

Phone Usage Data of India.

Algorithm used: Data Visualization , Exploratory Data Analysis



A

ADM Course Project Report in
partial fulfilment of the degree

Bachelor of Technology
in
Computer Science & Engineering

By

Name: V. Preetham

HT No :(2303a51387)

Name: M.Rushikesh

HT No :(2303a51399)

Name: A.VASANTHA
SHOBA RANI.

HT NO :(2303a51395)

Name: S.AKSHITH
REDDY

HT NO :(2303a51383)

Name: Ch. Bhanu Prakash

HT NO :(2303a51407)

Under the guidance of

Bediga Sharan
Assistant Professor

Submitted to

School of Computer Science and Artificial Intelligence

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**CERTIFICATE**

This is to certify that the **Applications of Data Mining– Course Project** Report entitled “**Phone Usage Data of India.**” is a record of bonafide work carried out by the student(s) “**V. Preetham, Ch. Bhanu Prakash , S. Akshith Reddy, A. Vasantha Shoba Rani, M. Rishikesh**” bearing **Hallticket No(s) 2303a51387 —, 2303a51407, 2303a51383, 2303a51395, 2303a51399**, during the academic year 2024-25 in partial fulfillment of the award of the degree of *Bachelor of Technology* in **Computer Science & Engineering** by the SR University, Warangal.

Supervisor

(Mr. Bediga Sharan)

Assistant Professor

Head of the Department

(Dr. M. Sheshikala)

Professor

ORGANIZATION OF REPORT

1. Title page
2. Certificate
3. Table of Contents
4. Abstract

OBJECTIVE OF THE PROJECT

1. DEFINITIONS OF THE ELEMENTS USED IN THE PROJECT
2. DESIGN
 1. SCREENS
3. IMPLEMENTATION
 1. CODE
4. RESULTSCREENS
5. CONCLUSION

ABSTRACT

This project aims to analyze mobile phone usage behavior among Indian users using a simulated dataset consisting of 17,686 entries. The dataset covers various demographic and behavioral metrics such as age, gender, location, phone brand, screen time, data usage, app installations, social media, and gaming patterns. Through detailed data preprocessing, visualization, and exploratory analysis, insights were derived about user preferences and digital habits. The study also highlights behavioral trends across gender and locations, such as time spent on social media, primary phone usage purposes, and e-commerce activities. This analysis can aid telecom companies, marketers, and policymakers in understanding mobile engagement patterns to design more user-focused services and digital initiatives.

OBJECTIVE OF THE PROJECT

to analyze and visualize mobile phone usage patterns among Indian users. By examining various attributes such as screen time, data usage, social media activity, and primary use of phones, the goal is to extract meaningful insights from the dataset to understand trends, behaviors, and correlations in digital usage across different demographics. This project aims to:

Identify common phone brands and operating systems.

Analyze digital habits across age, gender, and location.

Provide visual interpretations of user behavior to aid data-driven decision-making.

DEFINITIONS OF THE ELEMENTS USED

- **Dataset:** mobile phone usage behavior among Indian users using a simulated dataset consisting of 17,686 entries
- **Outcome:** Target variable indicating analysis of phone usage
- **Algorithms Used:**
 - Exploratory Data Analysis
 - Data Visualization Algorithms

- **Metrics:**

- Age
- Screen Time (hrs/day)
- Data Usage (GB/month)
- Social Media Time (hrs/day)
- E-commerce Spend (INR/month)

DESIGN:

Git hub url of Project->

V. Preetham : <https://github.com/Preetham1928/ADMPROJECT>

M.Rushikesh : <https://github.com/rishikesh715/ADM/blob/main/ADMPRO.ipynb>

Ch. Bhanu Prakash : <https://github.com/bhanusprakash/2303A51407-ADM/blob/main/ADMPRO.ipynb>

A.VASANTHA SHOBA RANI. : <https://github.com/2303A51395/adm>

S.AKSHITH REDDY :
<https://github.com/rishikesh715/ADM/blob/main/ADMPRO.ipynb>

7.1 SCREENS

As this is a backend machine learning model project, it does not involve front-end GUI screens. Instead, the outputs are presented through Jupyter/Colab notebooks with printed metrics and visual plots.

