```
In [1]: import pandas as pd
```

```
In [2]: data = pd.read_csv("phone_usage_india.csv")
        data
```

```
Out[2]:
                                                                                  Calls
         User                                  Phone           Data Usage      Duration
           ID  Age  Gender    Location         Brand       OS  (GB/month)    (mins/day)

    0  U00001   53    Male      Mumbai          Vivo  Android        23.9          37.9
    1  U00002   60   Other       Delhi        Realme      iOS        28.1          13.7
    2  U00003   37  Female   Ahmedabad         Nokia  Android        12.3          66.8
    3  U00004   32    Male        Pune       Samsung  Android        25.6         156.2
    4  U00005   16    Male      Mumbai        Xiaomi      iOS         2.5         236.2
  ...     ...  ...     ...         ...           ...      ...         ...           ...
17681  U17682   27   Other      Mumbai         Apple      iOS        36.6          90.4
17682  U17683   40  Female     Chennai          Oppo      iOS        12.9         243.0
17683  U17684   34  Female   Ahmedabad        Realme  Android        48.9          74.7
17684  U17685   22    Male   Hyderabad          Vivo  Android        25.6         105.6
17685  U17686   43   Other     Kolkata          Oppo      iOS        22.5         215.2

17686 rows × 15 columns
```

```
In [3]: data.isna()
```

Out[3]:

	User ID	Age	Gender	Location	Phone Brand	OS	Data Usage (GB/month)	Calls Duration (mins/day)	Number of Apps Installed
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
17681	False	False	False	False	False	False	False	False	False
17682	False	False	False	False	False	False	False	False	False
17683	False	False	False	False	False	False	False	False	False
17684	False	False	False	False	False	False	False	False	False
17685	False	False	False	False	False	False	False	False	False

17686 rows × 15 columns

In [4]:
```
data.isna().sum()
```

Out[4]:
```
User ID                         0
Age                             0
Gender                          0
Location                        0
Phone Brand                     0
OS                              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
```

In [5]:
```
data.columns
```

Out[5]:
```
Index(['User ID', 'Age', 'Gender', 'Location', 'Phone Brand', 'OS',
       '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)', 'Primary Use'],
      dtype='object')
```

In [6]:
```
data.describe()
```

Out[6]:

	Age	Data Usage (GB/month)	Calls Duration (mins/day)	Number of Apps Installed	Social Media Time (hrs/day)	E-comm Sp (INR/mo
count	17686.000000	17686.000000	17686.000000	17686.000000	17686.000000	17686.000
mean	37.584247	25.411257	151.405846	104.584869	3.252369	5075.707
std	13.338252	14.122167	84.923353	55.217097	1.590223	2871.604
min	15.000000	1.000000	5.000000	10.000000	0.500000	100.000
25%	26.000000	13.200000	77.325000	57.000000	1.900000	2587.500
50%	38.000000	25.300000	150.600000	104.000000	3.200000	5052.000
75%	49.000000	37.600000	223.900000	152.000000	4.600000	7606.000
max	60.000000	50.000000	300.000000	200.000000	6.000000	10000.000

In [7]:

```
data.duplicated()
```

Out[7]:

```
0        False
1        False
2        False
3        False
4        False
         ...  
17681    False
17682    False
17683    False
17684    False
17685    False
Length: 17686, dtype: bool
```

In [8]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17686 entries, 0 to 17685
Data columns (total 15 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   Data Usage (GB/month)        17686 non-null  float64
 7   Calls Duration (mins/day)    17686 non-null  float64
 8   Number of Apps Installed     17686 non-null  int64  
 9   Social Media Time (hrs/day)  17686 non-null  float64
 10  E-commerce Spend (INR/month) 17686 non-null  int64  
 11  Streaming Time (hrs/day)     17686 non-null  float64
 12  Gaming Time (hrs/day)        17686 non-null  float64
 13  Monthly Recharge Cost (INR)  17686 non-null  int64  
 14  Primary Use                  17686 non-null  object 
dtypes: float64(5), int64(4), object(6)
memory usage: 2.0+ MB
```

In [9]:
```
data.shape
```

Out[9]:
```
(17686, 15)
```

In [10]:
```
data.dtypes
```

Out[10]:
```
User ID                          object
Age                               int64
Gender                           object
Location                         object
Phone Brand                      object
OS                               object
Data Usage (GB/month)           float64
Calls Duration (mins/day)       float64
Number of Apps Installed          int64
Social Media Time (hrs/day)     float64
E-commerce Spend (INR/month)      int64
Streaming Time (hrs/day)        float64
Gaming Time (hrs/day)           float64
Monthly Recharge Cost (INR)       int64
Primary Use                      object
dtype: object
```

In [11]:
```
# Identify object-type columns
object_columns = data.select_dtypes(include=['object']).columns
object_columns
```

Out[11]:
```
Index(['User ID', 'Gender', 'Location', 'Phone Brand', 'OS', 'Primary Use'], dt
ype='object')
```

In [12]:
```
# Identify int-type columns
int_columns = data.select_dtypes(include=['int64']).columns
int_columns
```

Out[12]:
```
Index(['Age', 'Number of Apps Installed', 'E-commerce Spend (INR/month)',
       'Monthly Recharge Cost (INR)'],
      dtype='object')
```

In [13]:
```
# Identify float-type columns
float_columns = data.select_dtypes(include=['float64']).columns
float_columns
```

Out[13]:
```
Index(['Data Usage (GB/month)', 'Calls Duration (mins/day)',
       'Social Media Time (hrs/day)', 'Streaming Time (hrs/day)',
       'Gaming Time (hrs/day)'],
      dtype='object')
```

In [14]:
```
data['Gender'].value_counts()
```

Out[14]:
```
Gender
Female    5969
Other     5892
Male      5825
Name: count, dtype: int64
```

In [15]:
```
data['Location'].value_counts()
```

