



Flipkart
VIVO

Mobile Phone Brand Prediction

Using Machine Learning

SRI ATCHYA S

Introduction

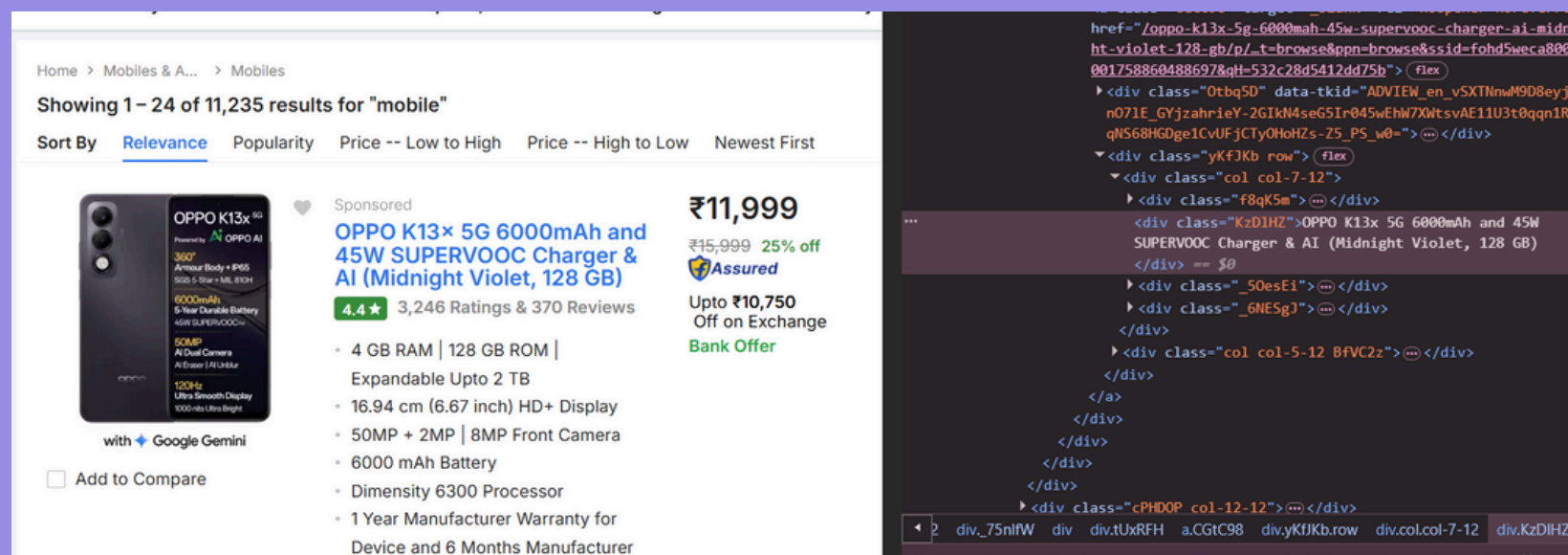
- Collected around 7,000 mobile phone records by web scraping Flipkart.
- Performed data cleaning and preprocessing to handle missing values, duplicates, and inconsistencies.
- Conducted Exploratory Data Analysis to study feature distributions and brand insights.
- Applied both supervised mode and unsupervised mode.
- Compared model performances, selected the best-performing model, and used it to predict mobile phone brands.



Web scrapping

- Collected data from Flipkart using Python libraries (BeautifulSoup / Requests / Selenium).
- Extracted details such as brand, price, storage, battery, display size, and camera.
- Scraped around 7,000 mobile phone records for analysis.
- Stored data into a structured CSV/Excel format for further processing.
- Automated scraping with loops & pagination to cover multiple product pages.

Websrapping



Data Cleaning



- **MISSING VALUES**

- Filled missing values with 0 for numerical columns.
- Used price correlations to estimate and fill missing values in camera, battery, and display features.

