



Bank Marketing Campaign project description

OutLine

1. Problem
2. Data
 - Data cleaning
 - EDA
3. Models (logistic regression and XGBClassifier)
4. Result
5. Recommendation



1.

Problem

Bank Marketing campaign



- ▷ *We want to know if the customer is willing to subscribe a term of deposit?*
- ▷ *What other information can we generate from the data?*



Process



2. Data set

Data set includes more than 41K row with about 21 column.

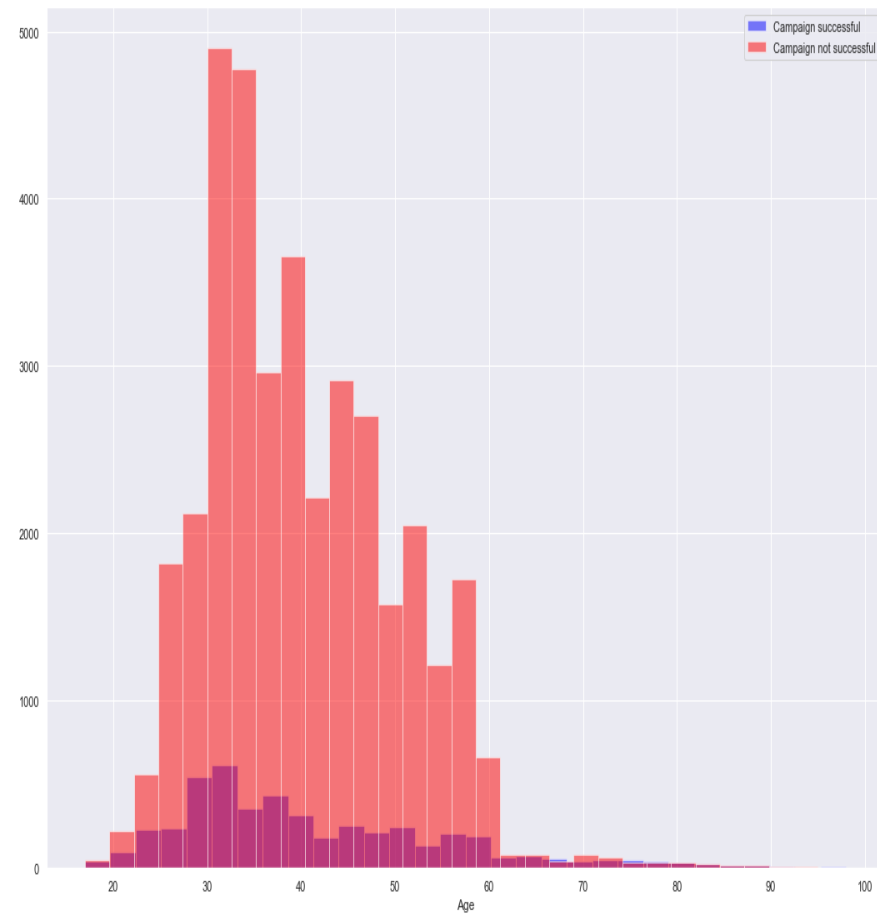
Explore to understand the categorical and numerical features

Data cleaning:

- remove duplicate values
- replace the outliers with mean value

What have we found in the data ?