```
Out[15]: Location
         Jaipur       1824
         Pune         1815
         Chennai      1790
         Kolkata      1789
         Bangalore    1786
         Ahmedabad    1785
         Delhi        1775
         Mumbai       1722
         Lucknow      1700
         Hyderabad    1700
         Name: count, dtype: int64
```

```
In [16]: data['Phone Brand'].value_counts()
```

```
Out[16]: Phone Brand
         Nokia           1816
         OnePlus         1807
         Xiaomi          1803
         Vivo            1797
         Apple           1775
         Samsung         1764
         Realme          1762
         Google Pixel    1729
         Motorola        1717
         Oppo            1716
         Name: count, dtype: int64
```

```
In [17]: data['OS'].value_counts()
```

```
Out[17]: OS
         Android    8851
         iOS        8835
         Name: count, dtype: int64
```

```
In [18]: data['Primary Use'].value_counts()
```

```
Out[18]: Primary Use
         Education        3601
         Gaming           3576
         Work             3557
         Social Media     3501
         Entertainment    3451
         Name: count, dtype: int64
```

```
In [ ]:
```

Data Understanding

- Given dataset having no missing values
- Given dataset having no duplicate values
- Given dataset (17686, 15) shape which means data represent 16 attribute and 17686 rows of data
- 16 Attribute divide three datatypes

1. Int

2. Float
3. Object

- `Int` datatype having columns:- 'Age', 'Number of Apps Installed', 'E-commerce Spend (INR/month)', 'Monthly Recharge Cost (INR)'
- `float` datatype having columns:- 'Data Usage (GB/month)', 'Calls Duration (mins/day)', 'Social Media Time (hrs/day)', 'Streaming Time (hrs/day)', 'Gaming Time (hrs/day)'
- `Object` datatype having columns:- 'User ID', 'Gender', 'Location', 'Phone Brand', 'OS', 'Primary Use'], dtype='object'
- data tells `Female` are mostly phones users in india
- `Jaipure` is the highest phones users city in india and `Lucknow and Hyderabad` comes last in the list
- `Nokia & OnePlus` most trusted brands by indias.
- Most of the indians utilize there phone because of `Education Purpose`

Utill now we understand the basic information from then, Now PowerBI Dashboard to Visualiy representage the data.

PowerBI Dashbord Steps & Responsibilities

Responsibilities

1. **Developed interactive Power BI dashboards** to visualize data to better understanding.
2. **Transformed raw data into meaningful insights** using DAX calculations and data modeling.
3. **Integrated multiple data sources** Python Jupyter Notebook for real-time and automated reporting.
4. **Optimized dashboard performance** by implementing filters, slicers, and efficient data queries.
5. **Collaborated with stakeholders** to design user-friendly dashboards for data-driven decision-making.

Steps

1. Crete a `Time Base Analysis Dashboard` to understand time spent history by individual
   - Create a Total time spent column to get a better idea of how much each individual spends on their mobile
   - Created Age_group column using the conditional column to divide ages by teen, adult, and old

    - Created graphs to understand each time spent zone based on Age, gender, Location, Game, Social media, streaming, and calls
    - Through Graphs Understand the hidden meaning of data
2. Crete an `Amount Base Analysis Dashboard` to understand the amount spent history by individual
    - Create a Total amount spent column to get a better idea of how much each individual spent on their mobile
    - Created graphs to understand each time spent zone based on Age, gender, Location, E-commerce and Recharge
    - Through Graphs Understand the hidden meaning of data.
3. Crete an `Individual Base Analysis Dashboard` to understand the each person spent time and money history.
    - Create duplicate columns of fe
    - Created graphs to understand each time spent zone based on Age, gender, Location, E-commerce and Recharge
    - Through Graphs Understand the hidden meaning of data.

```
In [ ]:
```