Miss. Anjali Singh

Phone Use India Analysis

Using Python

Introduction:

Smartphone usage in India has seen rapid growth, with millions of users relying on mobile devices for communication, entertainment, education, and business. Analysing mobile usage patterns helps understand consumer behaviour, market trends, and key factors influencing smartphone habits. This project aims to study phone usage data in India, uncovering insights into screen time, app usage, data consumption, and preferred activities.

Objective:

➢ **The objective of this project is to analyse phone usage patterns in India, focusing on:**

- Understanding user demographics (age, gender, location).
- Identifying popular phone brands and operating systems (Android vs. iOS).
- Examining daily screen time, data consumption, and app usage behaviour.
- Exploring user preferences for gaming, social media, and streaming services.
- Drawing insights that can help businesses, app developers, and telecom companies optimize services.

Project Objective:

➢ **This project aims to:**

- Analyse smartphone usage trends across different demographics.
- Compare Android and iOS market share in India.
- Identify gaming trends, including the most popular genres and the impact of age on gaming habits.
- Examine mobile data consumption and its relation to different user activities.
- Predict potential user behaviour based on their phone usage patterns

Dataset Description:

- **This dataset provides insights into mobile phone usage patterns across different demographics in India. It includes 17,686 records with 16 key attributes, covering:**

- User Demographics: Age, Gender, Location
- Device Information: Phone Brand, Operating System (OS)
- Usage Statistics: Screen Time (hrs/day), Data Usage (GB/month), Call Duration (mins/day)
- App Usage Behaviour: Social Media Time, Streaming Time, Gaming Time
- Financial Details: E-commerce Spend, Monthly Recharge Cost
- Primary Use of Smartphone: Identifying whether the user primarily uses the phone for social media, gaming, education, or entertainment

Tools and Technologies Used:

- **For this project, we utilized the following tools:**

- **Python**: Main programming language.
  - **Pandas**: Data manipulation.
  - **NumPy**: Numerical operations.
- **Visualization**:
  - **Matplotlib** & **Seaborn**: For data visualization.

Key Question to Address:

1. What is the age distribution of mobile users in India?
2. How does data usage vary by gender?
3. Which phone brands are the most popular?
4. How does the average monthly recharge cost vary by age group?
5. What are the correlations between different usage factors?
6. What are the primary uses of mobile phones?
7. What is the distribution of total screen time between social media usage and streaming?
8. How does call duration vary by gender across different locations?
9. How does average data consumption vary by age group?
10. How is the daily social media usage distributed among users?

Data Loading and Overview:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
# Loading or processing Dataset
df = pd.read_csv('phone_usage_india .csv')
```

I have used Python's pandas library to load the data for analysis. Below is a snapshot of the data to provide an overview of the dataset.

Study Data:

Dataset Preview:

```
df.head(10)
```

	User ID	Age	Gender	Location	Phone Brand	OS	Screen Time (hrs/day)	Data Usage (GB/month)	Calls Duration (mins/day)	Number of Apps Installed	Social Media Time (hrs/day)	E-commerce Spend (INR/month)	Streaming Time (hrs/day)	Gaming Time (hrs/day)
0	U00001	53	Male	Mumbai	Vivo	Android	3.7	23.9	37.9	104	3.9	469	5.2	4.1
1	U00002	60	Other	Delhi	Realme	iOS	9.2	28.1	13.7	169	2.8	4997	5.1	0.4
2	U00003	37	Female	Ahmedabad	Nokia	Android	4.5	12.3	66.8	96	3.0	2381	1.7	2.9
3	U00004	32	Male	Pune	Samsung	Android	11.0	25.6	156.2	146	5.2	1185	3.2	0.3
4	U00005	16	Male	Mumbai	Xiaomi	iOS	2.2	2.5	236.2	86	5.5	106	3.4	2.3
5	U00006	21	Male	Jaipur	Oppo	iOS	5.4	10.6	210.6	25	4.2	6285	0.6	4.8
6	U00007	57	Female	Lucknow	Apple	iOS	6.0	35.2	154.5	123	0.8	2653	2.9	2.3
7	U00008	56	Other	Kolkata	Realme	iOS	3.1	43.5	125.3	188	2.3	9767	5.2	5.0
8	U00009	46	Female	Kolkata	Oppo	Android	5.3	46.4	21.3	194	3.7	2870	6.1	2.8
9	U00010	44	Other	Kolkata	Apple	iOS	9.9	10.6	180.2	84	4.8	9193	7.6	0.4

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17686 entries, 0 to 17685
Data columns (total 16 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   User ID                       17686 non-null  object
 1   Age                           17686 non-null  int64
 2   Gender                        17686 non-null  object
 3   Location                      17686 non-null  object
 4   Phone Brand                   17686 non-null  object
 5   OS                            17686 non-null  object
 6   Screen Time (hrs/day)         17686 non-null  float64
 7   Data Usage (GB/month)         17686 non-null  float64
 8   Calls Duration (mins/day)     17686 non-null  float64
 9   Number of Apps Installed      17686 non-null  int64
 10  Social Media Time (hrs/day)   17686 non-null  float64
 11  E-commerce Spend (INR/month)  17686 non-null  int64
 12  Streaming Time (hrs/day)      17686 non-null  float64
 13  Gaming Time (hrs/day)         17686 non-null  float64
 14  Monthly Recharge Cost (INR)   17686 non-null  int64
 15  Primary Use                   17686 non-null  object
dtypes: float64(6), int64(4), object(6)
memory usage: 2.2+ MB
```

```
# display statistcal summary
df.describe()
```