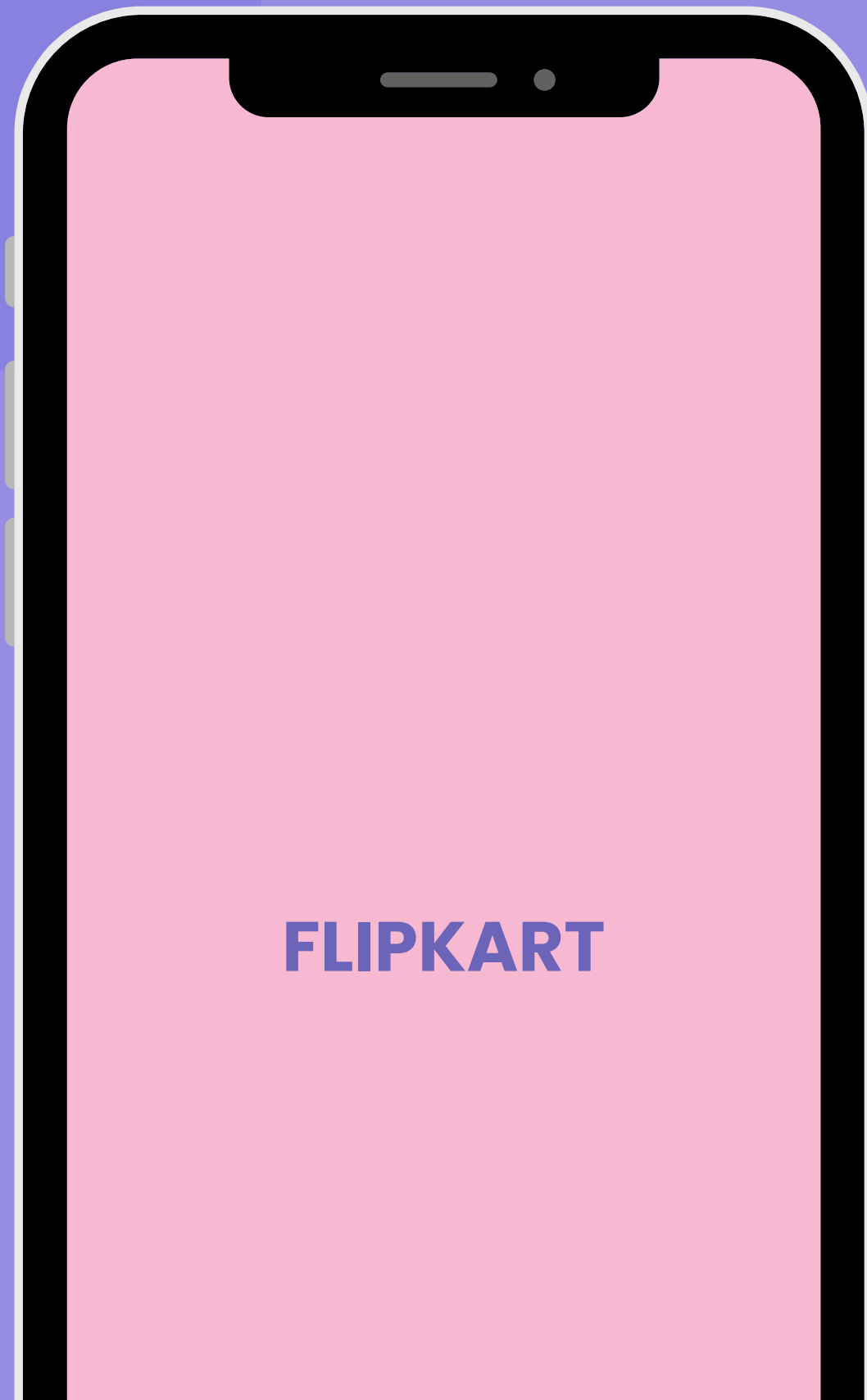
- **INCONSISTANT VALUE**

- Dropped rows where Price was NaN and converted float values to integers.
- Normalized inconsistent entries (e.g., changed 4685 → 1000) for better consistency.

- **OUTLIER FINDING**

- Identified outliers in features using boxplots.
- Chose to retain outliers instead of removing them to preserve real-world variations.

