

Data Science Toolbox Python Programming

**PROJECT REPORT**

**( Exploring the Impact of Phone Features on Price  )**

Submitted by:

Name: Shiv Narayan Dwivedi

Registration No: 12314985

Roll No: 57

Section: K23EL

Course Code: INT375

**Under the Guidance of:**

**Dr. Dhiraj Kapila**

**Department of Computer Science Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Shiv Narayan Dwivedi bearing Registration no 12314985 has completed INT375 project titled, **“**Exploring the Impact of Phone Features on Price ” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Acknowledgement**

I would like to express my sincere gratitude to Dr. Dhiraj Kapila for his valuable guidance, encouragement, and support throughout the duration of this project. I am also thankful to Lovely Professional University and the faculty of School of Computer Science & Engineering for providing the necessary infrastructure and resources. Lastly, I extend my thanks to all those who have directly or indirectly helped me in the successful completion of this project**.**

**DECLARATION**

I, Shiv Narayan Dwivedi student of Computer Science under CSE/IT Department at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:12-04-2025 Signature: Shiv Narayan Dwivedi

Registration No.12314985 Name: Shiv Narayan Dwivedi

**Table Of Content**

| **S. No.** | **Title** | **Page No.** |
| --- | --- | --- |
| **1.** | **Introduction** | **4** |
| **2.** | **Exploratory Data Analysis (EDA)** | **5** |
| **3.** | **Source of Dataset** | **6** |
| **4.** | **EDA Process** | **6** |
| **5.** | **Analysis on Dataset** | **6** |
| **5.1** | **General Data Exploration** | **7** |
| **5.2** | **Data Cleaning and Preprocessing** | **8** |
| **5.3** | **Correlation Analysis** | **8** |
| **5.4** | **Outlier Detection** | **9** |
| **5.5** | **Trend Analysis** | **10** |
| **5.6** | **Categorical Analysis - Pie and Bar Charts** | **11** |
| **5.7** | **Scatter and Pair Plots** | **13** |
| **6.** | **Conclusion** | **14** |
| **7.** | **Future Scope** | **15** |
| **8.** | **References** | **16** |

**1. Introduction:**

In the modern era, mobile phones have become an essential part of daily life. With the vast variety of smartphone models and brands available in the market, it becomes necessary to analyze their specifications to understand trends, performance metrics, and pricing. This project performs **Exploratory Data Analysis (EDA)** on a smartphone dataset using Python libraries like **NumPy**, **Pandas**, **Matplotlib**, and **Seaborn**. The analysis provides insights into features such as **RAM**, **Mobile Weight**, **Battery Capacity**, and **Launch Prices** across different countries.

**2. Exploratory Data Analysis (EDA):**

Exploratory Data Analysis (EDA) is a key step to understand the structure and patterns in the dataset before applying any advanced analytics. In this project, EDA was used to explore the features of mobile phones such as RAM, battery capacity, screen size, and prices in different countries. The process involved cleaning the data, converting necessary columns to numeric formats, and handling missing values. Various plots like bar charts, box plots, scatter plots, and violin plots were used to visualize trends and relationships among different features. This helped in identifying important patterns, detecting outliers, and gaining useful insights from the data.

This included:

* Loaded and cleaned the dataset
* Converted RAM, Battery, and Weight into numeric format
* Used seaborn and matplotlib for visualizations
* Detected missing data and removed or handled it
* Generated plots for company-wise RAM, price comparisons, weight vs. battery, etc.
* Used violin and box plots for outlier detection
* Observed overall trends and correlations

**3. Source of Dataset:**

The dataset used for this analysis was collected from publicly available mobile specification data. It contains detailed information about various mobile phone models from different companies, including features such as RAM, battery capacity, camera details, screen size, processor, and prices across multiple countries (India, USA, Pakistan, China, and Dubai). The dataset was structured in tabular form and imported into the analysis environment using CSV format.

Dataset Title **Exploring the Impact of Phone Features on Price**  
File Name: Mobiles Dataset (2025).csv  
Source: Kaggle

**4. EDA Process:**

The Exploratory Data Analysis (EDA) process involves several systematic steps to understand, clean, and visualize the dataset. Below is a breakdown of the EDA process followed in this project:

1. **Importing Libraries**
   * Imported essential Python libraries such as pandas, numpy, matplotlib, and seaborn for data handling and visualization.
2. **Loading the Dataset**
   * The dataset was loaded using pandas.read\_csv() and displayed to understand its basic structure and columns.
3. **Cleaning the Data**
   * Removed unnecessary spaces and characters from column names.
   * Converted data types of columns like RAM, Battery Capacity, and Mobile Weight to numeric.
   * Handled missing values using dropna() and fillna() functions.
4. **Understanding Data Structure**
   * Used functions like df.head(), df.info(), and df.describe() to get an overview of the data types, shape, and summary statistics.
5. **Visual Analysis**
   * Plotted boxplots, bar charts, violin plots, and regression plots to analyze relationships and trends.
   * Categorical columns like 'Company Name' were visualized using countplots and pie charts.
   * Numerical columns were examined using scatter plots and histograms.
6. **Correlation Check**
   * Used heatmaps to visualize correlation between numerical features.
7. **Outlier Detection**
   * Identified extreme values in RAM, battery, and weight using box and violin plots.