	Age	Screen Time (hrs/day)	Data Usage (GB/month)	Calls Duration (mins/day)	Number of Apps Installed	Social Media Time (hrs/day)	E-commerce Spend (INR/month)	Streaming Time (hrs/day)	Gaming Time (hrs/day)	Monthly Recharge Cost (INR)
count	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000
mean	37.584247	6.546376	25.411257	151.405846	104.584869	3.252369	5075.707848	4.250616	2.490874	1042.785367
std	13.338252	3.172677	14.122167	84.923353	55.217097	1.590223	2871.604841	2.155683	1.446003	552.502067
min	15.000000	1.000000	1.000000	5.000000	10.000000	0.500000	100.000000	0.500000	0.000000	100.000000
25%	26.000000	3.800000	13.200000	77.325000	57.000000	1.900000	2587.500000	2.400000	1.200000	561.000000
50%	38.000000	6.600000	25.300000	150.600000	104.000000	3.200000	5052.000000	4.200000	2.500000	1040.000000
75%	49.000000	9.300000	37.600000	223.900000	152.000000	4.600000	7606.000000	6.100000	3.700000	1521.750000
max	60.000000	12.000000	50.000000	300.000000	200.000000	6.000000	10000.000000	8.000000	5.000000	2000.000000

Data Cleaning:

```
#Finding duplicates;
df_cleaned = df.drop_duplicates()
print("original rows:",len(df))
print("Rows after removing duplicates:",len(df_cleaned))
```

```
original rows: 17686
Rows after removing duplicates: 17686
```

```
#checking for missing values

Missing_values= df.isnull().sum()
print("Missing values in each column:\n",Missing_values)
```

```
Missing values in each column:
 User ID                        0
Age                            0
Gender                         0
Location                       0
Phone Brand                    0
OS                             0
Screen Time (hrs/day)          0
Data Usage (GB/month)          0
Calls Duration (mins/day)      0
Number of Apps Installed       0
Social Media Time (hrs/day)    0
E-commerce Spend (INR/month)   0
Streaming Time (hrs/day)       0
Gaming Time (hrs/day)          0
Monthly Recharge Cost (INR)    0
Primary Use                    0
dtype: int64
```

Exploratory Data Analysis (EDA):

❖ Jupyter Notebook:

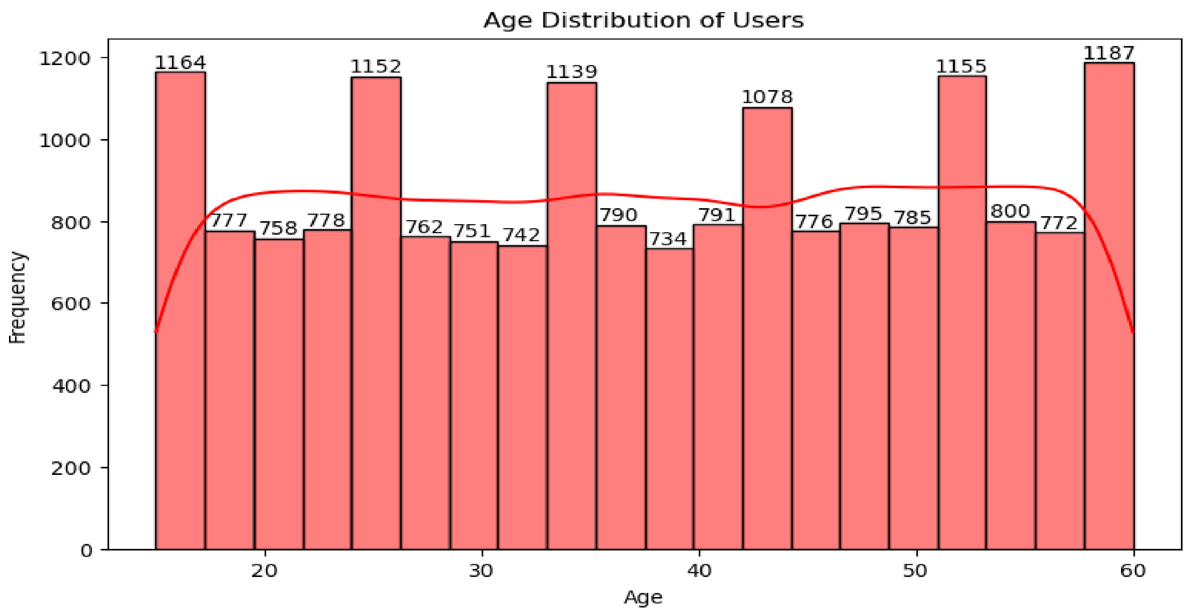
Used to write, run, and document Python code. It's an interactive environment where code, visualizations, and markdowns can be presented together, making it ideal for analysis. Seaborn and Matplotlib were used during EDA to create visualizations and explore data patterns.

1. Age Distribution:

Question: What is the age distribution of mobile users in India?

