

Sales and customer analysis

www.yojo.com

Group: Rigone, Bembhy, Iannotta

Data Preprocessing

Colors table

color	hue_deg	saturation_perc	light_perc
1-beige	60	56	91
2-black	0	0	0
3-blue	240	100	50
4-brown	0	59	41
5-burgundy	345	100	25
6-gray	0	0	50
7-green	120	100	25
8-navyblue	240	100	25
9-ofmanycolors			
10-olive	60	100	25
11-pink	350	100	88
12-red	0	100	50
13-violet	300	76	72
14-white	0	0	100

We decided not to add it because ...

`df.isnull().sum()`

year	0
month	0
day	0
country	0
category	0
color	0
position	0
photo	0
price	0
page	0
time	0
model	0
order	0
dtype:	int64

We search for null values and there were none.

`df.nunique()`

year	1
month	5
day	31
country	46
category	4
color	14
position	6
photo	2
price	20
page	5
time	20000
model	213
order	98
dtype:	int64

There was just one unique value for year = dropping

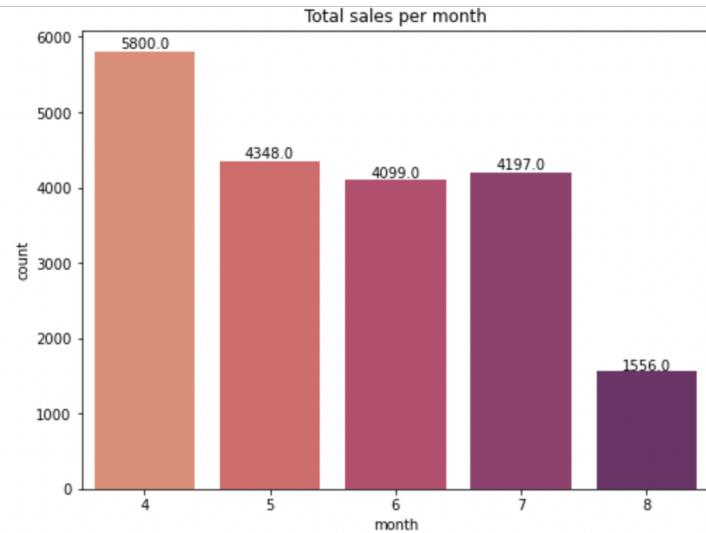
Split categorical and numerical features

We created a column for weekday

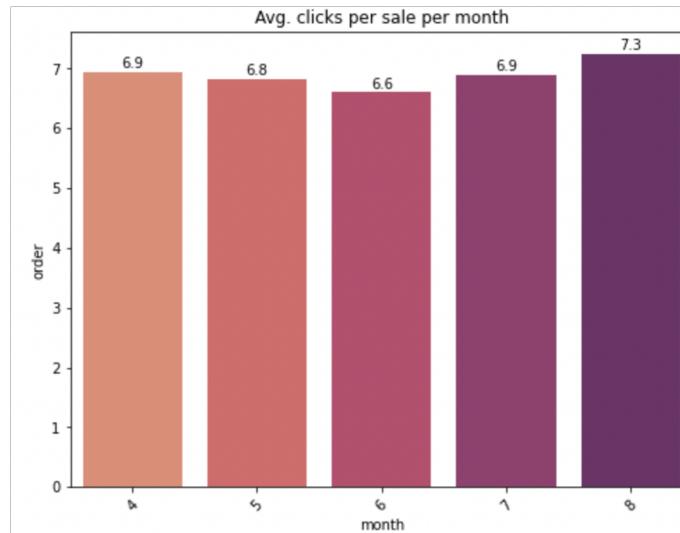
```
cat = ['month', 'day', 'country', 'category', 'color', 'position', 'photo', 'page', 'model', 'weekday']  
num = ['price', 'time', 'order']
```

Even though many of the categorical variables were integers, we considered them categorical for their nature.