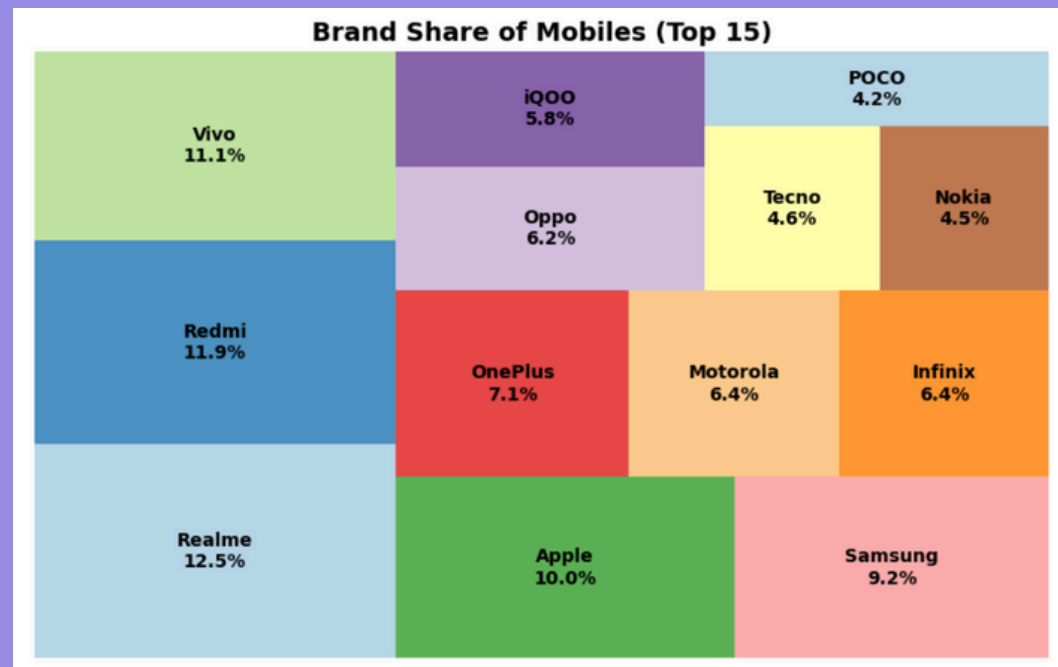
SQL DUMP AND ENCODING



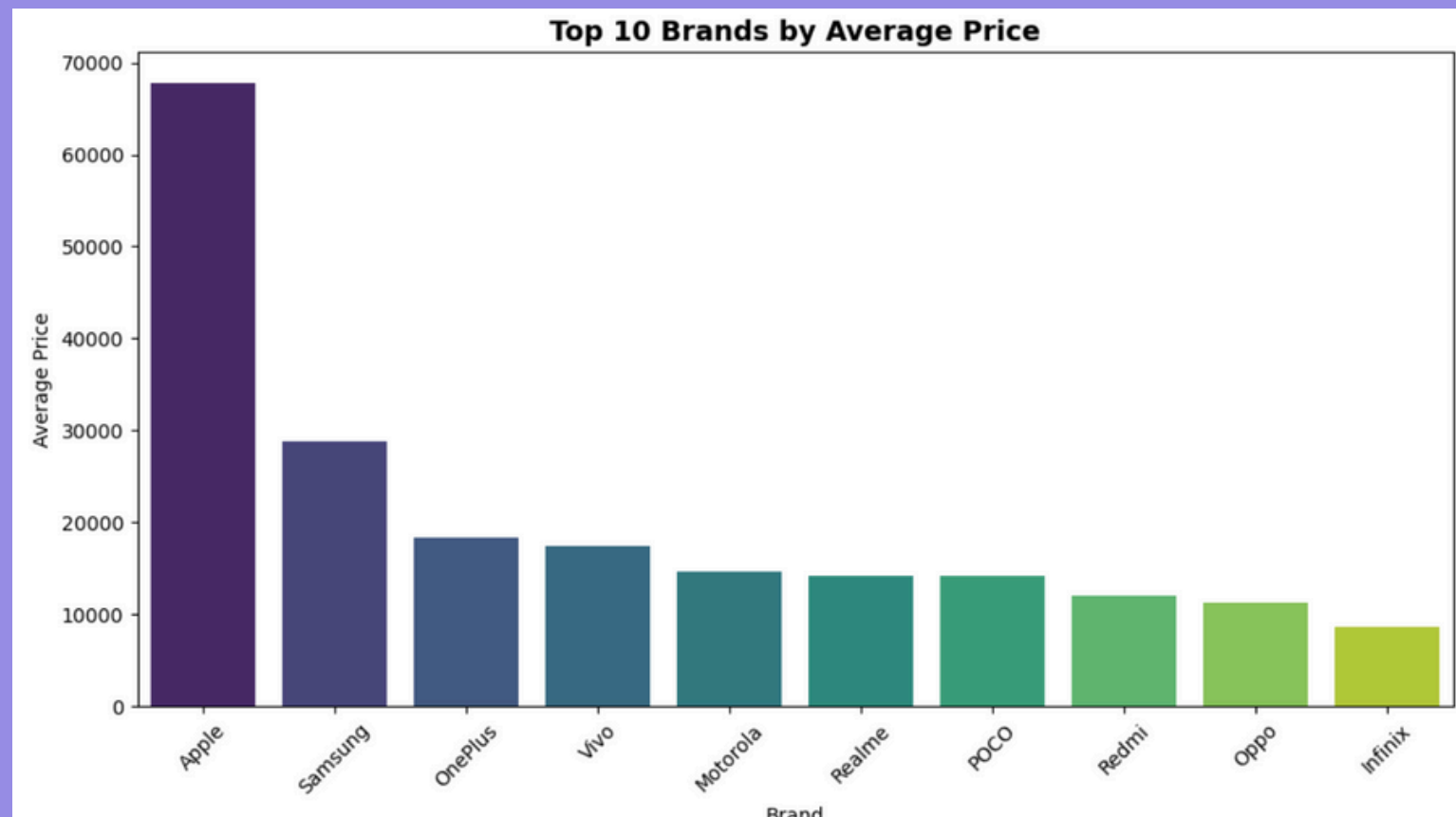
FLIPKART

- Exported into SQL database using `pandas.to_sql()` for structured storage and easy querying.
- Helps in data backup, scalability, and integration with other applications.
- Applied Label Encoding to convert categorical features into numeric values.
- Essential for algorithms that can't handle categorical data directly, ensuring the model can process all features.