- **Data Load & Overview:**

```
# Creating a visualization
fig, ax = plt.subplots(figsize=(7, 6))
ax = sns.histplot(df['Age'], bins=20, kde=True, color="Orange")
plt.title('Age Distribution of Users')
plt.xlabel('Age')
plt.ylabel('Frequency')
for container in ax.containers:
    ax.bar_label(container, fmt="%.0f")
plt.show()
```



- **Details:**

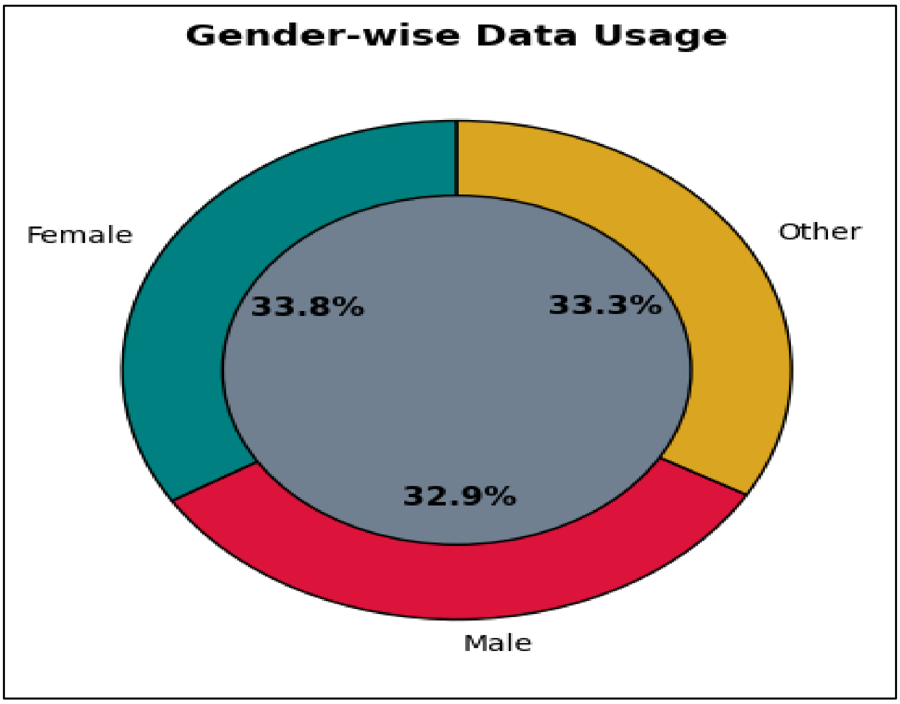
The histogram represents the **age distribution** of mobile phone users, ranging from **18 to 60 years**. The x-axis shows age, while the y-axis represents the number of users in each group. Peaks are observed around **18, 25, 30, 40, 50, and 60 years**, with some age groups exceeding **1100 users**. The distribution is fairly uniform, with slight fluctuations. The **KDE curve** (red line) highlights the overall trend. This graph effectively visualizes user demographics, helping identify key **target age groups** for mobile usage insights.

2. Gender-wise Data Usage:

Question: How does data usage vary by gender?

- **Data Load & Overview:**

```
# Grouping data based on specific columns
gender_data = df.groupby('Gender')['Data Usage (GB/month)'].sum()
colors = ['teal', 'crimson', 'goldenrod']
# Creating a visualization
fig, ax = plt.subplots(figsize=(5, 5))
ax.pie(gender_data, labels=gender_data.index, autopct='%1.1f%%', startangle=90,
    colors=colors, wedgeprops={'edgecolor': 'black'},
    textprops={'color': 'black', 'fontsize': 11, 'weight': 'bold'})
plt.gca().add_artist(plt.Circle((0, 0), 0.70, fc='slategray', edgecolor='black'))
plt.title("Gender-wise Data Usage", fontsize=13, fontweight="bold")
plt.show()
```



- **Details:**

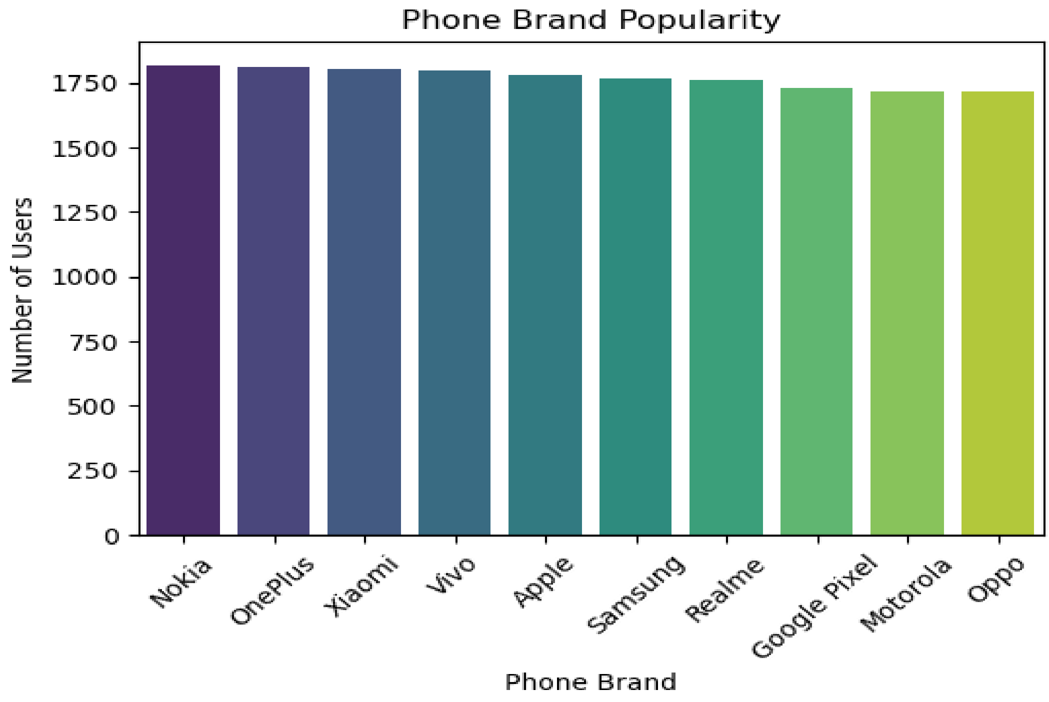
The donut chart represents monthly data consumption across different gender groups. The three segments, Female (33.8%), Male (32.9%), and Other (33.3%)—show that data usage is fairly evenly distributed among all genders. The inner gray circle enhances the visual clarity of the distribution. This chart effectively highlights that no single gender dominates mobile data consumption, indicating balanced digital engagement

3. Phone Brand Popularity:

Question: Which phone brands are the most popular?