Simple Implementation:

Load dataset

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
print("df shape: ")
print(f"{df.shape}\n")
```

```
print("df info: ")
print(f"{df.info()}\n")
```

```
print("df isnull sum: ")
print(f"{df.isnull().sum()}\n")
```

```
print("df describe: ")
print(f"{df.describe()}\n")
```

```
df shape:
(17686, 16)

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
```


CODE:

Exploratory Data

Analysis

```
# Age Distribution
# Get the top 10 most common ages
top_10_ages = df["Age"].value_counts().sort_values(ascending=False).head(10)

# Plot the distribution of the top 10 ages using a histogram with KDE
sns.displot(top_10_ages, kde=True, color="teal")
plt.show()

# Plot the top 10 ages as a bar plot
df["Age"].value_counts().sort_values(ascending=False).head(10).plot(kind='bar',
, figsize=(6, 4))
plt.title("Top 10 Age of users")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()

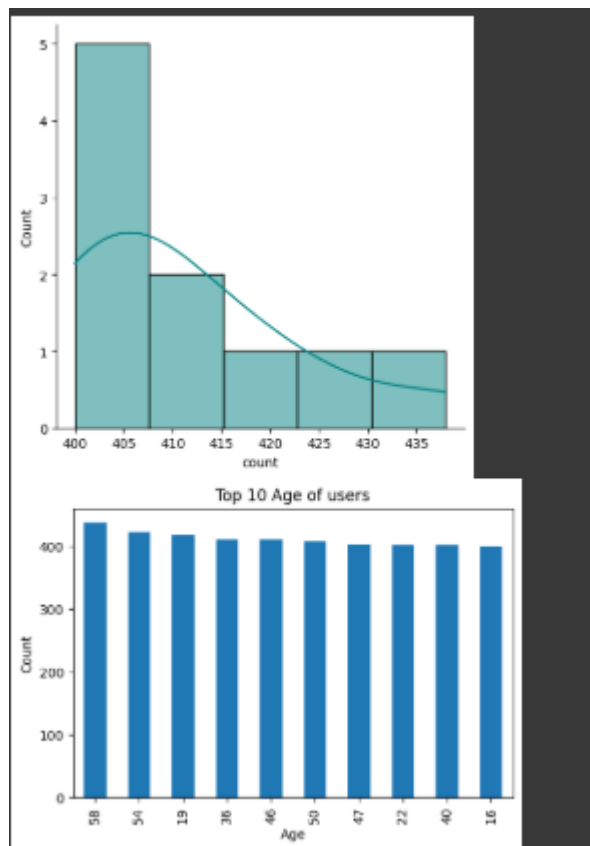
# Plot the gender counts as a bar plot
df["Gender"].value_counts().sort_values(ascending=False).head(10).plot(kind='bar',
, figsize=(6, 4), color="teal")
plt.title("Gender Counts")
plt.xlabel("Gender")
plt.ylabel("Count")
plt.show()

# Get the counts of the top 10 age groups for each gender
gender_age_counts =
df.groupby("Gender")["Age"].value_counts().sort_values(ascending=False).head(10).reset_index(name='Count')

# Plot the counts of the top 10 age groups for each gender as a bar plot
plt.figure(figsize=(6, 4))
sns.barplot(data=gender_age_counts, x='Gender', y='Count', hue='Age',
color='orange')
plt.title("Gender and Age Counts")
plt.show()
```

Output:

The Graphical Representation For Just Formate,



```
user_social_media = df.groupby("Gender")["Social Media Time  
(hrs/day)"].sum().sort_values(ascending=False).head(10)
```

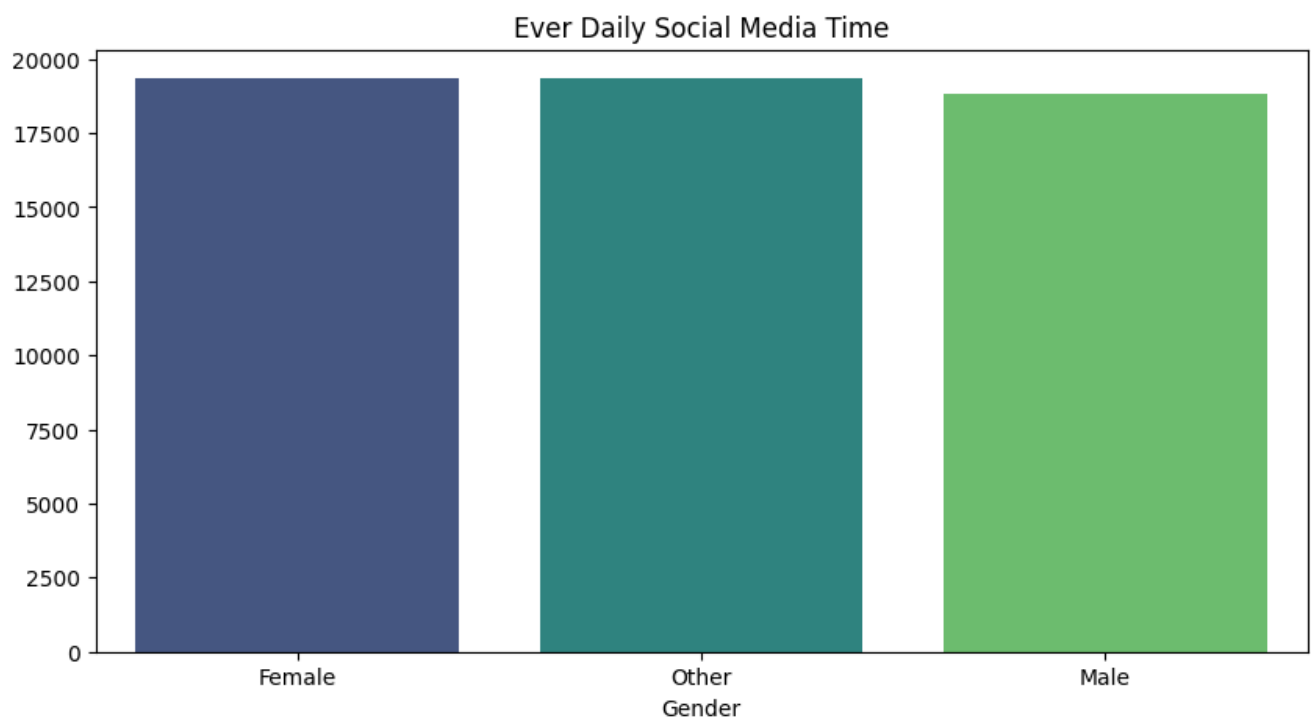
```
# Bar plot
```

