

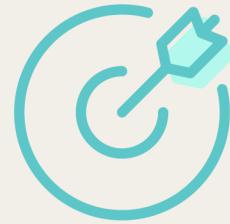
# Laptop

## Price Prediction with

## Lineer Regression

Zehra KEZER

# INTRODUCTION



## Goal

The problem we are trying to solve is to predict the prices of laptops using linear regression.



## Process

Scraping data from Trendyol and apply linear regression

# Methodology

## Data collection

- From Trendyol



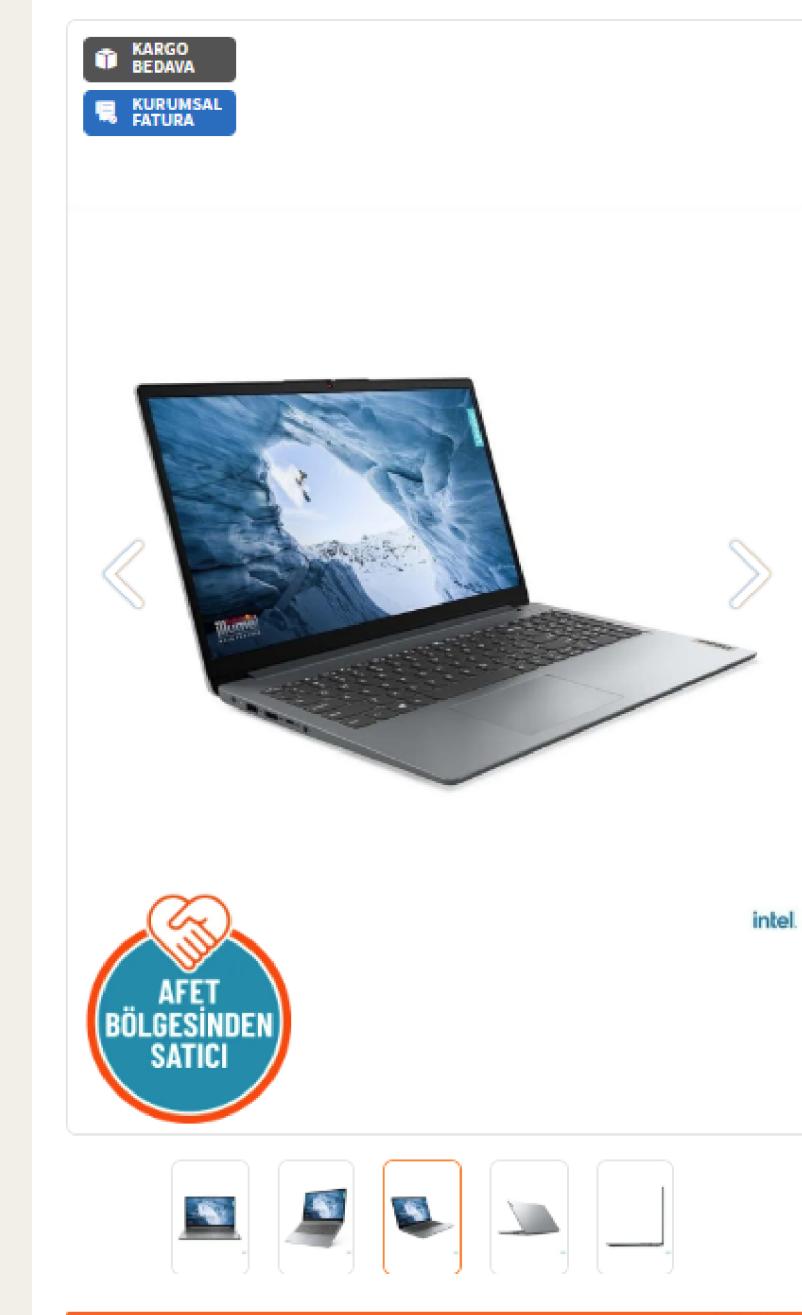
## Tools

- Visual studio code
- Python



## Librarys

- Beatifulsoup
- Pandas
- Matplotlib
- Seaborn
- Sklearn



Laptop kategorisinde [en çok satılan 2. ürün >](#)  
LENOVO Ideapad1/Celeron N4120/4GB Ram/128GB  
emmc /15.6"/Win 11 82V7005MTX  
Satıcı :DorukExpress

★★★★★ 334 Değerlendirme | 1069 Soru & Cevap

Peşin Fiyatına 3 Taksit!

5.738 TL

Koleksiyona Ekle

Son 2 Ürün

Ortunuz için uygun fiyatlara ek hizmet seçebilirsiniz

1 Yıl Ek Garanti   
**trendyolsigorta** 442 TL  
• Üretici garantisine ek - Onarım ve Değişim

Kapsamlı Onarım   
**trendyolsigorta** 442 TL  
• Ekran tamiri - Kaza

**Sepete Ekle**



12 saat 12 dakika içinde sipariş verirsen en geç yarın kargoda!

