EDA Analysis on Smartphones (2024)

This project performs an **Exploratory Data Analysis (EDA)** on smartphones released in 2024. The goal is to analyze various factors such as pricing, specifications, performance, and market trends. The dataset includes smartphones from different brands, their features, and consumer data to uncover valuable insights into the 2024 smartphone market.

Table of Contents

- Overview
- Dataset
- Installation
- Usage
- Results & Insights
- Contributing
- License

Overview

The smartphone industry in 2024 is seeing rapid advancements in hardware and software, with new features such as enhanced camera capabilities, faster processors, and longer battery life. This project analyzes a dataset of smartphones released in 2024 and aims to:

- Compare smartphone pricing trends across different brands and models.
- Examine key specifications such as RAM, storage, camera quality, and battery life.
- Analyze user ratings and reviews to determine consumer preferences.
- Visualize trends in pricing, performance, and market share for smartphones in 2024.

Dataset

The dataset used in this analysis includes the following key attributes for each smartphone model:

- **Brand**: The manufacturer of the smartphone.
- **Model**: The specific model or series of the smartphone.
- **Price**: Retail price of the smartphone (in USD).
- **Specifications**: Key features such as RAM, storage, camera resolution, processor type, and battery capacity.
- **Ratings**: User ratings and reviews.
- **Launch Date**: The release date of the smartphone.
- **Battery Life**: The estimated battery life in hours.
- **OS**: The operating system (Android, iOS, etc.).

Installation

To set up the project locally, follow these steps:

1. **Clone the repository**:

```
git clone https://github.com/your-username/EDA-Smartphones-2024.git
```

2. Navigate into the project directory:

cd EDA-Smartphones-2024

3. Install required dependencies: Ensure you have Python installed, then use the following command to install the necessary libraries:

pip install -r requirements.txt

Usage

To explore the dataset and run the analysis:

1. Open the Jupyter Notebook:

jupyter notebook notebooks/EDA_Smartphones_2024.ipynb

2. Follow the steps in the notebook to load the dataset, perform data cleaning, and visualize key trends.

You can run the notebook cell-by-cell to observe insights such as:

- Pricing trends by brand.
- Specification analysis (e.g., RAM, storage, camera performance).
- Performance comparison based on user ratings.

Results & Insights

1. Distribution of Numerical Features

- The histograms show the distribution of various smartphone features such as price, storage, battery capacity, RAM, etc.
- **Price and Price (USD)**: These histograms show a right-skewed distribution, meaning most phones have lower prices, with a few higher-priced models.
- **Storage**: Storage distribution indicates several common capacity clusters, likely corresponding to typical storage options (e.g., 64GB, 128GB, etc.).
- **RAM**: This feature has a few distinct clusters, suggesting that RAM is standardized at certain levels (e.g., 4GB, 8GB, etc.).
- **Battery**: The majority of batteries are between 3000 and 5000 mAh, which is typical for modern smartphones.
- **Display Size** and **PPI Density**: Display sizes appear to mostly cluster around typical smartphone screen sizes (6-7 inches), and PPI values indicate varying screen resolutions.
- **Other Features (NFC, Foldable, Year)**: Features like NFC and foldable screens are mostly binary, with most devices lacking foldable screens. The "Year" feature shows an increase in newer phones in the dataset.
- **Quantiles**: These quantiles (10, 50, 90) might represent price or other numerical feature breakpoints, indicating the distribution at specific percentile values.

2. RAM Distribution by Brand

- This box plot shows the RAM distribution for each brand.
- Brands like **Apple and Asus** have models with a wider range of RAM capacities, while other brands have more standardized RAM values.

- **Infinix** and **Cubot** generally have lower RAM values compared to others, suggesting they may focus on budget models.
- Brands like **Xiaomi** and **Samsung** show a moderate spread, with some outliers indicating higher RAM capacities in specific models.

3. Battery vs. Price Scatter Plot

- The scatter plot illustrates the relationship between battery capacity and price.
- There is a general trend where phones with higher battery capacity tend to be in the mid-to-higher price ranges, although some high-capacity batteries exist in lower-priced models.
- Outliers exist, especially at the extreme high battery capacities (e.g., around 10,000 mAh), but these do not necessarily correlate with very high prices.

Summary

- **Price** and **Battery** are not linearly related, though higher battery phones tend to cost more.
- **RAM** distributions vary significantly by brand, suggesting different brand focuses on performance.
- Common features such as **storage** and **display size** are clustered, indicating standard options across the market.

Contributing

We welcome contributions! If you'd like to improve the project, follow these steps:

1. Fork the repository.
2. Create a new branch (`git checkout -b feature-branch`).
3. Make your changes and commit them (`git commit -am 'Add new feature'`).
4. Push your branch to GitHub (`git push origin feature-branch`).
5. Open a pull request with a description of your changes.

Feel free to open an issue if you encounter any bugs or have feature requests.