Relation between age of
clients and success of the
campain

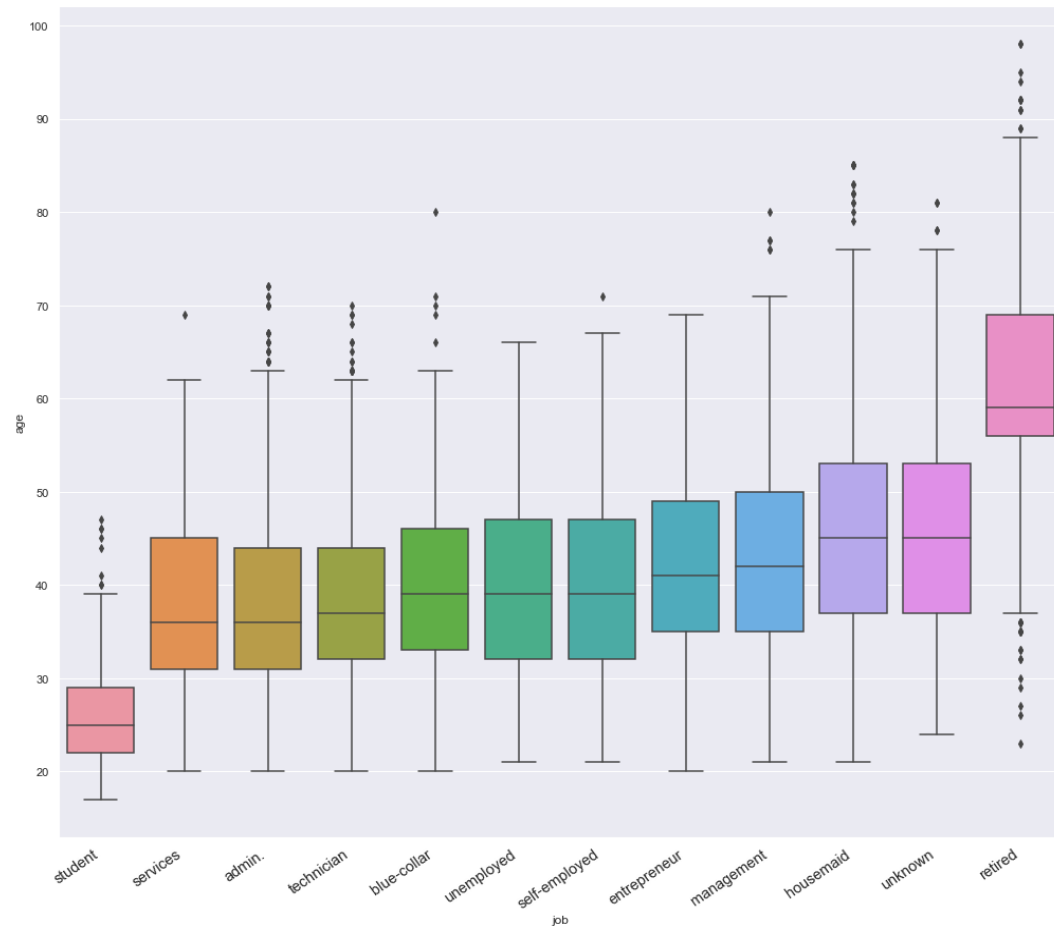


Observation:

Campaign seem to be most successful among younger (< 20 years old) and older clients (>60 years old)

What have we found in the data ?

