



AIFY LABS

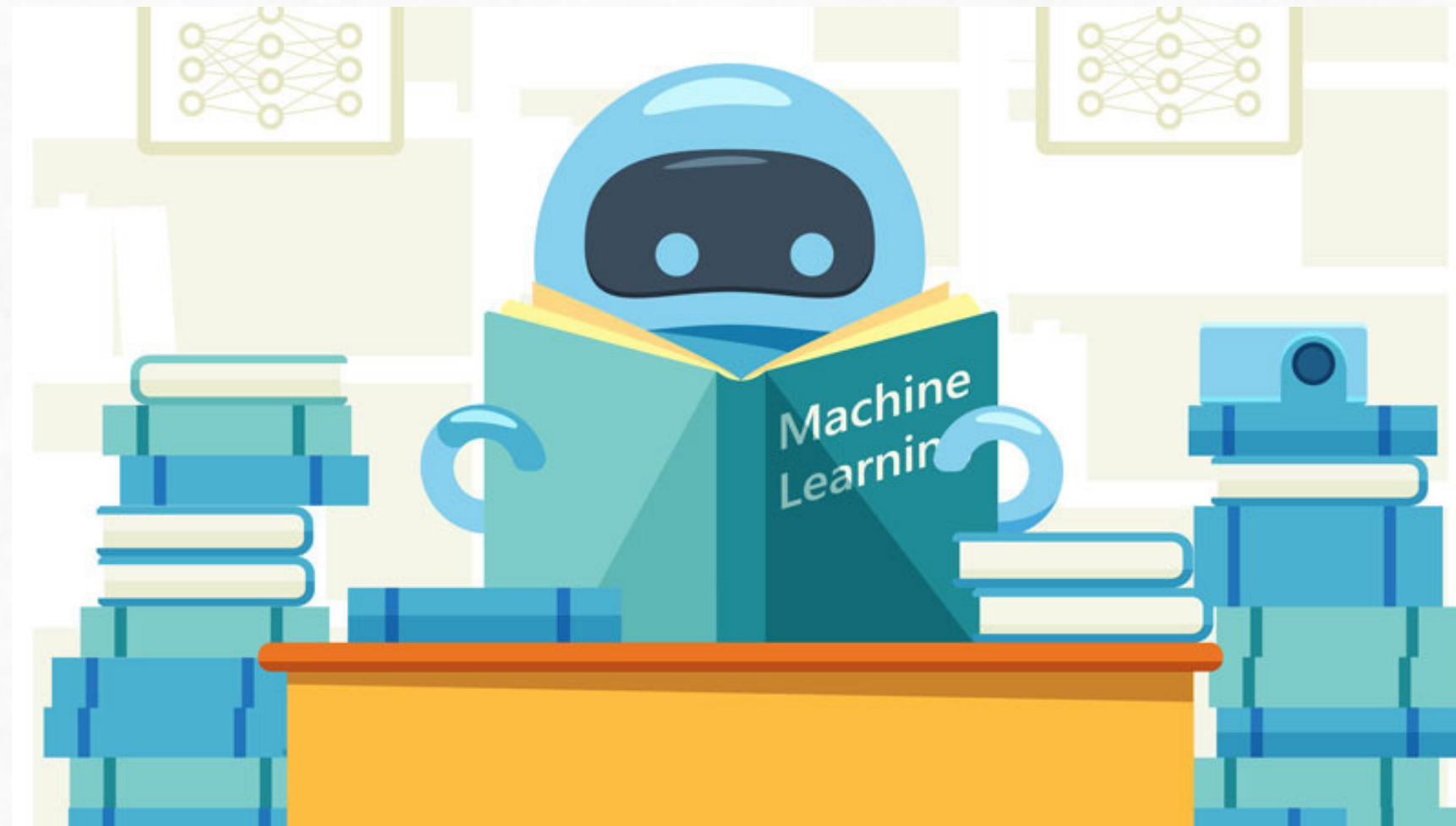
ML CASE - CHURN

BY PABLO ROLDÁN



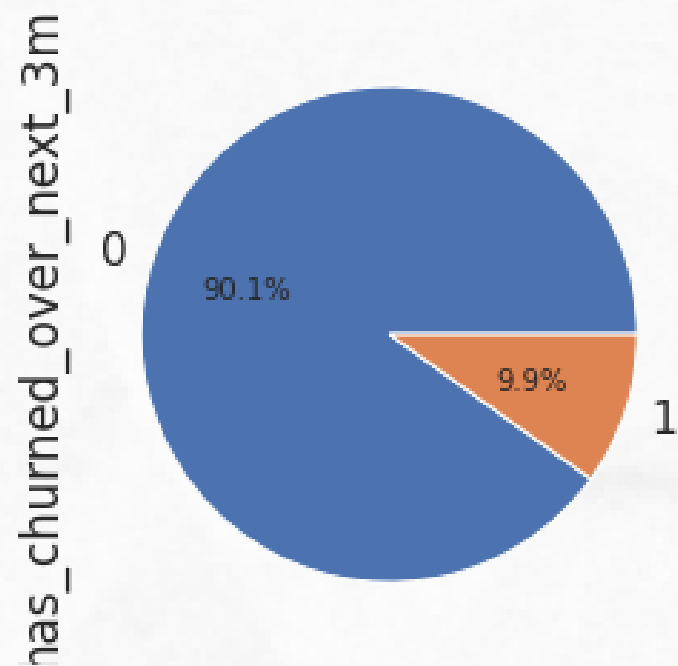
What is the task?

- Predict churn
- 20% discount measure



Our Data

- Current Churn rate



- Customer data

	contract_id	company_category	campaing_code_elect	channel	elect_cons_at_12m	
2684	46a6beccf1aa6f9d5b7985750544d569	ibkiiwcxiccxpoedpweiuxwbxbuewbxm	NaN	NaN	293097	
2017	c2af3ffb4d304807914b34ce318559b4	NaN	NaN	NaN	26070	...