- **Load Data & Overview:**

```
# Creating a visualization
fig, ax = plt.subplots(figsize=(7, 6))
ax = sns.countplot(x='Phone Brand', data=df, palette='viridis', order=df['Phone Brand'].value_counts().index)
ax.set_title('Phone Brand Popularity')
ax.set_xlabel('Phone Brand')
ax.set_ylabel('Number of Users')
plt.xticks(rotation=45)
plt.show()
```



❖ **Details:**

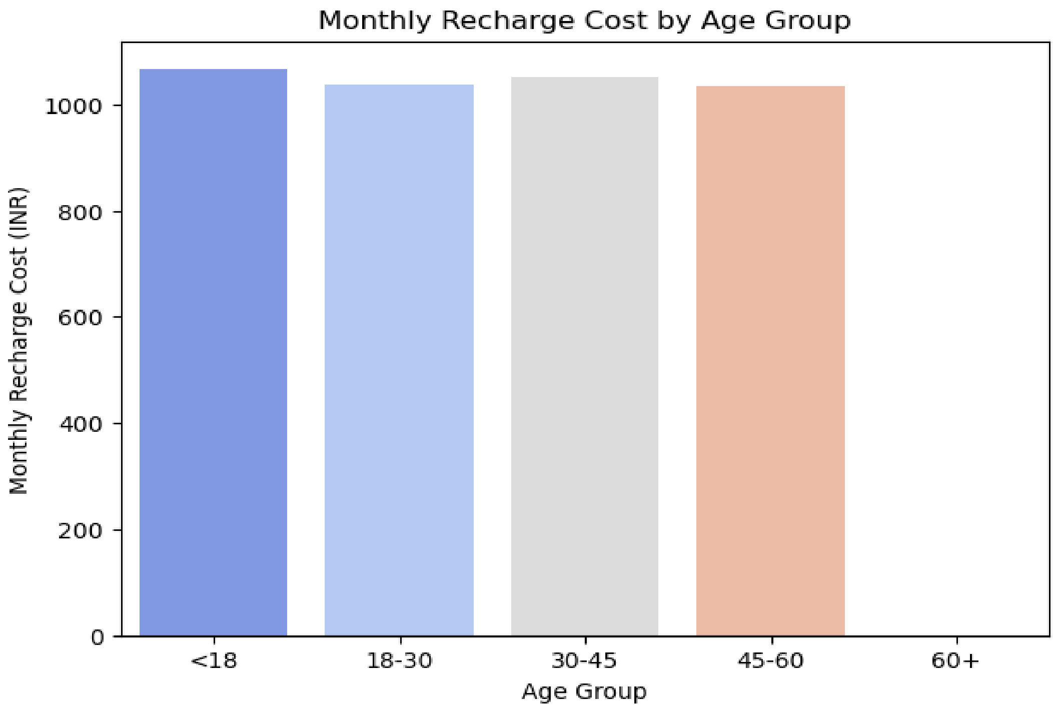
The **bar chart** represents the popularity of different **phone brands** based on the **number of users**. The x-axis lists various brands like **Nokia, OnePlus, Xiaomi, Vivo, Apple, Samsung, etc.**, while the y-axis shows the **user count**. The bars are sorted in descending order, indicating that all brands have a similar number of users. This visualization helps understand **brand preference trends** among mobile users.

4. Monthly Recharge Cost by Age Group:

Question: How does the average monthly recharge cost vary by age group?

❖ **Load Data & Overview:**

```
age_bins = [0, 18, 30, 45, 60, 100]

labels = ['<18', '18-30', '30-45', '45-60', '60+']

df['Age Group'] = pd.cut(df['Age'], bins=age_bins, labels=labels)

fig, ax = plt.subplots(figsize=(7, 5))

# Creating a visualization

ax = sns.barplot(data=df, x='Age Group', y='Monthly Recharge Cost (INR)',
errorbar=None,

palette=sns.color_palette("coolwarm", len(labels)))

ax.set_title('Monthly Recharge Cost by Age Group')

ax.set_xlabel('Age Group')

ax.set_ylabel('Monthly Recharge Cost (INR)')

plt.show()
```



