WEB SCRAPING AND LINEAR REGRESSION

By Turki & Ahmed

GOAL

- our goal is to build a regression model for car prices based on the data collected from carvago. com. We had 5959 rows and 11 columns.

Web Scraping

Exploratory Data Analysis

Modeling

Communicate Findings and Results.

PLAN

TOOLS USED

- Python
- Pandas
- NumPy
- BeautifulSoup
- Matplot
- Seaborn
- Sklearn
- Patsy
- request

In [103]:
```
df
```

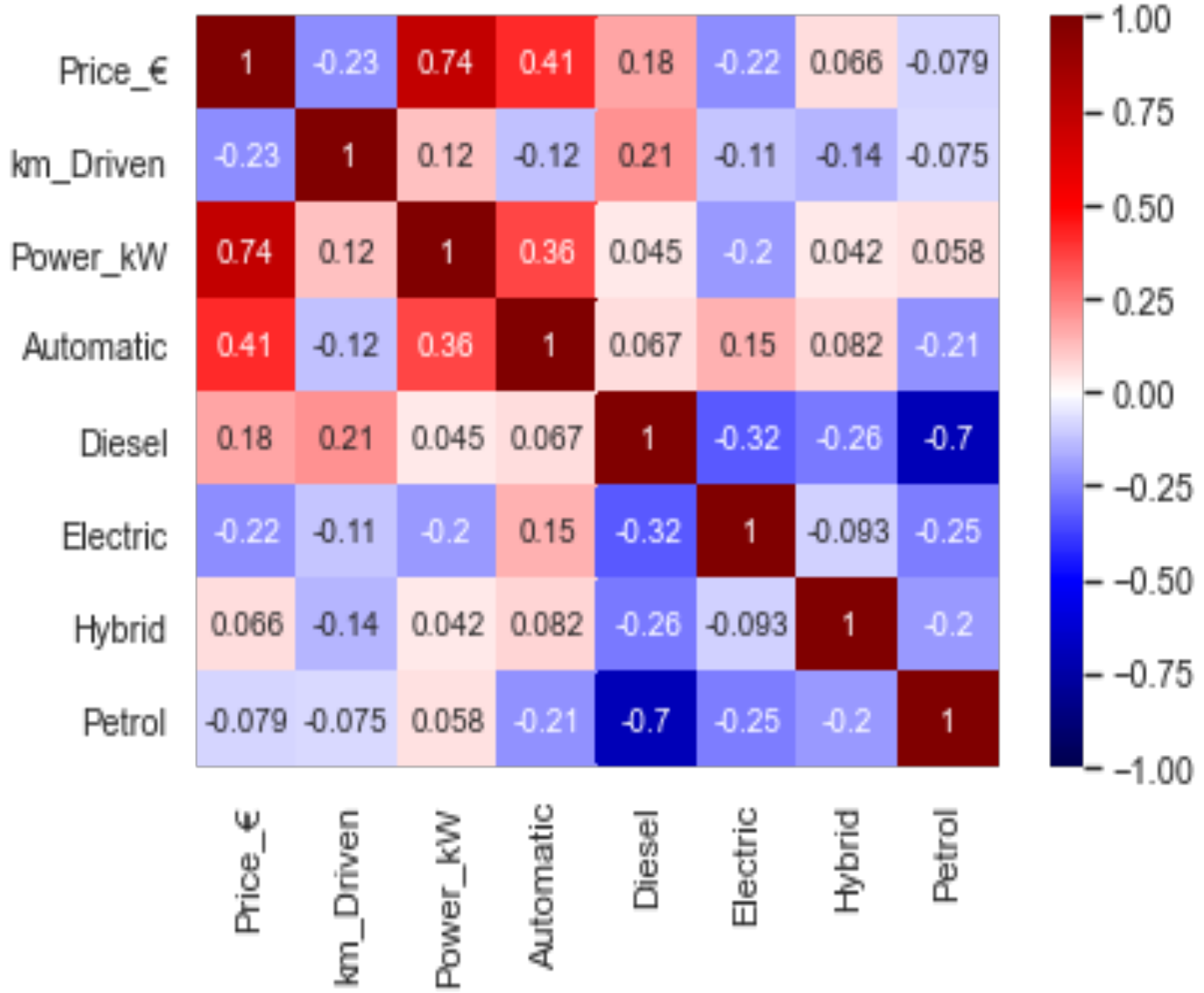
Out[103]:

	Car_Name	Price_€	km_Driven	First_Registration	Power_kW	Automatic	Manual	Diesel	Electric	Hybrid	Petrol
0	Volkswagen Caddy 2.0 TDI 75 kW	10699	134572	2/2018	75	0	1	1	0	0	0
1	BMW 120 141 kW	44449	5413	4/2021	141	1	0	0	0	0	1
2	BMW 640 250 kW	52899	27656	7/2019	250	1	0	0	0	0	1
3	Renault ZOE 70 kW without battery	9599	30200	12/2018	70	1	0	0	1	0	0
4	Mercedes-Benz E 63 AMG E T S 4Matic MCT 430 kW	47849	25000	7/2014	430	1	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...
5955	Renault Kadjar TCe 140 GPF 103 kW	21399	7990	9/2020	103	0	1	0	0	0	1
5956	Mazda CX-5 165 121 kW	28499	4000	3/2021	121	0	1	0	0	0	1
5957	Volkswagen T-Roc 1.5 TSI 110 kW	29599	500	11/2020	110	1	0	0	0	0	1
5958	Kia Optima 2.0 GDI Plug-In Hybrid Spirit 113 kW	31699	10500	9/2019	113	1	0	0	0	1	0
5959	Audi A4 Avant 45 TFSI S tronic sport 180 kW	32249	24679	3/2019	180	1	0	0	0	0	1