- Historic Customer data

	contract_id	reference_date	price_energy_1st_period	price_energy_2nd_period	price_energy_3rd_period	price_power_1st_period	price_power_2nd_period	price_power_3rd_period
7149	3934e65220d612324e33b86d5ce6d27a	2015-09-01	0.105169	0.078244	0.092591	36.735692	18.466007	12.349485

Merging data

- Overkill

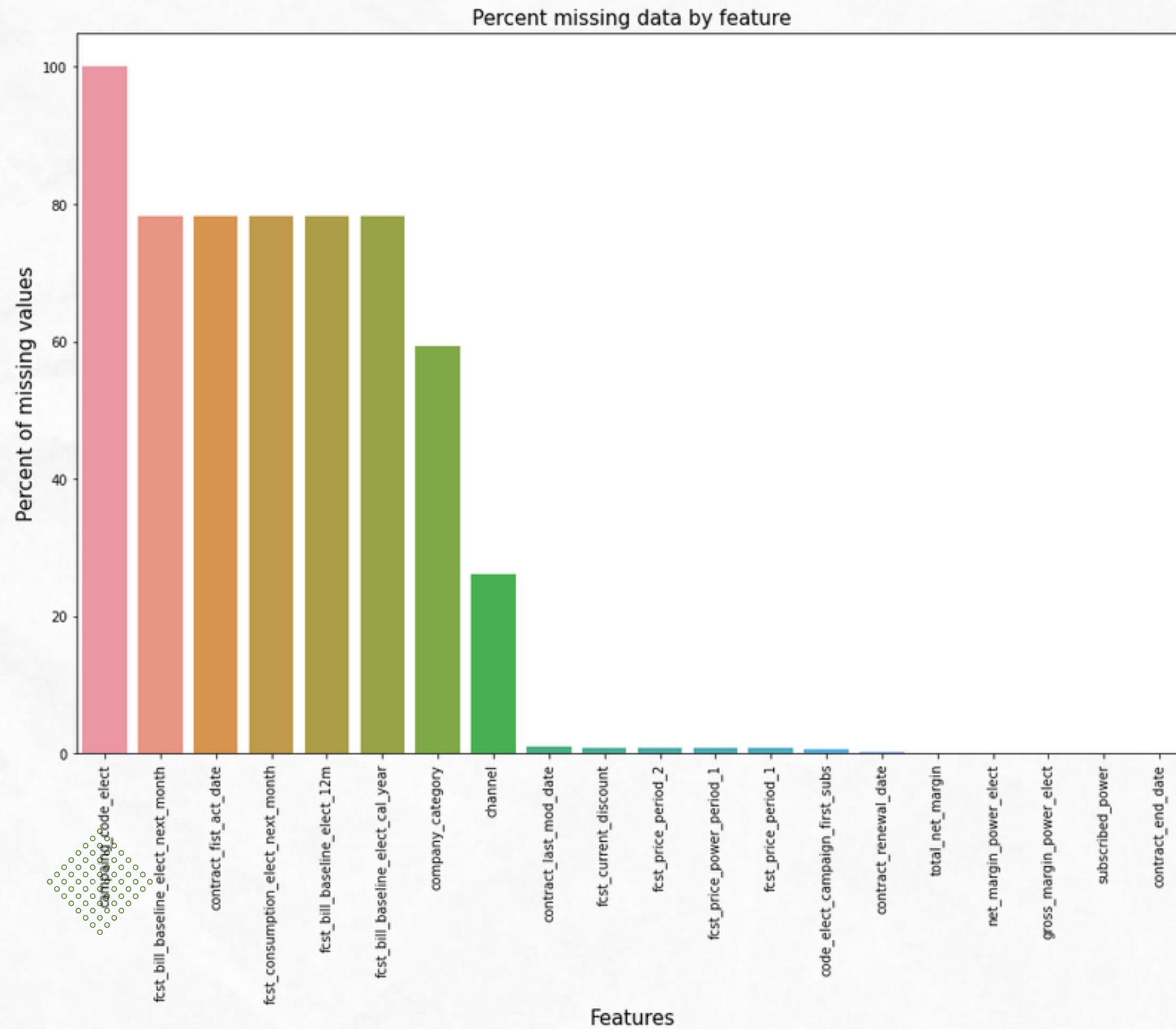
	01_price_energy_1st_period	01_price_energy_2nd_period	01_price_energy_3rd_period	01_price_power_1st_period	01_price_power_2nd_period	01_price_power_3rd_period	02_price_energy_1st_period	...
0	0.136157	0.000000	0.0	39.926856	0.0	0.0	0.136157	
1	0.134560	0.000000	0.0	39.926856	0.0	0.0	0.134560	
2	0.163324	0.067422	0.0	39.926856	0.0	0.0	0.163324	
3	0.136157	0.000000	0.0	39.926856	0.0	0.0	0.136157	
4	0.137101	0.000000	0.0	39.926856	0.0	0.0	0.137101	

- Price aggregation

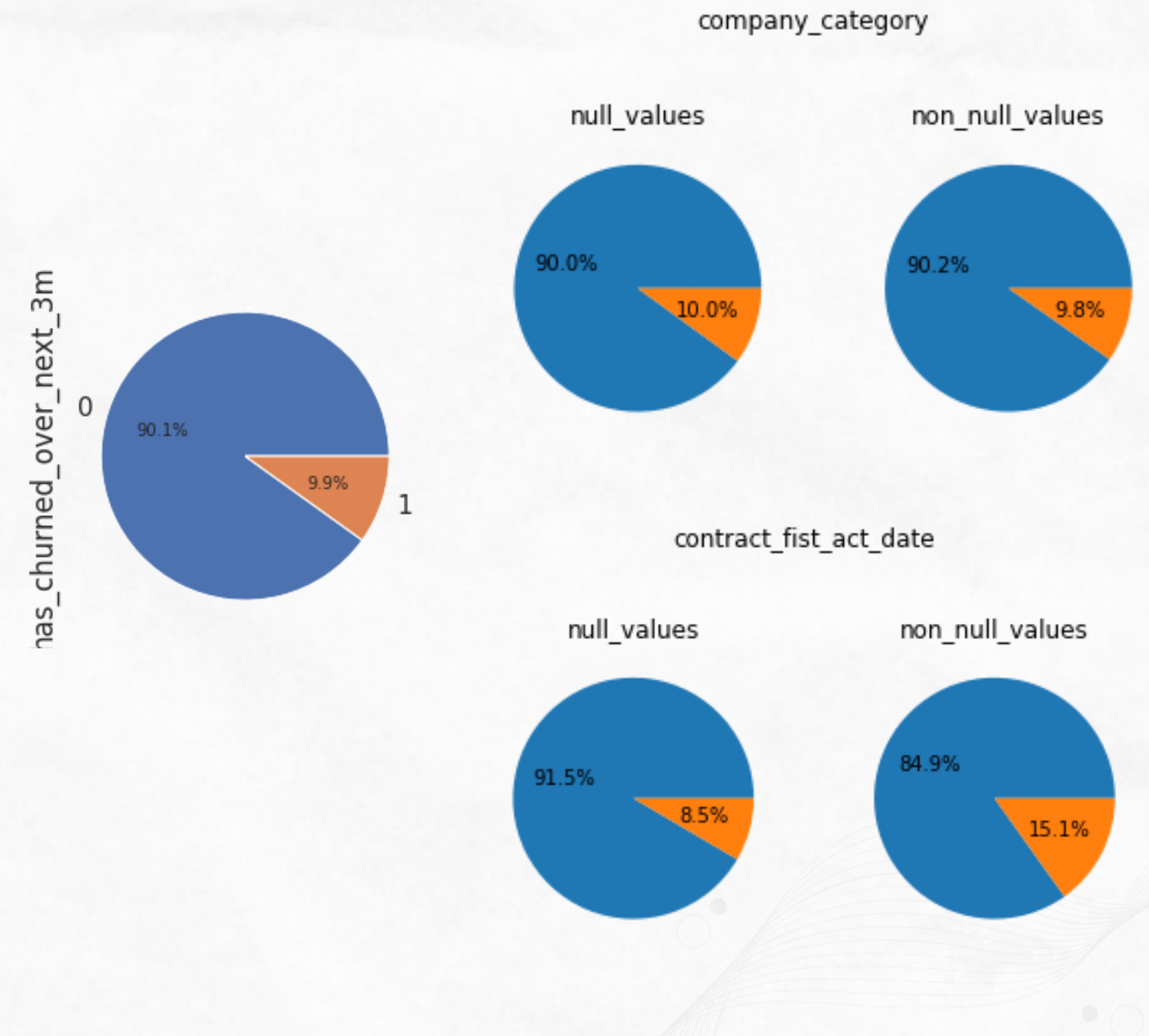
	01_price_energy	01_price_power	02_price_energy	02_price_power	03_price_energy	03_price_power	04_price_energy	...
0	0.136157	39.926856	0.136157	39.926856	0.136157	39.926856	0.137643	
1	0.134560	39.926856	0.134560	39.926856	0.134560	39.926856	0.134560	
2	0.160066	39.926856	0.160066	39.926856	0.160066	39.926856	0.160624	
3	0.136157	39.926856	0.136157	39.926856	0.136157	39.926856	0.137101	
4	0.137101	39.926856	0.137101	39.926856	0.137101	39.926856	0.137101	

