# Cimri.com

SmartPhone
Price Prediction
Project



FURKAN ŞÜKRÜ KÜTÜK UĞUR SELİM ÖZEN



#### Problem

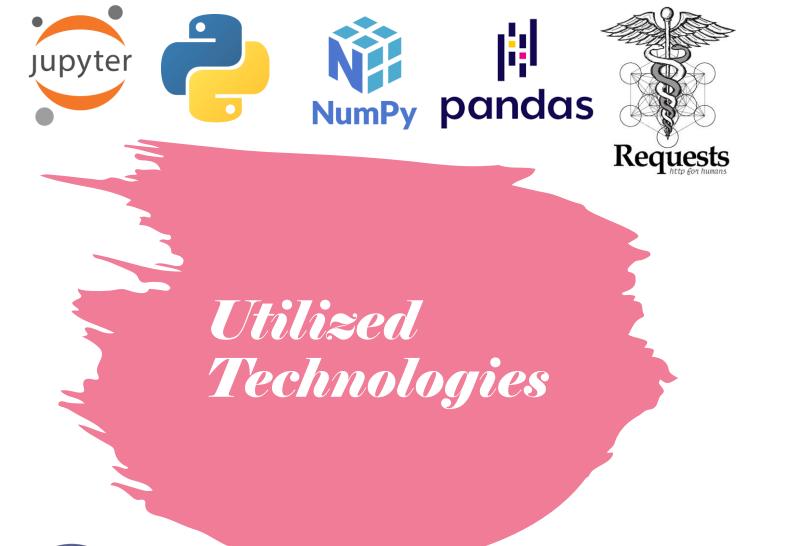
"
We are data analytics consulting team that serve on marketplace domains.

Our client requests building a prediction model for actual smartphone prices."

#### Methodology

- Data Collection with Web Scraping
- Data Cleaning and Tranformation
- Exploratory Data Analysis (EDA)
- Feature Selection and Modeling
- Model Prediction
- Interpreting Results





#### Data Overview

Observation: 1282

• Feature : **18** 

Target : PRICE

Data Size : 318 KB



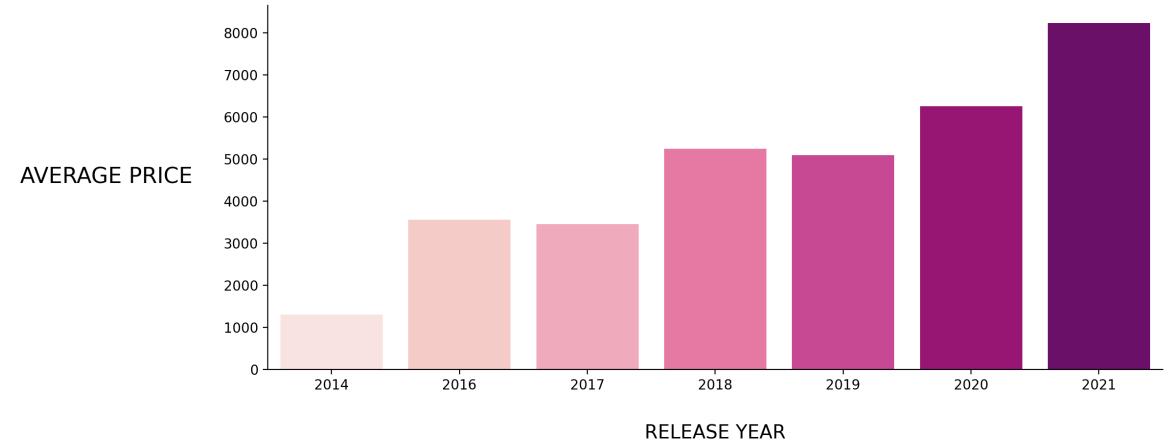
matpletlib Beautiful oup leaven



# Web Scraping Process

	• getAndParseURL()
2	• getPageLinks()
3	• getProductLinks()
4	• getDataFrame()
5	• getResult()