```
plt.figure(figsize=(10, 5))  
sns.barplot(x=user_social_media.index, y=user_social_media.values,  
palette="viridis")  
plt.title("Ever Daily Social Media Time")  
plt.show()
```

```
# Pie chart
```

```
plt.figure(figsize=(10, 5))  
plt.pie(user_social_media.values, labels=user_social_media.index,  
autopct='%1.1f%%', startangle=140, colors=sns.color_palette("viridis"))  
plt.title("Ever Daily Social Media Time")  
plt.show()
```

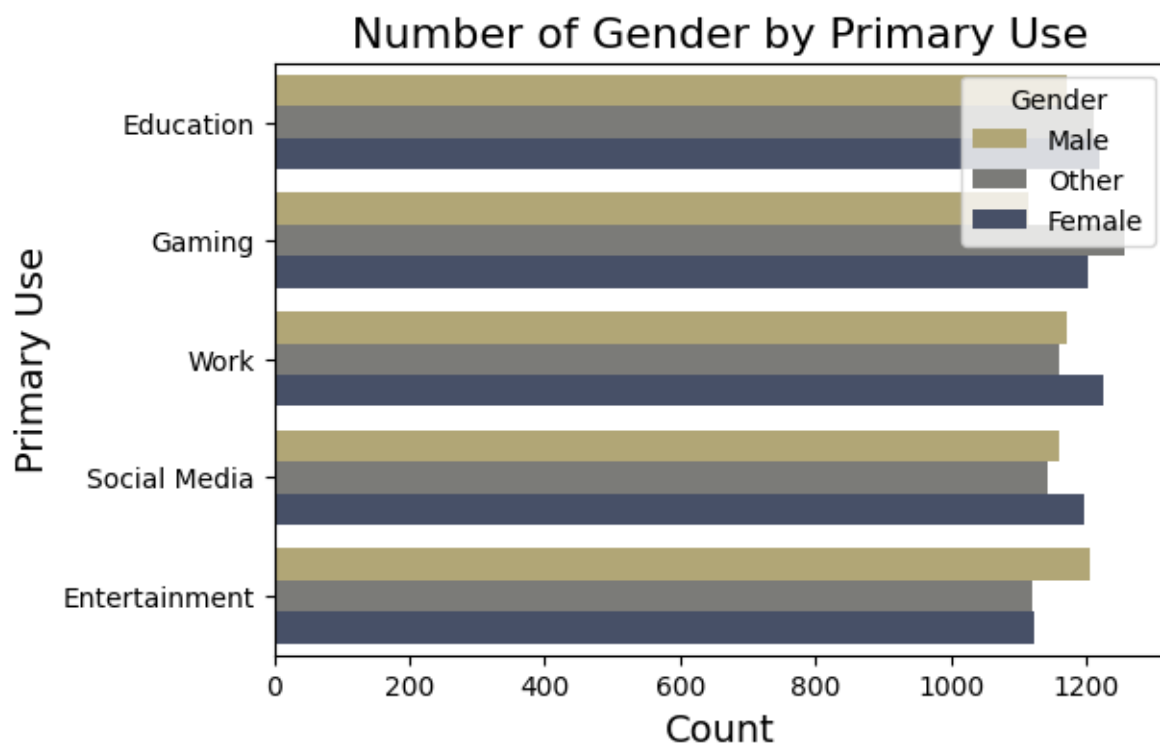
Output:



```
# Plot the number of users by primary use and age
plt.figure(figsize=(6, 4))
sns.countplot(data=df, y="Primary Use", order=df["Primary Use"].value_counts().index, hue="Age", palette="cividis_r")
plt.title("Number of Age by Primary Use", fontsize=16)
plt.xlabel("Count", fontsize=14)
plt.ylabel("Primary Use", fontsize=14)
plt.show()
```

```
# Plot the number of users by primary use and gender
plt.figure(figsize=(6, 4))
sns.countplot(data=df, y="Primary Use", order=df["Primary Use"].value_counts().index, hue="Gender", palette="cividis_r")
plt.title("Number of Gender by Primary Use", fontsize=16)
plt.xlabel("Count", fontsize=14)
plt.ylabel("Primary Use", fontsize=14)
plt.show()
```

OUTPUT:



Visualizing distributions of numeric columns

```
numeric_cols = df.select_dtypes(include=np.number).columns.tolist() #
```

Define numeric_cols here

```
for col in numeric_cols:
```

```
    plt.figure(figsize=(8, 5))
```

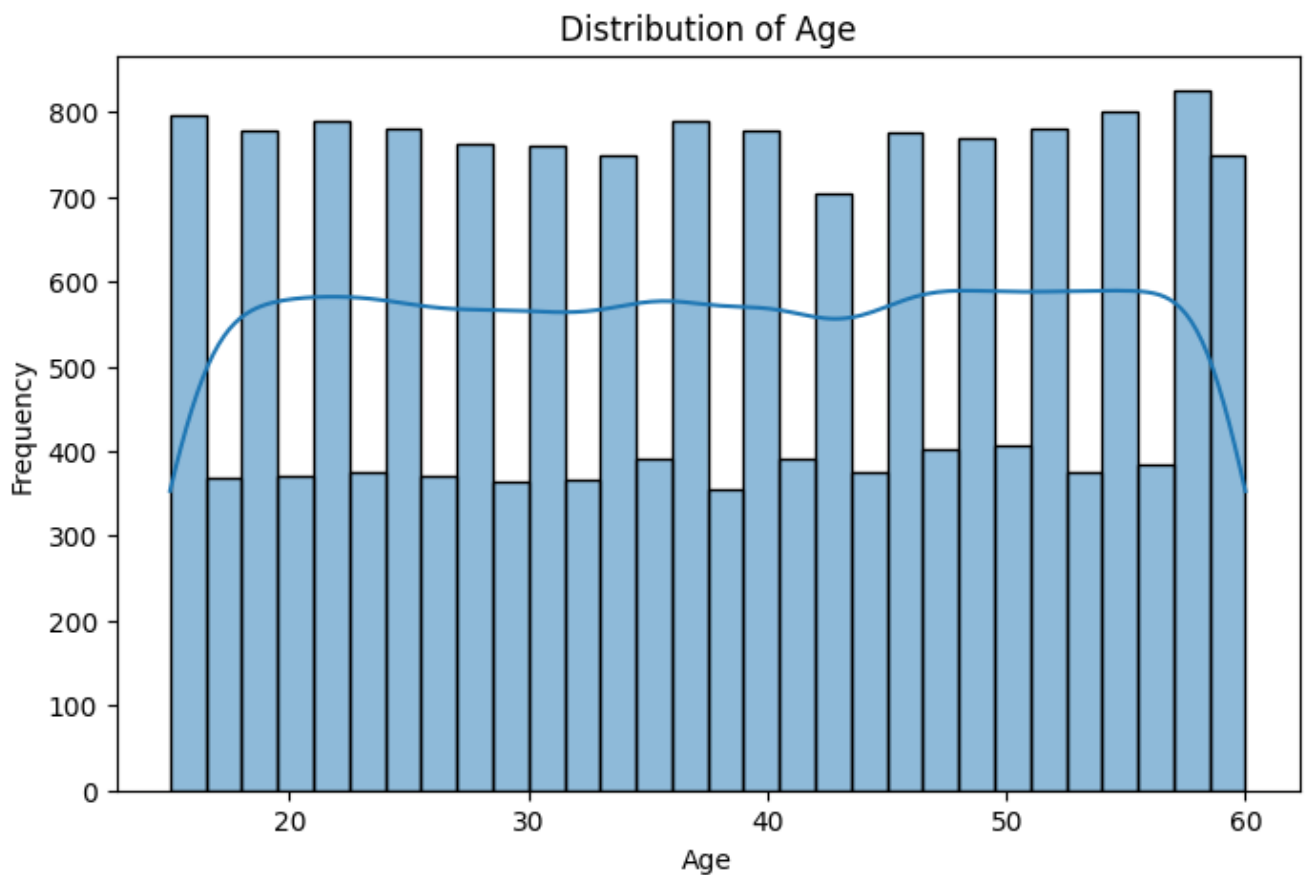
```
    sns.histplot(df[col], kde=True, bins=30)
```