31887 favori

Bilgisayar Rehberi >

Öne Çıkan Bilgiler

İşlemci Tipi	:	Intel Celeron	Ram (Sistem Belleği)	:	4 GB
İşletim Sistemi	:	Windows	Ekran Kartı	:	Intel UHD Graph...
Hard Disk Kapasitesi	:	128 GB	Garanti Tipi	:	Resmi Distribütö...

# Explore the Data

- Handling missing values
- Some incorrectly scraping data
- Unnecessary columns
- Removing duplicates
- Encoding categorical variables

MARKA	PRICE	İŞLEMCI	SİSTEM	SSD	EKRAN	RAM	ÇÖZÜNÜRLÜK	EKRANHIZ	EKRANKARTI	EKRANHAFIZA	İŞLEMCİMODEL	ÇEKİRDEKSAYISI	KAPASİTE	HARDDİSK	BOYUT
ASUS	11779.0	AMD Ryzen 7	Free Dos	512	AMD Radeon Graphics	16	1920	Belirtilmemiş	Dahili	Paylaşımılı	3700U	4	512	HDD	15.6
LENOVO	5400.0	Intel Celeron	Bu ürününden en fazla 1 adet sipariş verilebilir...	128	Intel UHD Graphics 600	4	1366	Belirtilmemiş	Dahili	Paylaşımılı	N4020	2	128	HDD	14
ASUS	12996.0	AMD Ryzen 7	Free Dos	512	AMD Radeon Graphics	24	1920	Belirtilmemiş	Dahili	Paylaşımılı	3700U	4	512	1	15.6
Apple	18599.0	Bu ürününden en fazla 1 adet sipariş verilebilir...	Bu ürününden en fazla 1 adet sipariş verilebilir...	Bu	Bu	Bu	2560	Belirtilmemiş	Bu	Belirtilmemiş	Bu	8	Bu	256	Bu
ASUS	14844.0	Intel Core i5	Free Dos	512	Nvidia GeForce GTX 1650	8	1920	144	Harici	4	10300H	4	512	HDD	15.6

