# Data Import and Preprocessing

I imported the necessary libraries for my project:

**from sklearn.preprocessing import LabelEncoder**

**from sklearn.cluster import KMeans**

**import matplotlib.pyplot as plt**

**import pandas as pd**

**import os**

Next, I created a directory named "Graphs" to store the visualizations:

**if not os.path.exists("Graphs"):**

**os.makedirs("Graphs")**

Then, I imported the dataset and performed some initial data exploration:

**data = pd.read\_csv("userbehaviour.csv")**

**null\_values = data.isnull().sum()**

**print("Null values in the dataset:\n", null\_values)**

**column\_info = data.info()**

**print("\nColumn information:\n", column\_info)**

**descriptive\_stats = data.describe()**

**print("\nDescriptive statistics:\n", descriptive\_stats)**

# Analysis of Screen Time and Spending Capacity

I analyzed the screen time and spending capacity data:

**highest\_screen\_time = data['Average Screen Time'].max()**

**print("Highest screen time:", highest\_screen\_time)**

**lowest\_screen\_time = data['Average Screen Time'].min()**

**print("Lowest screen time:", lowest\_screen\_time)**

**average\_screen\_time = data['Average Screen Time'].mean()**

**print("Average screen time:", average\_screen\_time)**

**highest\_amount\_spent = data['Average Spent on App (INR)'].max()**

**print("Highest amount spent:", highest\_amount\_spent)**

**lowest\_amount\_spent = data['Average Spent on App (INR)'].min()**

**print("Lowest amount spent:", lowest\_amount\_spent)**

**average\_amount\_spent = data['Average Spent on App (INR)'].mean()**

**print("Average amount spent:", average\_amount\_spent)**

# Relationship Analysis - Active Users vs. Uninstalled Users

I divided the users into installed and uninstalled categories and visualized their spending capacity against screen time:

**installed\_users = data[data['Status'] == 'Installed']**

**uninstalled\_users = data[data['Status'] == 'Uninstalled']**

**plt.figure(figsize=(12, 8))**

**plt.scatter(installed\_users['Average Screen Time'], installed\_users['Average Spent on App (INR)'],**

**s=installed\_users['Ratings']\*10, c='blue', alpha=0.5, label='Installed')**

**plt.scatter(uninstalled\_users['Average Screen Time'], uninstalled\_users['Average Spent on App (INR)'],**

**s=uninstalled\_users['Ratings']\*10, c='red', alpha=0.5, label='Uninstalled')**

**plt.ylabel('Average Spent on App (INR)')**

**plt.xlabel('Average Screen Time')**

**plt.title('Spending Capacity vs Screen Time')**

**plt.legend()**

**plt.grid(True)**

**plt.savefig('Graphs/Spending Capacity vs Screen Time.png')**

**plt.show()**

# Relationship Analysis - Ratings vs. Screen Time

I further explored the relationship between ratings and screen time for installed and uninstalled apps:

**installed\_apps = data[data['Status'] == 'Installed']**

**uninstalled\_apps = data[data['Status'] == 'Uninstalled']**

**plt.figure(figsize=(12, 8))**

**plt.scatter(installed\_apps['Average Screen Time'], installed\_apps['Ratings'],**

**s=installed\_apps['Average Spent on App (INR)']\*.4, c='blue', alpha=0.5, label='Installed Apps')**

**plt.scatter(uninstalled\_apps['Average Screen Time'], uninstalled\_apps['Ratings'],**

**s=uninstalled\_apps['Average Spent on App (INR)']\*.4, c='red', alpha=0.5, label='Uninstalled Apps')**

**plt.xlabel('Average Screen Time')**

**plt.ylabel('Ratings')**

**plt.title('Ratings vs Screen Time Bubble Graph')**

**plt.legend()**

**plt.grid(True)**

**plt.savefig('Graphs/Ratings vs Screen Time.png')**

plt.show()

# User Segmentation with K-means Clustering

Finally, I performed K-means clustering to segment users into different groups based on their behavior:

**label\_encoder = LabelEncoder()**

**data['Status'] = label\_encoder.fit\_transform(data['Status'])**

**kmeans = KMeans(n\_clusters=3, random\_state=0).fit(data.drop(['userid'], axis=1))**

**data['Cluster'] = kmeans.labels\_**

**colors = {0: 'blue', 1: 'green', 2: 'red'}**

**cluster\_labels = {0: 'Retained', 1: 'Churn', 2: 'Needs Attention'}**

**plt.figure(figsize=(12, 8))**

**for cluster\_num, color in colors.items():**

**cluster\_data = data[data['Cluster'] == cluster\_num]**

**plt.scatter(cluster\_data['Last Visited Minutes'], cluster\_data['Average Spent on App (INR)'], c=color, label=cluster\_labels[cluster\_num], s=50)**

**plt.ylabel('Average Spent on App (INR)')**

**plt.xlabel('Last Visited Minutes')**

**plt.title('User Segmentation with K-means Clustering')**

**plt.legend()**

**plt.savefig('Graphs/User Segmentation with K-means Clustering.png')**

**plt.show()**