```
    plt.title(f"Distribution of {col}")
```

```
    plt.xlabel(col)
```

```
    plt.ylabel("Frequency")
```

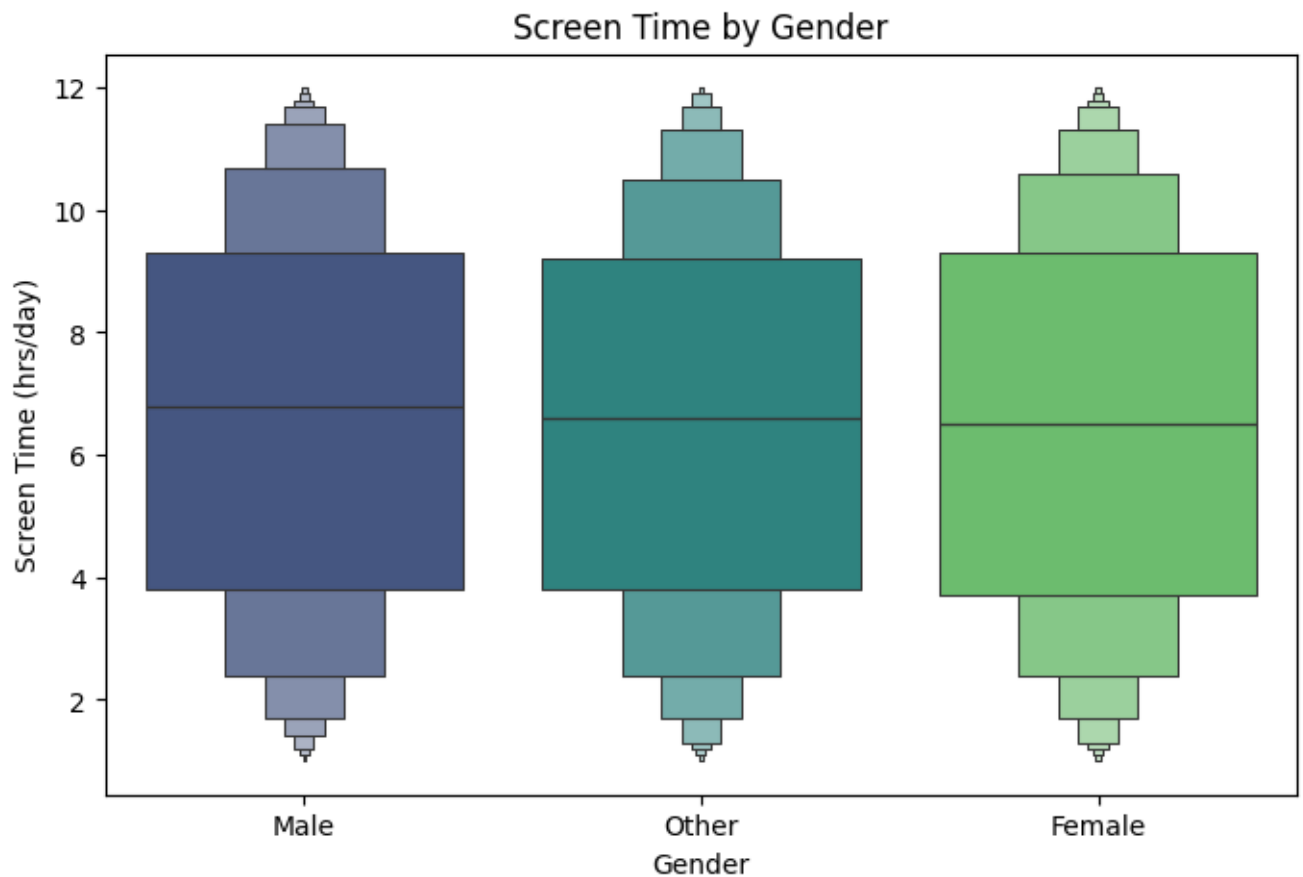
```
    plt.show()
```



```
# Calculate the total screen time by gender and plot using a boxen plot
screen_time = df.groupby("Gender")["Screen Time
(hrs/day)"].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(8, 5))
sns.boxenplot(x="Gender", y="Screen Time (hrs/day)", data=df,
palette="viridis")
plt.title("Screen Time by Gender")
plt.show()
```

