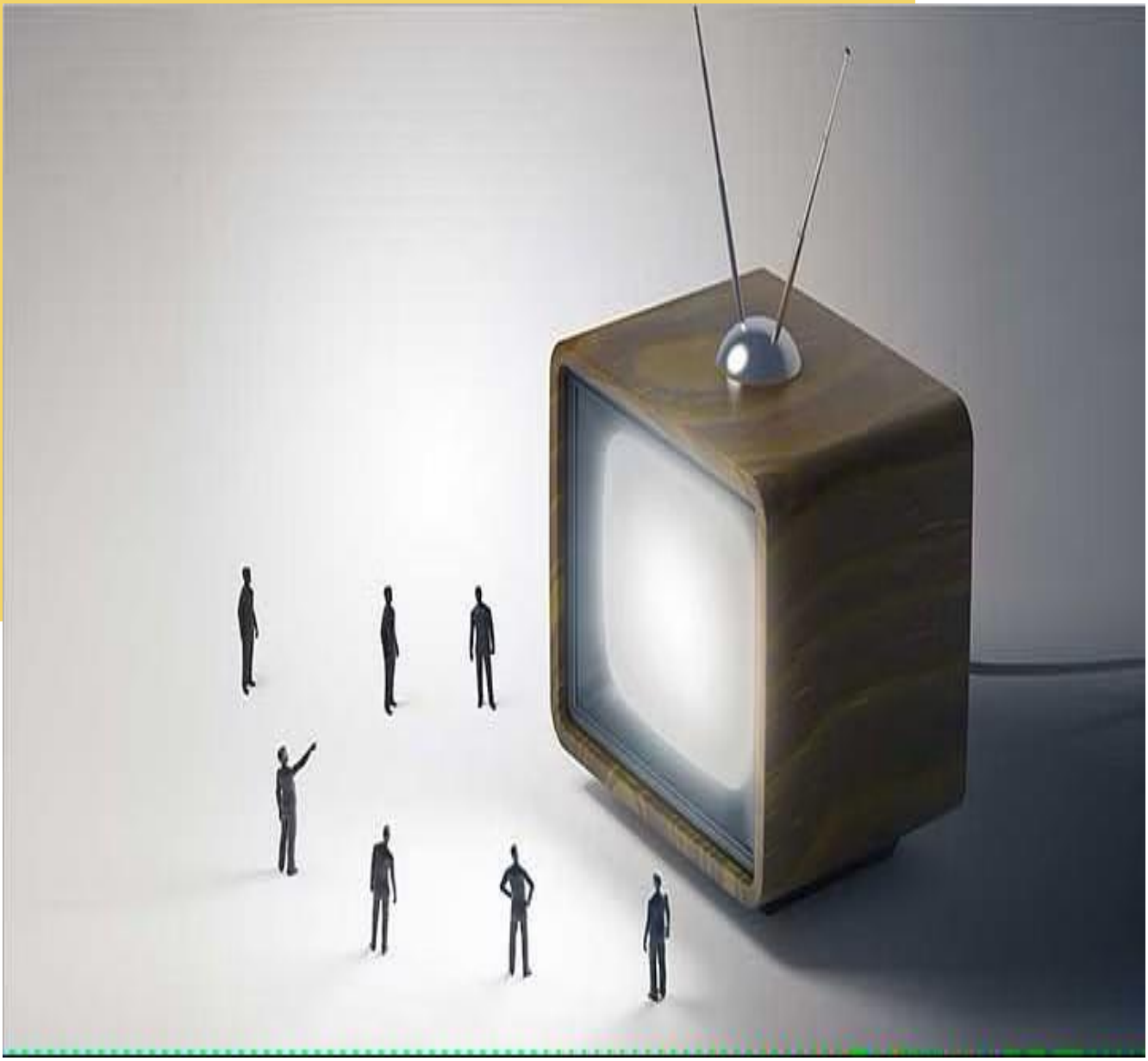


Television Price Prediction

Regression Analysis

Diwakar Babu K 22BM6JP16

Priyanshu 22BM6JP37

Noor Md Sohail 22BM6JP32

K Yashwanth 22BM6JP61

INTRODUCTION

With the rapid advancement of technology, televisions have become an essential part of our lives, providing entertainment and information. However, finding the right television at an affordable price can be a daunting task, considering the wide range of options available in the market. In such a scenario, web scraping coupled with the Beautiful Soup library presents an effective solution for analyzing television prices and making informed purchasing decisions.

Web scraping involves extracting data from websites, and by utilizing Beautiful Soup, a Python package for web scraping, we can efficiently gather information on television prices from popular online marketplaces like Flipkart. In this project, we aim to analyze television prices by scraping data from Flipkart using Beautiful Soup. By collecting data on various television models, their specifications, and their corresponding prices, we can gain insights into the current market trends and price ranges.