❖ **Details:**

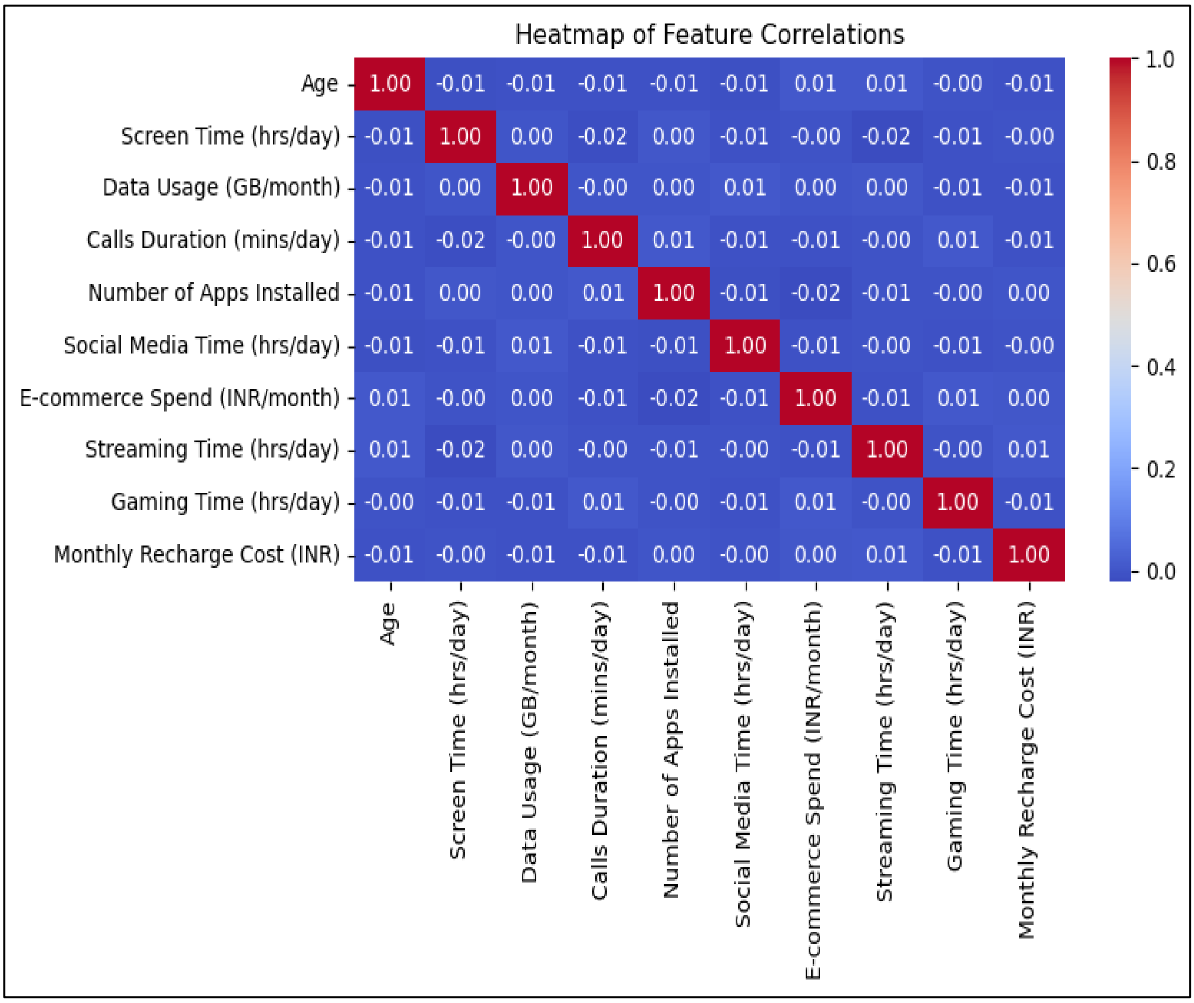
This **bar chart** represents the **monthly recharge cost (INR) across different age groups**. The x-axis categorizes users into age groups (**<18, 18-30, 30-45, 45-60, 60+**), while the y-axis shows their **average monthly recharge expenditure**. The costs appear fairly consistent across all age groups, with slight variations. This visualization helps understand **how different age groups spend on mobile recharges**.

5. **Heatmap of Correlations:**

Question: What are the correlations between different usage factors?

❖ **Load Data & Overview:**

```
# Creating a visualization
fig, ax = plt.subplots(figsize=(8,4 ))
numeric_columns = df.select_dtypes(include=[np.number])
correlation = numeric_columns.corr()
ax = sns.heatmap(correlation, annot=True, fmt='.2f', cmap='coolwarm')
ax.set_title('Heatmap of Feature Correlations')
plt.show()
```



❖ **Details: -**

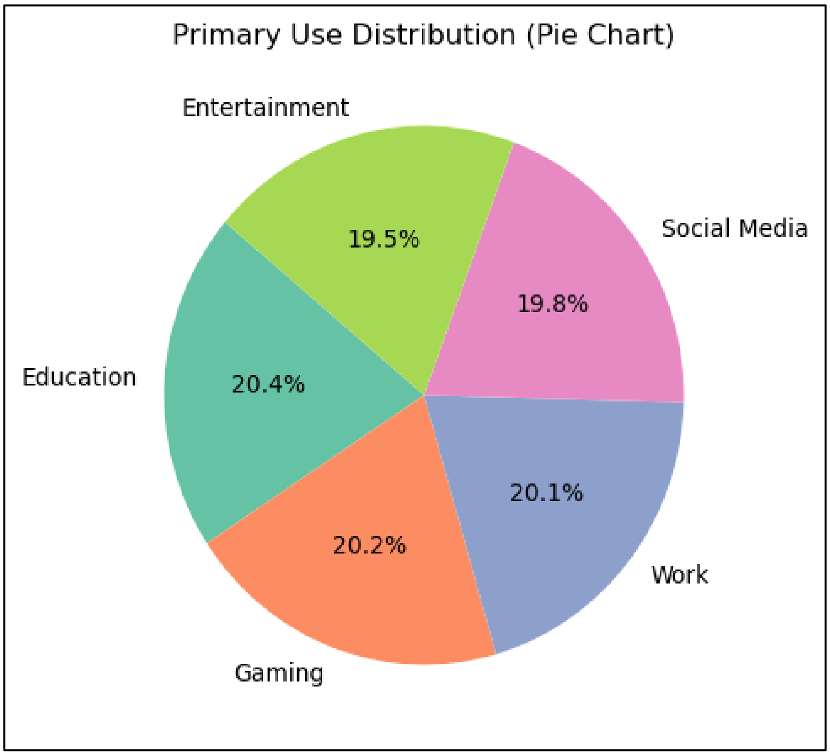
This **heatmap** represents the **correlation** between different numerical features in the dataset. The **colour intensity** indicates the strength of relationships between variables. A value **close to 1** (red) shows a strong positive correlation, while **values near 0** (blue) indicate little to no correlation. The heatmap helps identify **patterns** and **dependencies** among features like screen time, data usage, social media time, and monthly recharge cost.