Features Analysis: Month



Count of orders



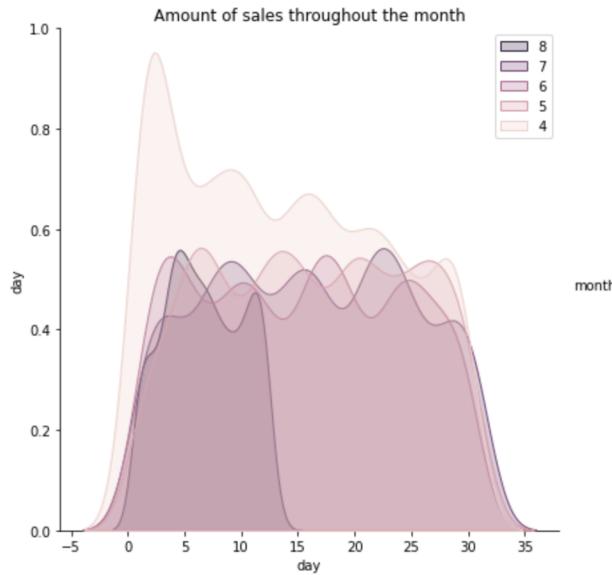
Average clicks

Insights

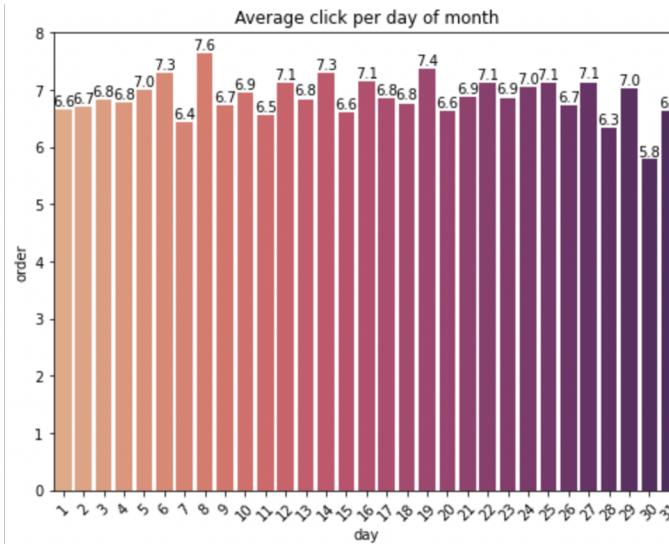
- April highest sales
- Next 5 months similar
- No very clear correlation between sales and clicks

Note: for the month of August we have the orders for just 13 days.

Features Analysis: *Day of the month*



Sales per month



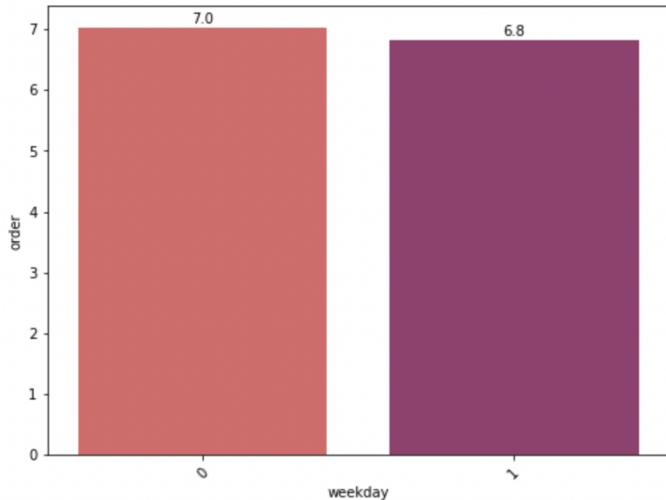
Average clicks

Note:

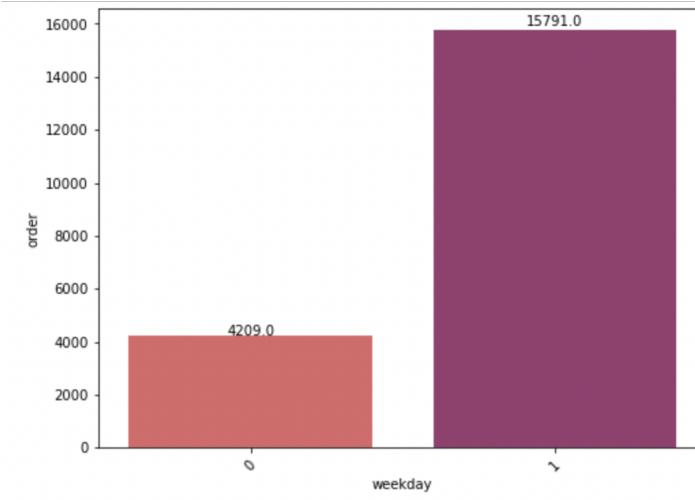
Insights

- Fluctuations across the different days of the months
- No specific pattern for day of average click on the days of week

