

Analysis of Laptop Prices dataset

Kacper Pieniażek

Wroclaw University of Technology

June 22, 2023

Dataset analysis

	laptop_ID	Company	Product	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_in_euros
0	1	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	1339.69
1	2	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	898.94
2	3	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel Core i5 T200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	575.00
3	4	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	2537.45
4	5	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	1803.60

#	Column	Non-Null	Count	Dtype
0	laptop_ID	1303	non-null	int64
1	Company	1303	non-null	object
2	Product	1303	non-null	object
3	TypeName	1303	non-null	object
4	Inches	1303	non-null	float64
5	ScreenResolution	1303	non-null	object
6	Cpu	1303	non-null	object
7	Ram	1303	non-null	object
8	Memory	1303	non-null	object
9	Gpu	1303	non-null	object
10	OpSys	1303	non-null	object
11	Weight	1303	non-null	object
12	Price_in_euros	1303	non-null	float64

Dataset analysis

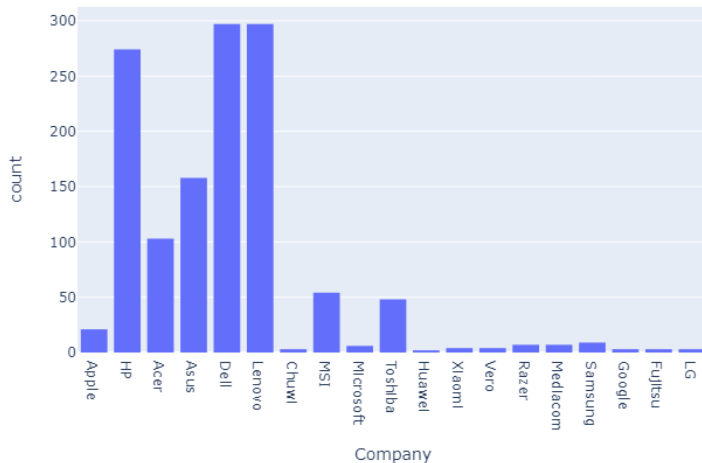
The goal is to:

- ▶ Analyse the dataset in search of some interesting correlations between features
- ▶ Do hypothesis testing
- ▶ Visualize analysis

Product analysis

Let's begin with something simple and look into amount of laptops that each brand produced.

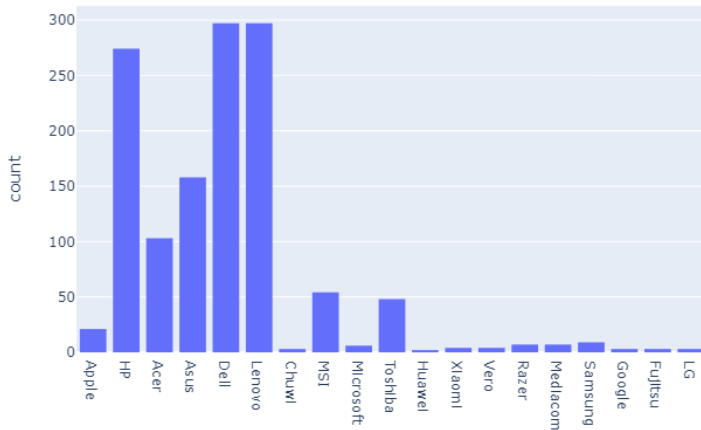
Amount of laptops per brand



Product analysis

We can easily see that Dell, Lenovo and HP are dominating other brands when it comes to amount of different laptops. Let's keep that in mind for later.

