SmartPrice Predictor: AI-Powered Smart Phone Price Predictor

Bharadwaj Srikaa

Date: 22-05-2024

Abstract: SmartPrice Predictor is an innovative application designed to forecast smartphone prices using advanced segmentation machine learning models. By analyzing various features such as brand, specifications, user reviews, market trends, and historical pricing data, SmartPrice Predictor provides accurate and reliable price predictions for a wide range of smartphones. This application is tailored to assist consumers, retailers, and manufacturers in making informed decisions regarding smartphone purchases, pricing strategies, and market positioning. The intuitive interface and powerful algorithms ensure that users receive precise pricing insights, enhancing their overall purchasing and marketing experience.

Step 1: Prototype Selection

Problem Statement

Mobiles, first invented in 1992 and launched in 1994 by the techno-giant IBM, have become an integral part of the lives of human beings. Today, these technical devices serve a multitude of purposes - calling, video calls, texts, internet, mailing, playing games, taking pictures, shopping etc. Due to these very purposes, the buyers often take many parameters into consideration such as brand, processor, memory size (internal & external), camera, battery backup, among others. However, one parameter that is generally not considered is the price. As such, the main objective of this report is to introduce a system to cross-validate the price of a mobile phone based on its features.

Market/Customer/Business Need Assessment

Price is the most important side of shopping. Customers are very often interested in knowing the price of the item they wish to buy. Likewise, they are also interested in knowing whether the item is worth the price or not given its features. Hence, the type of service proposed here will enable the common man to have an estimate of the price of a mobile before making a purchase.

Target Specifications

The service will be essential for almost everyone in predicting the mobile price by means of:

Brand

Display Size

RAM Capacity

Internal Memory (ROM)

Type of Android

Resolution

Mobile weight

Battery Support (in mAh)

External Search

- ➔ Dataset
- ➔ Business Models.
- ➔ Financial Modelling.
- ➔ Market Trends

Cleaning The data:

Dataset before Cleaning it:

	mobile_name	mobile_price	mobile_color	dual_sim	disp_size	resolution	os	num_cores	mp_speed	int_memory	ram	p_cam
0	Realme Narzo 10A (So White, 32 GB)	₹8,999	So White	Dual Sim	16.56 cm (6.52 inch)	1600 x 720 Pixels	Android 10	Octa Core	2 GHz	32 GB	3 GB	12MP + 2MP + 2MP
1	Realme Narzo 10A (So Blue, 32 GB)	₹8,999	So Blue	Dual Sim	16.56 cm (6.52 inch)	1600 x 720 Pixels	Android 10	Octa Core	2 GHz	32 GB	3 GB	12MP + 2MP + 2MP
2	Realme Narzo 10A (So Blue, 64 GB)	₹9,999	So Blue	Dual Sim	16.56 cm (6.52 inch)	1600 x 720 Pixels	Android 10	Octa Core	2 GHz	64 GB	4 GB	12MP + 2MP + 2MP
3	Realme Narzo 10A (So White, 64 GB)	₹9,999	So White	Dual Sim	16.56 cm (6.52 inch)	1600 x 720 Pixels	Android 10	Octa Core	2 GHz	64 GB	4 GB	12MP + 2MP + 2MP
4	Realme Narzo 10 (That Green, 128 GB)	₹11,999	That Green	Dual Sim	16.51 cm (6.5 inch)	1600 x 720 Pixels	Android 10	Octa Core	2 GHz	128 GB	4 GB	48MP + 8MP + 2MP + 2MP

Dataset after cleaning it:

	mobile_price	disp_size	resolution	os	num_cores	mp_speed	int_memory	ram	battery_power	mob_width	...	Orange	Purple	
0	8999	16.56	1152000	10	8	2.0	32	3	5000	75.0	...	0	0	
1	8999	16.56	1152000	10	8	2.0	32	3	5000	75.0	...	0	0	
2	9999	16.56	1152000	10	8	2.0	64	4	5000	75.0	...	0	0	
3	9999	16.56	1152000	10	8	2.0	64	4	5000	75.0	...	0	0	
4	11999	16.51	1152000	10	8	2.0	128	4	5000	75.4	...	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
454	13490	14.48	1036800	7	8	2.5	32	3	3180	73.1	...	0	0	
455	9490	14.48	1036800	7	8	2.5	16	2	3180	73.1	...	0	0	
456	22990	15.82	2462400	8	8	2.0	64	4	3400	75.3	...	0	0	
457	13990	16.59	2527200	9	8	2.1	128	4	4020	76.1	...	0	0	
458	13990	16.59	2527200	9	8	2.1	128	4	4020	76.1	...	0	1	

459 rows × 54 columns

Why cleaning the dataset is very important before performing any analysis on it?