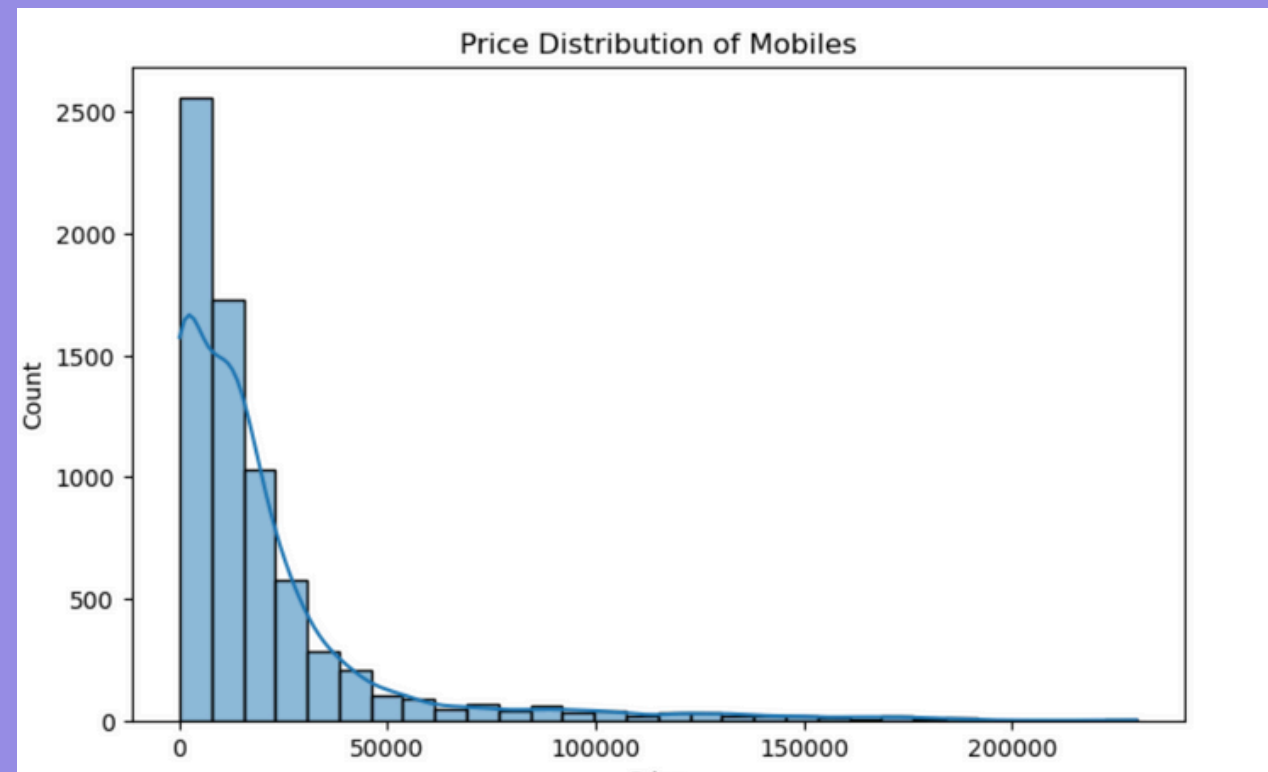
EDA



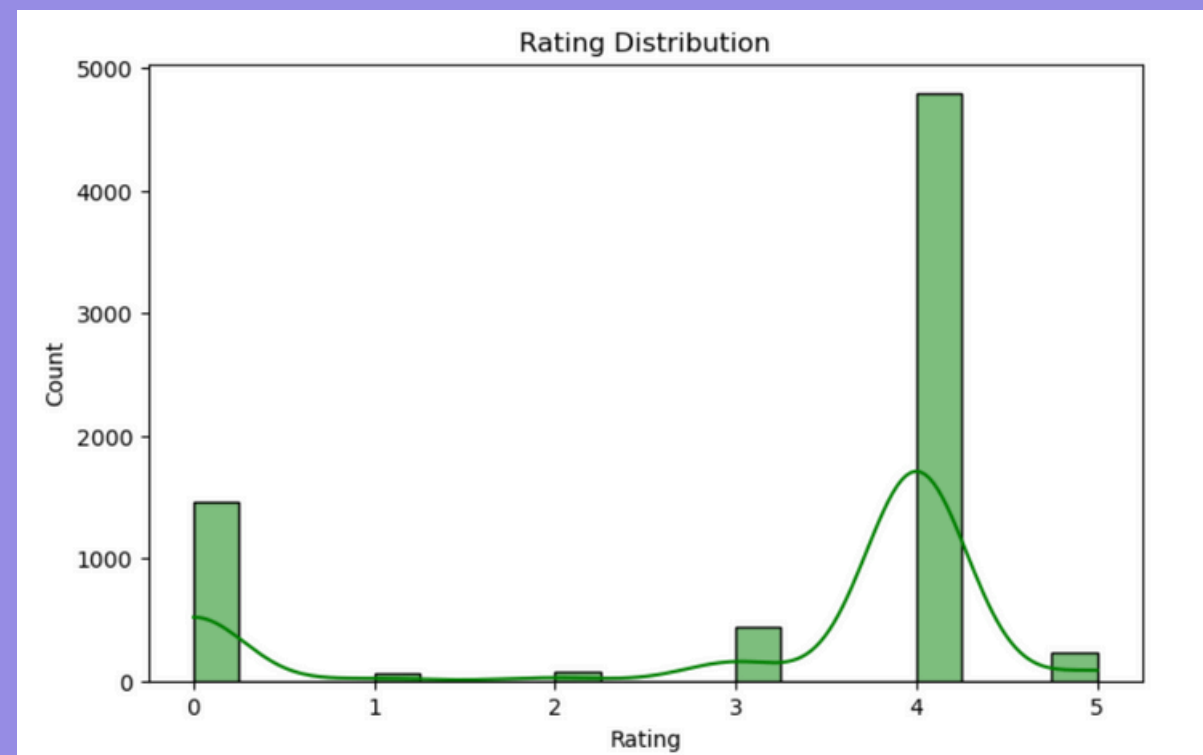
- Realme , Redmi , and Vivo dominate the mobile market among the top 15 brands.
- Apple and Samsung also hold strong positions.