```
# Plot the total e-commerce spend by gender using a bar plot
plt.figure(figsize=(8, 5))
sns.barplot(x="Gender", y="E-commerce Spend (INR/month)", data=df,
palette="viridis")
plt.title("E-commerce Spend by Gender")
plt.show()
```

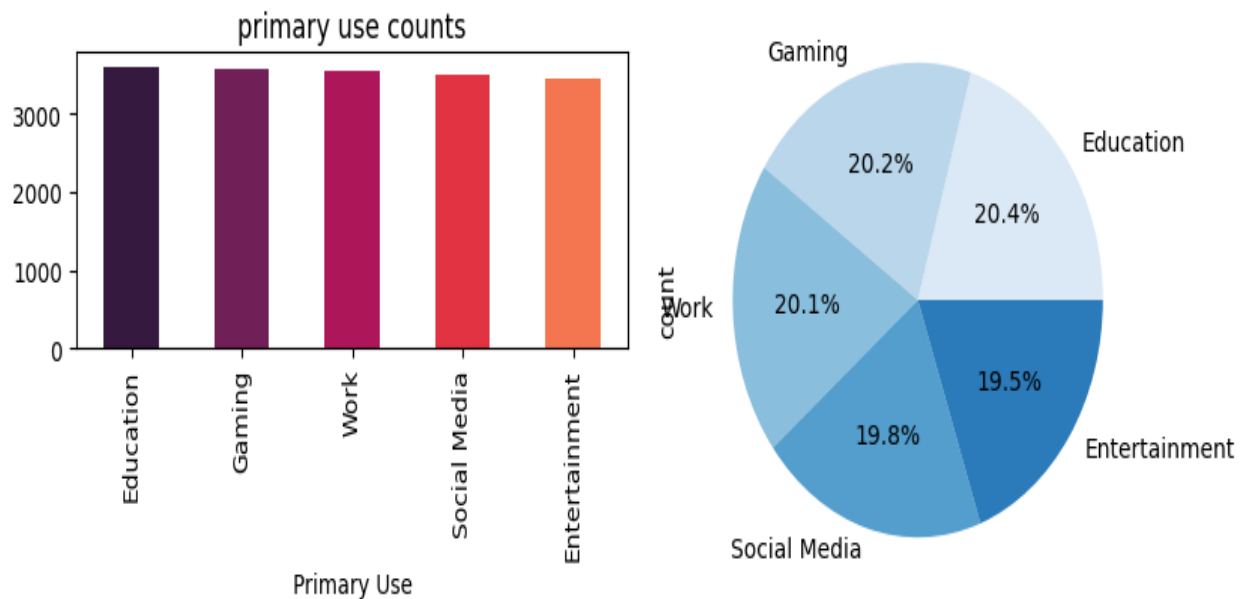
```
# Calculate the total gaming time by gender and plot using a boxen plot
Gaming_time = df.groupby("Gender")["Gaming Time
(hrs/day)"].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(8, 5))
sns.boxenplot(x="Gender", y="Gaming Time (hrs/day)", data=df,
palette="viridis")
plt.title("Gaming Time by Gender")
plt.show()
```



primary use counts

```
df["Primary
Use"].value_counts().sort_values(ascending=False).head(10).plot(kind='bar',
figsize=(5,2), color=sns.color_palette("rocket"))
plt.title("primary use counts")
plt.show()
```

```
df["Primary
Use"].value_counts().sort_values(ascending=False).head(10).plot(kind='pie',
figsize=(6, 4), autopct='%1.1f%%', colors=sns.color_palette("Blues"))
plt.show()
```



Plot the number of users by phone brand