6. Primary Use Distribution:

Question: What are the primary uses of mobile phones?

- **Load Data & Overview:**

```
# Creating a visualization
fig, ax = plt.subplots(figsize=(5, 5))
primary_use_counts = df['Primary Use'].value_counts()
ax.pie(primary_use_counts, labels=primary_use_counts.index,
autopct='%1.1f%%',
startangle=140, colors=sns.color_palette('Set2',
len(primary_use_counts)))
ax.set_title('Primary Use Distribution (Pie Chart)')
plt.show()
```



- **Details:**

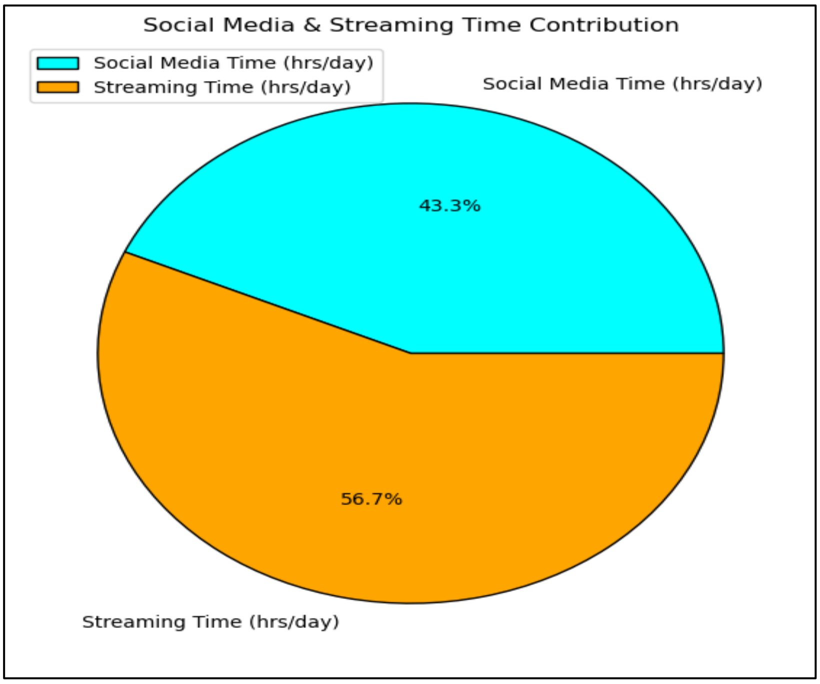
This **pie chart** represents the percentage distribution of different **primary uses** of mobile devices. The categories include **Education (20.4%)**, **Gaming (20.2%)**, **Work (20.1%)**, **social media (19.8%)**, and **Entertainment (19.5%)**. The values are **evenly distributed**, indicating that users engage in a balanced mix of activities on their devices.

7. Social Media Time vs. Streaming Time:

Question: What is the distribution of total screen time between social media usage and streaming?

- **Load Data & Overview:**

```
plt.figure(figsize=(6,6))

time_data = df[['Social Media Time (hrs/day)', 'Streaming Time
(hrs/day)']].mean()

# Creating a visualization

time_data.plot.pie(autopct='%1.1f%%', colors=['cyan', 'orange'],
wedgeprops={'edgecolor': 'black'})

plt.title('Social Media & Streaming Time Contribution')

plt.ylabel('')

plt.legend(time_data.index, loc="best")

plt.show()
```



- **Details:**

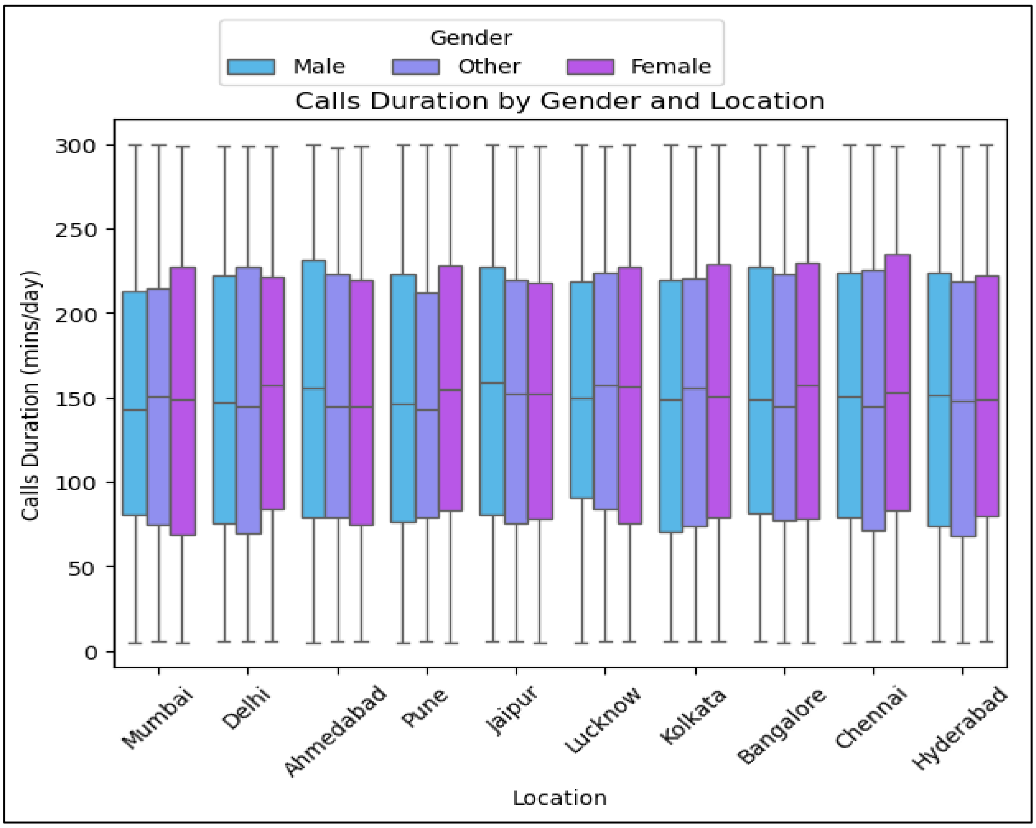
This **pie chart** visualizes the average contribution of **social media Time (43.3%)** and **Streaming Time (56.7%)** in users' daily screen usage. **Streaming Time dominates**, indicating that users spend more time watching content compared to social media engagement.