Age and Job relation



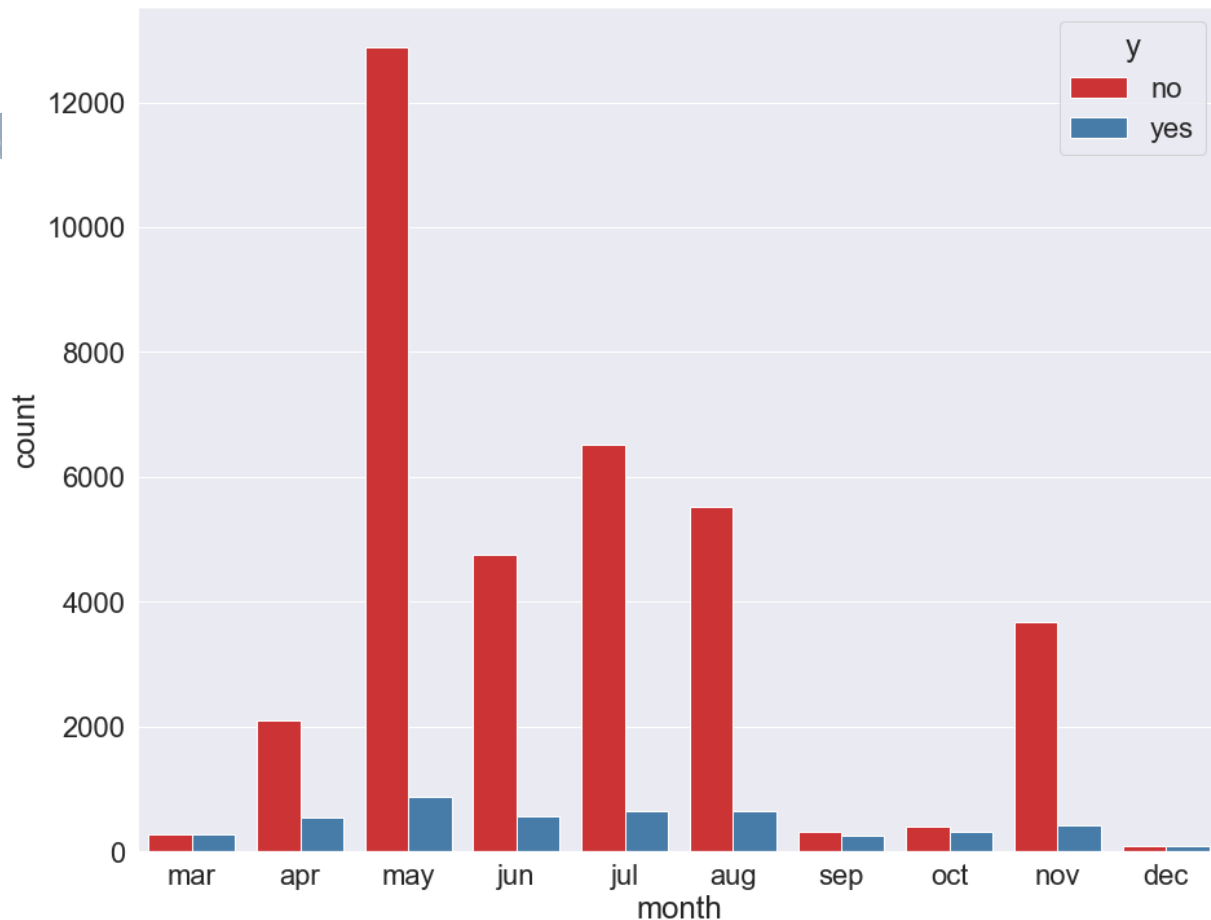
Observation:

Students tend to be younger while retirees are the oldest on average.

Note that there are apparent outliers among both professions (young retirees and older students).

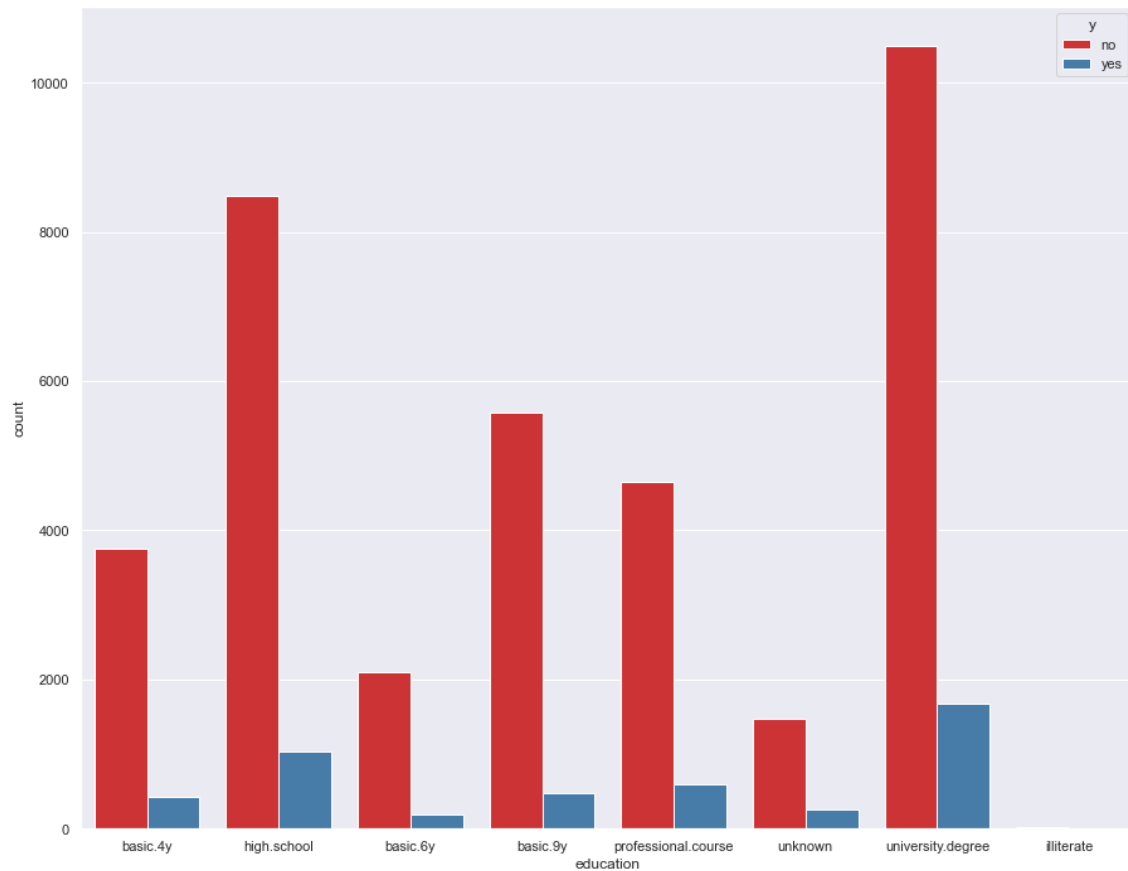
What have we found
in the data ?

