TV TELEVISION PRICE PREDICTION AND CLUSTERING USING WEB SCRAPING AND MACHINE LEARNING

SRI HARI HARAN CR
DADS BATCH JUNE
CAPSTONE PROHECT

Introduction to Web Scraping

Predicting trends with data analysis for better decision-making. This presentation delivers a clear overview on Flipkart TV data insights.

What is Web Scraping?

Automatically extracting data from websites for analysis.

Why Flipkart TVs?

Large variety, dynamic pricing, valuable trend data.

Tools Used

Python libraries like Beautiful Soup and Scrapy.

WEB SCRAPING LAPTOP DATA FROM FLIPKART

TV TELEVISION DETAILS INCLUDING PRICES WERE SCRAPED FROM FLIPKART USING BEAUTIFULSOUP TO CREATE A STRUCTURED DATASET FOR ANALYSIS AND PREDICTION.

DATASET

Number Of Rows	561
Number Of Columns	7

COLUMNS : PRODUCT NAME, PRICE, REVIEW, RATING REVIEW , SUPPORTE APPS , OS, SOUND SYSTEM

DATA PREPROCESSING

- **STANDARDIZE DATA TYPES**
- **HANDLE MISSING VALUES**
- **REMOVE DUPLICATES**
- **OUTLIER DETECTION AND REMOVAL**

Data Scraping in Flipkart TV Data

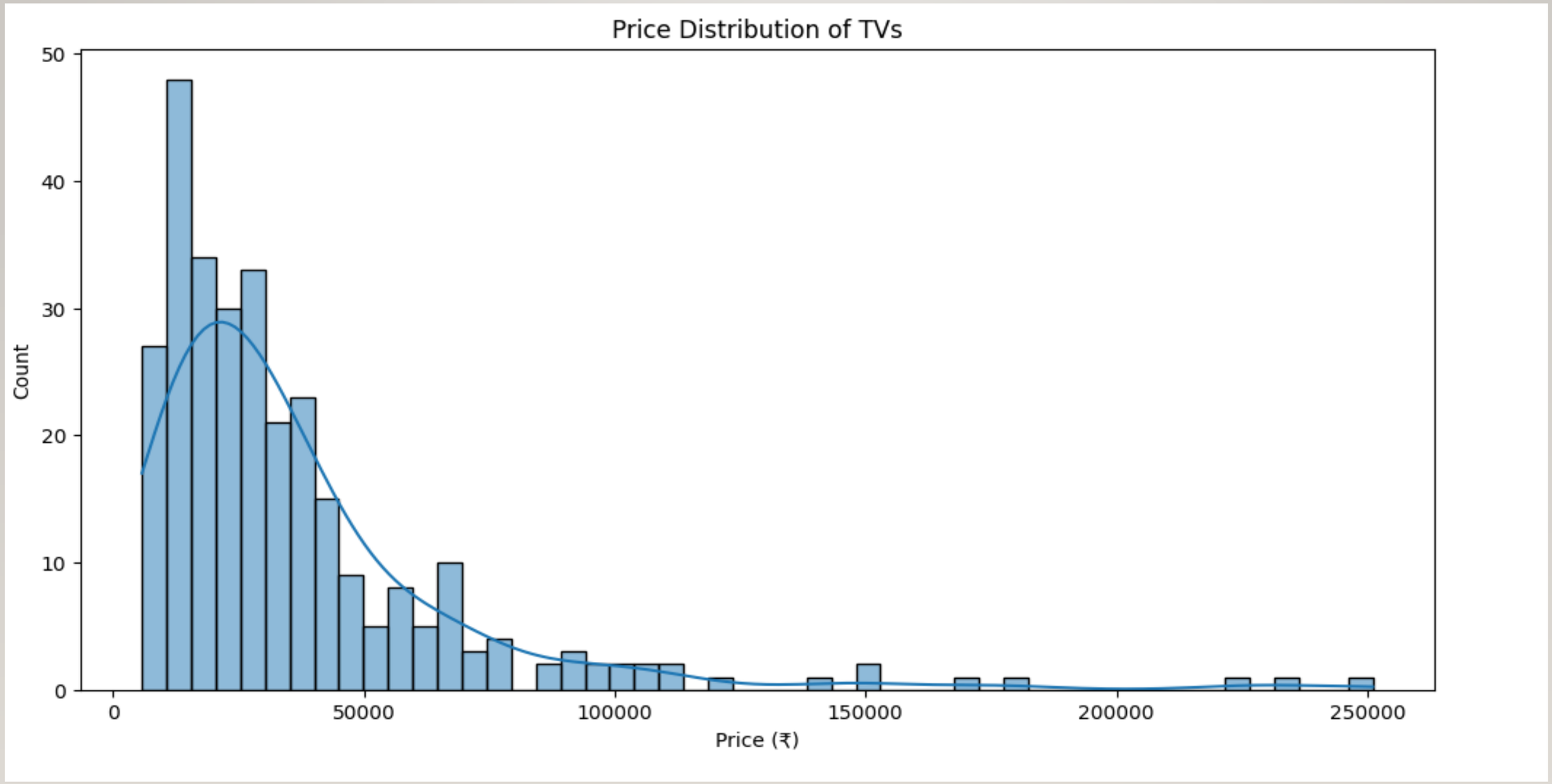
Key Data Points

- Product name
- Price
- Rating
- Reviews
- Supported Apps
- OS
- Resolution
- Sound system