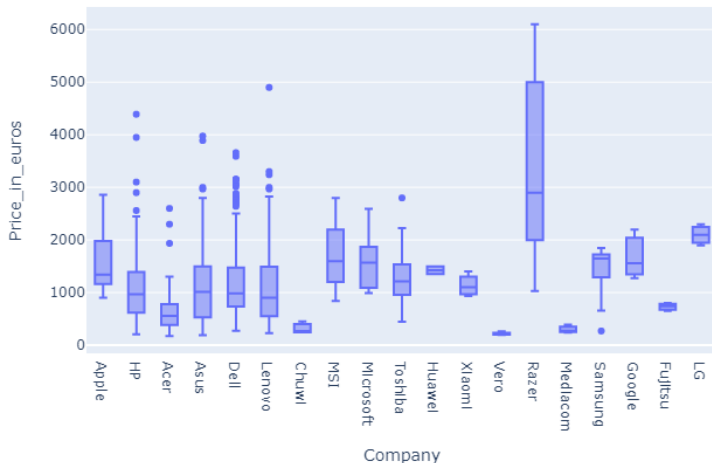
Amount of laptops per brand



Product analysis

Now let's look into what prices of each brand.

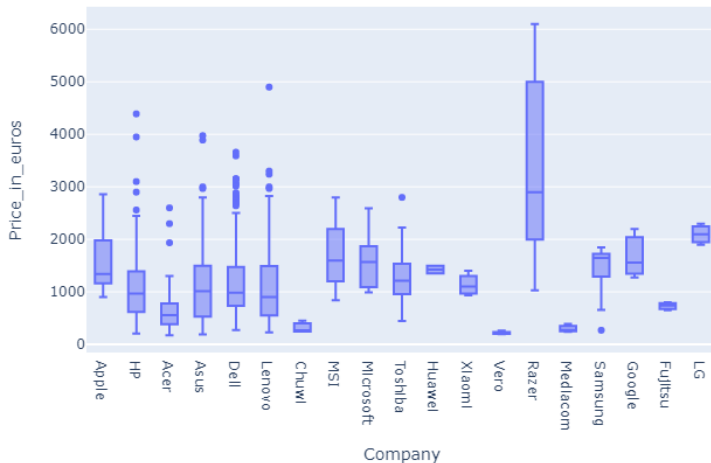
Laptop prices by brand



Product analysis

Razer heavily dominates upper fence and median values. They are the most expensive brand by far, even compared to Apple.

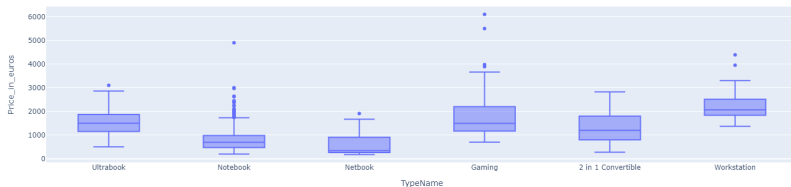
Laptop prices by brand



Product analysis

Razer has mostly products tagged as "Gaming", so it's expected to see that category as the most expensive as well.

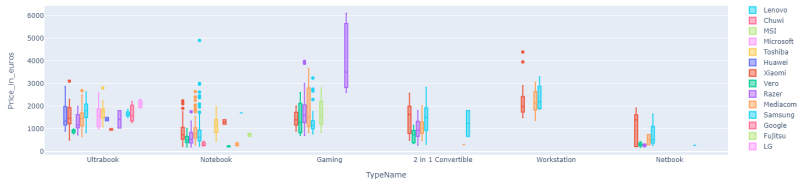
Laptop prices by type



Product analysis

As can be seen here, Razer's non-gaming products tend to be around average values for other brands. So most of their high-valued merchandise is in fact from Gaming category.