Cleaning the dataset is crucial before performing any analysis because it ensures the accuracy and reliability of the results. Raw data often contains errors, inconsistencies, missing values, and duplicates that can skew analysis and lead to incorrect conclusions. By cleaning the dataset, we eliminate noise and irrelevant information, allowing for more precise and meaningful insights. This process enhances the quality of the data, making it suitable for robust statistical analysis and machine learning models, ultimately leading to more reliable, valid, and actionable outcomes.

Exploratory Data Analysis (EDA):

What is Exploratory Data Analysis?

Exploratory Data Analysis (EDA) is a critical step in the data analysis process that involves summarizing and visualizing the main characteristics of a dataset, often using graphical techniques. EDA helps analysts understand the underlying patterns, trends, and relationships within the data, identify anomalies, and formulate hypotheses. By employing various statistical tools and visualizations such as histograms, box plots, scatter plots, and correlation matrices, EDA provides insights that guide further data processing and analysis. This preliminary exploration is essential for making informed decisions about which modeling techniques to apply and for ensuring the integrity and quality of the analytical results.

These are the results of Exploratory Data Analysis performed on my dataset.

Heatmap:
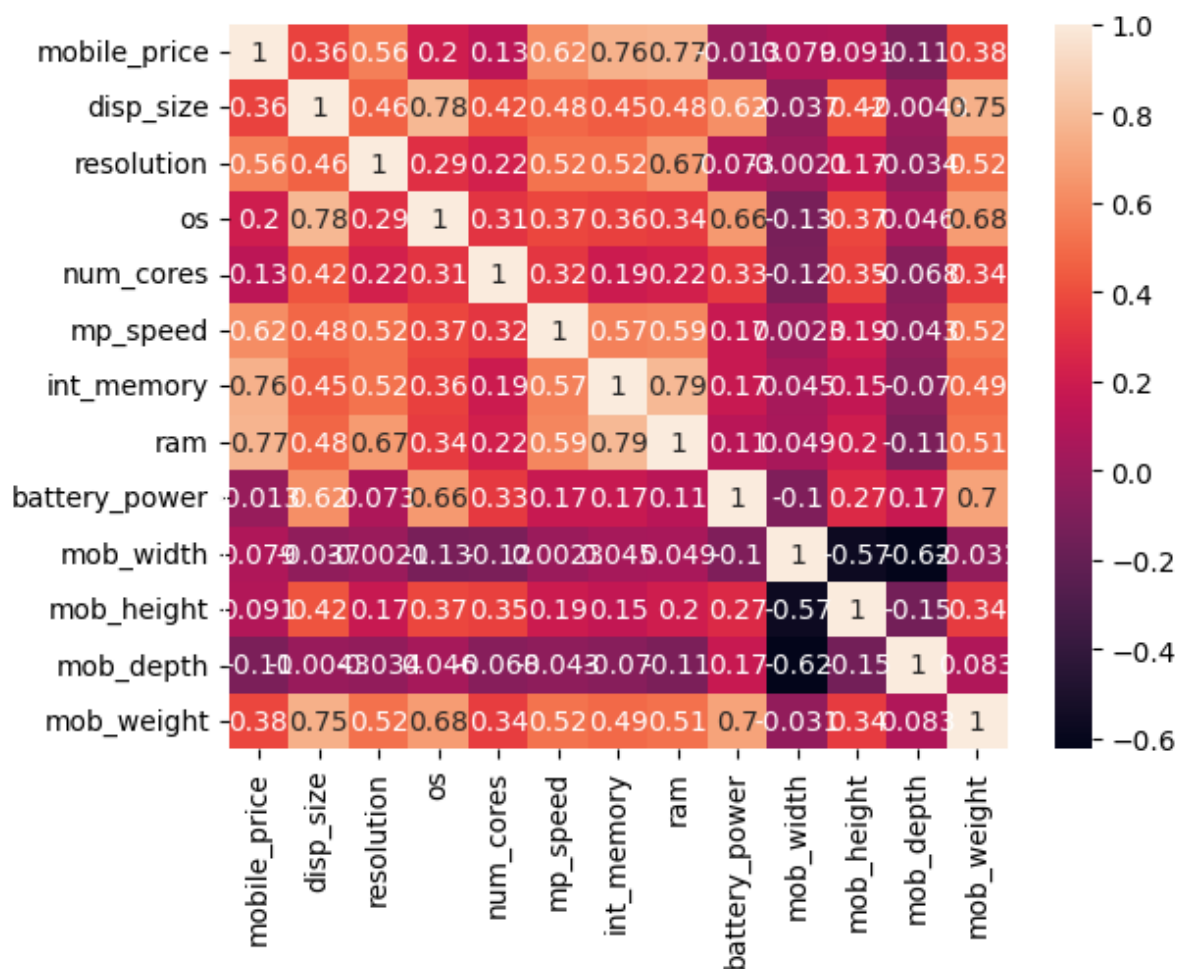
Pairplot Graph:
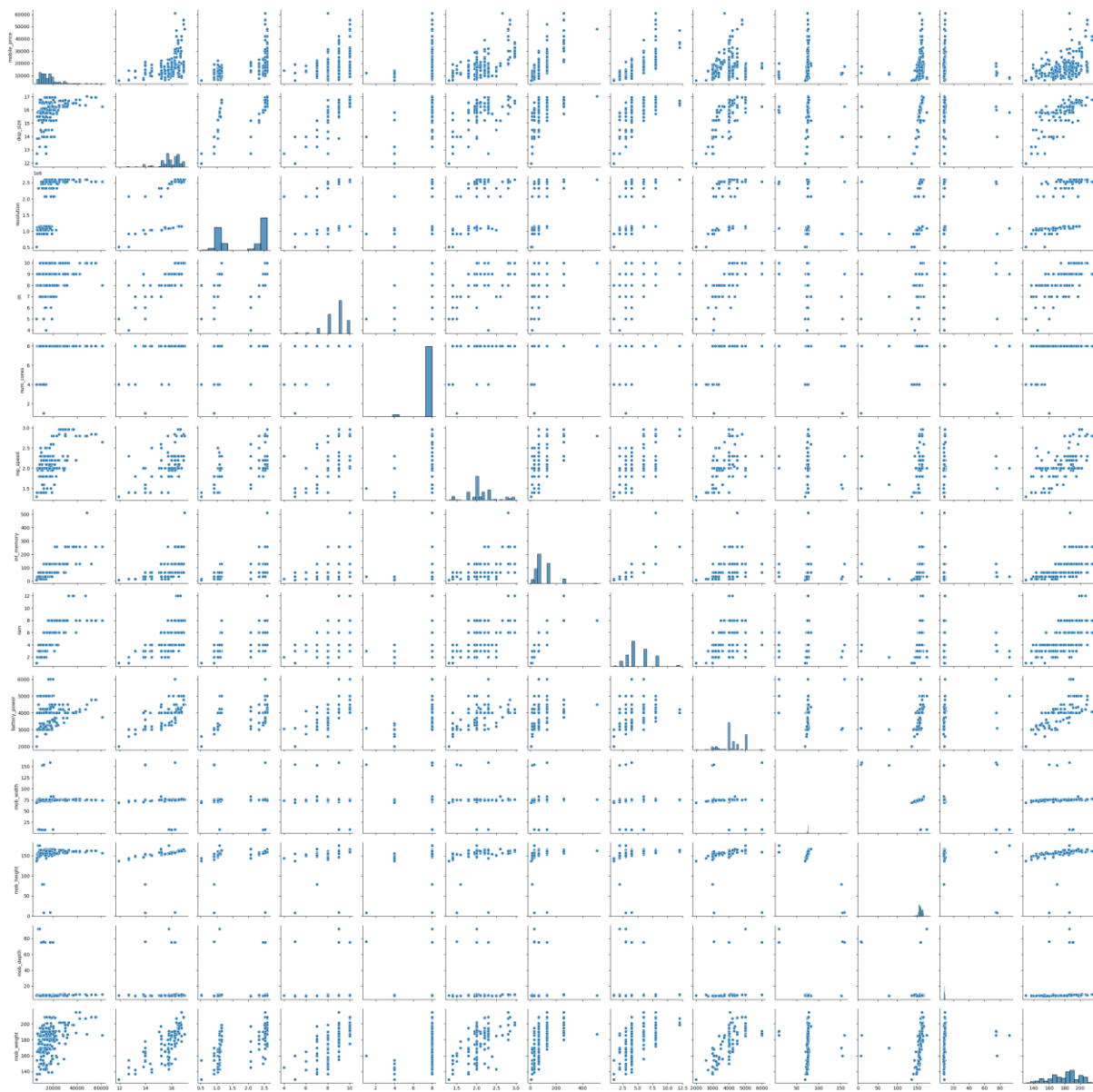
In data analysis, a plot graph provides a visual representation of the data, making it easier to identify patterns, trends, and relationships that might not be immediately apparent from raw data alone. Plot graphs can reveal distributions, highlight outliers, show correlations between variables, and illustrate changes over time. For instance, scatter plots can display the relationship between two continuous variables, bar