Campaigns are more
success in May and July



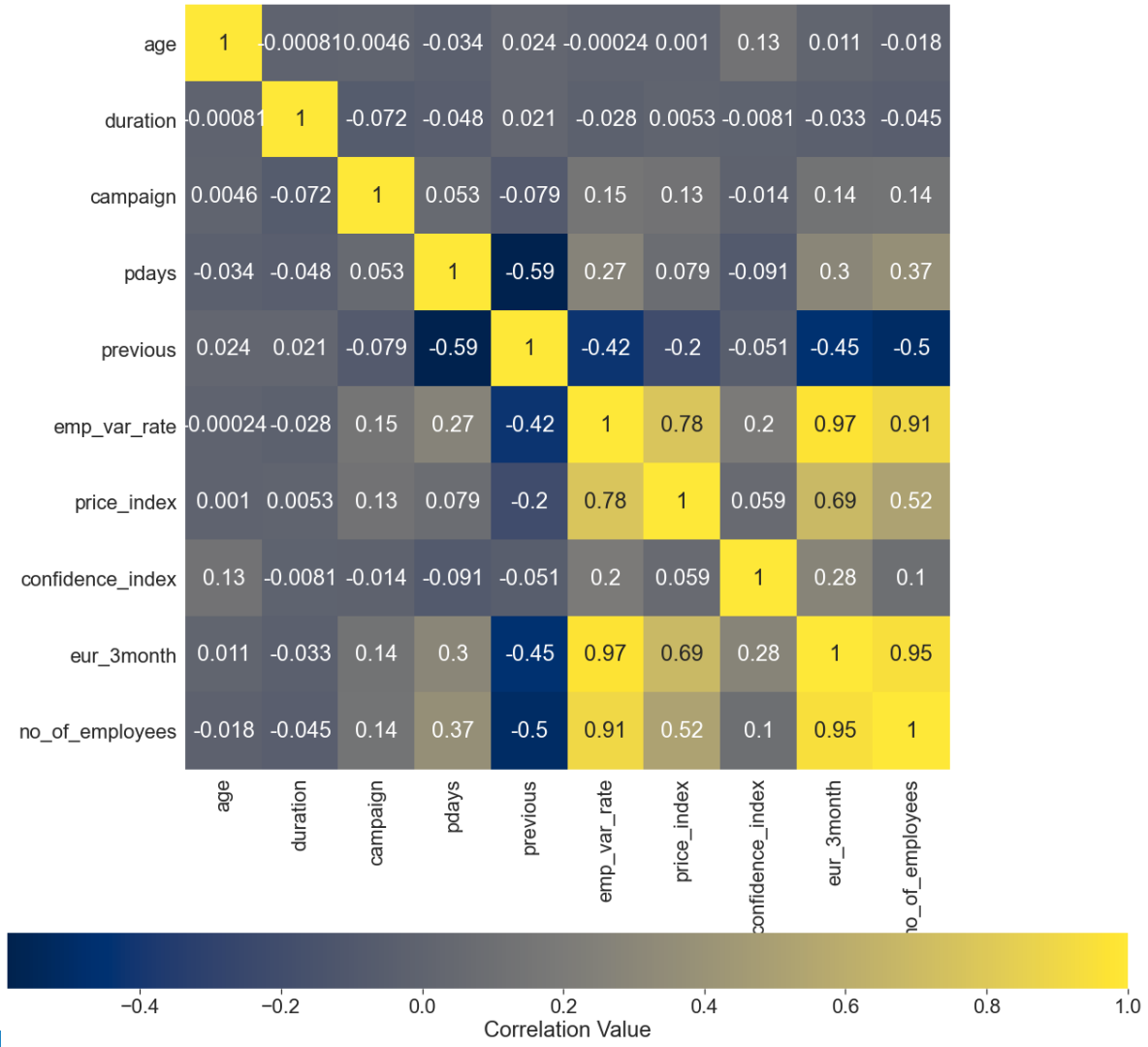
What have we found in the data ?