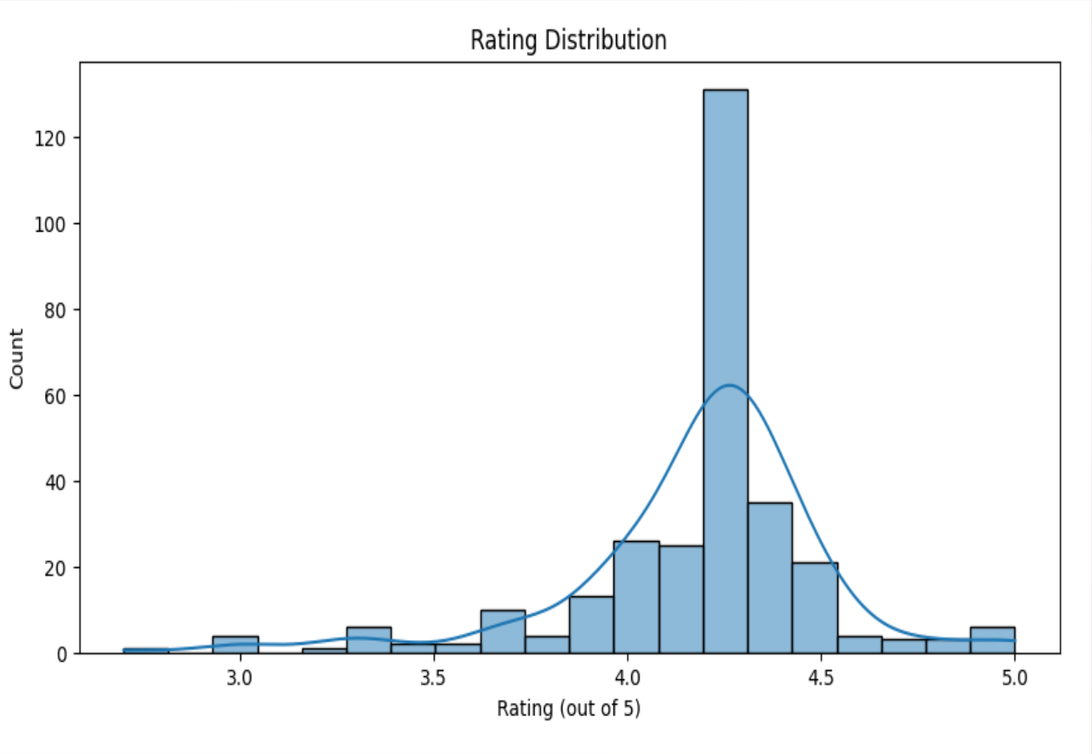
Data Management

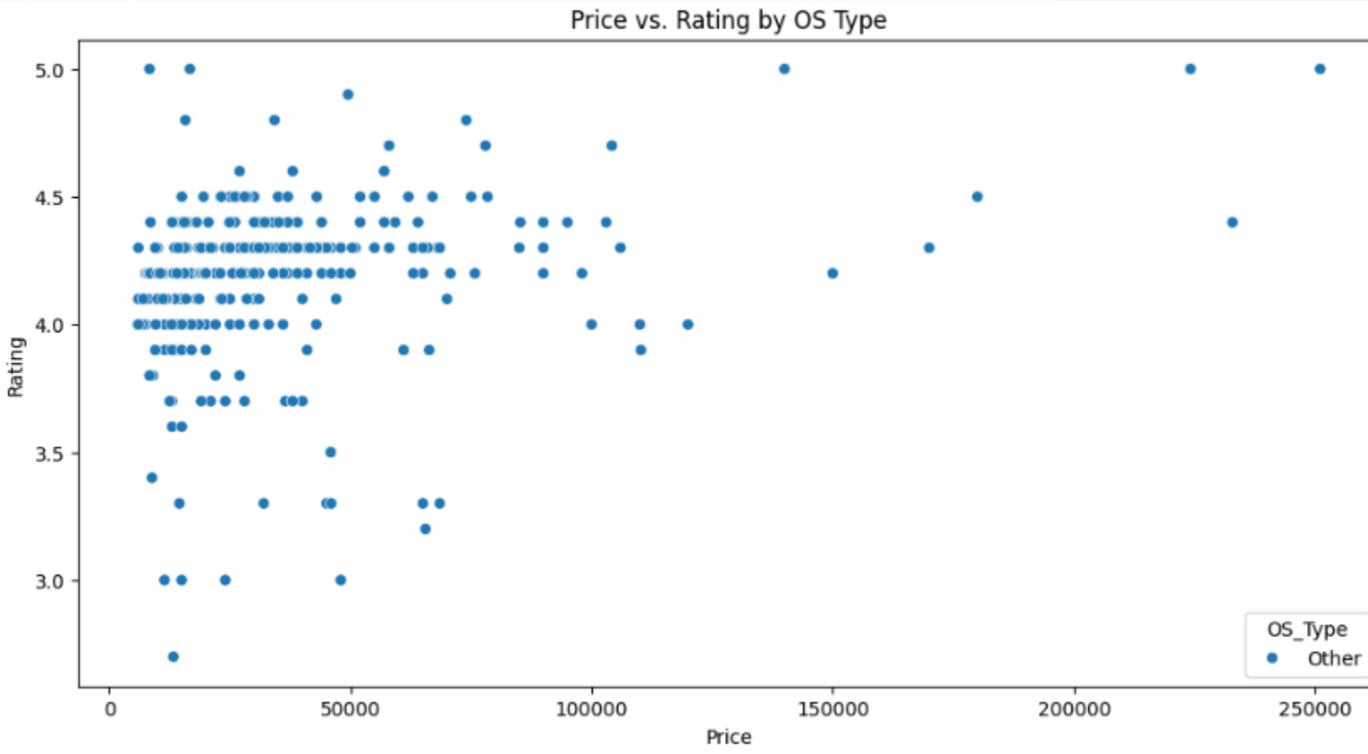
- Cleaning and formatting inconsistencies
- Storing in CSV
- Collected Data are total 561