#### EDA-I

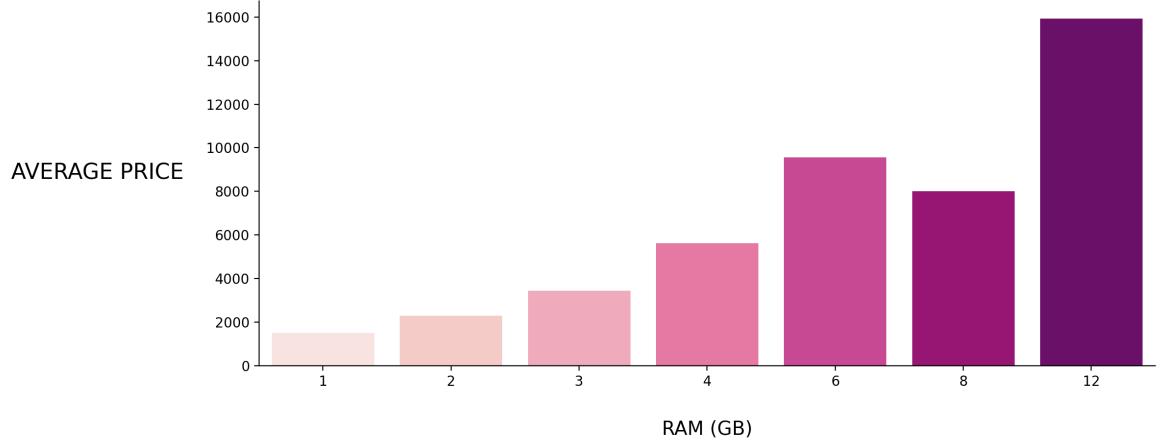




#### EDA - 2

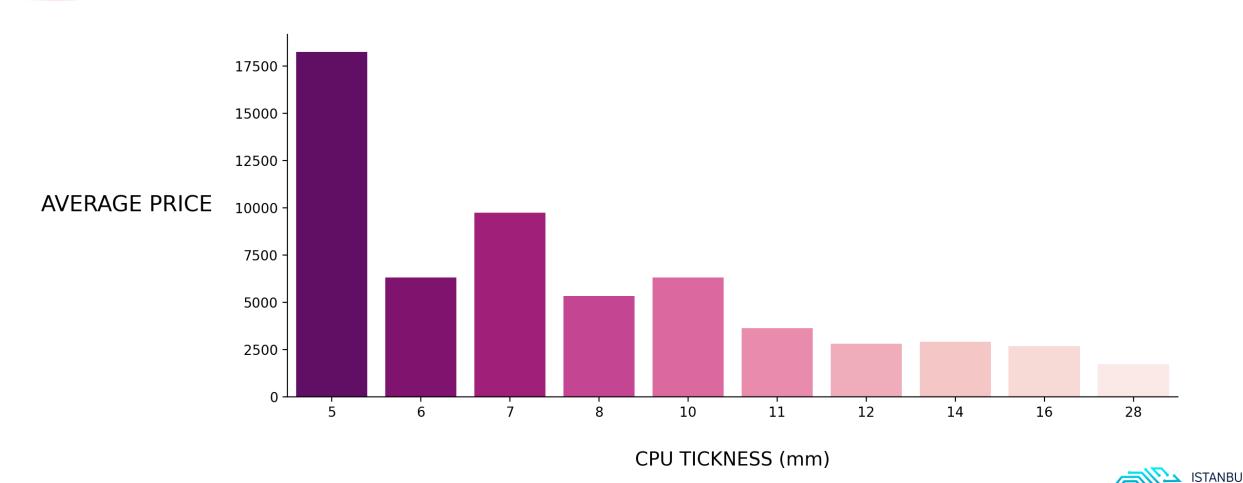


#### EDA - 3

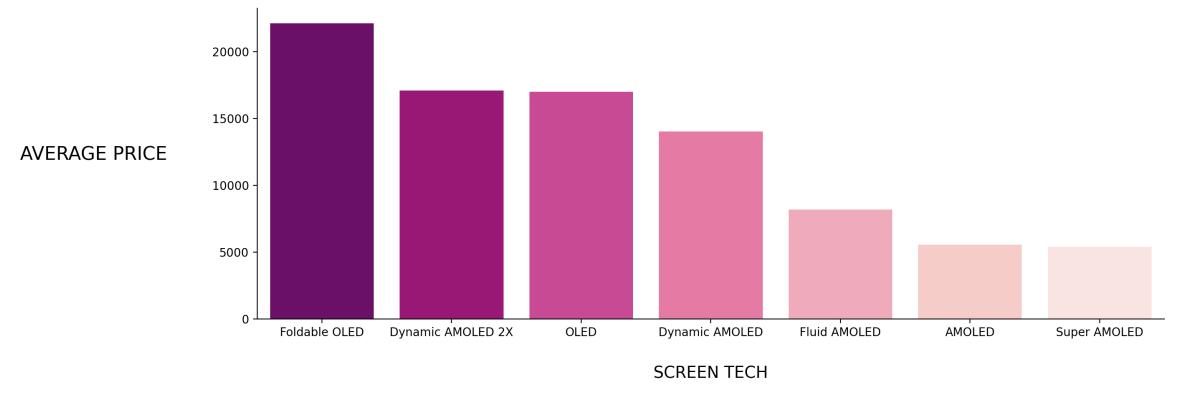




## EDA-4

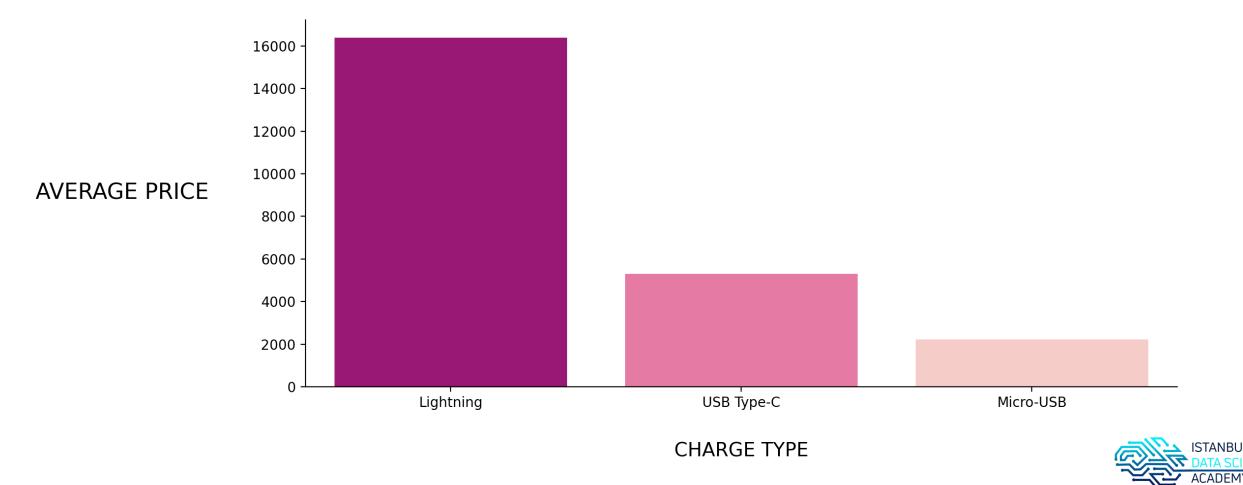


### EDA-5





### EDA - 6



#### Correlation Heatmap

PRICE	1.00	0.81	0.73	0.40	0.40	0.31	0.20	0.11	0.04	-0.05	-0.07	-0.08	-0.16	-0.30	-0.32	-0.57
CPU_Freq	0.81	1.00	0.58	0.54	0.49	0.23	0.33	0.25	0.12	-0.13	0.07	-0.11	-0.28	-0.19	-0.32	-0.80
Storage_Capacity	0.73		1.00	0.37	0.39	0.29	0.28	0.17	0.13	-0.01	0.07	-0.02	-0.10	-0.11	-0.19	-0.45
Screen_Body_Ratio	0.40	0.54	0.37	1.00	0.64	0.45		0.43	0.72	-0.11	0.42	0.17	-0.21	0.40	0.03	-0.60
RAM	0.40	0.49	0.39	0.64	1.00	0.44	0.36	0.61		-0.07		0.16	-0.17	0.28	0.01	-0.52
Weight	0.31	0.23	0.29	0.45	0.44	1.00	0.35	0.14	0.70	-0.07	0.30	0.51	-0.21	0.54	0.44	-0.30
Release_Year	0.20	0.33	0.28		0.36	0.35	1.00	0.13		-0.21	0.34	0.14	-0.36	0.46	0.19	-0.37