Campaign success by client's level of education

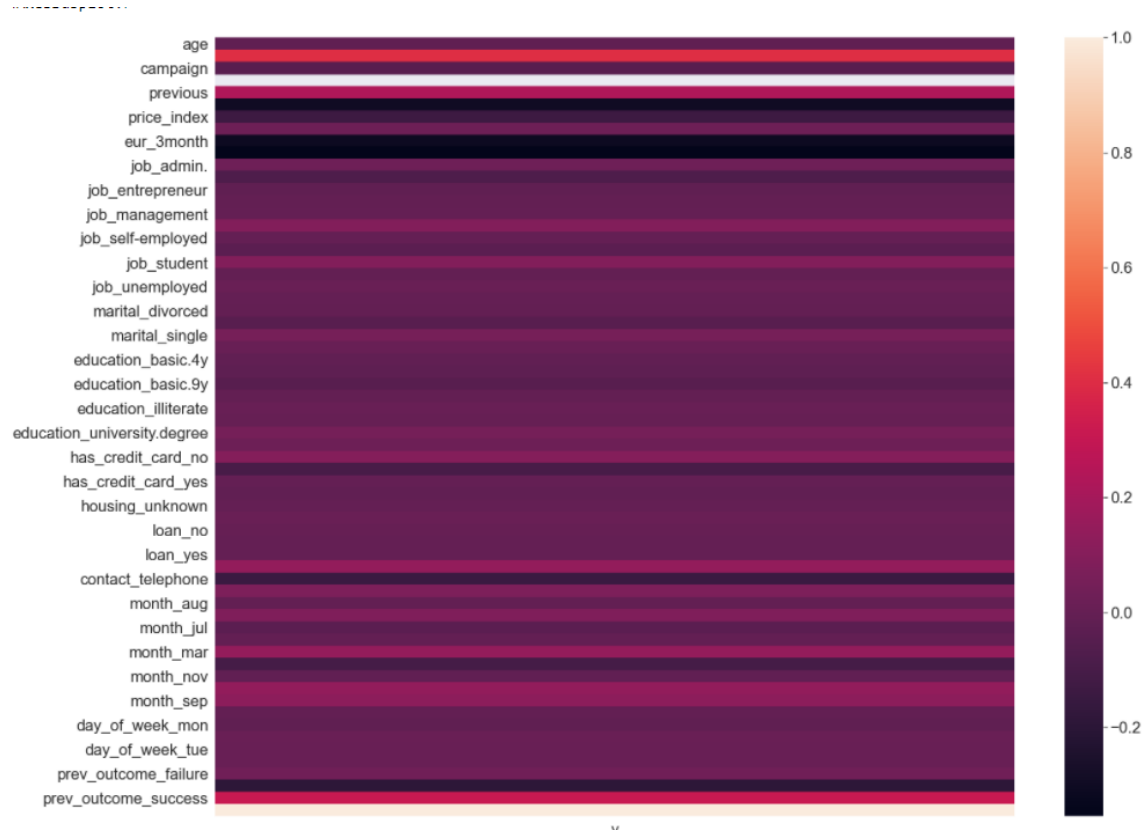


And more futures were added in the EDA
code..