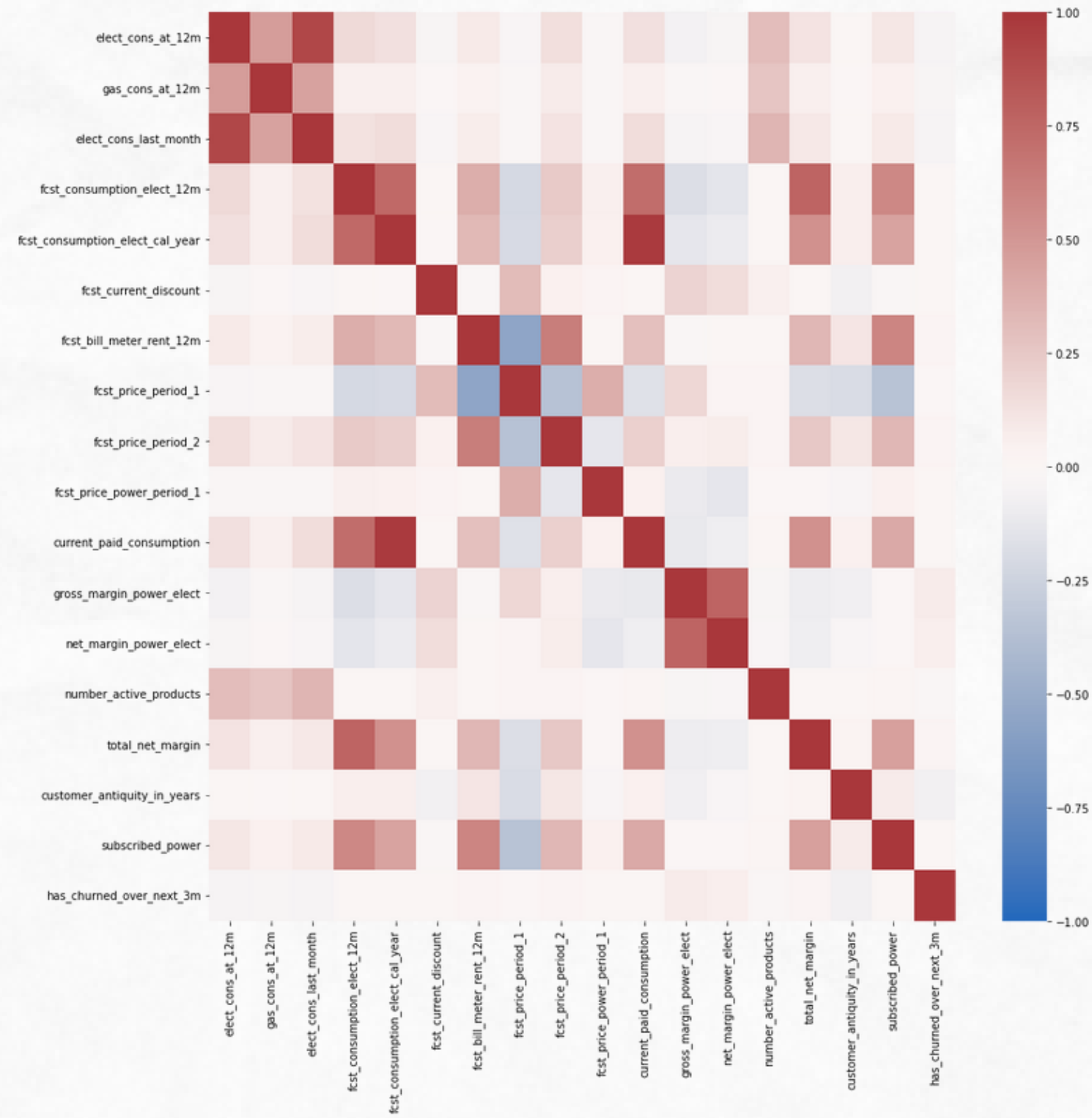
Cleaning data

- Percent missing data



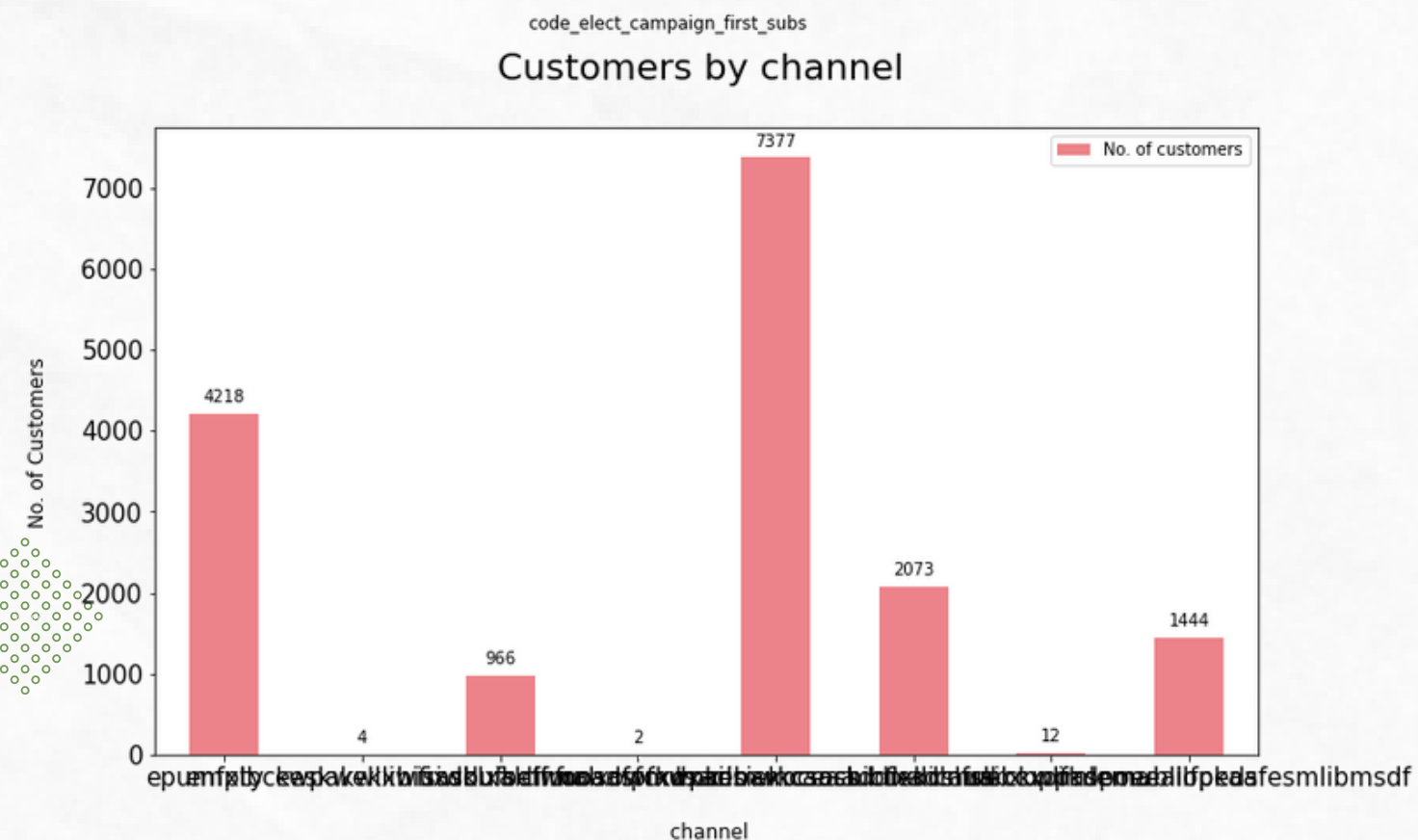