```
plt.figure(figsize=(8, 4))
sns.countplot(data=df, y="Phone Brand", order=df["Phone Brand"].value_counts().index, palette="inferno")
plt.title("Most Popular Phone Brands", fontsize=16)
plt.xlabel("Count", fontsize=14)
plt.ylabel("Phone Brand", fontsize=14)
plt.show()
```

Plot the number of users by phone brand and gender

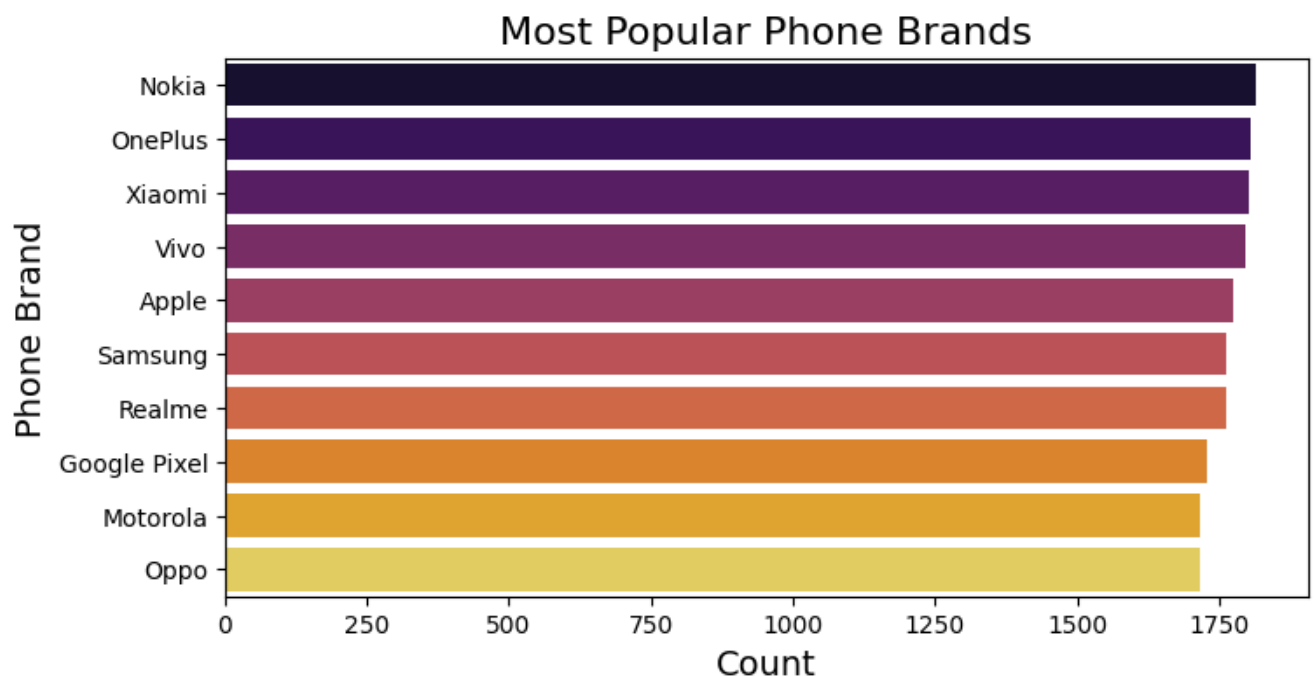
```
plt.figure(figsize=(8, 4))
sns.countplot(data=df, y="Phone Brand", order=df["Phone Brand"].value_counts().index, hue="Gender", palette="inferno")
plt.title("Most Popular Phone Brands by Gender", fontsize=16)
plt.xlabel("Count", fontsize=14)
plt.ylabel("Phone Brand", fontsize=14)
plt.show()
```

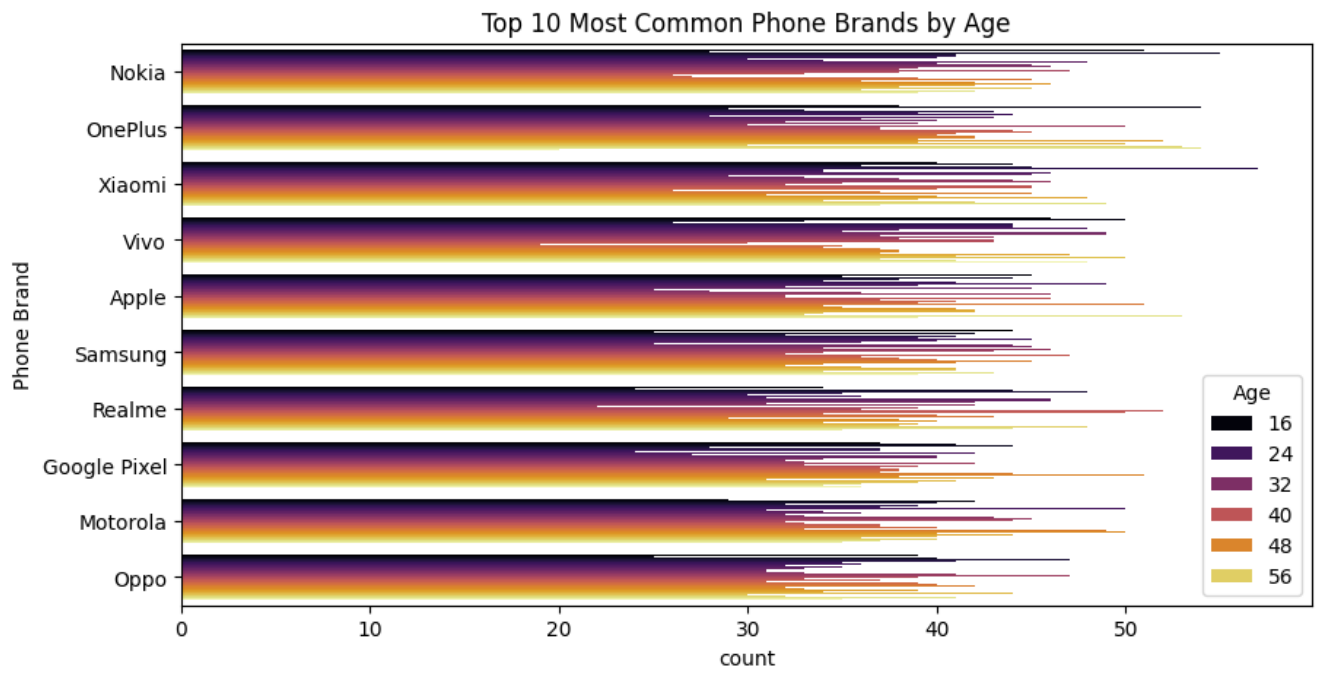