CORRELATION MATRIX



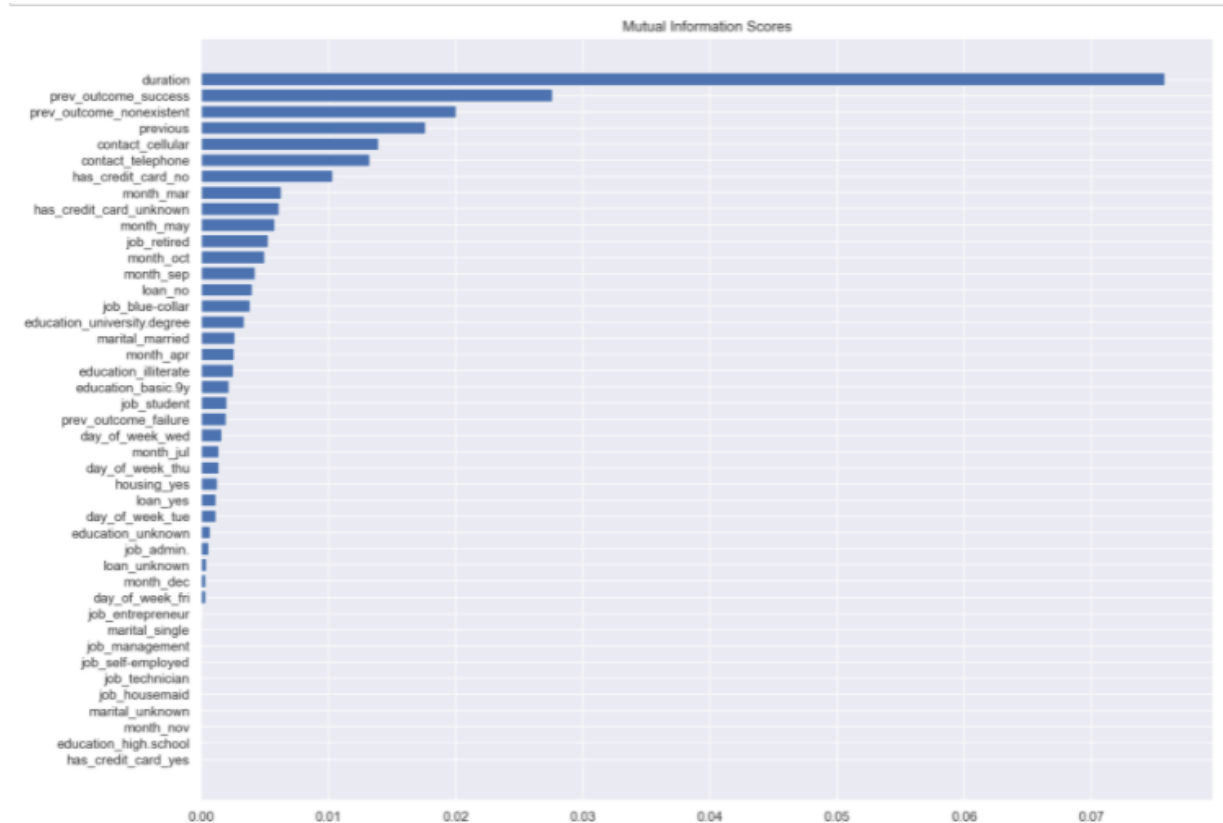
How much do the other columns correlate with the target column "y"?