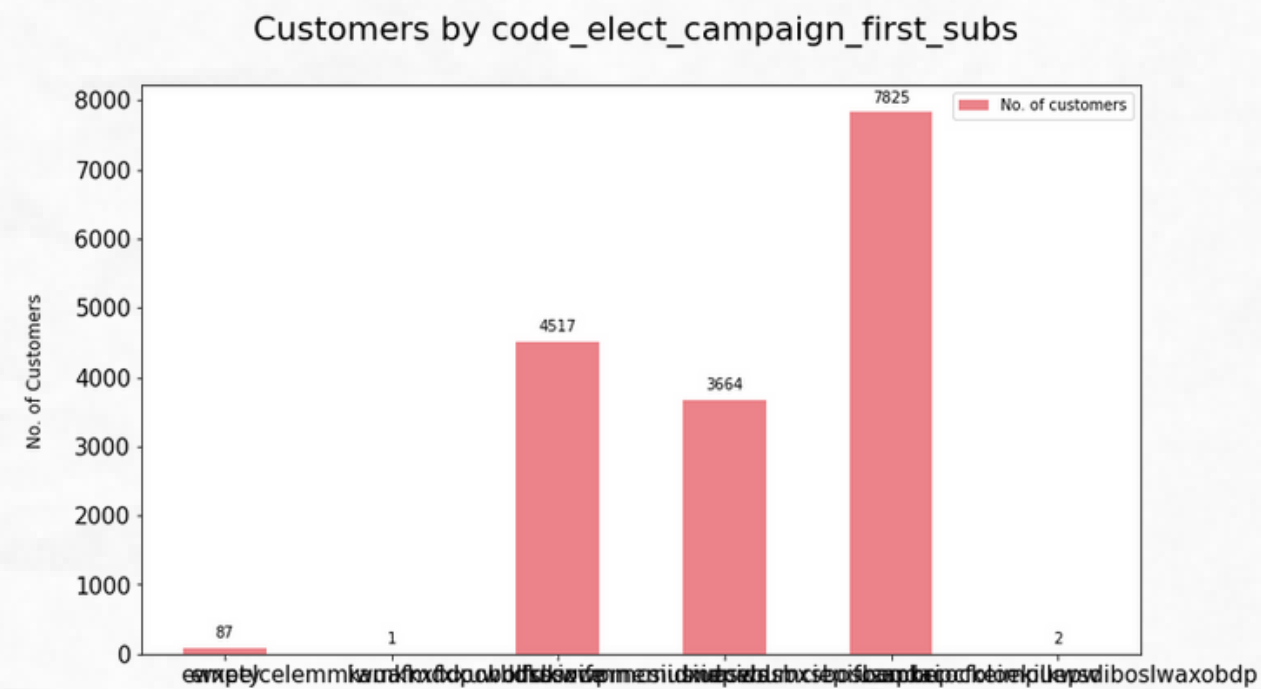
- Were they important?





Exploratory Data Analysis (EDA)

