

Marketing Report

GROUP 5

Nguyen Minh Anh

Ha Long Giang

Pham Duc Hieu Thanh

Nguyen Linh Chi

Table of Contents

1 Introduction

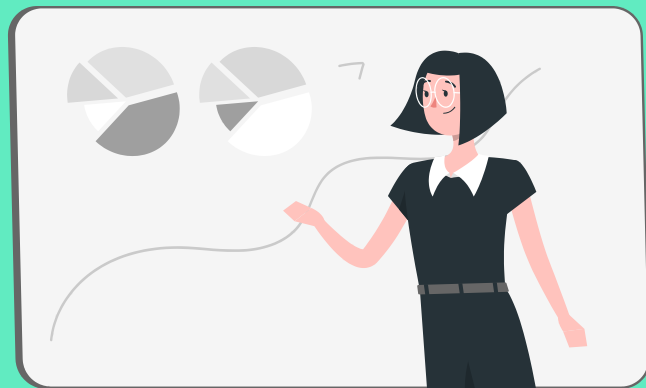
3 Model

2 Insight

4 Conclusion

1

INTRO



Overview

Number of customer

41K

Customers

Total conversions

4640

Conversion

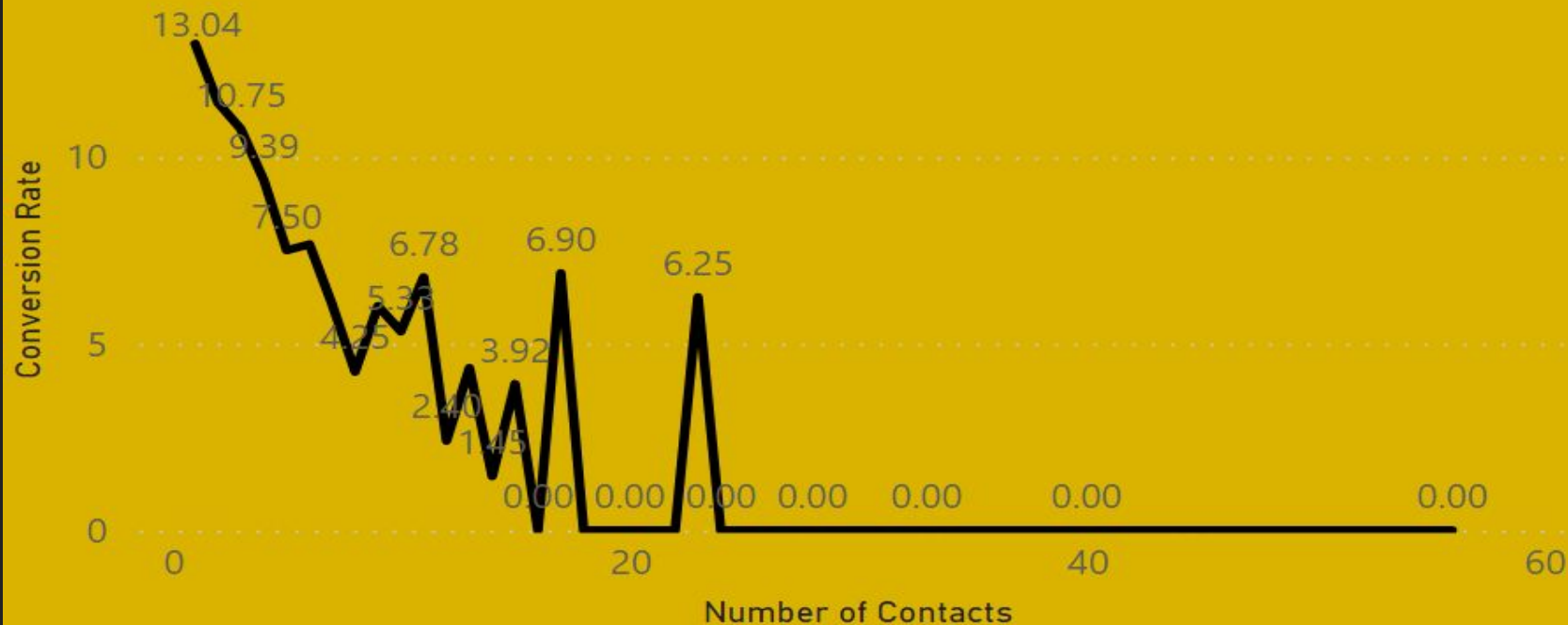
Conversion rate
11.27%



237.95%

Compare with pre campaign

Conversion Rates by Campaign



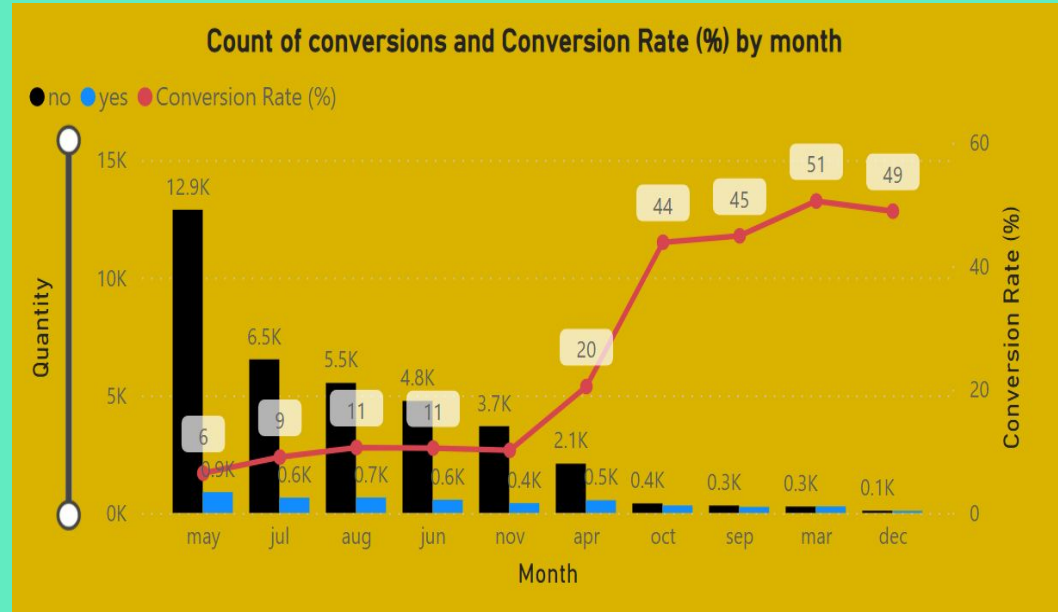
2

INSIGN

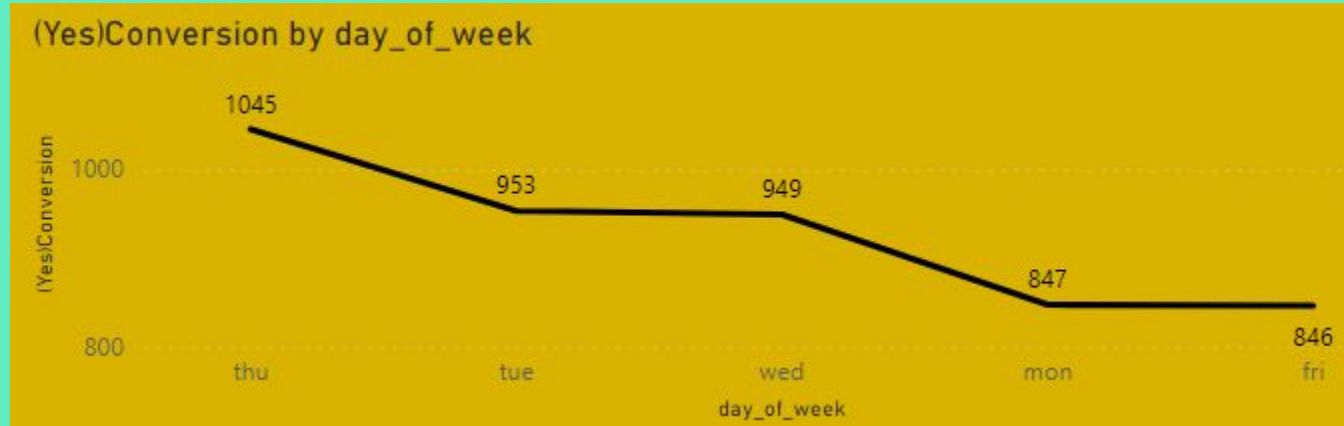


Conversion rate by month