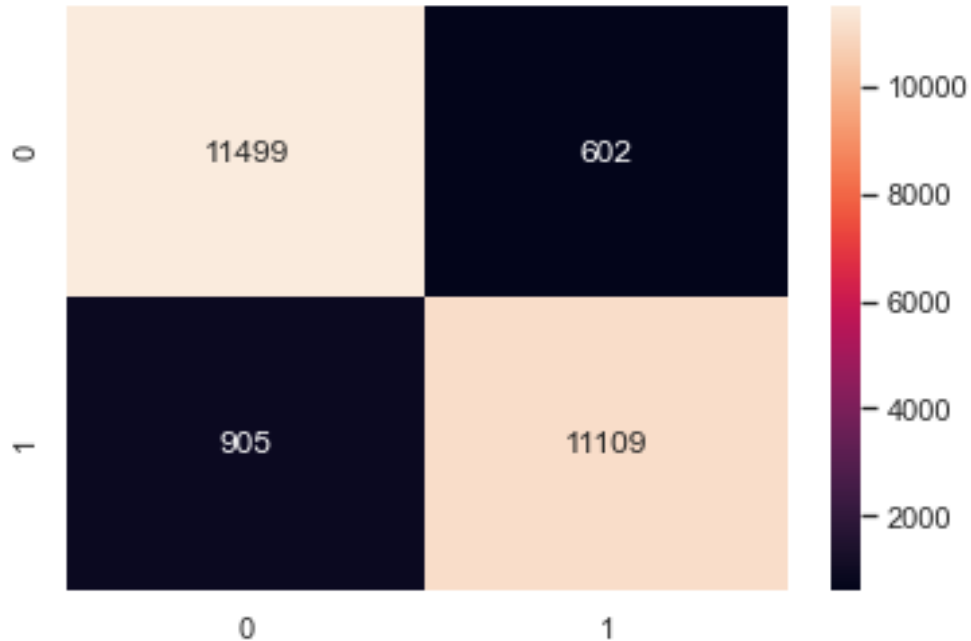
Mutual information score

We have a number of features that are highly informative and several that don't seem to be informative at all (at least by themselves).

We will try dropping features with 0.0 MI scores.



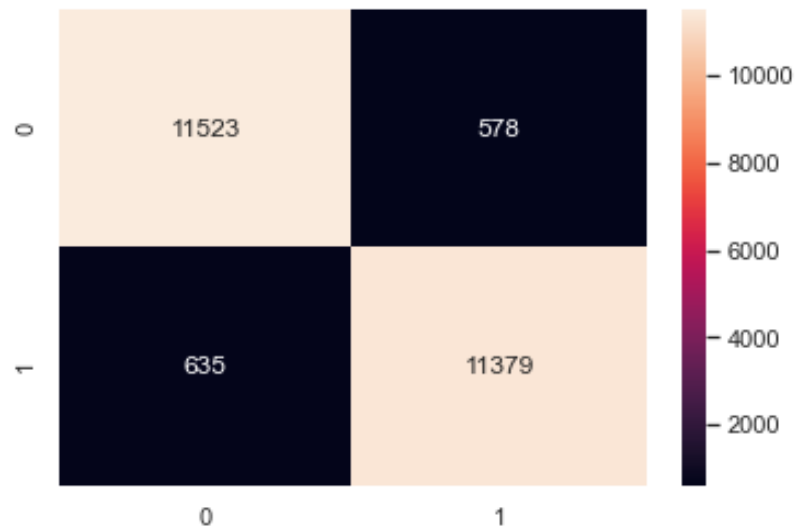
3. Models: Logistic regression



The accuracy is: 0.937

3. Models: XGBClassifier

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| no | 0.95 | 0.95 | 0.95 | 12101 |
| yes | 0.95 | 0.95 | 0.95 | 12014 |
| accuracy | | | 0.95 | 24115 |
| macro avg | 0.95 | 0.95 | 0.95 | 24115 |
| weighted avg | 0.95 | 0.95 | 0.95 | 24115 |



Future Recommendation

- Other models and evaluate results.(ANN)
- Focus on other aspect of the data.

Thanks!

Any questions?

Sarah Alzighibi