- Categorical Data



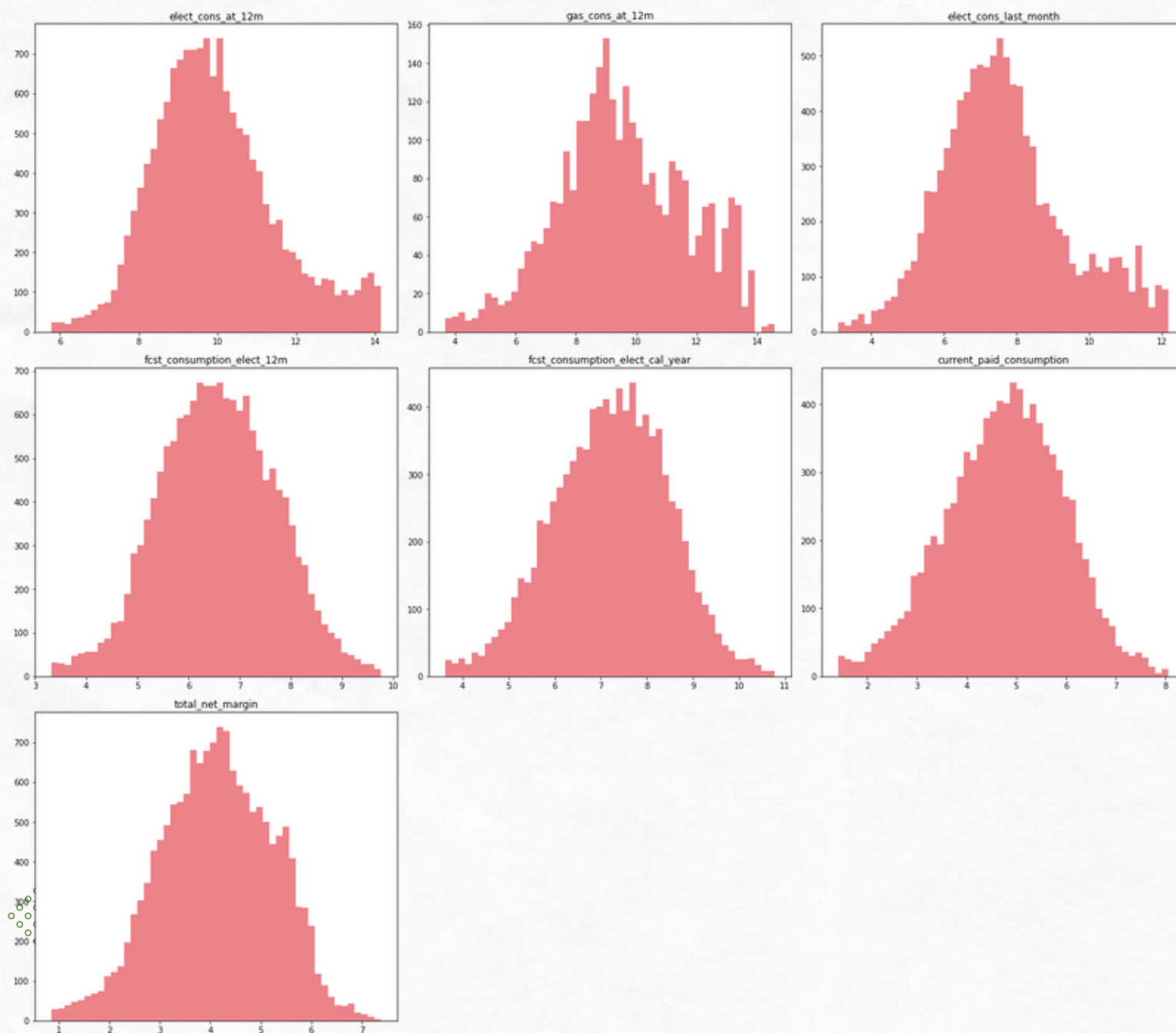
- Numerical data



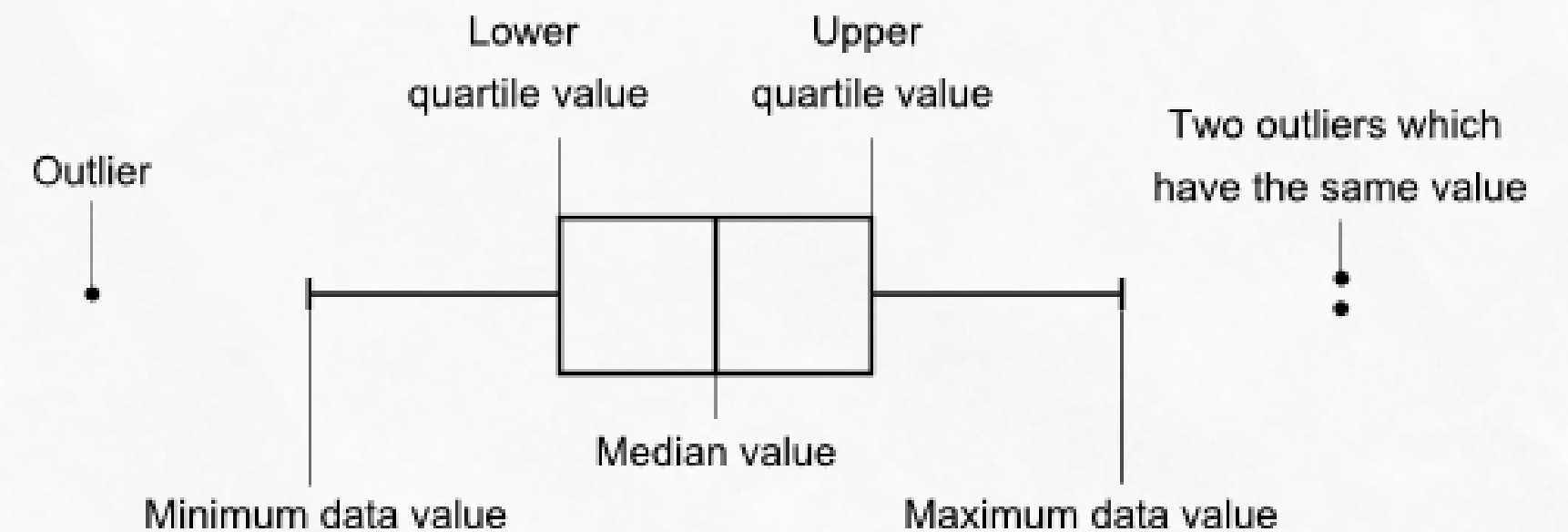
Outliers

- Exponential Data

Histograms of Numerical Columns (without outliers)