Features Analysis: *Day of the month*



Average click weekday

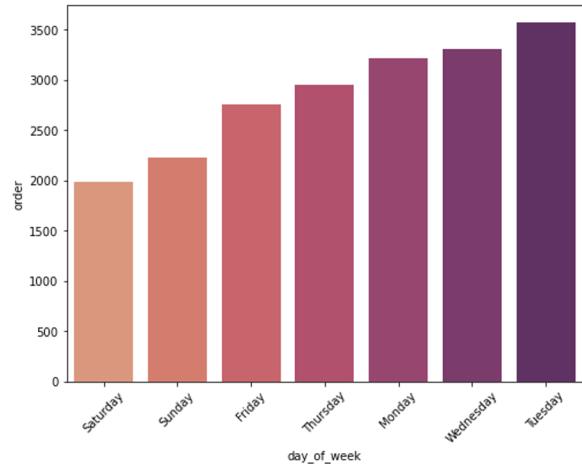


Sales weekday

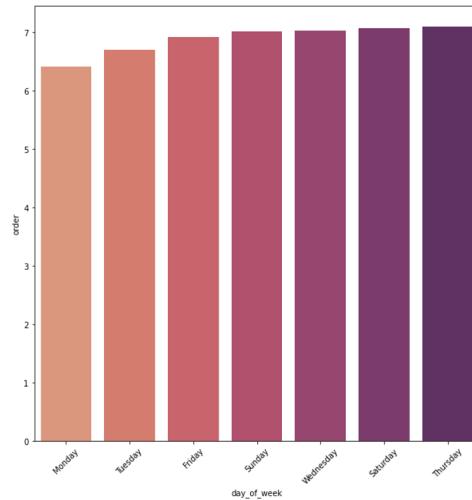
Insights

- Average click weekday doesn't show any specific difference (not good for predicting)
- We see that the ecommerce sales much more on weekdays

Features Analysis: *Day of the week*



Count of orders



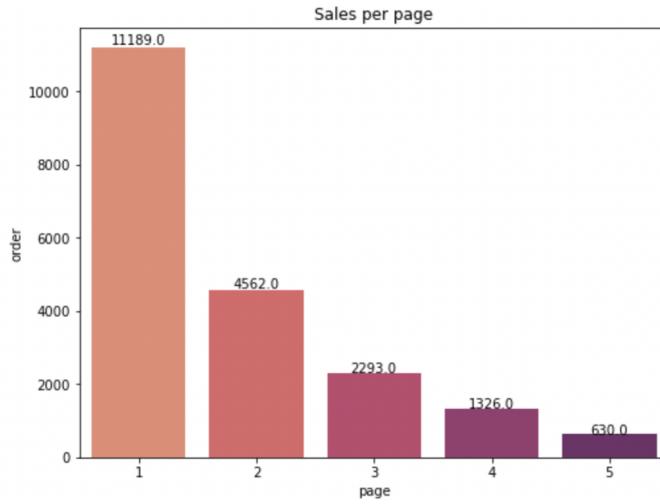
Average clicks

Note:

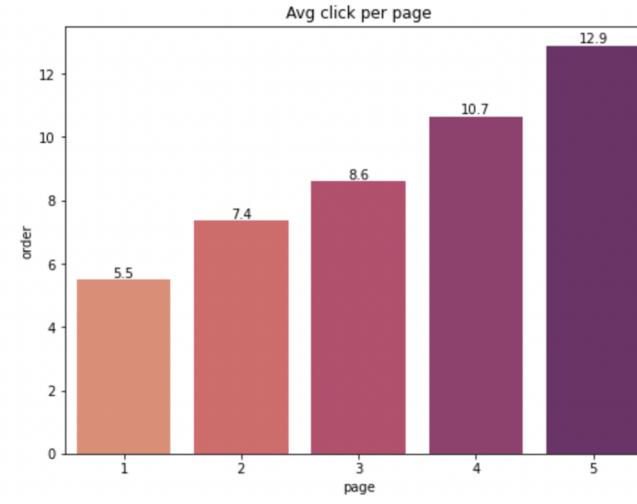
Insights

- Most of the sales are made during the week
- It is slightly easier for customers to buy after the weekend

Features Analysis: Page



Count of orders



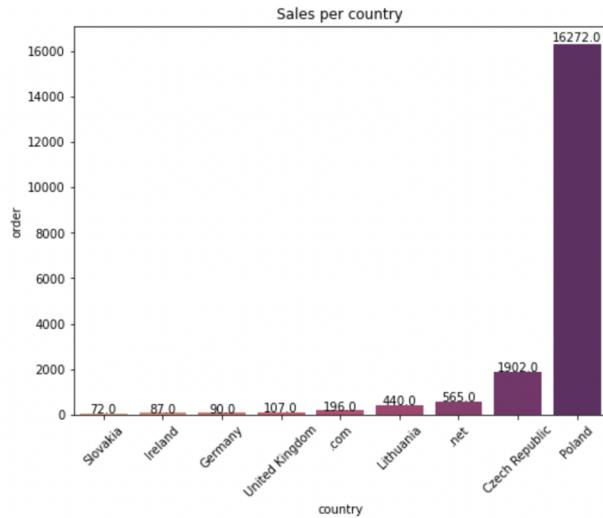
Average clicks

Insights

- Products on first pages sell much more than the ones on the last
- Products on first pages sell with less clicks than the ones on the last pages
- Good predictor of clicks