- Apple dominates in pricing.
- Samsung is the next tier.
- OnePlus and Vivo are mid-premium brands:



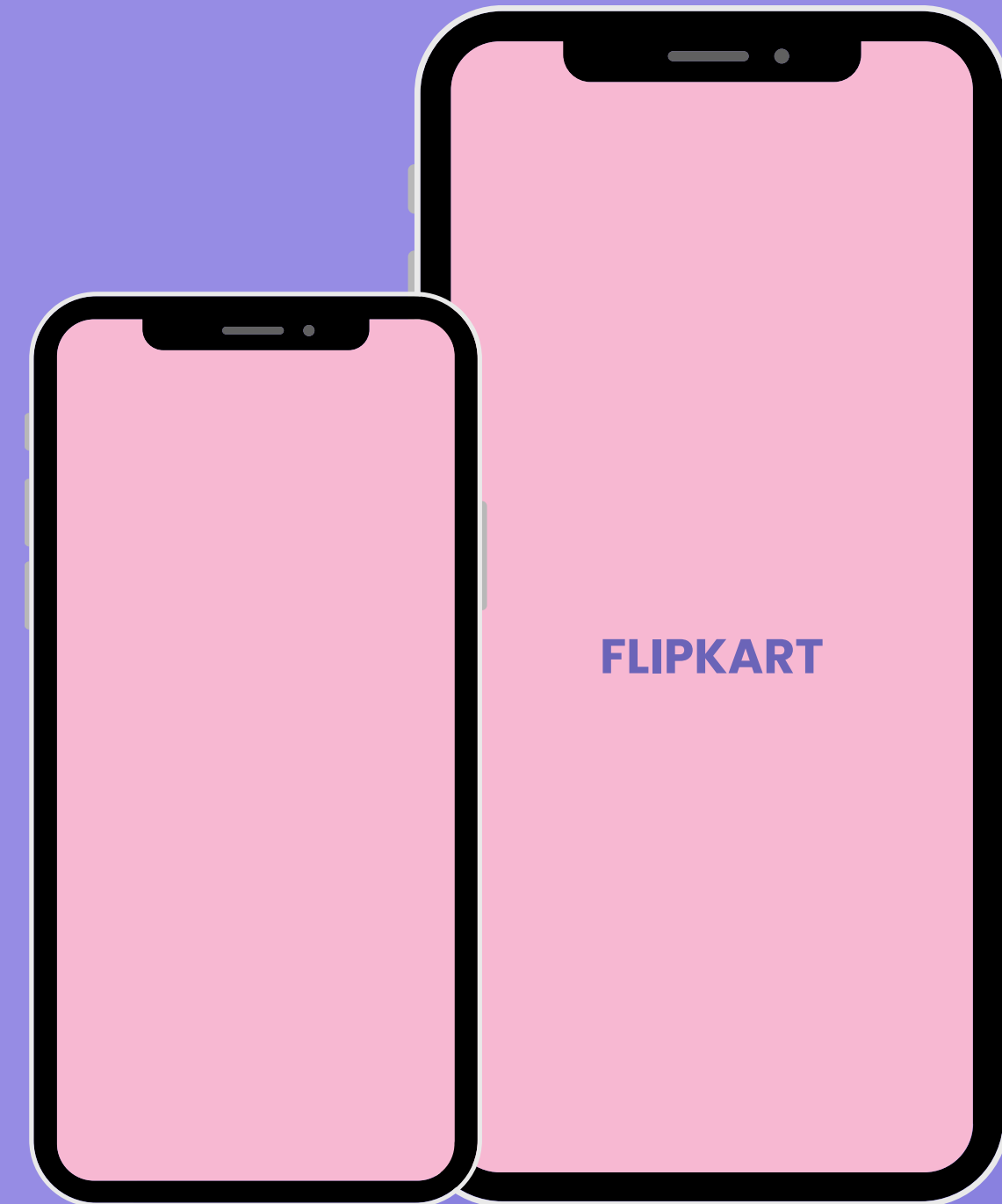
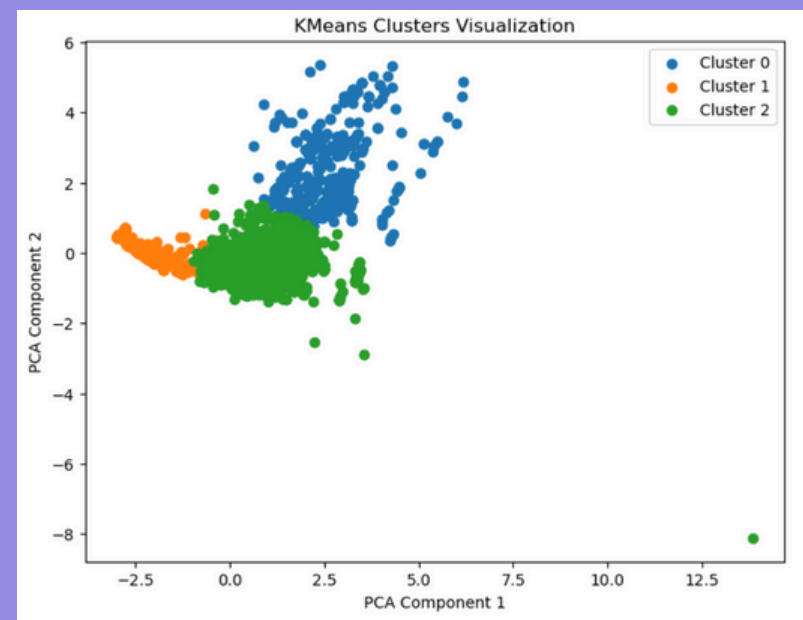
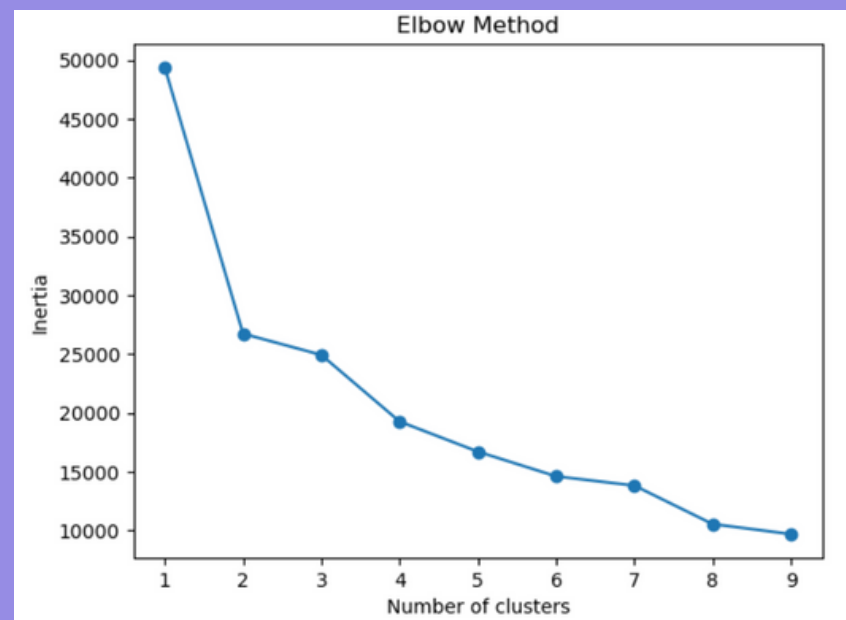
- Most mobiles are priced below ₹20,000, showing strong demand in the budget segment.
- The distribution is right-skewed, with fewer mobiles in the premium range above ₹50,000.
- A small but noticeable tail extends beyond ₹1,00,000, representing high-end flagship models.



- The majority of mobiles are rated 4 stars, indicating overall good customer satisfaction.
- A significant count at 0 rating suggests missing or unrated products in the dataset.
- Very few mobiles have ratings between 1 to 3 or exactly 5, showing ratings are polarized towards 4.