8. Calls Duration by Gender and Location:

Question: How does call duration vary by gender across different locations?

- **Load Data & Overview:**

```
fig, ax = plt.subplots(figsize=(7, 5))
# Creating a visualization
ax = sns.boxplot(data=df, x='Location',
y='Calls Duration (mins/day)', hue='Gender', palette='cool')
ax.set_title('Calls Duration by Gender and Location')
ax.set_xlabel('Location')
ax.set_ylabel('Calls Duration (mins/day)')
plt.xticks(rotation=45)
plt.legend(title='Gender', bbox_to_anchor=(0.4, 1.20), loc='upper center', ncol=3)
plt.show()
```



- **Details:**

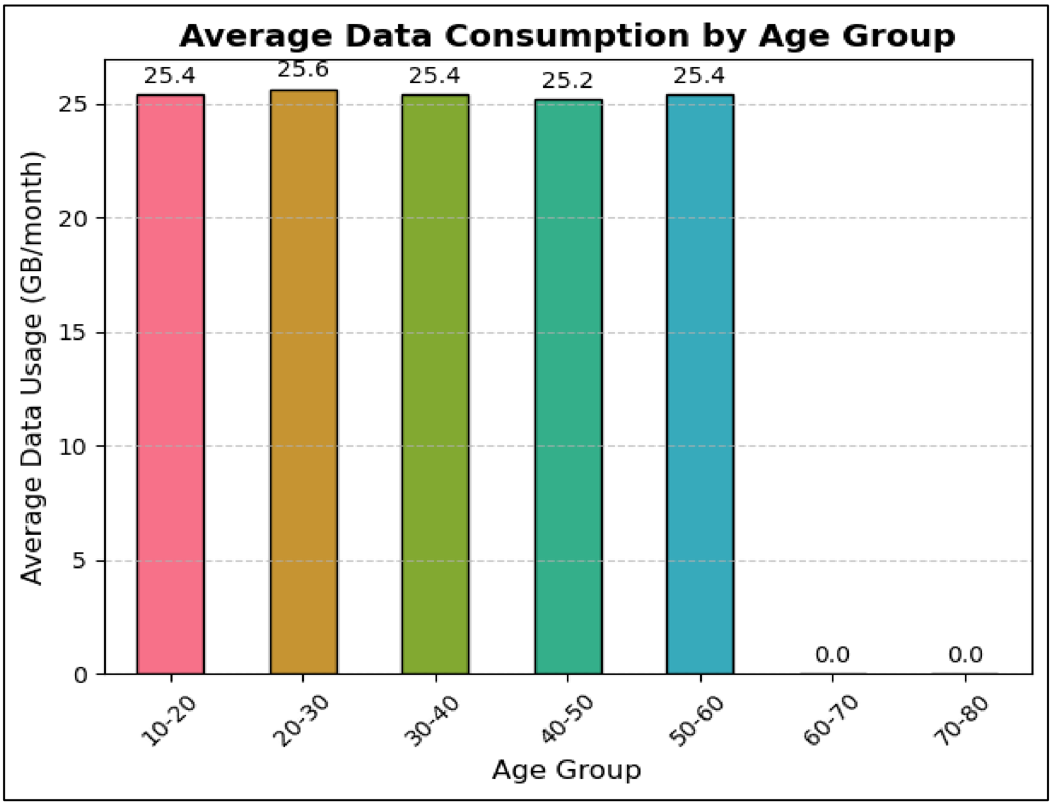
This **box plot** visualizes **daily call duration (mins)** across different **locations** for **Male, Female, and other** genders. The **median call duration** is fairly consistent across cities, with some variation. The **spread (IQR)** suggests that some users have significantly higher call times, but no extreme outliers.

9. Average Data Consumption by Age Group:

Question: How does average data consumption vary by age group?