charts can compare categorical data, histograms can show the distribution of a single variable, and line graphs can track changes over intervals. By translating complex data into a visual format, plot graphs enhance understanding and interpretation, facilitate comparisons, and support data-driven decision-making.
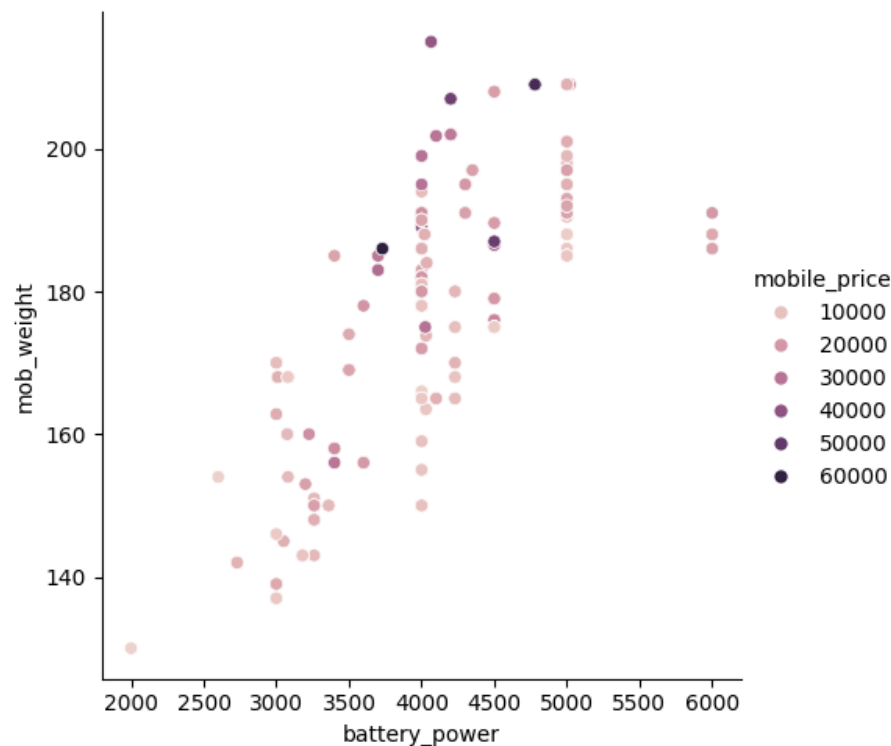
Relationship plot between Display size, Resolution and Mobile Price:
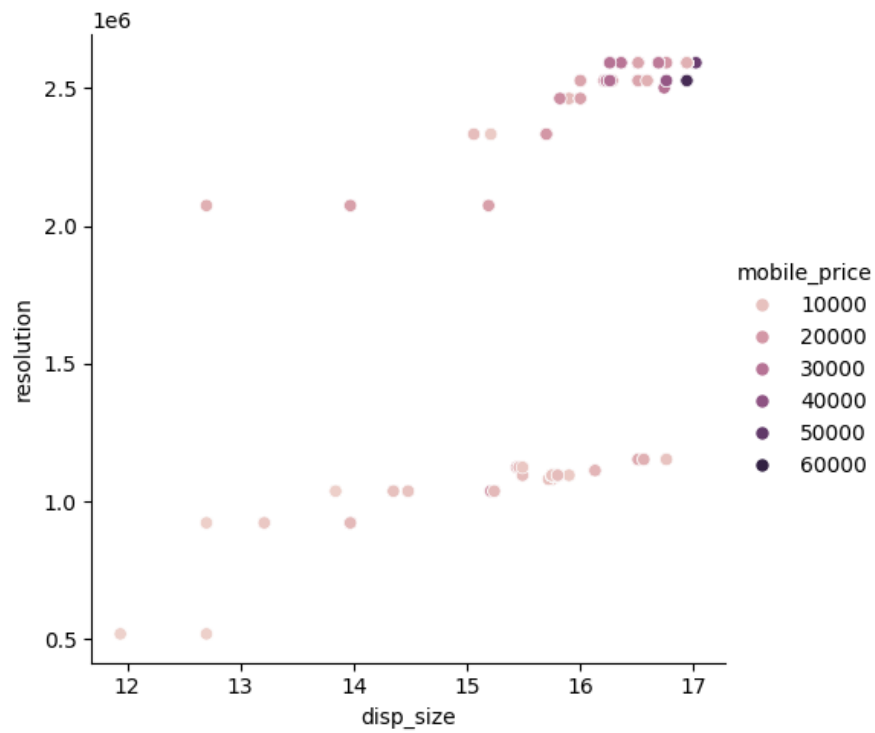
Relationship plot between Display size, Mobile Weight, and Mobile Price:
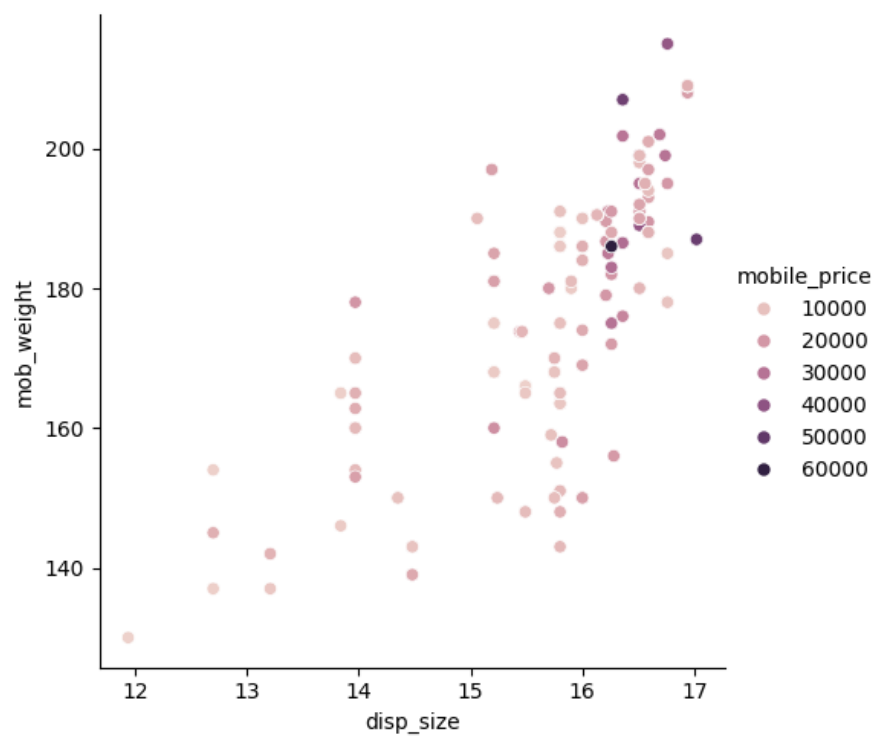
Relationship plot between Battery Power, Mobile Weight, and Mobile Price:


Distribution Plot:

In data analysis, a displot (distribution plot) provides valuable information about the distribution of a single continuous variable. It combines aspects of a histogram and a kernel density estimate (KDE) plot, allowing analysts to visualize the frequency distribution and the underlying probability density function of the data. Key insights from a displot include the shape of the distribution (e.g., normal, skewed, bimodal), the central tendency, variability, presence of outliers, and overall spread of the data. By understanding these characteristics, analysts can make informed decisions about further statistical analysis and appropriate modeling techniques.

Additionally, a displot can reveal patterns and insights that might not be apparent through summary statistics alone. It helps in identifying data distribution anomalies such as skewness, kurtosis, or multimodality, which can indicate the presence of different subpopulations within the dataset. By overlaying a KDE curve on the histogram, the displot smooths out the data to provide a clearer picture of the distribution's overall shape, making it easier to spot trends and deviations. This visualization is especially useful in comparing distributions across different subsets of data, facilitating a deeper understanding of the relationships and differences within the dataset. Overall, the displot is an essential tool in exploratory data analysis for comprehensively assessing the distributional properties of a variable.
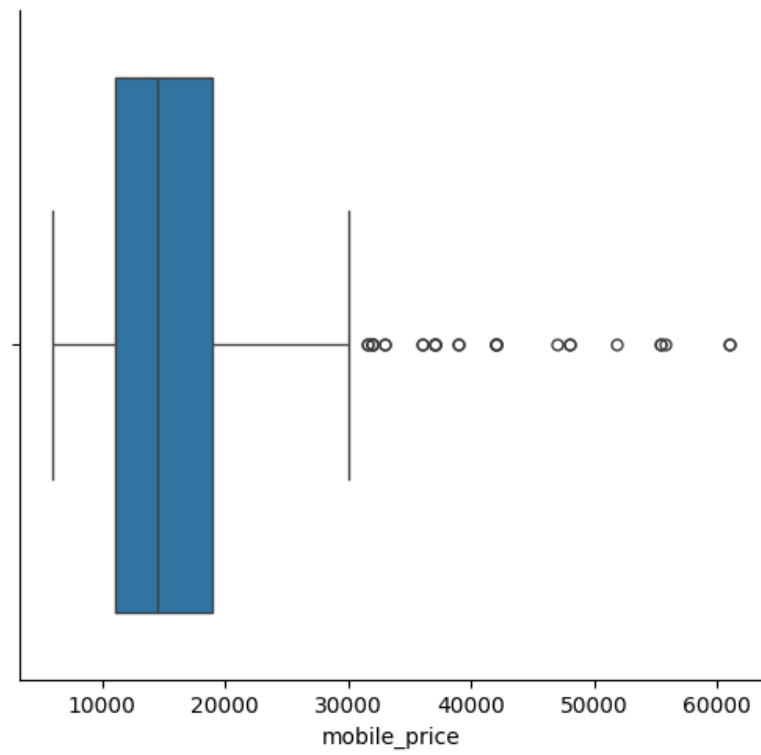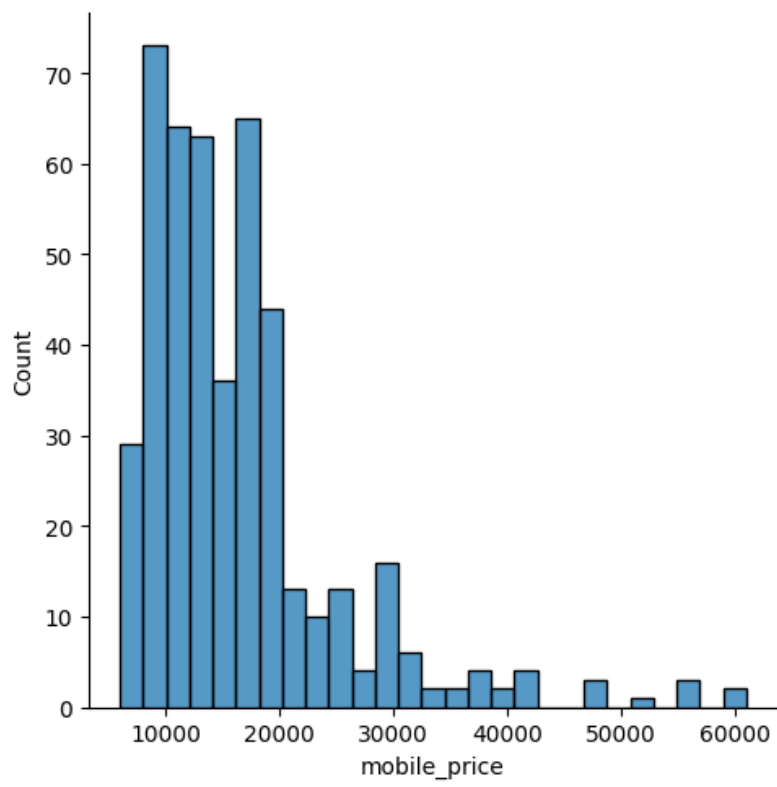
Principal Component Analysis:
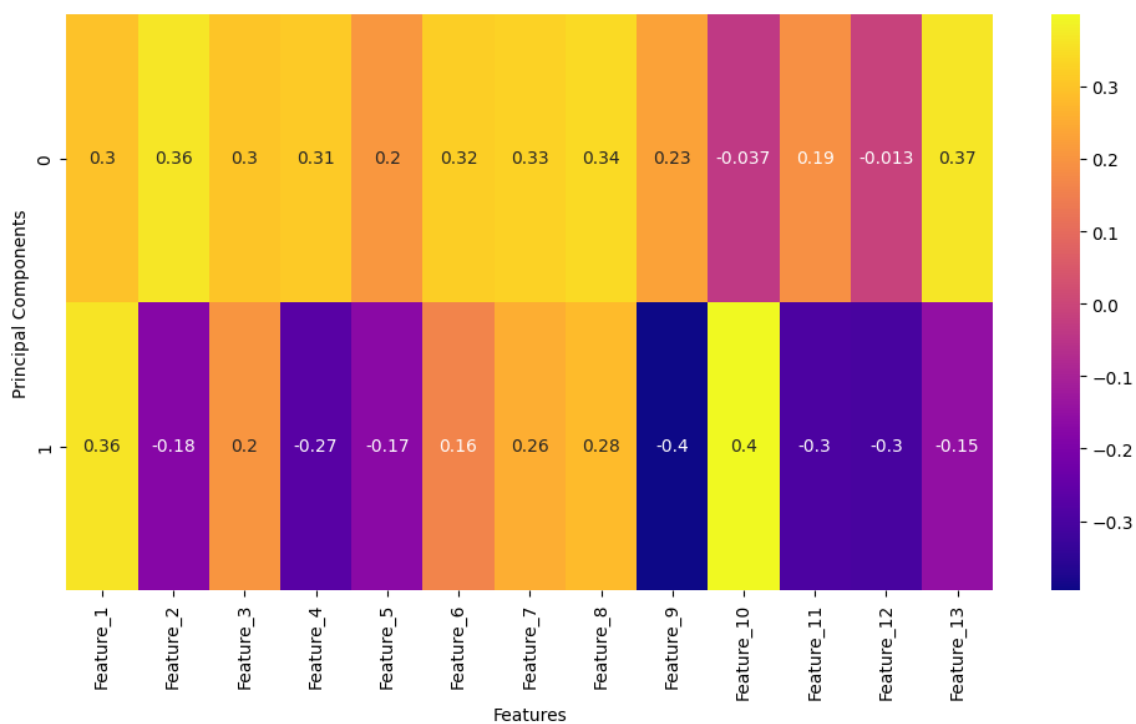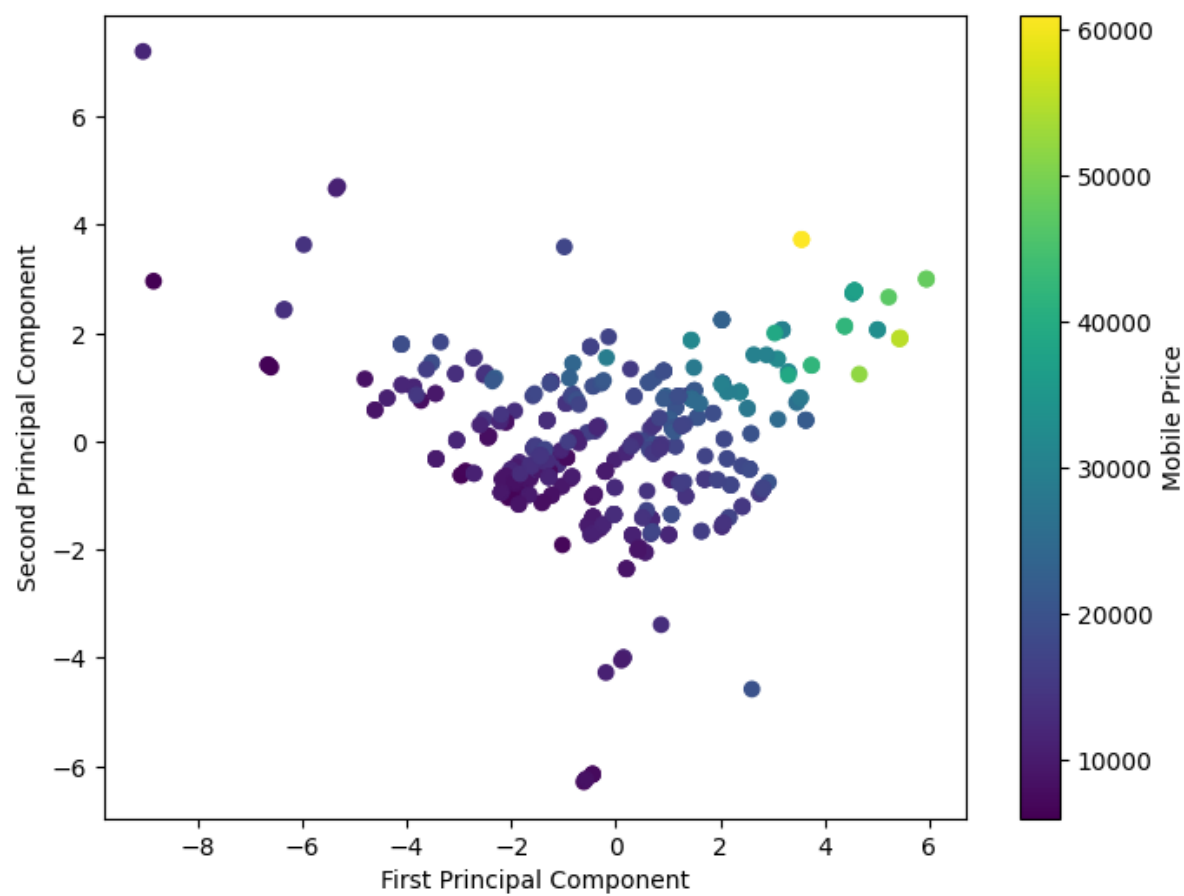
Applicable Regulations:

General Data Protection Regulation (GDPR): Applies to the processing of personal data of individuals within the European Union. It mandates strict data privacy and security measures, requiring user consent, data minimization, and the right to access and delete personal data.

California Consumer Privacy Act (CCPA): Provides California residents with the right to know what personal data is being collected, the purpose for its collection, and the right to access, delete, and opt-out of the sale of their personal data.

Federal Trade Commission (FTC) Regulations: In the United States, the FTC enforces rules against unfair or deceptive practices, including those related to data privacy and security.

Implement robust data security measures to protect against breaches and unauthorized access, complying with standards such as ISO/IEC 27001 for information security management.

Be aware of emerging regulations and guidelines aimed at ensuring the fairness, transparency, and accountability of AI systems. This includes avoiding biases in data and algorithms and providing explanations for automated decisions.

Business Opportunity:

Personalized Recommendations: By providing accurate price predictions, consumers can make informed decisions about the best time to purchase a smartphone, maximizing savings and value.

Competitive Pricing Insights: Consumers can use the app to compare predicted prices across different brands and models, enhancing their purchasing strategy.

Retailers and Manufacturers: Smartphone retailers and manufacturers can use the application's insights to understand market trends, optimize pricing strategies, and manage inventory more effectively.

Competitor Analysis: Businesses can gain insights into competitor pricing strategies and market positioning, helping them stay competitive.

Targeted Advertising: The application can partner with smartphone manufacturers and retailers to offer targeted advertisements based on user preferences and predicted price drops.
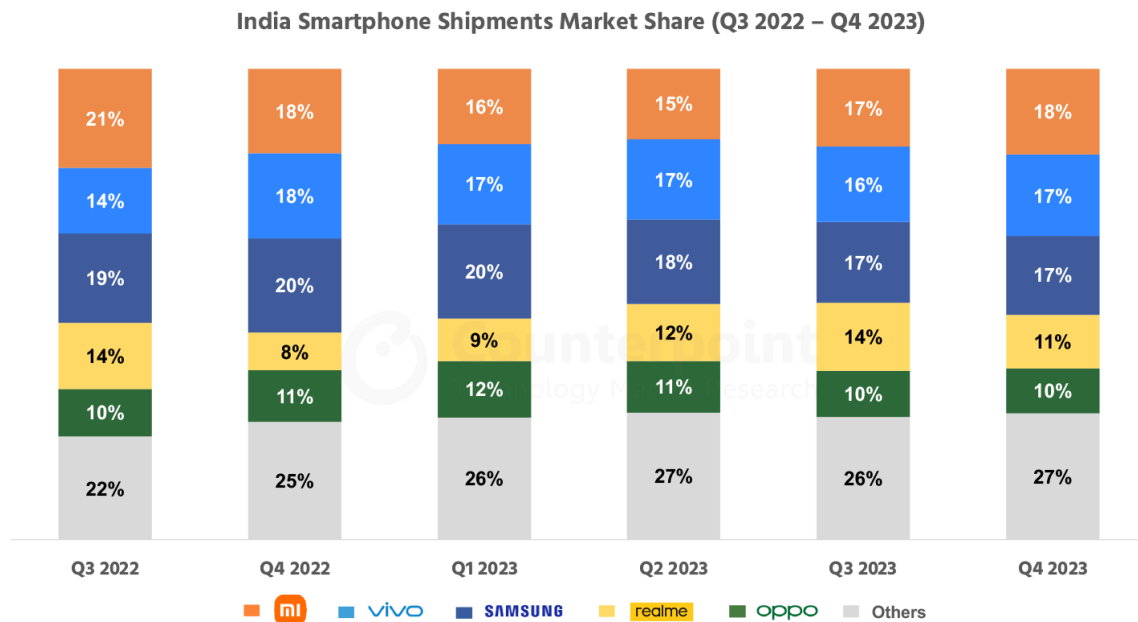
Affiliate Marketing: Monetize the app through affiliate marketing by directing users to purchase smartphones through partnered retailers, earning commissions on sales.

Premium Features: Offer a subscription model for advanced features such as detailed market trend reports, personalized alerts for price drops, and early access to predictions for new smartphone models.

Business Intelligence: Provide premium business intelligence reports to industry players, including trend analysis and predictive analytics.

Broader Product Range: Expand the predictive model to include other consumer electronics or high-demand products, creating a comprehensive price prediction platform.

Statistics of smartphone market in India

India Smartphone Shipments Market Share (Q3 2022 – Q4 2023)

	Q3 2022	Q4 2022	Q1 2023	Q2 2023	Q3 2023	Q4 2023
mi	21%	18%	16%	15%	17%	18%
vivo	14%	18%	17%	17%	16%	17%
SAMSUNG	19%	20%	20%	18%	17%	17%
realme	14%	8%	9%	12%	14%	11%
OPPO	10%	11%	12%	11%	10%	10%
Others	22%	25%	26%	27%	26%	27%

Financial Equation

The above diagram indicates how the smartphone companies in India have been growing for the past 2 years. Even with the impact of COVID, the industry has continued to grow. So if we are following the above trend it would be advisable to price our service around Rs.9000.

Once the customer base increases, we can either increase the price or reduce the duration for which our product will be available. Let's assume that the duration of developing the ML model takes about 3 weeks and the cost for producing the model is the salary of the members the team.

Let there be three ML engineers and two full stack web developer. Let the salary of the ML engineers be 'a' and the full stack web developer be 'b'. So, the total cost $c = 3*a + 2*b$.

So, the profit or financial equation will look like this $y = 9000*x(t) – (3*a + 2*b)$

Here $x(t)$ is a function that represents the growth of the customer base and y is the profit.