UNSUPERVISED LEARNING

- K MEANS CLUSTER



SUPERVISED LEARNING

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine(SVM)
- k-Nearest Neighbors (KNN)
- XGBoost



FLIPKART

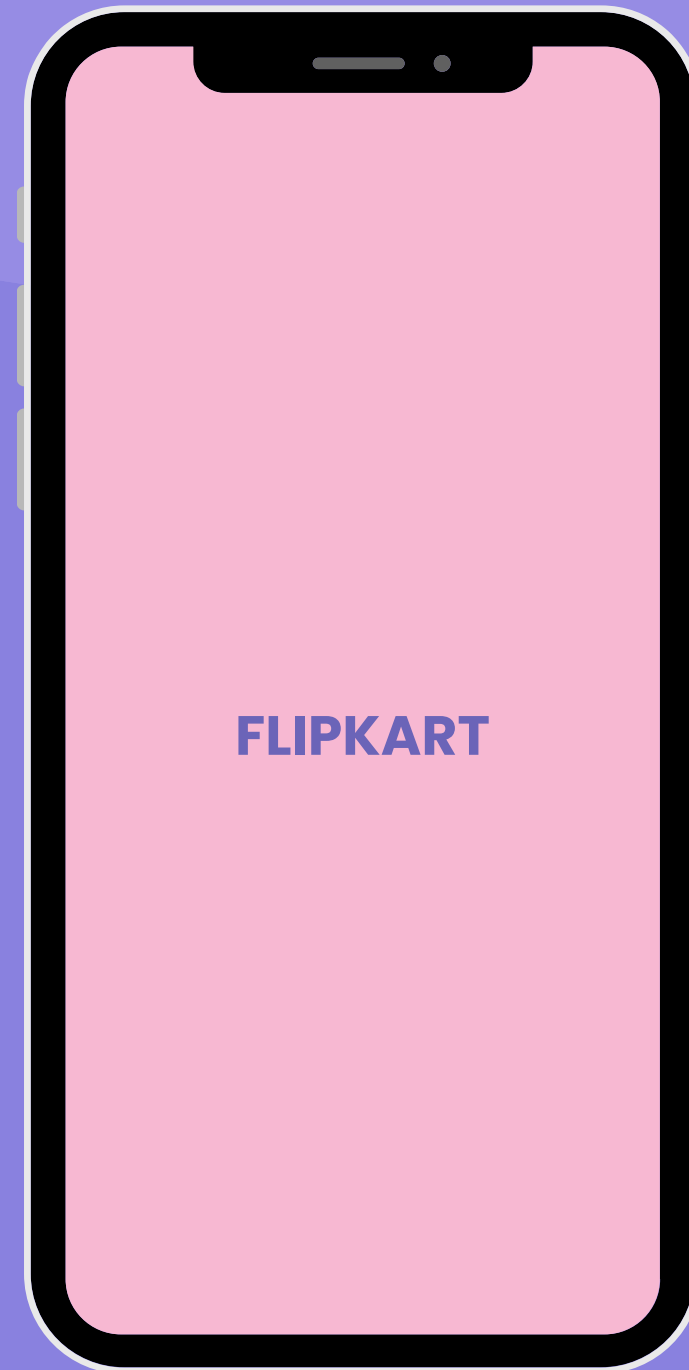
Hyperparameter tuning and cross validation

- **CROSS VALIDATION**

- Performed cross-validation (CV) across multiple supervised learning models to ensure robust evaluation.
- Identified the best-performing model by comparing metrics such as accuracy, precision, recall, and F1-score.

- **HYPERPARAMETER TUNNING**

Conducted randomized search to tune hyperparameters and improve XGBoost model performance.



CONCLUSION

A stylized illustration of a smartphone with a black frame and a pink screen. The word "FLIPKART" is displayed in blue capital letters on the screen. The phone is positioned on the left side of the slide, with a light purple background behind it.

FLIPKART

- Applied supervised learning models with cross-validation to identify the best-performing algorithm.
- Hyperparameter tuning with randomized search improved the XGBoost model performance.
- Final model achieved strong results on key metrics, making it suitable for deployment and decision-making.

THANK YOU