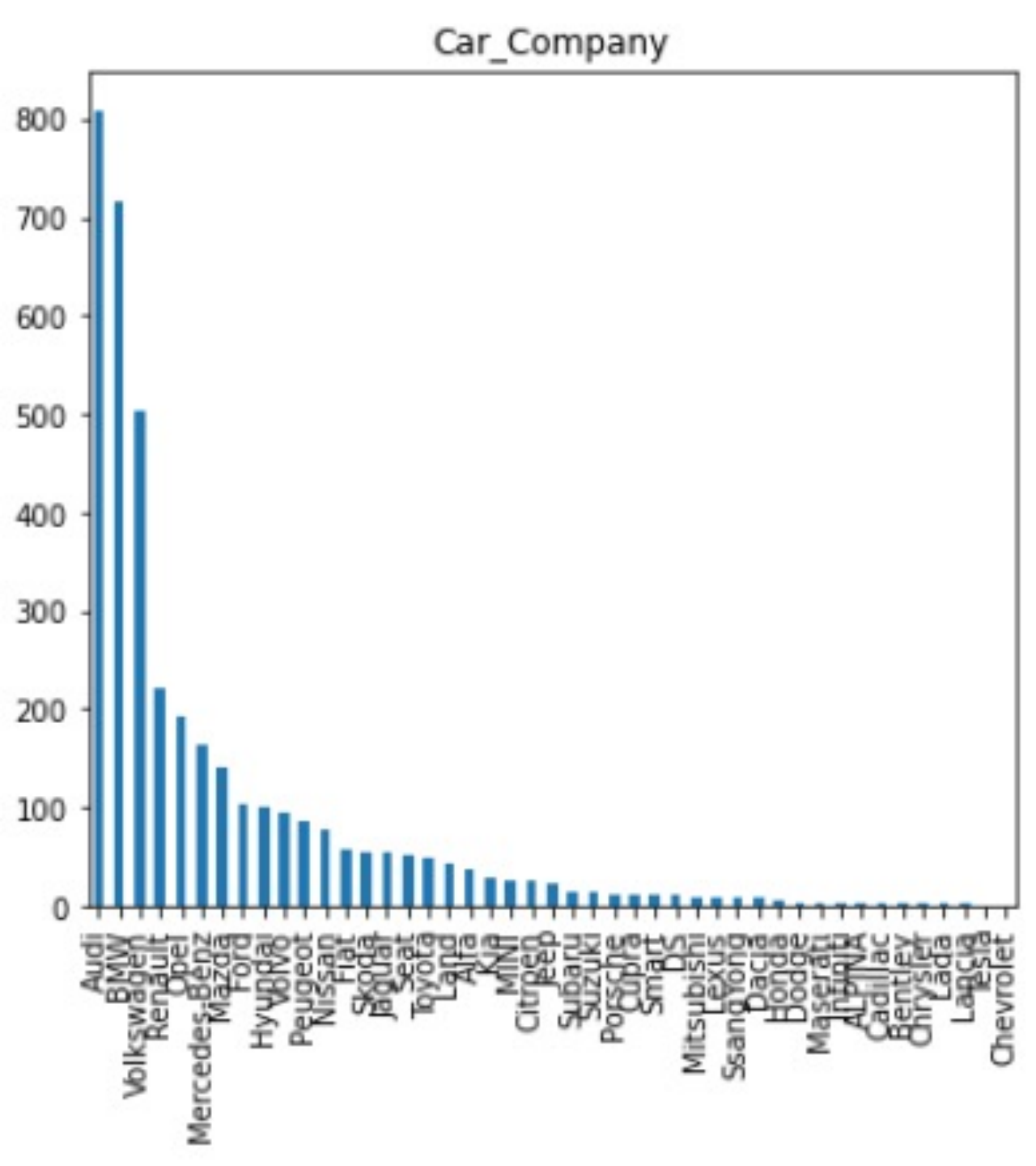
OUR DATASET

WITH TRANSMISSION AND FUEL CONVERTED TO DUMMY VARIABLES

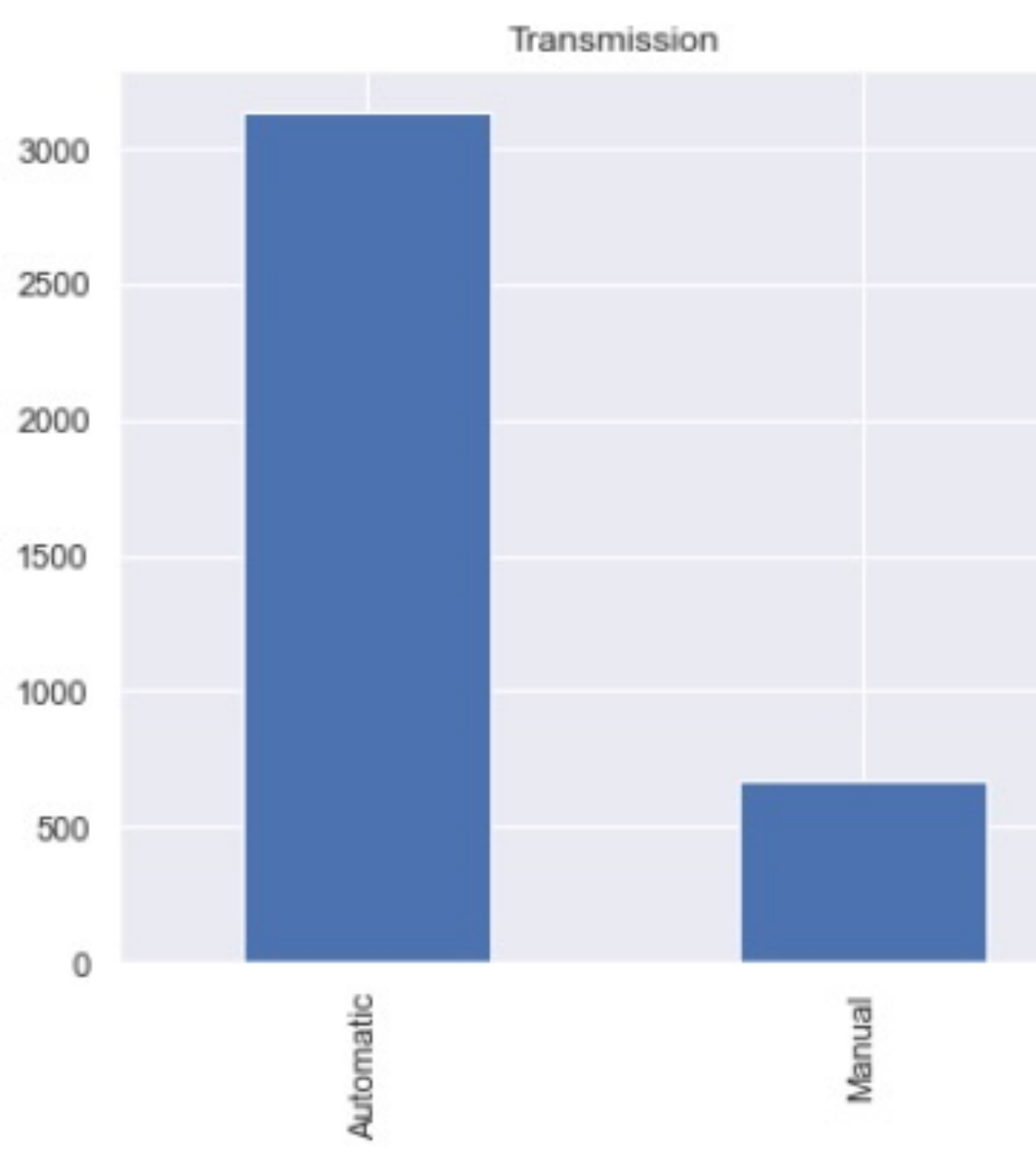
CORRELATION



CAR COMPANIES VS NUMBERS



TRANSMISSION NUMBERS IN OUR DATAFRAME



REGRESSION & FINDINGS

Linear Regression train R^2: 0.699

Ridge Regression train R^2: 0.699

Degree 2 polynomial regression train R^2: 0.699

Degree 2 polynomial regression test R^2: 0.705

THANK YOU FOR LISTENING