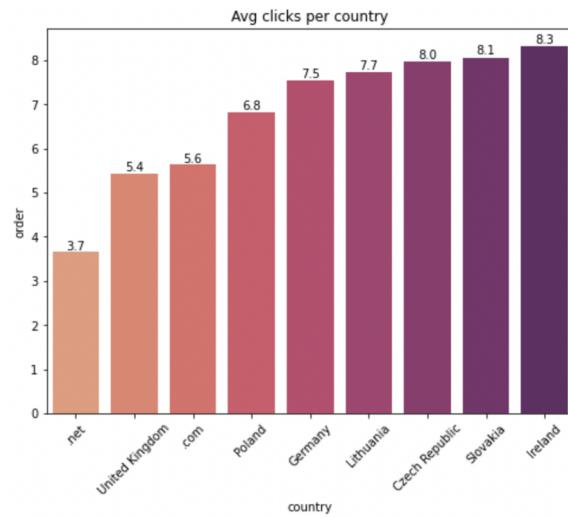
Note:

Features Analysis: Countries



Count of orders

For countries with at least 50 orders

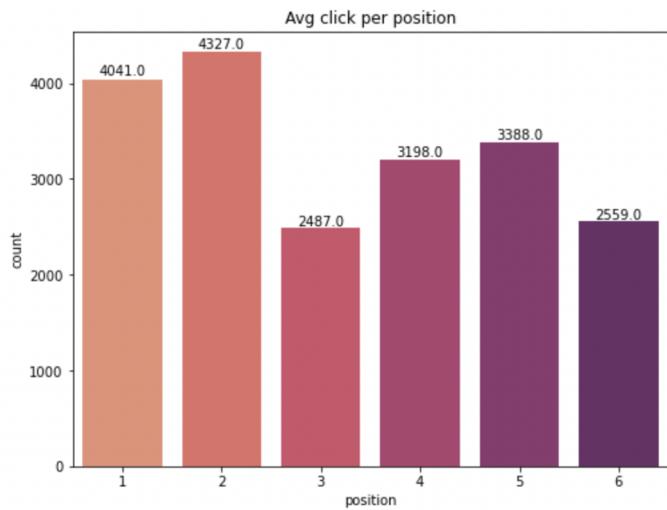


Average clicks

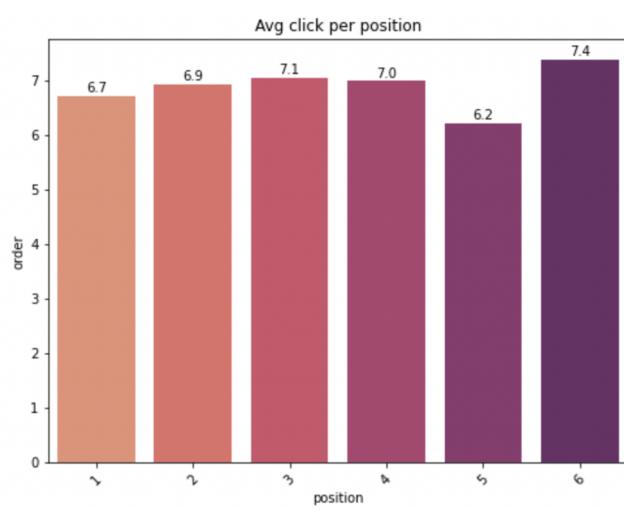
country	order	count
Ireland	8.310345	87
Slovakia	8.055556	72
Czech Republic	7.965300	1902
Lithuania	7.718182	440
Germany	7.544444	90
Poland	6.820551	16272
.com	5.637755	196
United Kingdom	5.429907	107
.net	3.672566	565

Note: After this analysis we transformed Poland and Czech into 1 other countries into 0.