The process starts by scraping Flipkart's website and extracting relevant data such as brand, screen size, resolution, connectivity options, and other specifications. We then perform data cleaning and preprocessing to ensure the accuracy and quality of the collected information.

Next, we apply data analysis techniques to explore the relationships between television features and their prices. By visualizing the data, we can identify patterns and trends that influence the pricing of televisions. This analysis can provide valuable insights into the factors that contribute to price variations, helping consumers understand the pricing dynamics of different television models.

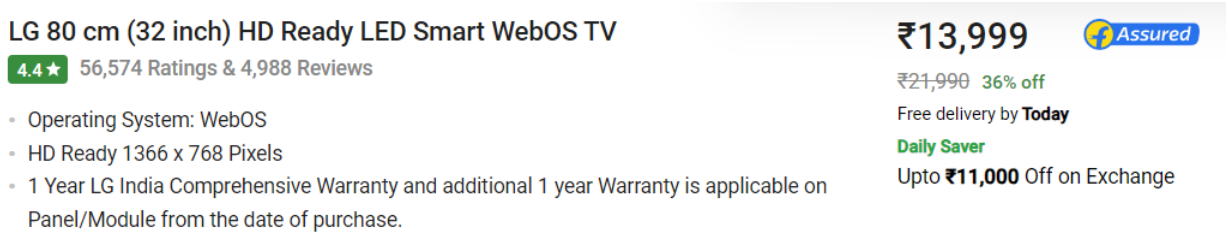
By leveraging web scraping and data analysis, consumers can make informed decisions about their television purchases. Whether they are seeking a specific brand, a particular screen size, or a combination of features that align with their preferences and budget, this analysis will enable them to compare prices and choose the best option available.

In conclusion, web scraping using Beautiful Soup provides a powerful method for analyzing television prices on platforms like Flipkart. By extracting and analyzing data on television models and their prices, consumers can gain valuable insights into market trends and make well-informed decisions when purchasing a television.

METHODOLOGY

- Extracting the data points using beautiful soup

  We have extracted close to 1600 television data consisting of various features ranging from various specifications to customer ratings. It consists of unstructured test data as shown below