- **Load Data & Overview:**

```
plt.figure(figsize=(7,5))
age_group = pd.cut(df['Age'], bins=[10, 20, 30, 40, 50, 60, 70, 80],
                   labels=["10-20", "20-30", "30-40", "40-50", "50-60", "60-70", "70-80"])
colors = sns.color_palette("husl", len(age_group.cat.categories))
# Creating a visualization
ax = df.groupby(age_group)['Data Usage (GB/month)'].mean().plot(kind='bar', color=colors,
edgecolor='black')
plt.title('Average Data Consumption by Age Group', fontsize=14, fontweight='bold')
plt.xlabel('Age Group', fontsize=12)
plt.ylabel('Average Data Usage (GB/month)', fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
for container in ax.containers:
    ax.bar_label(container, fmt="%.1f", padding=3, fontsize=10)
plt.show()
```



- ❖ **Details:**

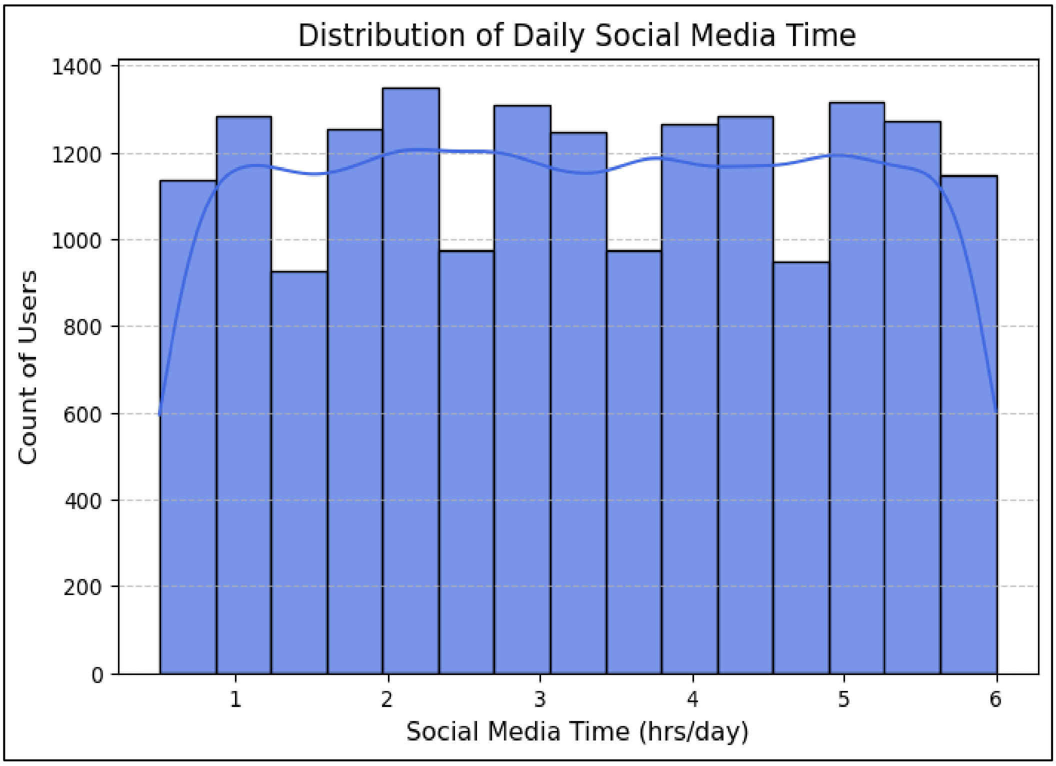
This **bar chart** displays the **average monthly data usage (GB)** across different **age groups**. The **20-30 age group** has the highest consumption (**25.6 GB/month**), while other age groups (10-60) have similar usage (~25 GB). **60+ age groups show no data usage**, indicating minimal or no digital engagement.

10. Daily social media usage distributed among users:

Question: How is the daily social media usage distributed among users?

- **Load Data & Overview:**

```
plt.figure(figsize=(8, 5))
# Creating a visualization
sns.histplot(df['Social Media Time (hrs/day)'], bins=15, kde=True,
color='royalblue', edgecolor='black', alpha=0.7)
plt.title('Distribution of Daily Social Media Time', fontsize=14)
plt.xlabel('Social Media Time (hrs/day)', fontsize=12)
plt.ylabel('Count of Users', fontsize=12)
# Adding grid for better readability
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```



❖ **Details:**

This **histogram** shows how many users spend different amounts of time on **social media daily**. The data is grouped into **15 bins**, with a **KDE (Kernel Density Estimation) curve** overlaid to show the distribution trend. Most users spend between **1 to 5 hours per day** on social media, with peaks around **2 and 5 hours**. The **gradual decline at higher hours** suggests fewer users spend excessive time online.

Conclusion:

This project provided an in-depth analysis of mobile phone usage trends across different demographics in India. By leveraging data visualization and statistical analysis, we explored key aspects such as call duration, data consumption, and social media usage. The findings highlight significant variations based on age, gender, and location, offering valuable insights into user behaviour. This study serves as a foundation for understanding evolving mobile usage trends and their broader implications on digital connectivity and communication habits.

Future Scope:

As mobile technology continues to evolve, this study can be expanded in several ways:

- **Integration of Emerging Technologies:** The impact of **5G, AI-driven analytics, and IoT** on mobile usage can be explored in future research.
- **Behavioral Analysis with Machine Learning:** Predictive models can be developed to forecast trends in data consumption and user behaviour.
- **Impact of Digital Policies:** Studying the effects of government regulations, such as data privacy laws and internet accessibility policies, on mobile usage trends.

Final Thoughts:

This project not only enhances our understanding of mobile usage in India but also provides valuable insights for telecom companies, policymakers, and digital marketers. With the rapid digital transformation, continuous research and data-driven decision-making will play a crucial role in shaping the future of mobile communication in the country.