Features Analysis: Position



Count of orders

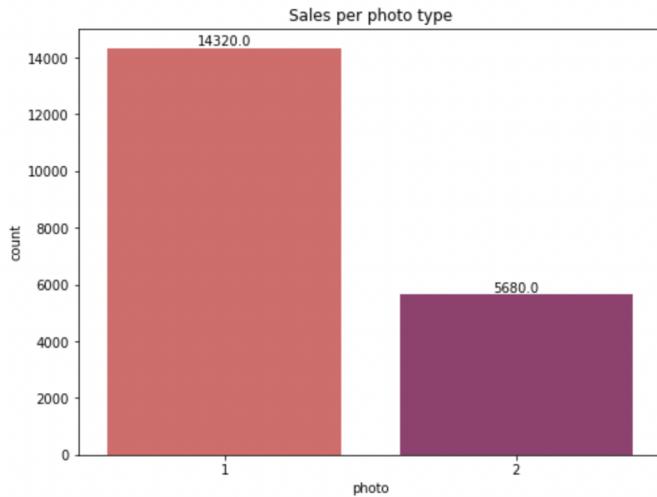


Average clicks

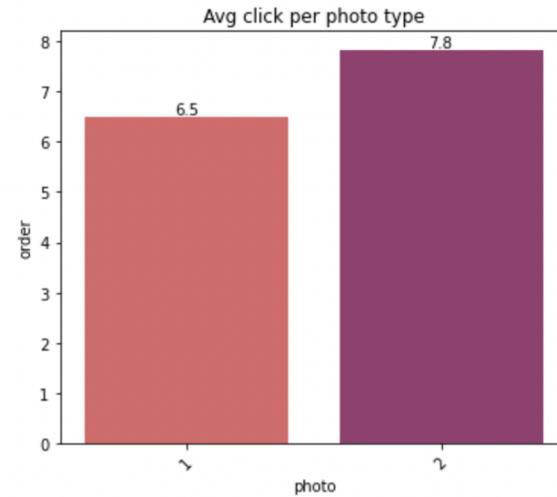
Insights

- Amount of sales is very related to the position of the product
- Avg clicks is not as much related but still the pattern appears
- Could be a good predictor of clicks

Features Analysis: *Photo*



Count of orders

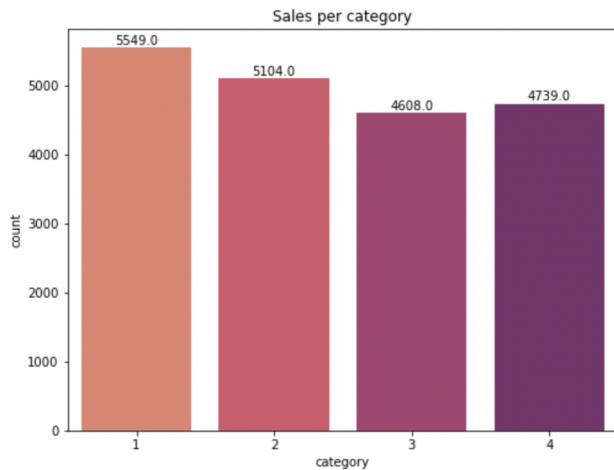


Average clicks

Insights

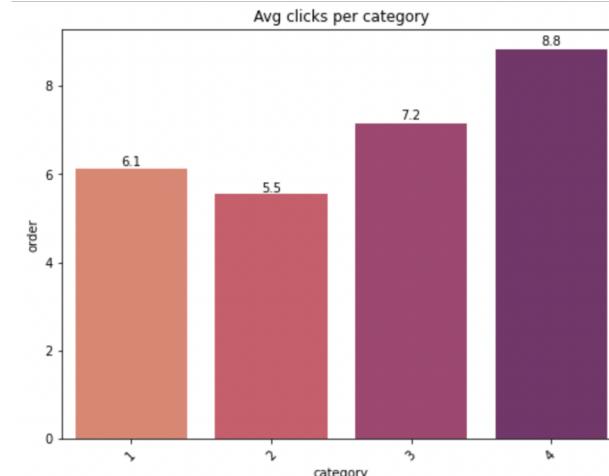
- There are more sales of face photos than profile (but we don't know total products of each)
- Avg clicks shows a significant difference between the two types
- Good predictor