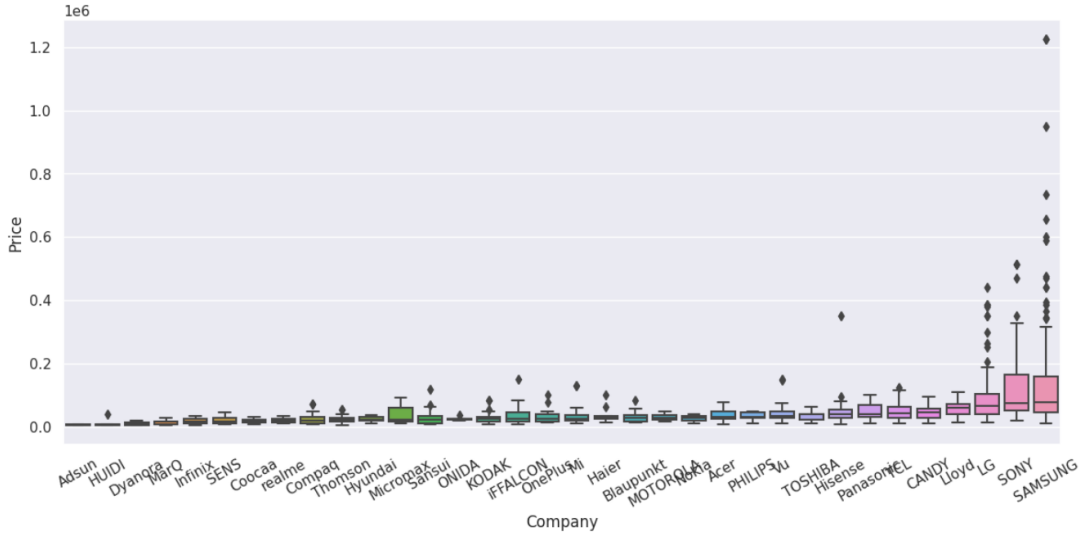


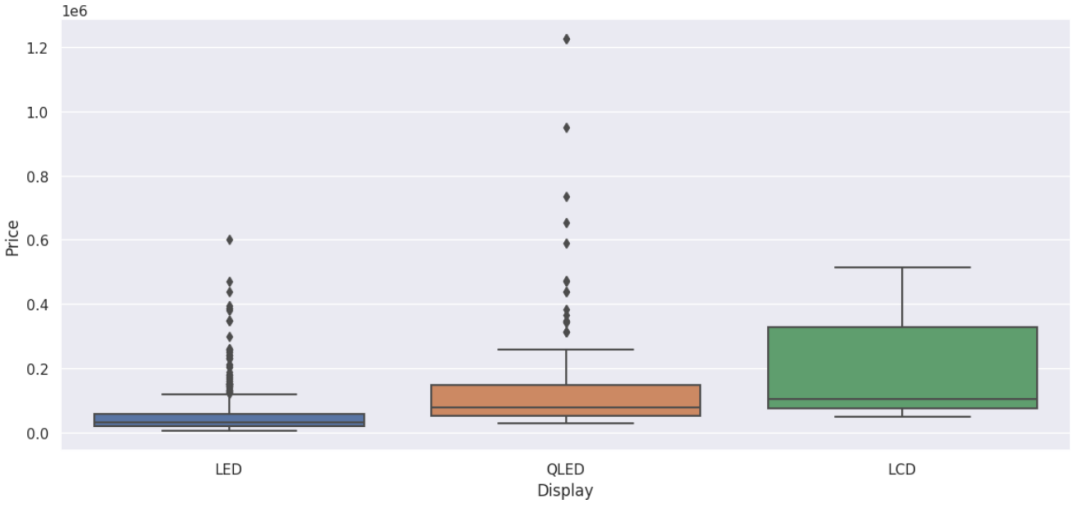
- Pre-processing the unstructured data to extract the relevant features for modeling

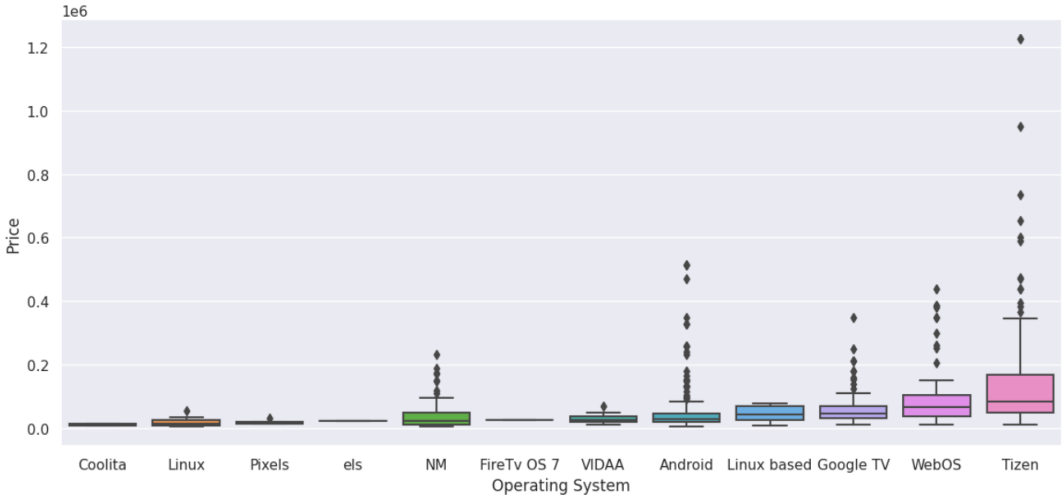
  We have extracted the following features Company, Display, Size, Operating System, Product Warranty, Resolution, Pixels, Price and Rating.

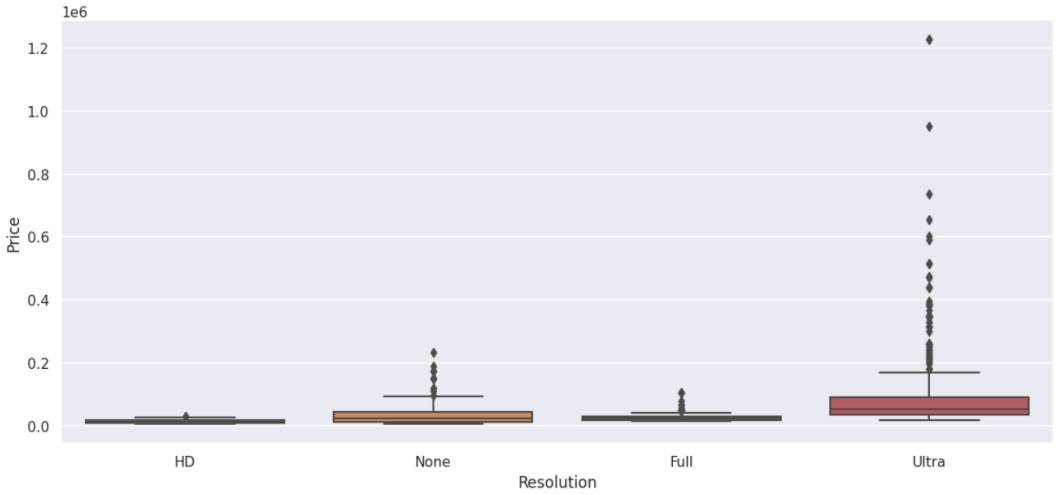
Company	Display	Size	Operating Syste	Product Warra	Resolutio	Pixels	4K	Price	Rating
SONY	LED	65	Android	1	Ultra	3840 x 2160	Yes	₹ 4,69,999	5
LG	LED	65	WebOS	1	Ultra	3840 x 2160	Yes	₹ 3,49,990	3.8
SAMSUNG	QLED	55	Tizen	1	Ultra	3840 x 2160	Yes	₹ 3,44,900	4.2
SAMSUNG	QLED	55	Tizen	1	Ultra	3840 x 2160	Yes	₹ 3,44,900	4.2
SAMSUNG	QLED	65	Tizen	1	Ultra	3840 x 2160	Yes	₹ 3,41,900	2.3
SONY	LED	65	Android	1	Ultra	3840 x 2160	Yes	₹ 2,59,999	4.3
SONY	LED	65	Android	1	Ultra	3840 x 2160	Yes	₹ 2,59,999	4.3
SAMSUNG	QLED	65	Tizen	1	Ultra	3840 x 2160	Yes	₹ 2,58,999	3
SAMSUNG	QLED	65	Tizen	1	Ultra	3840 x 2160	Yes	₹ 2,58,999	3