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1176 entries, 0 to 1175
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   TITLE            1176 non-null    object 
 1   MARKA           1176 non-null    object 
 2   PRICE            1024 non-null    float64
 3   ISLEMCI          1176 non-null    object 
 4   SİSTEM           1176 non-null    object 
 5   SSD              1176 non-null    object 
 6   EKRAN            1176 non-null    object 
 7   RAM              1176 non-null    object 
 8   ÇÖZÜNÜRLÜK      1176 non-null    object 
 9   EKRANHIZ         1089 non-null    object 
 10  EKRANKARTI       1176 non-null    object 
 11  EKRANHAFIZA      1161 non-null    object 
 12  İŞLEMCİMODEL     1176 non-null    object 
 13  ÇEKİRDEKSAYISI   1130 non-null    object 
 14  KAPASİTE          1176 non-null    object 
 15  HARDDİSK          1092 non-null    object 
 16  BOYUT             1176 non-null    object 
 17  EKRANBELLEKTİPİ  1176 non-null    object 
dtypes: float64(1), object(17)
memory usage: 165.5+ KB
```

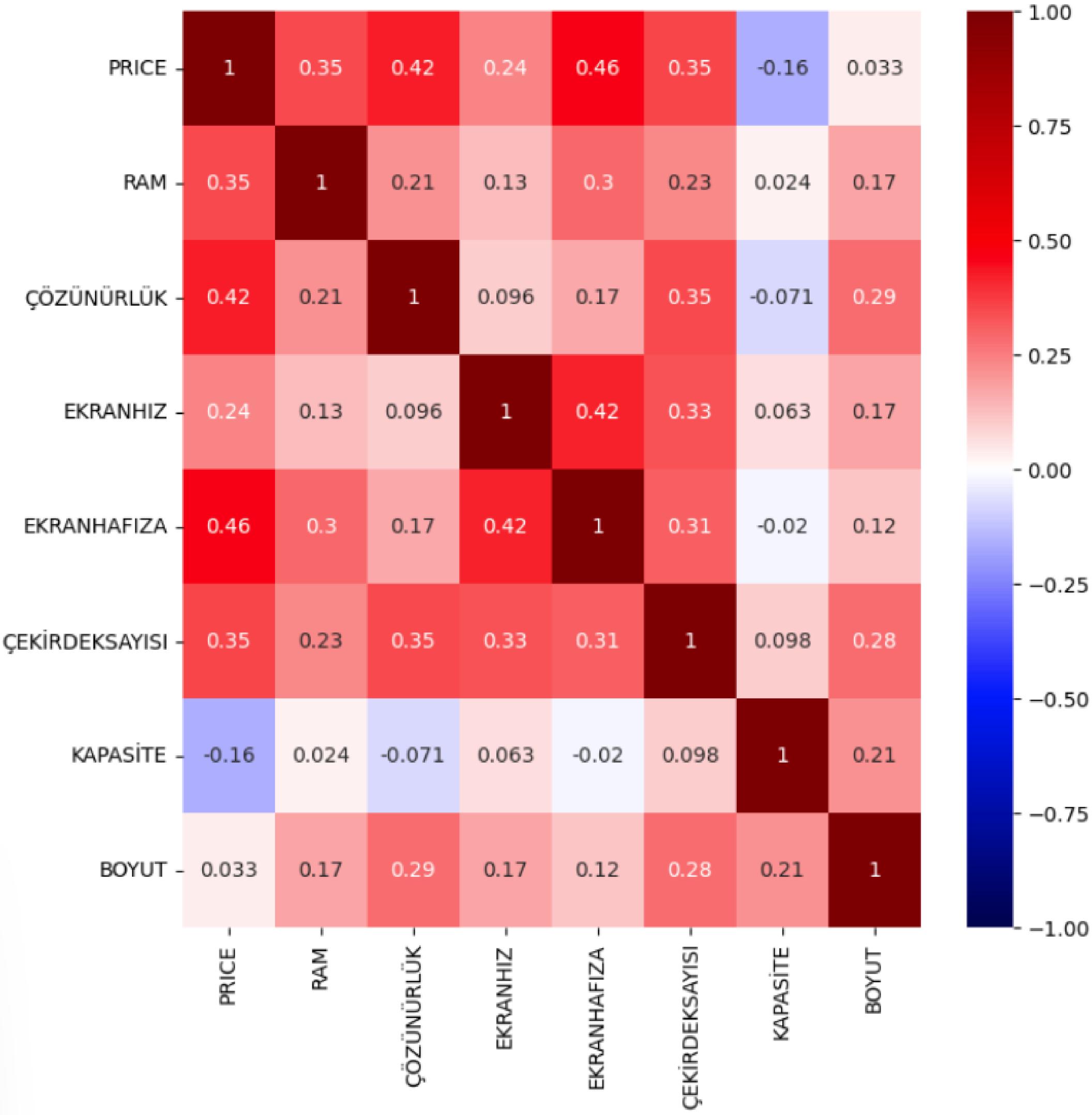
To better understand the relationship between the variables in our Laptop Price Prediction dataset, we performed a correlation analysis.

Positive correlation:

- ÇÖZÜNÜRLÜK
- RAM
- EKRANHIZ
- EKRANHAFIZA
- ÇEKİRDEKSAYISI

No correlation:

- KAPASİTE
- BOYUT



# Model Training

**Target Variable:** price

**Predictors:** ÇÖZÜNÜRLÜK, RAM, EKRANHIZ, EKRANHAFIZA, ÇEKİRDEKSAYISI, ISLEMCI, SISTEM, EKRANBELLEKTİPİ, EKRAN, İŞLEMCI MODEL

Splitting the data:

```
x Train: (567, 50)  
x Validation: (190, 50)  
x test: (190, 50)
```

Select model:

```
Linear Regression val R^2: 0.751  
Ridge Regression val R^2: 0.751  
Degree 2 polynomial regression val R^2: -77.645
```

	R-squared
Train	0.8116666611817
Validation	0.75059746629548

	Mean Squared Error
Train	0.065214851927784
Validation	0.08033659224481

MSE of 0.06521485192778449 indicates that linear regression model is making accurate predictions on average and has a good fit to the data

# RESULT

## Most important features:

- ISLEMCI\_Intel Celeron
- ISLEMCI\_Intel Core i7
- ISLEMCI\_Intel Core i9
- EKRANBELLEKTİPİ\_GDDR6
- MARKA\_Apple

**R-squared**      **Mean Squared Error**

**Test**      0.771888110727568

0.0867150591062441

	Prediction	Price
	682	24387.365593
	1158	12532.762960
	145	21525.402387
	88	10770.194791
	39	5647.795367
	...	...
	829	11660.157806
	869	26512.648177
	1111	12731.427484
	626	22374.732643
	1080	11496.196645

# Conclusion

Our model achieved a Train R-squared score of 0.81, which indicates that it has a good fit to the training data and can make accurate predictions on new, unseen data.

We found that the most important independent variables for predicting laptop prices are processor, brand and screen size.

we should include additional arguments for improvement in our model

**Thanks for listening**