Features Analysis: Category



Count of orders

1. Trousers
2. Skirts
3. Blouses
4. Sale

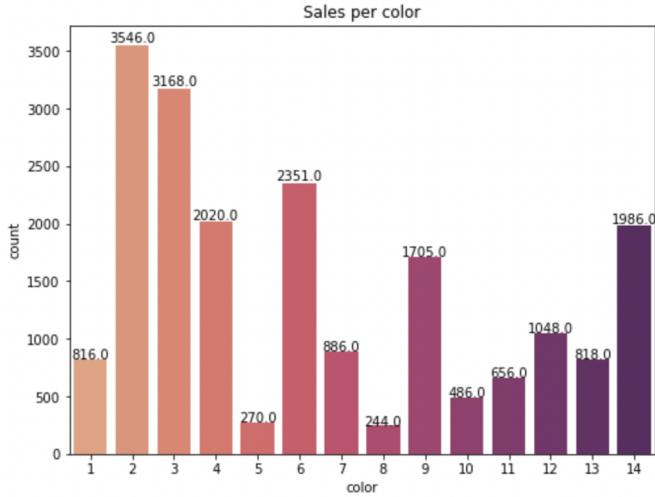


Average clicks

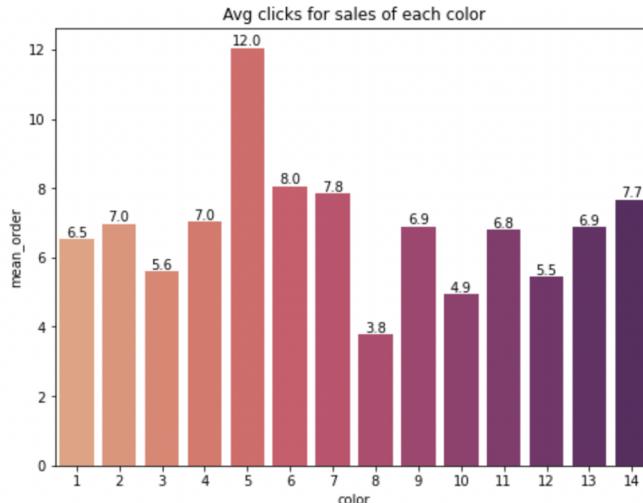
Insights

- Trousers are the top sellers
- Sales need more clicks, followed by blouses
- Skirts and trousers sell with less clicks
- Good predictor

Features Analysis: Color



Count of orders



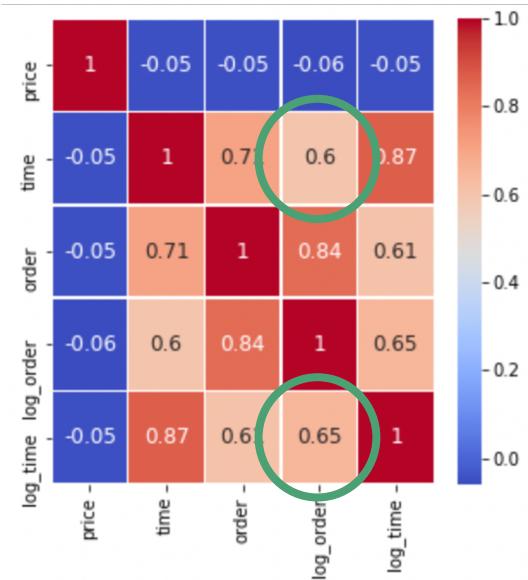
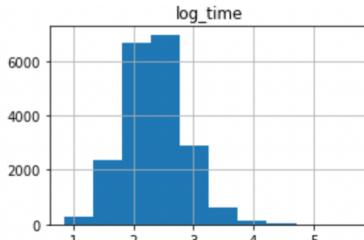
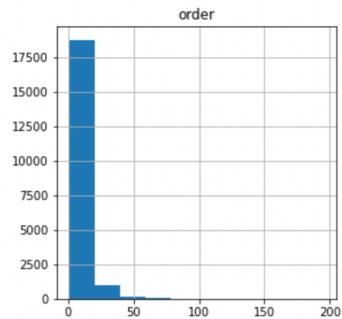
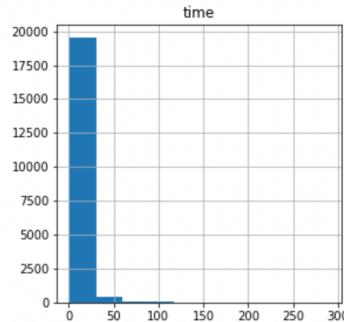
Average clicks

Insights

- Some colors sell faster and more than others but no specific correlation between sales and clicks
- We are going to add the variable to the model because of the high variations in the clicks

Note:

Features Analysis: Numerical



- Doing logs of exponential distributions
- Detecting high correlation variables which will be good predictors

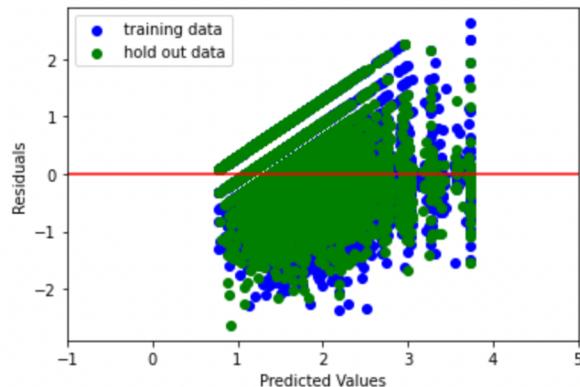
Features Analysis: *Other actions*

- Dropping “model” column
- Concat and divide into train and test
- Standardizing numerical values