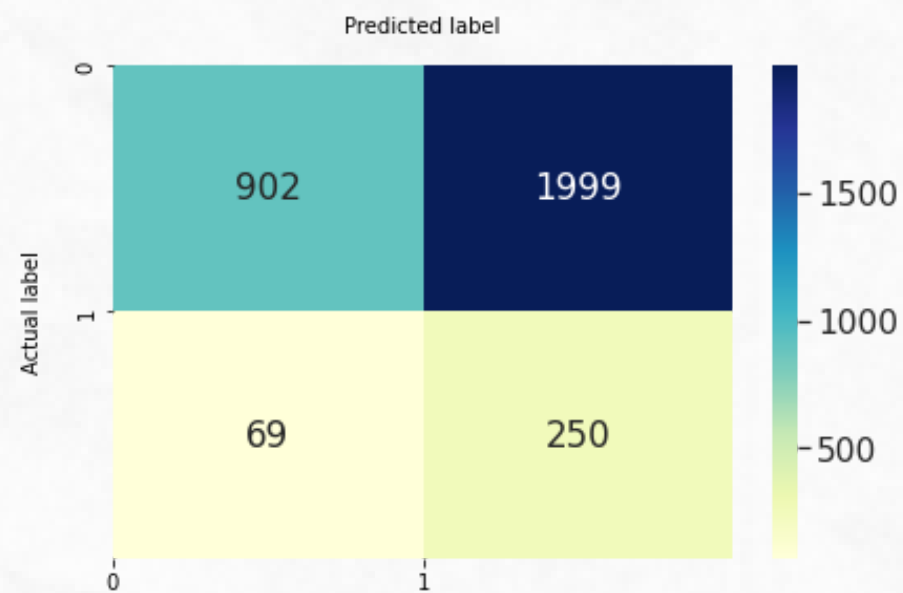
- Numerical Data



Modelling

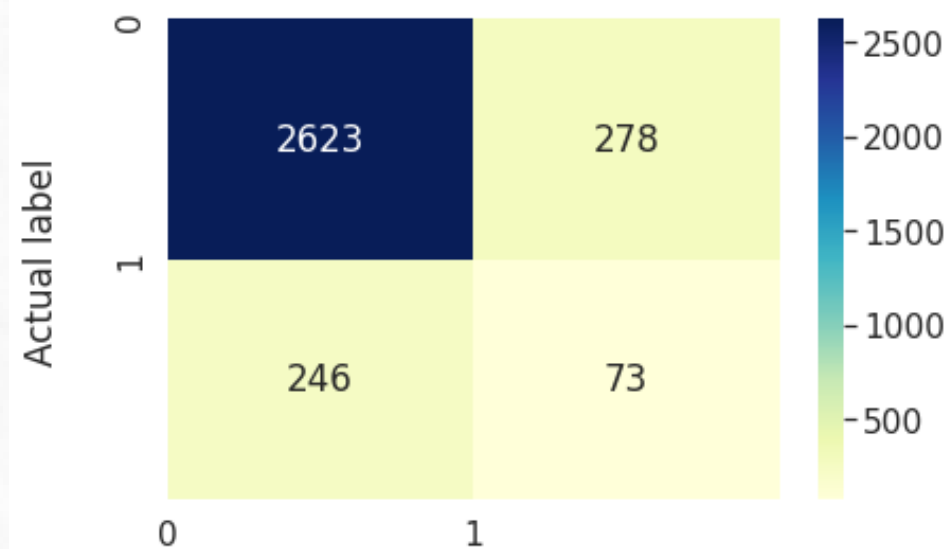
- Logistic Regression

Confusion matrix Logistic Regression



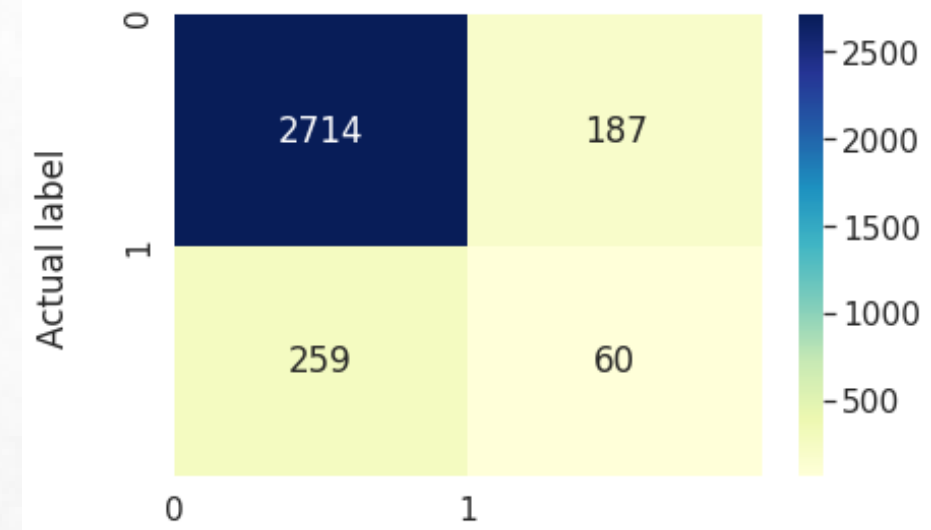
- Decision Tree Classifier

Confusion matrix Decision Tree Classifier



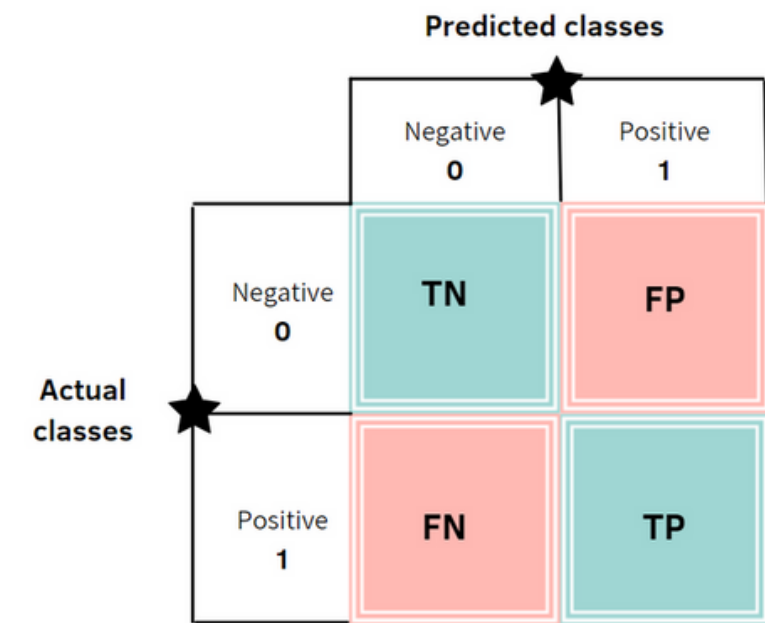
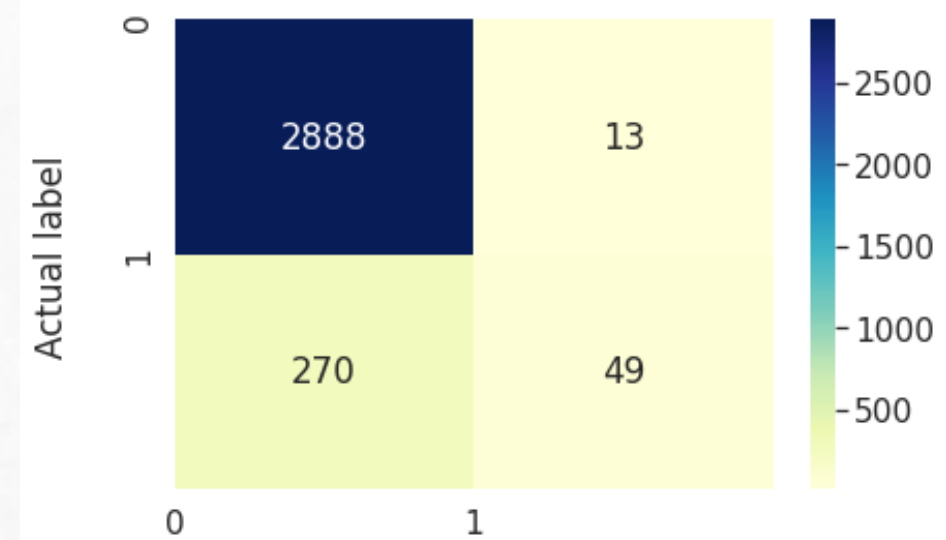
- KNN

