**Name :- Tanuja Kamlakar Gaikwad**

**PRN:- 202401030028**

**Roll no:-CC-77**

**Assignment no 1**

**Mobile Sets Dataset Analysis using Pandas**

**1. Introduction**

In this project, we perform data analysis on a mobile sets dataset using **Python** and the **Pandas** library.  
The dataset contains features such as brand name, price, RAM, storage capacity, battery life, and camera specifications of various mobile phones.  
The goal of this analysis is to understand market trends, identify high-performing brands, and explore key features that influence mobile prices and popularity.

**2. Tasks**

The main tasks performed are:

* Loading and inspecting the dataset.
* Cleaning the dataset (handling missing values).
* Analyzing important features like brand, price, RAM, storage, battery, and camera.
* Visualizing data distributions and feature correlations.
* Answering interesting business questions like:
  + Which brands are most common?
  + What is the price distribution?
  + Which phones have the best battery life?

**3.Pandas Code Solutions**

# Import necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

df = pd.read\_csv('mobiles.csv')

# Display the first few records

print(df.head())

# Get an overview of the dataset

print(df.info())

print(df.describe())

# Check for missing values

print(df.isnull().sum())

# Drop missing values (if any)

df.dropna(inplace=True)

# Analyze the number of phones per brand

brand\_counts = df['Brand'].value\_counts()

print(brand\_counts)

# Find the average price per brand

average\_price = df.groupby('Brand')['Price'].mean()

print(average\_price)

# Plot the distribution of mobile phone prices

sns.histplot(df['Price'], bins=30, kde=True)

plt.title('Price Distribution of Mobiles')

plt.xlabel('Price')

plt.ylabel('Number of Phones')

plt.show()

# Analyze the correlation between numerical features

correlation\_matrix = df.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Feature Correlation Heatmap')

plt.show()

# Find the top 10 mobiles with the largest battery capacity

top\_battery = df.sort\_values('Battery', ascending=False).head(10)

print(top\_battery)

# Find affordable mobiles with more than 6GB RAM

affordable\_high\_ram = df[(df['RAM'] >= 6) & (df['Price'] < 20000)]

print(affordable\_high\_ram)

# Save the cleaned dataset

df.to\_csv('mobiles\_cleaned.csv', index=False)