VISUALIZATION

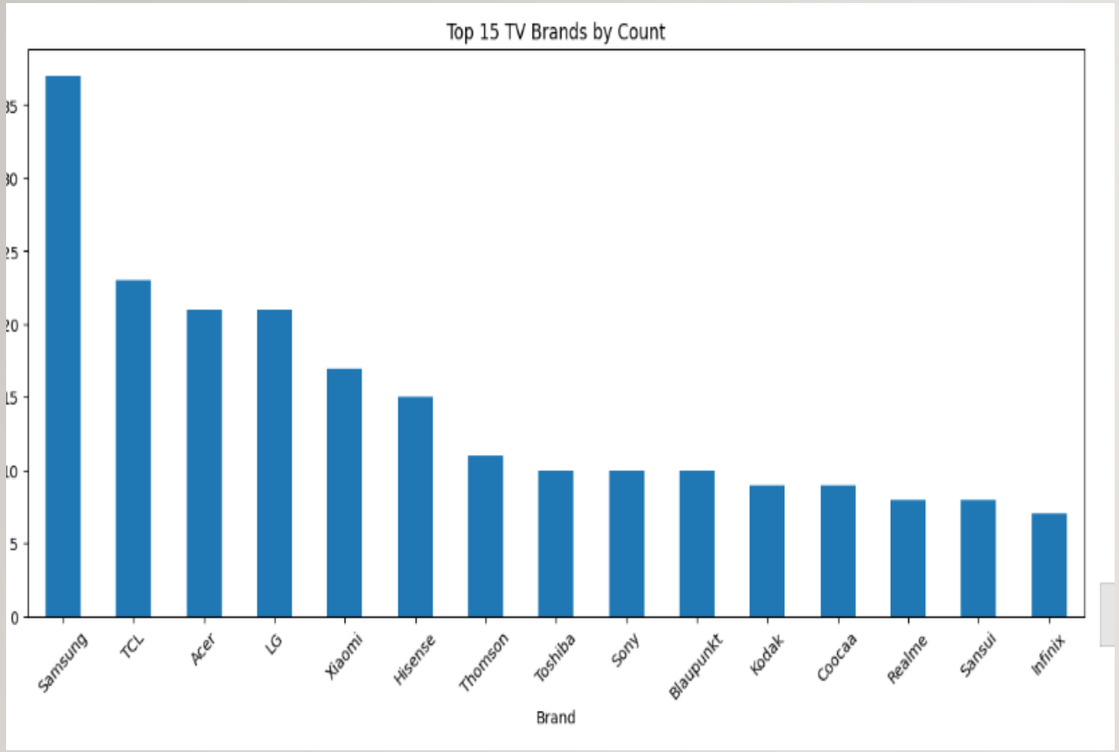


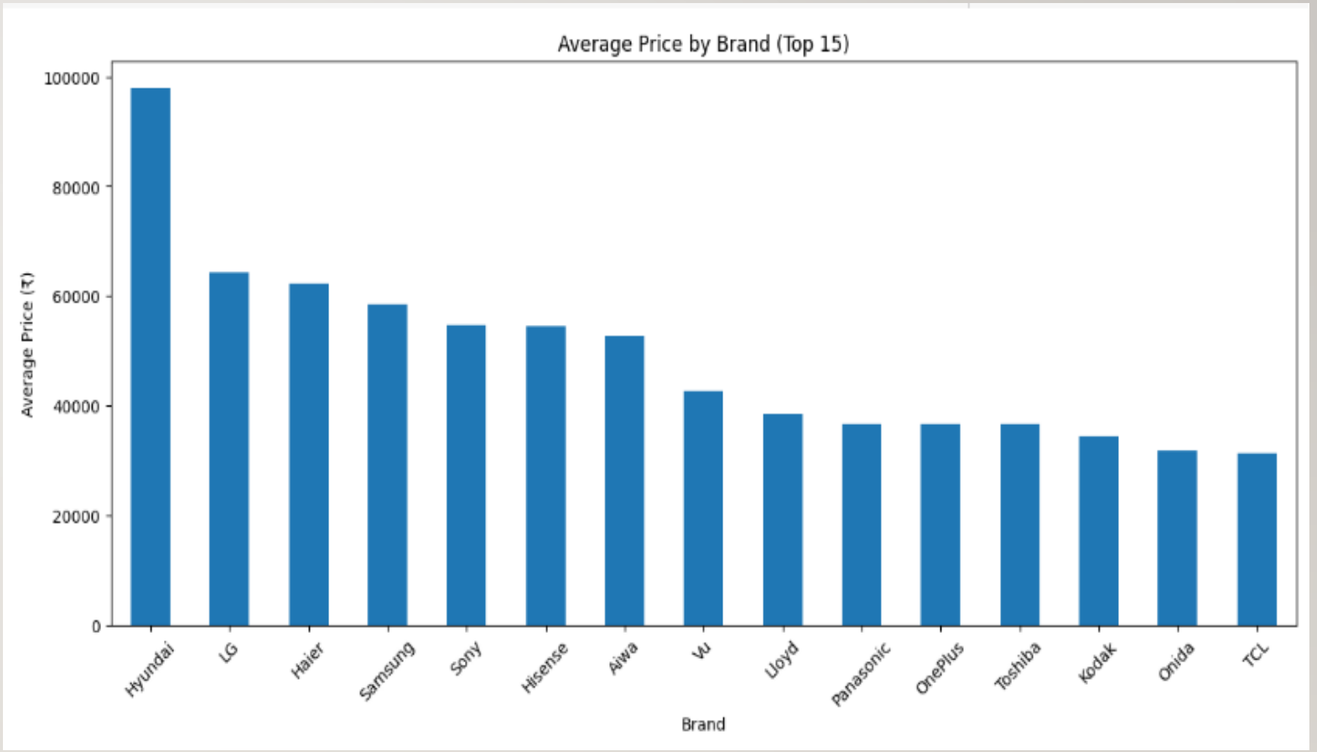
VISUALIZATION



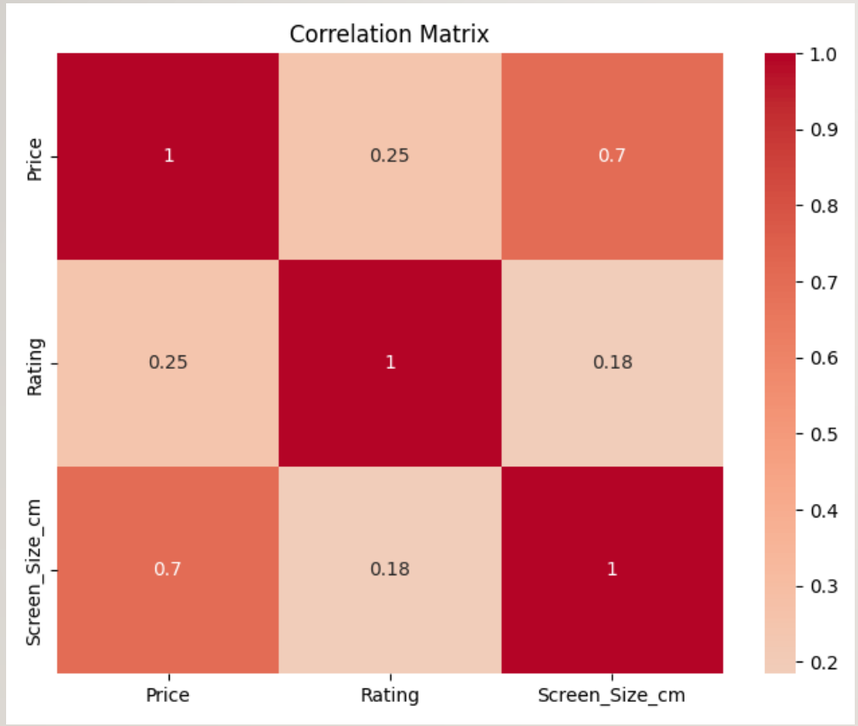


VISUALIZATION



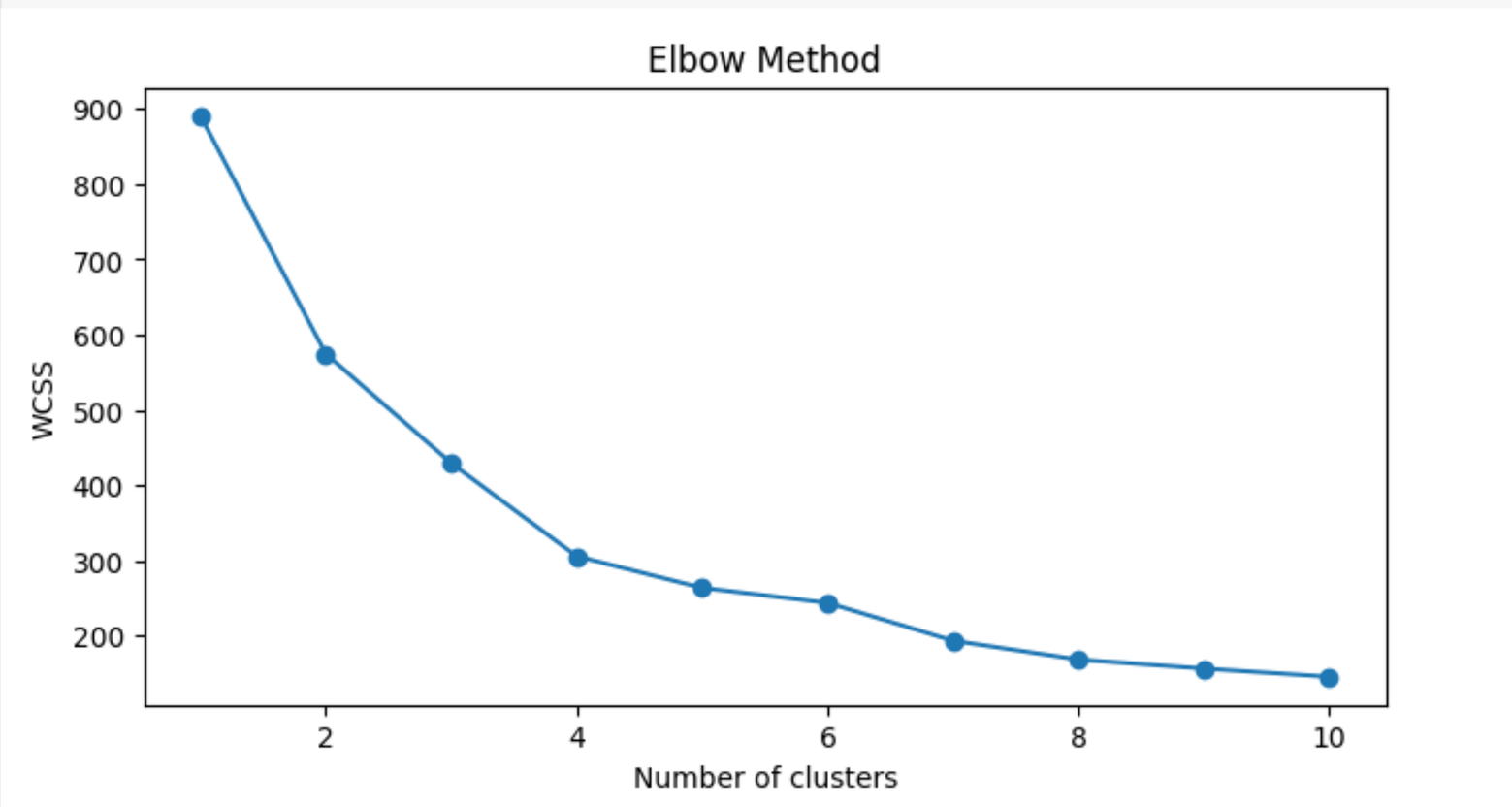


Correlation



UNSUPERVISED MACHINE LEARNING

K-MEANS CLUSTERING



Supervised Learning: Predicting TV Prices

Algorithms	Linear Regression, Random Forest
Training Data	Scraped feature-price pairs
Metrics	Mean Squared Error, R-squared