Confusion matrix KNN



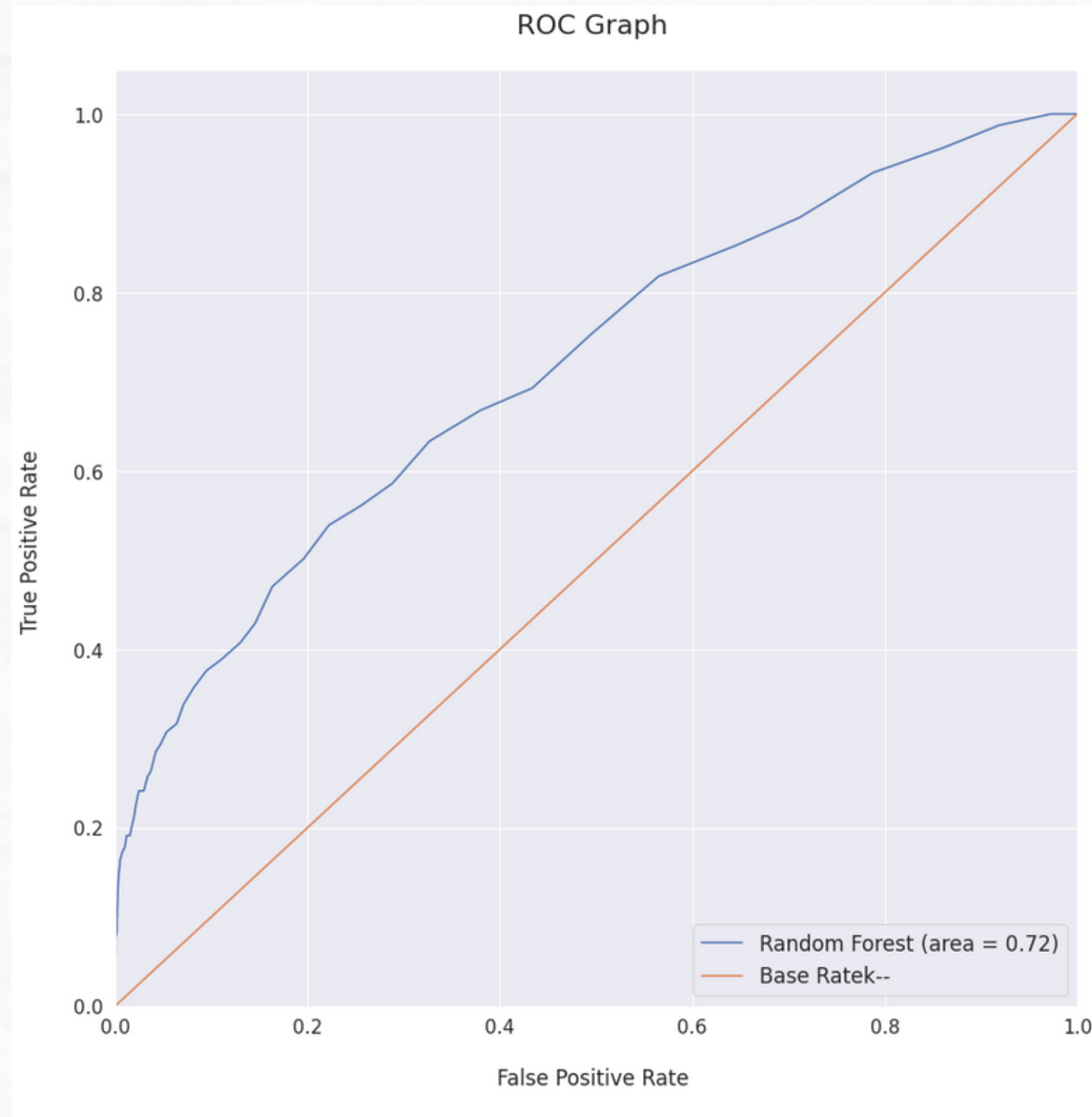
- Random Forest

Confusion matrix Random Forest



Metrics

- ROC Curve



- Brier Score

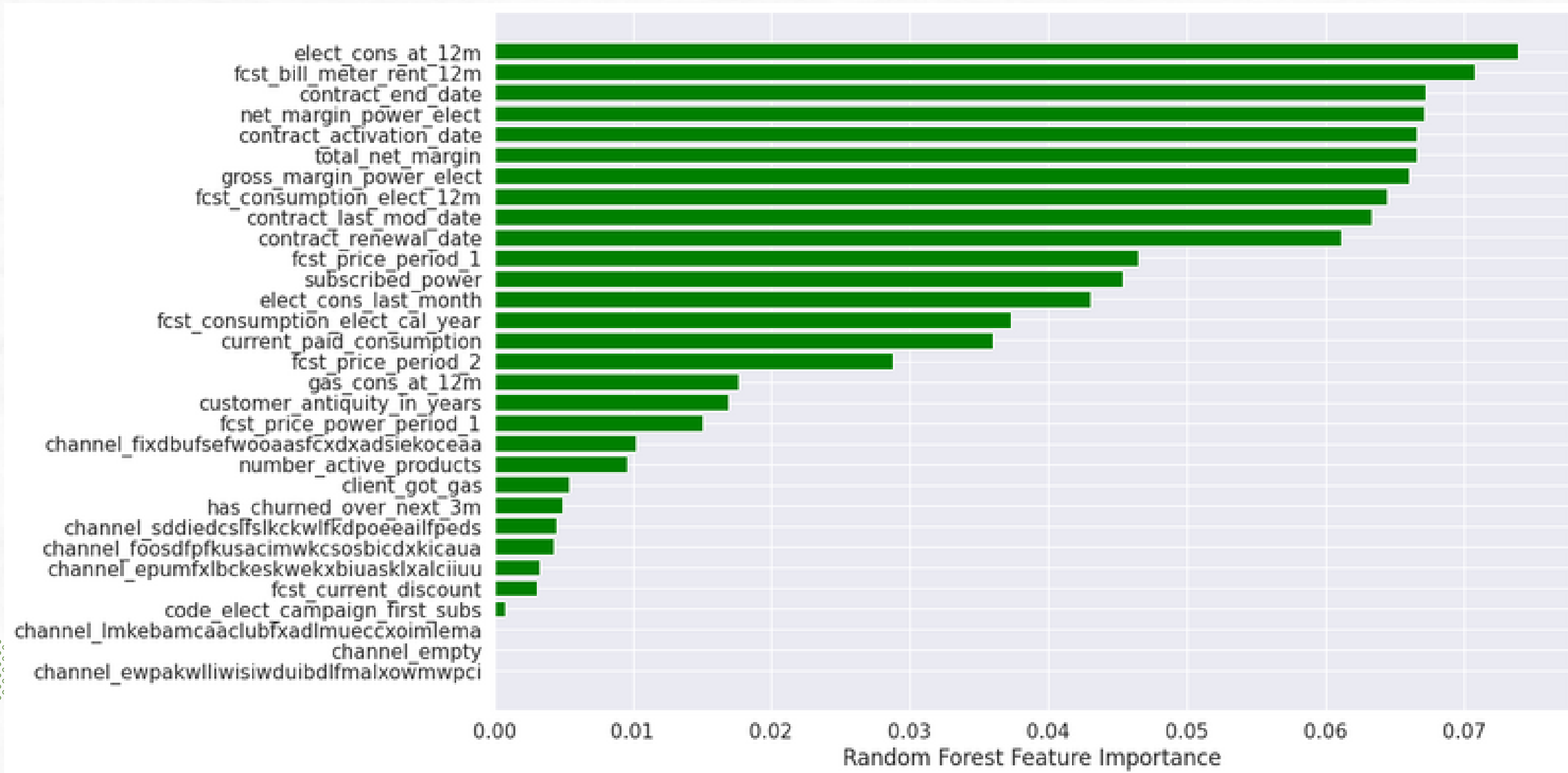
0.07850649068322982

Predicted test

	contract_id	does_churn	prob_of_churn
2725	c0b7dbcae838c20005a08a5a93d10ba4	1.0	0.40
2327	b2326a939e35c9930cd47152321b09fd	1.0	0.40
1768	af746cde31bc82e7262c592a85e0adf7	0.0	0.36
1254	25ac33d516bc673cad5bdc234661bcda	0.0	0.36