Selfie_Camera_Resolution	0.11	0.25	0.17	0.43	0.61	0.14	0.13	1.00	0.32	-0.06		0.02	-0.10	0.15	-0.08	-0.35
Screen_Size	0.04	0.12	0.13	0.72		0.70	0.51	0.32	1.00	-0.08	0.46	0.45	-0.17	0.73	0.36	-0.32
index	-0.05	-0.13	-0.01	-0.11	-0.07	-0.07	-0.21	-0.06	-0.08	1.00	-0.10	-0.03	0.55	-0.05	-0.01	0.15
Camera_Resolution	-0.07	0.07	0.07	0.42	0.58	0.30	0.34	0.51	0.46	-0.10	1.00	0.15	-0.18	0.49	0.21	-0.28
Width	-0.08	-0.11	-0.02	0.17	0.16	0.51	0.14	0.02	0.45	-0.03	0.15	1.00	-0.04	0.38	0.17	-0.03
level_0	-0.16	-0.28	-0.10	-0.21	-0.17	-0.21	-0.36	-0.10	-0.17		-0.18	-0.04	1.00	-0.10	0.00	0.27
Battery_Size	-0.30	-0.19	-0.11	0.40	0.28	0.54	0.46	0.15	0.73	-0.05	0.49	0.38	-0.10	1.00	0.53	-0.08
Tickness	-0.32	-0.32	-0.19	0.03	0.01	0.44	0.19	-0.08	0.36	-0.01	0.21	0.17	0.00	0.53	1.00	0.16
CPU_Tickness	-0.57	-0.80	-0.45	-0.60	-0.52	-0.30	-0.37	-0.35	-0.32	0.15	-0.28	-0.03	0.27	-0.08	0.16	1.00
