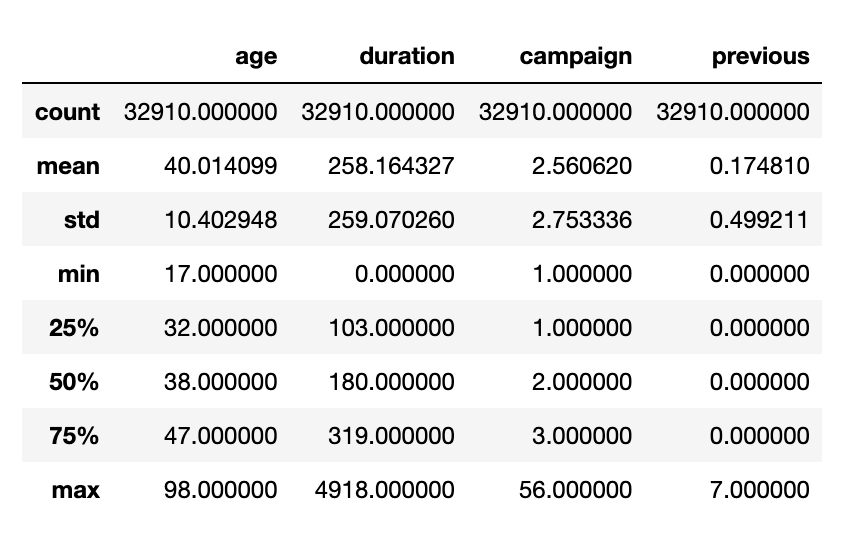
EDA-

1. Many categorical values. Few continuous values.

Table

Description automatically generated

1. Percentile distribution of data-( duration may have some outliers).



3. positive class is approx 10 %, imbalanced data treatment can be done.

Graphical user interface, text, application

Description automatically generated

4. Duration has some outliers-( must be treated for regression based models)

Chart, calendar

Description automatically generated

5. 25-32 more chances of saying yes compare to 30-40.

A picture containing chart

Description automatically generated

6. Retired people and students are more prone to say yes.

Timeline

Description automatically generated with medium confidence

7- similarly singles have more chances of saying yes.

8-we may club 3 categories of basic education-

Chart

Description automatically generated

9- similarly there are unknowns in default, housing and loan variables, should be investigated ( Data Quality Issue)

10-contacting on telephone is less likely to take product.

11- relative acceptances are more in Mar-April and Sep-Oct.

12- High call duration has more chances of product acceptance-

A picture containing diagram

Description automatically generated

13- success rate is more if poutcome was success.

14- correlation analysis-

A picture containing table

Description automatically generated

14- most of the call are not lasting long ( long calls are leading to conversion) and age group is 25-45 mostly whereas retired have most chances of conversion.

### Models Building and Feature Engineering-

1. Running the basic catboost model

* iterations=1000, learning\_rate=.001, depth=8

train-

Table

Description automatically generated

Test-

Table

Description automatically generated with medium confidence

1. Exhaustive Grid Search for Catboost-

*grid = {'max\_depth': [8,10,14],'n\_estimators':[2000, 1500],'l2\_leaf\_reg':[5,15,30,100], 'border\_count':[10,20,50,254]}*

*cv = 3,train\_size=0.8*

*Chart

Description automatically generated*

Performance on test data-

Table

Description automatically generated

1. Variable Importance- from Catboost

Chart, bar chart

Description automatically generated

1. Variable importance – Shap Values

Chart, bar chart

Description automatically generated

1. Shap Values force plot- effect of different values of a variables

Chart

Description automatically generated

6- one hot encoding and data scaling of input data with standard scaling and running multiple classifier algos using cross\_val\_score from sklearn–

Text

Description automatically generated

After doing hyperparameter tuning of LR-

The mean accuracy of the model is: 0.9070987938495535

* *Model can be improved further by doing error analysis.*

*Other things to be checked-*

* *other encoding techniques ordered encoding for education, target encoding.*
* *Class imbalance treatment.*
* *Outlier removal from few variables.*
* *Merging few cardinal values like education*
* *Data binning for age.*

### Building Deep Learning Models

1. Gaining initial understanding of architecture- ( Adam optmiser)

Text

Description automatically generated

Chart, line chart

Description automatically generated

Simple overfitting the model-

Adding dropout layer-

Graphical user interface, text, application, email

Description automatically generated

Chart, line chart

Description automatically generated

1. Building simpler NN models-

Text

Description automatically generatedA picture containing chart

Description automatically generated

Chart

Description automatically generated

Accuracy- 90 ( 0.9062701463699341 sth)

1. Keras Tuning to find best parameters-

Text

Description automatically generated

Training best model -patience = 10 ( 0.8940244913101196- Accuracy)

Chart, line chart

Description automatically generated

Training best model patience = 50

Chart, line chart

Description automatically generated

Accuracy is less than 90%.

Things to be checked-

*Embedding layers can be used improve on DL models. More Hyperparamets tuning can be done.*

*More architecture can be explored.*