743	37d7077da5a390bd087630b59516d3e0	0.0	0.35
...
1423	e57002278064e9c82c466e7098765382	0.0	0.12
643	cd281baf585b10aef7ab4e705f9abb7a	0.0	0.12
190	4431f949af70489fc8087c16a100cc71	0.0	0.12
3067	2149d160d765bf198ac62e7520c0c60b	0.0	0.11
1645	44c1f402a87e671799cea460b15b6c62	0.0	0.11

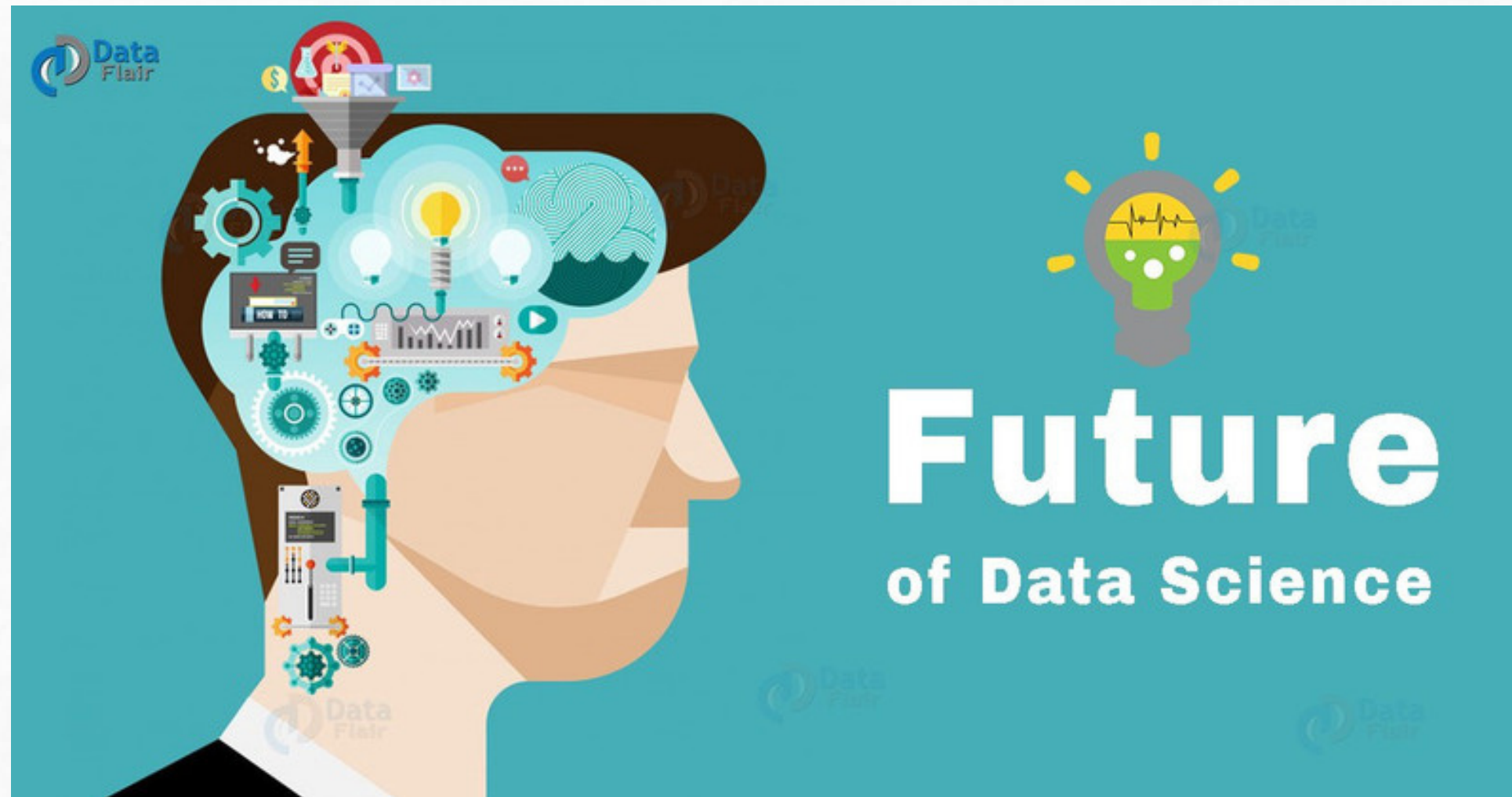
20% discount measure

- Feature Importance



Conclusion

- Is it good enough?
- Further improvements



THANK YOU

FOR YOUR ATTENTION