This process helped ensure the data was clean, well-structured, and ready for further insights and modeling.

**5. Analysis on Dataset:**

The dataset provides detailed specifications of mobile phones launched by various companies across different countries. The analysis focused on both **quantitative** and **categorical** features such as RAM, battery capacity, mobile weight, screen size, and launched prices.

Different types of visualizations and summary statistics were used to uncover hidden trends, variations across brands, and comparative price differences. Here’s a breakdown of the analysis:

**5.1 General Data Exploration:**

**5.1 General Data Exploration**

The first step in analyzing the dataset was to understand its structure. This involves checking the number of rows, columns, data types, and any missing values. We also explored the top 5 records of the dataset to gain a quick overview of the data.

A screenshot of a computer

AI-generated content may be incorrect.

**5.2 Data Cleaning and Preprocessing:**

n this step, the dataset was cleaned by performing the following tasks:

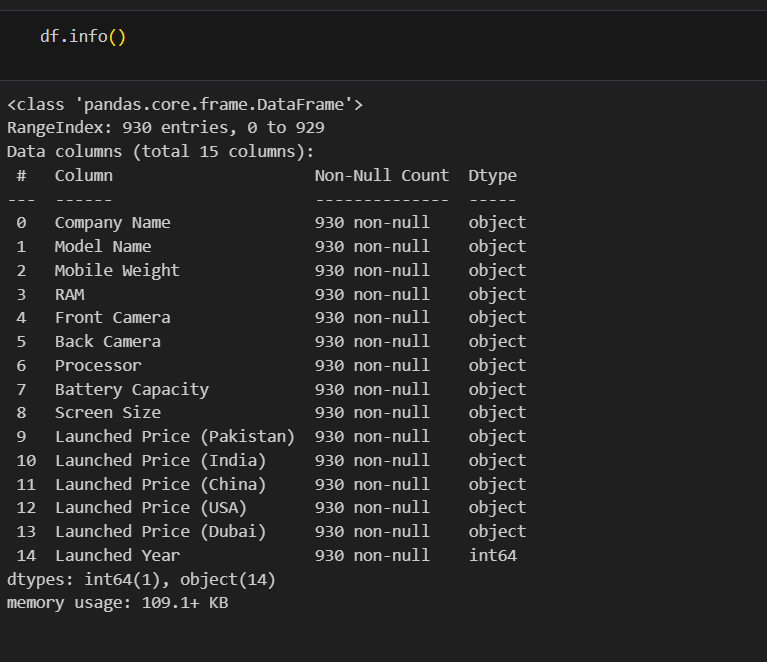
Removing Extra Spaces and Newlines: Stray spaces or newlines in column names were removed.

A screenshot of a computer program

AI-generated content may be incorrect.

Converting Data Types: Columns like 'RAM', 'Mobile Weight', and 'Battery Capacity' were converted to numeric values to ensure the data can be analyzed effectively.

Handling Missing Values: Rows with missing values in critical columns such as 'Mobile Weight' and 'Battery Capacity' were either removed or filled with appropriate values.



**5.3 Correlation Analysis:**

Correlation analysis helps to understand the relationship between different numerical features in the dataset. For example, we might expect that heavier phones tend to have larger battery capacities.

1. Code :

A computer screen with text on it

AI-generated content may be incorrect.

II. Visualisation:

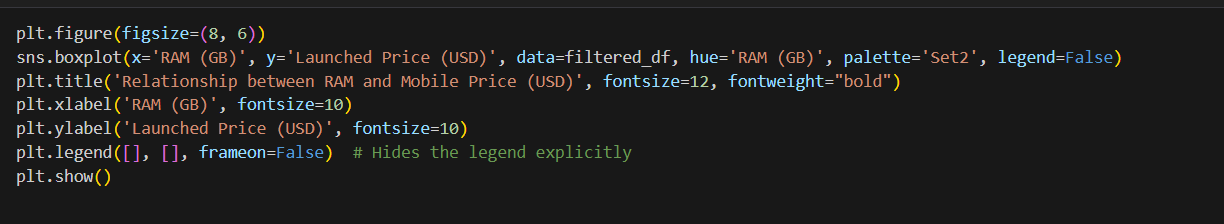
A graph of a company

AI-generated content may be incorrect.

**5.4 Outlier Detection**

Outliers are data points that significantly differ from other observations. These can distort statistical analyses and model training. Box plots and violin plots are useful to detect outliers in features like RAM, mobile weight, and battery capacity..

