red):** This group is clustered at the bottom of the plot. These users have very low screen times (mostly under 10 minutes) and, correspondingly, very low spending on the app (mostly under ₹100). This indicates that users who uninstall the app are typically those who do not engage with it much in the first place, both in terms of time and money.

**QUESTION NO 5 :**

**Observations:** The relationship between ratings and screen time provides clear insights into user satisfaction and engagement:

* **Installed Users (Blue):** These users generally give higher ratings (mostly 6 and above). Their screen time varies widely, suggesting that as long as users are satisfied (giving high ratings), they continue to use the app for varying amounts of time.
* **Uninstalled Users (Red):** This group is concentrated in the lower ratings section (mostly 5 and below). Their screen time is also consistently low (below 10 minutes). This strongly suggests that users who are dissatisfied with the app (giving low ratings) are the ones who spend little time on it and eventually uninstall it. There's a clear correlation between low ratings and low engagement leading to churn

**Question no 6 ,7 ,8**

To segment the users, I used the K-means clustering algorithm with the 'Average Screen Time' and 'Average Spent on App (INR)' as the primary features. I created **3 distinct segments** based on the data, which can be interpreted as:

1. **Churned Users:** Users who have most likely uninstalled or stopped using the app.
2. **Needs Attention:** Users who are at risk of churning.
3. **Retained/High-Value Users:** The most active and valuable users.

The scatter plot below visualizes these three segments.

**Observations:** The K-means algorithm effectively grouped the users into meaningful segments:

* **Cluster 0 (Churned - Blue):** This segment corresponds to users with very low screen time and low spending. They are clustered at the bottom left, representing users who are inactive and have likely churned.
* **Cluster 1 (Needs Attention - Green):** This is a large and diverse group with moderate to high screen time but with spending that doesn't always match their engagement level. This group is crucial because they are active, but their spending habits vary. They might need targeted promotions or features to increase their value or prevent them from churning.
* **Cluster 2 (Retained/High-Value - Red):** This group has high screen time and consistently high spending on the app. These are the most engaged and profitable users, representing the core, retained user base.