	PRICE	CPU_Freq	Storage_Capacity	Screen_Body_Ratio	RAM	Weight	Release_Year	Camera_Resolution	Screen_Size	index	Camera_Resolution	Width	level_0	Battery_Size	Tickness	CPU_Tickness

- 0.8 - 0.6 - 0.4 - 0.2 - 0.0 - -0.2

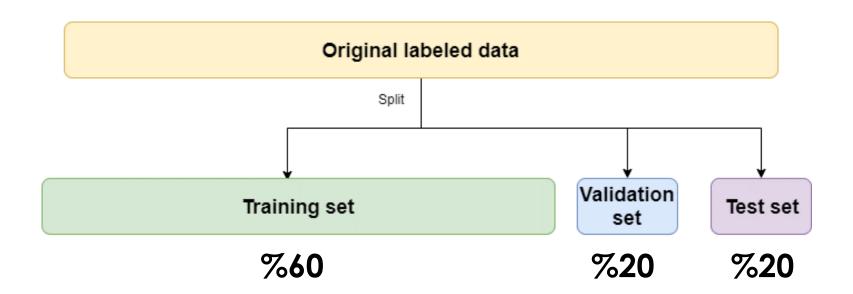
- -0.4

- -0.6



#### Data Modeling & Feature Selection

- Sorting dataset by product release year for testing built model with the most up-to-date data.
- One-Hot Encoded Features
- Screen\_Tech , CPU\_Model , GPU\_Model , Operating\_System , Charge\_Type





#### Model Prediction Test Results

Model Algorithm	R2-Score (%60 Train)	R2-Score (%80 Train)	MSE (%60 Train)	MSE (%80 Train)	10-Fold Cross Validation
Linear Regression	0.923	0.863	5461971	9759771	0.917
Ridge Regression	0.941	0.954	4206589	3261140	0.950
Lasso Regression	0.942	0.950	4757696	3565501	0.948
Elasticnet Regression	0.825	0.849	12468972	10791699	0.821





