SYNOPSIS

Student Name: Chauhan Vandana Ramdayal

Student Id: AF0411629

Topic: Analysis of Mobile

Dataset Overview:
The dataset contains detailed information about various mobile phones, including features, prices, and other specifications. Each row represents an individual phone, and the columns provide the following details:

- **Phone_name**: The name of the phone.
- **Brands**: The brand of the phone.
- **Price**: The price of the phone.
- **Internal_Storage**: The phone's internal storage capacity.
- **Operating_System_Type**: The type of operating system (e.g., Android, iOS).
- **USB_Type**: The type of USB port used.
- **5G_Availability**: Whether the phone supports 5G.
- **Selfie_Camera**: The quality of the selfie camera (megapixels).
- **RAM_Storage**: The phone's RAM capacity.
- **Country_of_Origin**: The country from which the phone originates.
- **Battery_Capacity**: The battery capacity of the phone.
- **Price_Range**: The price range category (e.g., low, medium, high).
- **Battery_Capacity_Range**: The range of battery capacities.
- **Total_Mobile**: The total number of mobiles in the dataset.

Problem Statement:

1. What are the features and prices of all mobile phones in the dataset?
2. What are the 5 most expensive phones and their prices?
3. What are the 5 cheapest phones and their prices?
4. List the top 5 Samsung phones with prices and all features.
5. List Android phones and the top 5 highest-priced Android phones.
6. List Android phones and the top 5 lowest-priced Android phones.
7. List iOS phones and the top 5 highest-priced iOS phones.
8. List iOS phones and the top 5 lowest-priced iOS phones.
9. Write a query to find phones that support 5G, and list the top 5 phones with 5G support.
10. Find the total price of all mobiles, grouped by brand name.

Data Preprocessing Steps:

1. **Data Cleaning**: Handle missing values and ensure data consistency.
2. **Normalization**: Normalize numeric features like Price and Battery Capacity if necessary.
3. **Encoding Categorical Variables**: Convert categorical features (e.g., Operating_System_Type, Brand) into numeric form using one-hot encoding or label encoding.

Implementation Process:

1. **Data Ingestion**: Load the dataset into MySQL Workbench for SQL queries and analysis.
2. **Preprocessing**: Clean and prepare the data for analysis.
3. **SQL Querying**: Use SQL to filter, aggregate, and extract information as per the problem statements.
4. **Visualization**: Use Power BI for interactive dashboards and Python libraries (e.g., Matplotlib, Seaborn) for data visualization, including bar charts, scatter plots, and pie charts.
5. **Reporting**: Summarize findings in reports or dashboards, providing insights into pricing trends, brand performance, and feature distribution.

Technologies:

- **MySQL Workbench**: For data querying and management.
- **Power BI**: For interactive visualizations and dashboards.
- **Python (Pandas, Matplotlib, Seaborn)**: For data manipulation and visualizations.

Software Requirements:

- Operating Systems: Windows
- IDE: MySQL Workbench for SQL queries, Jupyter Notebook (for Python), or Power BI for visualizations.

Hardware Requirements:

- RAM: Minimum 8GB (recommended 16GB for handling large datasets).
- Processor: Minimum Intel i5, recommended Intel i7 for faster data processing.
- Storage: SSD recommended (at least 256GB for large datasets).

This analysis will help in understanding mobile phone pricing trends, customer preferences based on operating systems, and the impact of 5G support, among other factors, providing actionable insights for businesses and consumers alike.