Goal	Accurate price prediction based on features

MODEL TRAINING

Trained multiple supervised learning models:

- **Logistic Regression**
- **K-Nearest Neighbors (KNN)**
- **Support Vector Machine (SVM)**
- **Decision Tree, Random Forest**
- **XGBoost Classifier**

Model	Accuracy
XGBOOST	96.7%
Random Forest	90%
KNN	82%
SVM	82%
Logistic Regression	78%

Hyperparameter Tuning: Optimizing Model Performance

1

What Are Hyperparameters?

Settings that guide the learning algorithm's behavior.

2

Tuning Techniques

- Grid Search
- Random Search

3

Example

Adjusting tree count in Random Forest model.

4

Impact

Enhanced accuracy and model reliability for price prediction.

Results & Insights: Trends in Flipkart TV Sales

Popular Sizes

43-inch and 55-inch TVs dominate sales.

Brand Performance

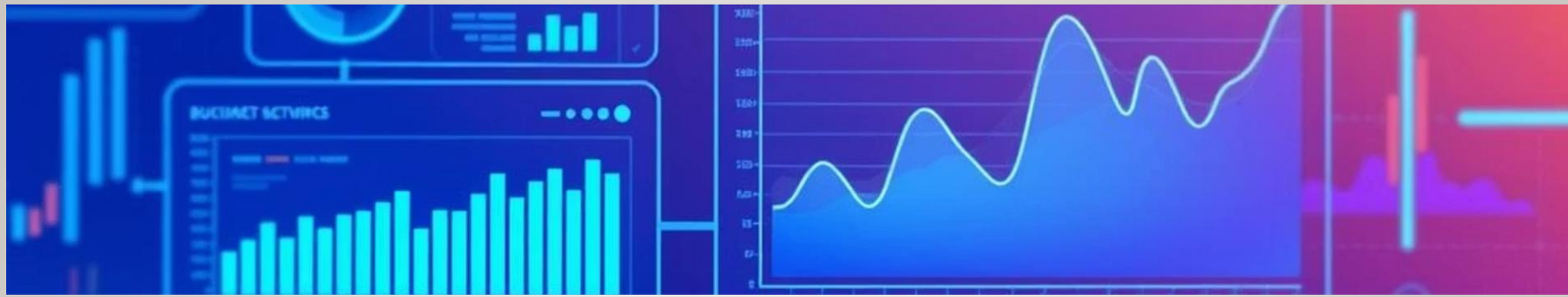
Leading market share: Samsung, Xiaomi.

Price Sensitivity

Price changes strongly affect sales volume.

Feature Demand

Smart TVs and 4K resolution are highly preferred.



Conclusion & Future Applications

Summary

Combining scraping and ML generates valuable TV market insights.

Applications

Supports dynamic pricing, inventory, and targeted marketing.

Future Work

Include reviews and competitor analysis for deeper insights.

Ethical Use

Maintain respect for privacy and fair data application.