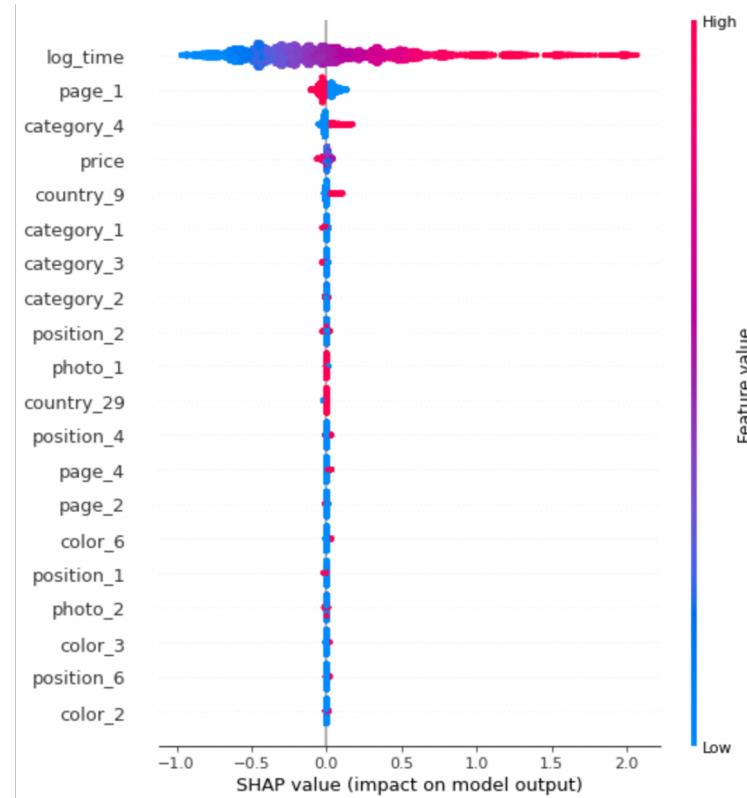
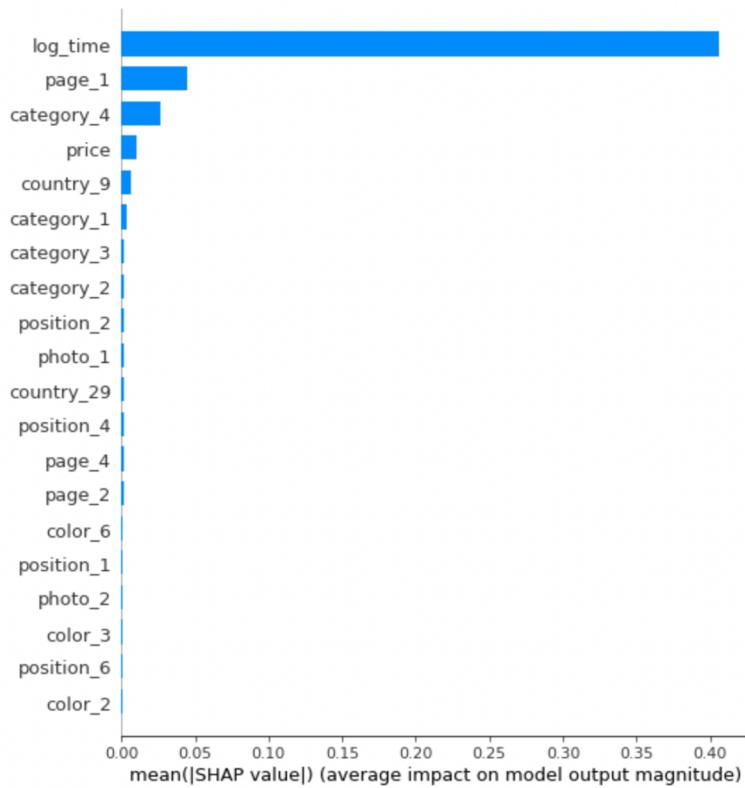
Modelling: our best find, *Random Forest*

```
Best score: -0.474879 using {'criterion': 'mse', 'min_samples_leaf': 55, 'n_estimators': 500, 'random_state': 42}
```

MAE	train 0.460	(3.422689)	test 0.479	(3.583695)
MSE	train 0.338		test 0.365	
RMSE	train 0.582		test 0.604	
r2	train 0.478		test 0.442	



Interpretation



Clustering

Three different clusters:

1. Dark purple “The saver”

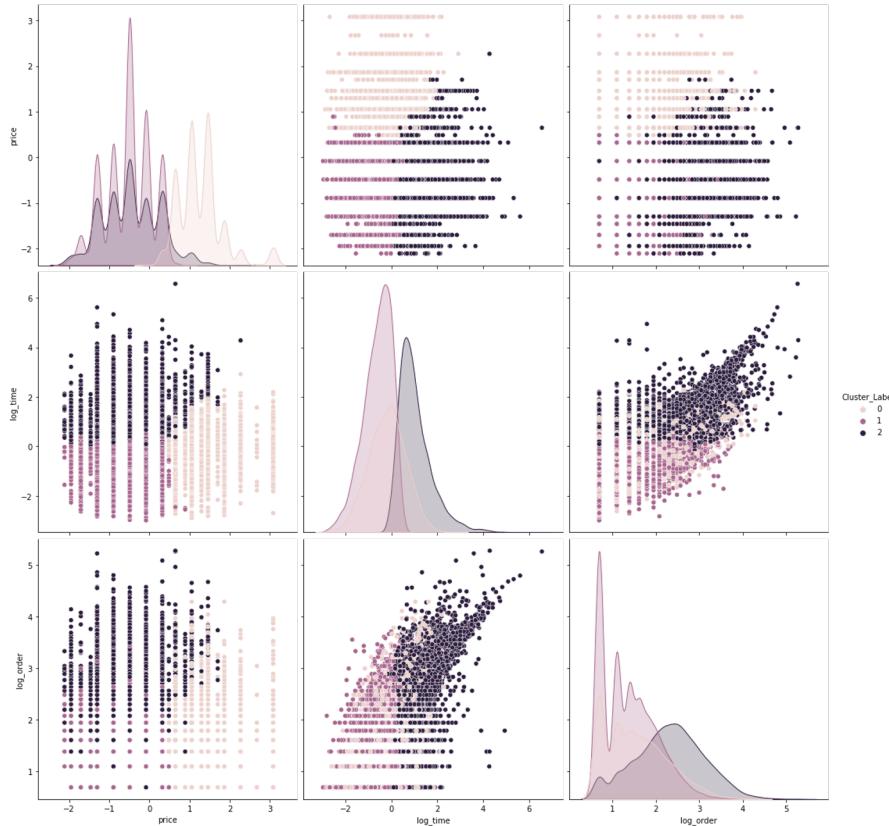
A lot of time spent on the website, a lot of clicks but buying cheaper than average items.

1. Purple “The hustly”

Spends not much time on the website, and buys cheap items. Might also buy in a couple clicks.

1. Pink “The big spender”

Does not spend a lot of time on the website, but buys pricey items.



Business Insights: Sales analysis

Revenue over months across the best selling countries

Country	Date				
	April	May	June	July	August
.com	1.986	1.770	2.449	2.304	406
.net	8.930	7.585	3.836	3.406	1.606
Czech Republic	31.023	23.320	17.519	12.258	3.725
Germany	1.093	695	895	995	248
Ireland	971	685	1.020	848	281
Lithuania	6.739	2.691	5.259	3.649	764
Poland	195.761	151.604	144.674	158.780	61.060
Slovakia	1.336	661	678	317	319
United Kingdom	1.053	824	1.024	1.464	458

Orders over months across the best selling countries

Country	Date				
	April	May	June	July	August
.com	47	37	54	49	9
.net	196	172	87	76	34
Czech Republic	678	496	382	265	81
Germany	24	14	22	25	5
Ireland	22	18	22	19	6
Lithuania	158	61	120	84	17
Poland	4.539	3.464	3.315	3.590	1.364
Slovakia	28	15	15	7	7
United Kingdom	25	20	21	32	9

Average price paid over months across the best selling countries

Country	Date				
	April	May	June	July	August
.com	42.26	47.84	45.35	47.02	45.11
.net	45.56	44.10	44.09	44.82	47.24
Czech Republic	45.76	47.02	45.86	46.26	45.99
Germany	45.54	49.64	40.68	39.80	49.60
Ireland	44.14	38.06	46.36	44.63	46.83
Lithuania	42.65	44.11	43.83	43.44	44.94
Poland	43.13	43.77	43.64	44.23	44.77
Slovakia	47.71	44.07	45.20	45.29	45.57
United Kingdom	42.12	41.20	48.76	45.75	50.89

Business Ideas: What to do now

- Expand our presence in other markets.
- Push to place picture, type of picture and colors which sell more than others on the website.
- Increase advertising.
- Analyse data on not-buying visitors.

THANK YOU!