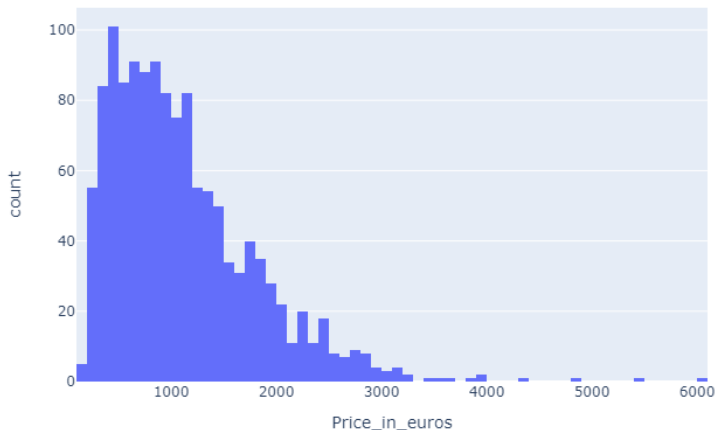
Laptop prices by brand and type



Prices analysis

Now let's look into prices distribution and test our hypothesis that prices are normally distributed.

Laptop prices distribution



Prices analysis

```
4 result = shapiro(df['Price_in_euros'])
5 alpha = 0.05
6
7 if result.pvalue < alpha:
8     print("Prices are not normally distributed")
9 else:
10    print("Prices are normally distributed")
11
12
```

✓ 0.0s

Prices are not normally distributed

Prices analysis

Another hypothesis we could check is to compare mean prices of two popular brands using t-test. Our hypothesis is that Lenovo and Dell have the same mean price

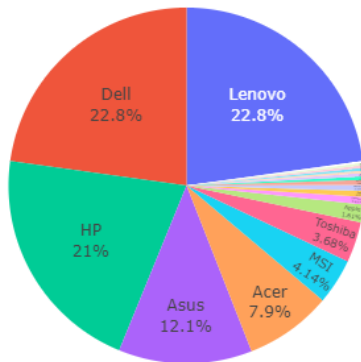
```
3 # H0: mean price of Lenovo laptops = mean price of Dell laptops
4
5 df_hp = df[df['Company'] == 'Lenovo']
6 df_dell = df[df['Company'] == 'Dell']
7
8 result = ttest_ind(df_hp['Price_in_euros'], df_dell['Price_in_euros'])
9 print('p-value:', result.pvalue)
10
11 alpha = 0.05
12
13 if result.pvalue < alpha:
14     print("Mean price of Lenovo laptops is not equal to mean price of Dell laptops")
15 else:
16     print("Mean price of Lenovo laptops is equal to mean price of Dell laptops")
17
✓ 0.0s
```

p-value: 0.07702514743218304

Mean price of Lenovo laptops is equal to mean price of Dell laptops

Company analysis

Coming back to previous charts, let's change histogram of amounts of laptops produced by each company into a pie chart.

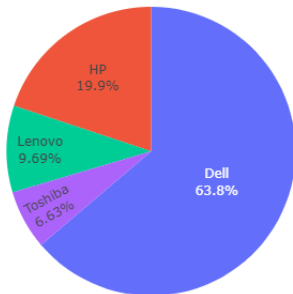


Company analysis

Now let's see what happens if we consider that each product can be released in multiple variants (for example, with different memory, graphics card)

Company analysis

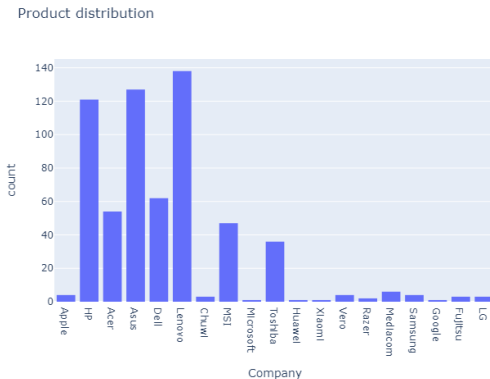
Here we consider top 20 laptops grouped by their product name and brand.



We can see Dell heavily dominates, which means they released the most variants of the same laptops.

Company analysis

Now if we remember that Dell was among companies with the most products released, what happens when we only consider "unique" models, without their different variations?