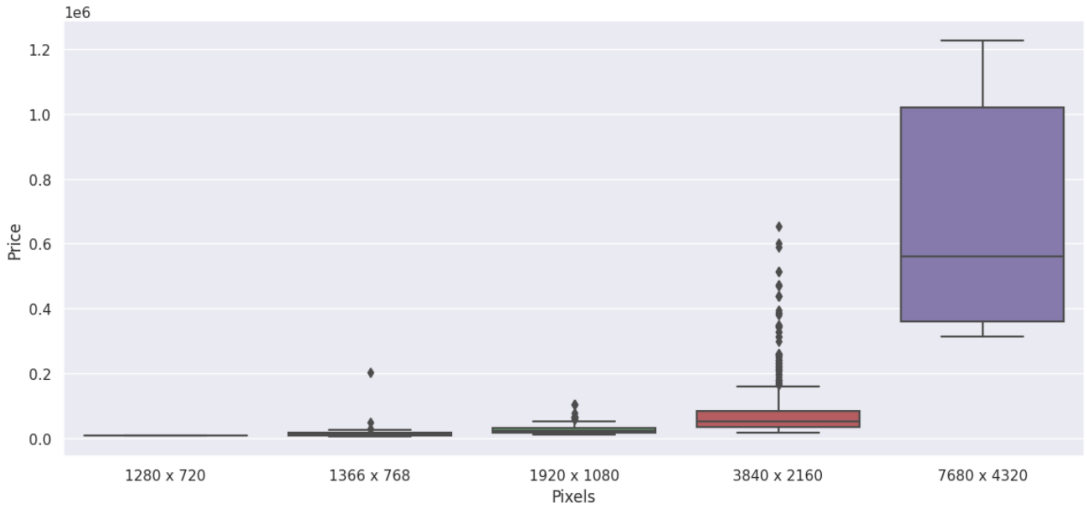
- Following are the relations seen between prices and various parameters


1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Price
Adsun
HUIDI
Dyanora
MarQ
Infinix
SENS
Coocaa
realme
Compaq
Thomson
Hyundai
Micromax
Sansui
ONIDA
KODAK
iFFALCON
OnePlus
Mi
Haier
Blaupunkt
MOTOROLA
Nokia
Acer
PHILIPS
Vu
TOSHIBA
Hisense
Panasonic
TCL
CANDY
Lloyd
LG
SONY
SAMSUNG
Company


1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Price
LED
QLED
LCD
Display


1e6
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Price
Coolita
Linux
Pixels
els
NM
FireTv OS 7
VIDAA
Android
Linux based
Google TV
WebOS
Tizen
Operating System





- Encoding for the categorical variables

  As most of the variables are categorical variables, care is taken to see which variables are ordinal and nominal in nature and are encoded accordingly.

- Modelling using Regression Techniques

  Below are the results from various regression techniques. For the best model, we have taken an ensemble of the top five performing regressors and generated a voting regressor. For measuring the model fitting, we have used MAPE metric.

	Model Name	Train MAPE	Test MAPE
0	linear	0.460033	0.481073
1	ridgecv	0.457734	0.484401
2	knn	0.249628	0.29533
3	rf	0.11257	0.216068
4	lgb	0.173864	0.223883
5	dart	0.200437	0.233425
6	huber	0.28974	0.324934
7	ard	0.463929	0.470384
8	hgb	0.164236	0.208209
9	gb	0.175782	0.226822
10	xgb	0.783414	0.754062
11	Voting Regressor	0.154474	0.203362