1. Code:



II. Visualisation:

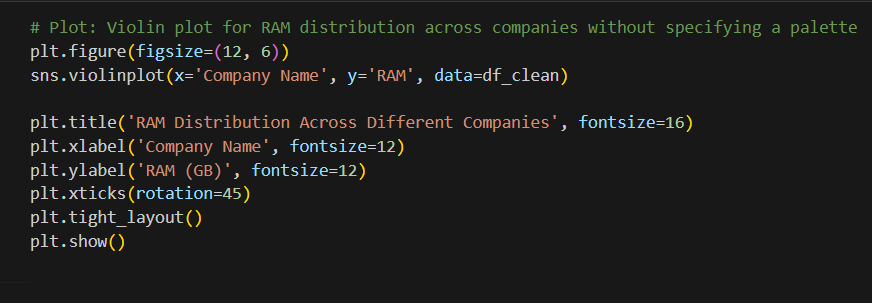
A graph of different colored boxes

AI-generated content may be incorrect.

**5.5 Trend Analysis**

Trend analysis helps us identify patterns in the data over different features. For example, we can analyze how battery capacity increases with mobile weight, or how the RAM distribution changes across companies.

1. Code:



1. Visualisation:-

A diagram of a diagram

AI-generated content may be incorrect.

**5.6 Categorical Analysis - Pie and Bar Charts**

Categorical features like 'Company Name' and 'Processor' can be analyzed using pie charts and bar plots to understand the frequency distribution of different categories.

I.Code

A computer screen shot of a program

AI-generated content may be incorrect.

II. Visualisation:

A pie chart with numbers and text

AI-generated content may be incorrect.

**5.7 Scatter and Pair Plots:**

Scatter plots and pair plots are used to visualize the relationships between multiple numerical variables in the dataset. These plots help us detect multi-dimensional trends and outliers simultaneously.A screen shot of a computer code

AI-generated content may be incorrect.

Visualisation:

A graph with different colored dots

AI-generated content may be incorrect.

A diagram of a company

AI-generated content may be incorrect.

**6. Conclusion:**

In this analysis, we explored a dataset containing various mobile phone specifications across different companies. The primary goal was to understand the relationships between features such as mobile weight, battery capacity, RAM, and other attributes, and identify trends and patterns within the data.

Key findings from the analysis include:

1. **Data Distribution**: The distribution of mobile weight, RAM, and battery capacity shows varied patterns across different companies. Some companies focus on lightweight models, while others prioritize larger batteries and higher RAM capacities.
2. **Correlation Insights**: A noticeable correlation between mobile weight and battery capacity was found, which is expected as heavier phones often come with larger batteries. Additionally, some companies tend to offer higher RAM configurations across multiple models.
3. **Outliers and Variations**: Outliers were detected in features like RAM, where certain mobile models had unusually high or low RAM configurations. This indicates that a few companies offer flagship models with significantly higher RAM capacities compared to others.
4. **Company Analysis**: By visualizing the number of models offered by each company, we could see that some companies have a more diversified product portfolio, while others focus on a smaller range of models.
5. **Trends and Patterns**: Scatter plots and regression analysis showed interesting trends, such as the relationship between mobile weight and battery capacity, and how mobile weight often correlates with a higher battery capacity.

In conclusion, the dataset provides valuable insights into how mobile phone specifications vary across different companies and how some features (like weight and battery capacity) are interdependent. This analysis can be useful for companies and consumers alike in understanding market trends, identifying opportunities for product differentiation, and making informed decisions when purchasing a mobile phone.

**7. Future Scope:**

While this analysis provides valuable insights into mobile phone specifications, there are several areas that can be explored further:

* **Predictive Modeling**: Using machine learning to predict mobile prices or other features based on existing data.
* **Price Trend Analysis**: Analyzing mobile price trends across regions to identify patterns and pricing strategies.
* **Sentiment Analysis**: Applying NLP techniques to analyze consumer reviews for better understanding of product satisfaction.
* **Feature Engineering**: Creating new metrics like battery-to-weight ratio to provide deeper insights into mobile efficiency.
* **Comparative Analysis of Brands**: More detailed comparisons between brands to understand competition on key features.
* **Exploring New Features**: Including new data like 5G support, camera configurations, and foldable designs to keep up with market trends.

### 8. References:

 **Kaggle. (n.d.).** Mobile Phone Specifications Dataset. Retrieved from https://www.kaggle.com/datasets

* This dataset was used as the primary source of information for the analysis, containing details of various mobile phones from different companies.

 **Seaborn Documentation. (n.d.).** Seaborn: Statistical Data Visualization. Retrieved from https://seaborn.pydata.org/

* Seaborn was used for creating various visualizations, including box plots, violin plots, and scatter plots, to analyze the data.

 **Matplotlib Documentation. (n.d.).** Matplotlib: Visualization with Python. Retrieved from <https://matplotlib.org/>

* Matplotlib was used for additional customizations in plotting and visualizing trends and distributions in the data.

 **Pandas Documentation. (n.d.).** Pandas: Data Structures and Data Analysis. Retrieved from https://pandas.pydata.org/

* Pandas was used for data manipulation, cleaning, and processing, allowing for a deeper exploration of the dataset.