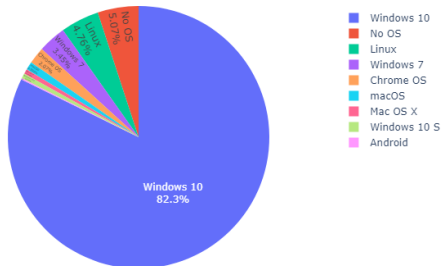


Dell's chart is now much lower, which means they heavily re-iterate on the same laptop models.

OS analysis

Finally, let's see what operating systems are shipped with each laptop

Operating System Distribution



Suprisingly enough, there is an Android laptop in the dataset.

OS analysis

```
1 df_android = df[df['OpSys'] == 'Android']  
2  
3 df_android.head()
```

✓ 0.0s

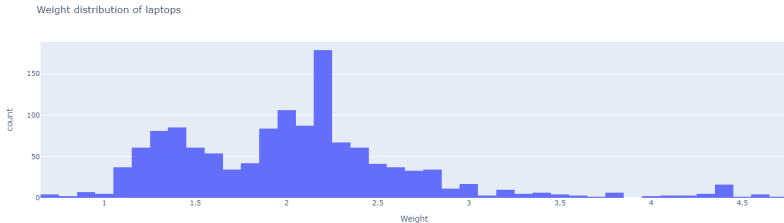
Python

	laptop_ID	Company	Product	Type/name	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_in_euros
50	51	Lenovo	Yoga Book	2 in 1 Convertible	10.1	IPS Panel Touchscreen 1920x1200	Intel Atom x5-Z8550 1.44GHz	4GB	64GB Flash Storage	Intel HD Graphics 400	Android	0.69kg	319.0
1114	1129	Lenovo	Yoga Book	2 in 1 Convertible	10.1	IPS Panel Touchscreen 1920x1200	Intel Atom x5-Z8550 1.44GHz	4GB	64GB Flash Storage	Intel HD Graphics 400	Android	0.69kg	549.0

It's Lenovo's Yoga model, and it's very light. Probably one of the lightest, let's look into that.

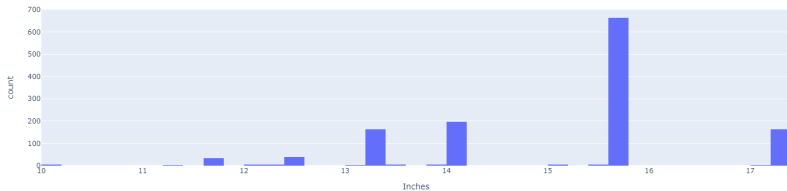
Weight analysis

We can see a spike around 2.1 kg, which our first guess is that it relates to screen size. So we expect fairly big screen sizes to be the most frequent.

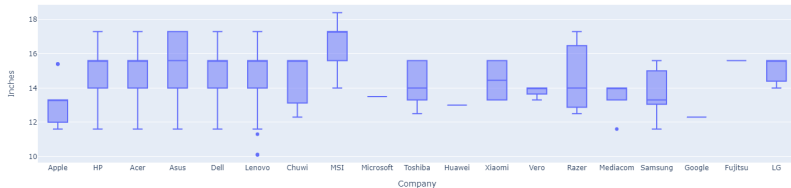


Weight analysis

Inches distribution of laptops



Weight distribution of Android laptops by brand



Hypothesis testing

Let's assume that prices are correlated to screen resolution and test that hypothesis.

```
3 result = pearsonr(df['Price_in_euros'], df['Inches'])
4 print('p-value:', result.pvalue)
5
6 alpha = 0.05
7
8 if result.pvalue < alpha:
9 |     print("There is a correlation between price and screen resolution")
10 else:
11 |     print("There is no correlation between price and screen resolution")
12
```

✓ 0.0s

p-value: 0.013808550905012235

There is a correlation between price and screen resolution

Thanks for your attention.

A repository with code for this presentation can be found here:

`github.com/dntAtMe/laptop-prices-dataset-analysis`

Dataset

<https://www.kaggle.com/datasets/gyanprakashkushwaha/laptop-price-prediction-cleaned-dataset>