Predicting the PRICE of items listed on Vinted.

Introduction

Vinted is a Lithuanian online marketplace and community that allows users to sell, buy, and swap secondhand clothing, accessories and even some homeware. It has a built-in price indicator however it doesn't show what it's based on and doesn't seem correct. Therefore, I decided to create my own.
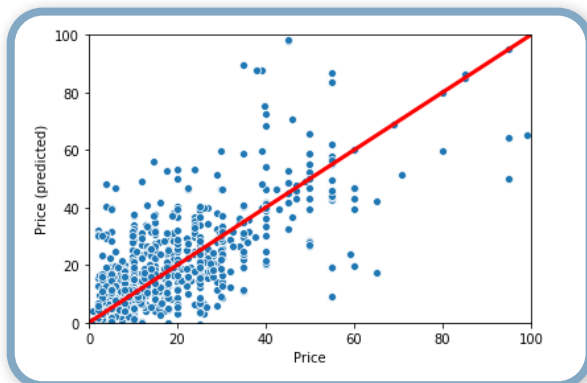
01. Data Cleaning

Data was scraped using the following API: https://github.com/Gertje823/Vinted-Scraper, and adjusted adding a few more variables, such as likes and views. On top of that, the data was cleaned and transformed to create the variable: Amount of Images and Ratio (from views and likes). This resulted in the following dataset:

ID	Category	Price	Size	State	Brand	Colors	Views	Interested	Amount_Of_Pictures	Ratio
841147011	Dress	29.0	M / 38 / 10	Very good	Vintage	Beige	114	20	3	5.700000
841212517	Suit	30.0	S / 36 / 8	Very good	Vintage	Black	231	26	4	8.884615
836855778	Sweater	15.0	M / 38 / 10	Very good	Vintage	White	83	20	1	4.150000
836637612	Jacket	165.0	M / 38 / 10	Very good	genuine leather	Black	534	27	8	19.777778
841235285	Jacket	60.0	M / 38 / 10	Very good	Dolce & Gabbana	Light blue	94	9	6	10.444444

02. Data Analysis

Of all the categories of the items that were scraped:

Highest mean:	Lowest mean:
€ 59.62	€ 8.83
Category: Coat	Category: Accessory
Highest mode:	Lowest mode:
€ 55.0	€ 3.0
Category: Jacket	Category: Jewellery

Another interesting insight was that every distribution between the price and the state of the item, was right skewed. This means that for every state there are more items that are priced lower in perspective. However, each distribution appears to be bimodal which means there is also a significant amount of items that are priced at a higher amount.
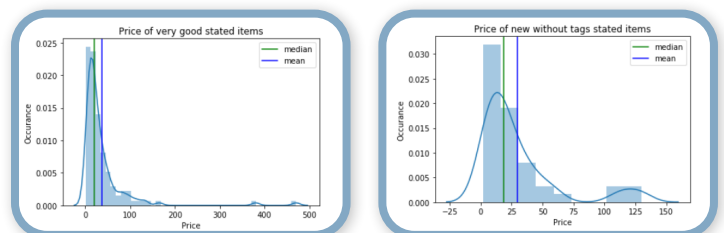
Also interesting was the case of fast fashion vs. slow fashion where brand such as Bershka and Zara where compared to Vintage items and how the price is distributed:

03. Data Model

The model of linear regression was used to predict the price. This plot shows that predictions are okay when the price is lower because the dots are further apart the higher the price. We can also conclude the model usually overshoots; so predicts the price too high.
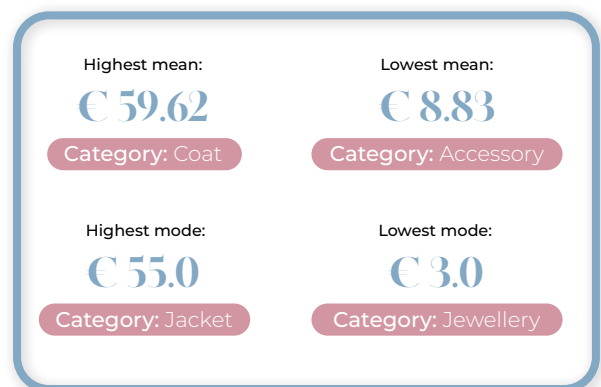

R2: 0.56   **RMSE:** 20.6