```
# Get the top 10 most common phone brands
top_10_phone_brands = df["Phone Brand"].value_counts().head(10).index

# Filter the DataFrame to include only the top 10 phone brands
filtered_df = df[df["Phone Brand"].isin(top_10_phone_brands)]

# Create the count plot
plt.figure(figsize=(10, 5))
sns.countplot(data=filtered_df, y="Phone Brand",
order=top_10_phone_brands, hue="Age", palette="inferno")
plt.title("Top 10 Most Common Phone Brands by Age")
plt.show()
```

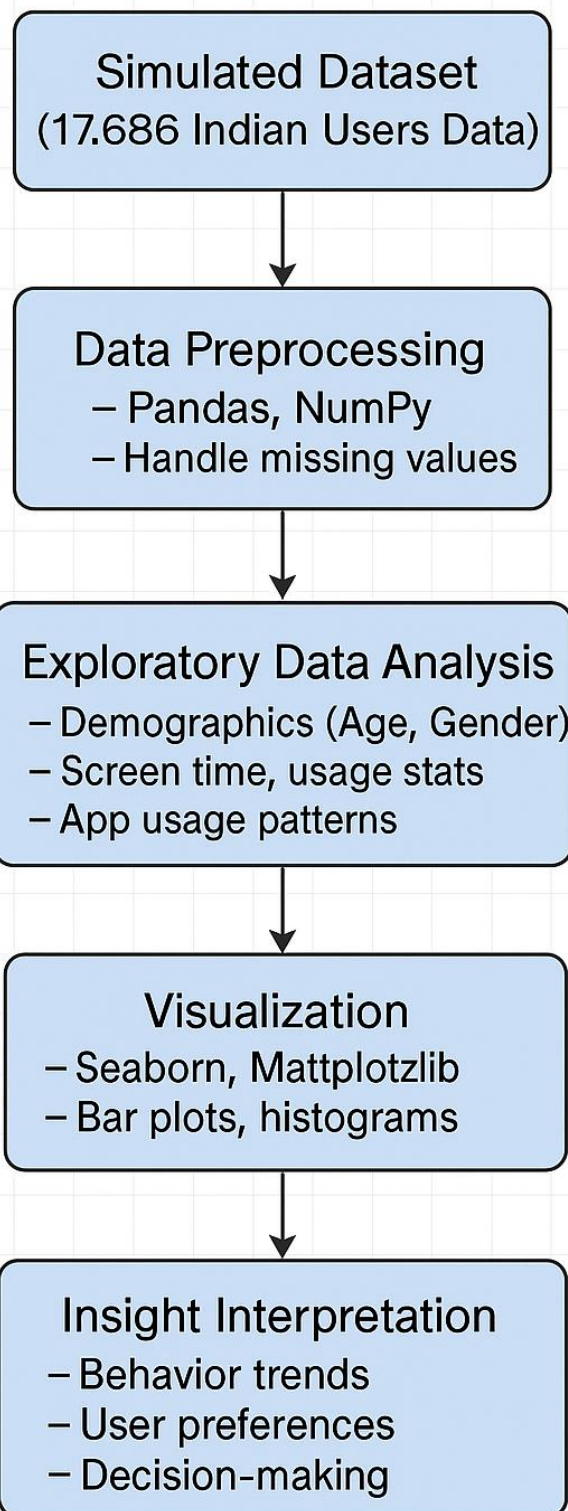




CONCLUSION:

This project successfully analyzed the mobile phone usage patterns in India using a comprehensive dataset. Through data visualization, it was found that usage habits vary significantly across different age groups, genders, and locations. The project can help mobile service providers and digital marketers better understand their user base and create more personalized services.

BLOCK DIAGRAM:



Thank You