- The campaign lasts from April to September. It can be seen that the campaign reaches the most customers(13.8k) and brings the most conversions(0.9k) in May and then the months of June, July, and August.
- Although reaching many customers, the effectiveness of these months is not high, with the rate of 6%, 9%, 11%, 11% respectively.
- In contrast to May, which is the months of March, 9, October, December with the number of customers reaching only from 0.2k to 0.7k customers, the effect is very good with the highest rate up to 51%.

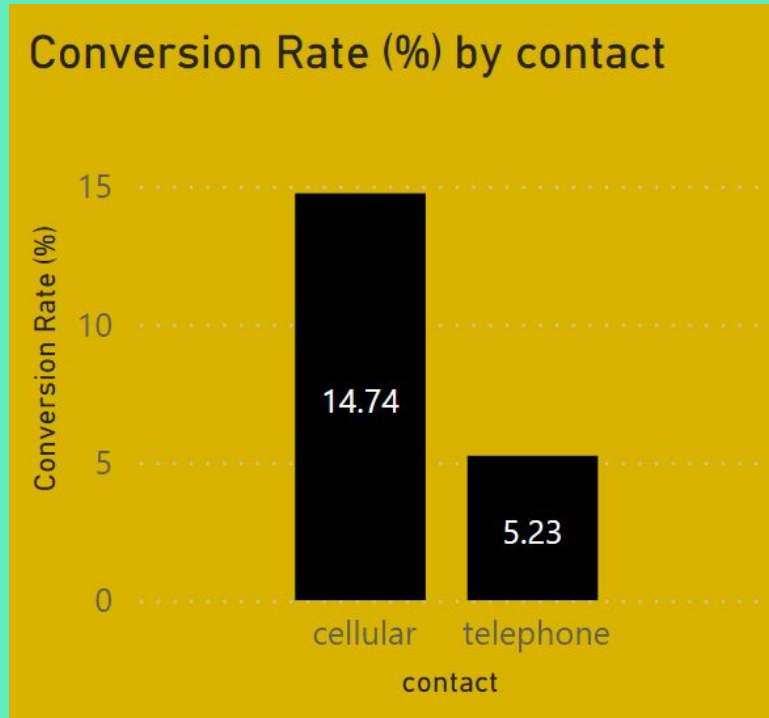


Which day we should contact our customer?



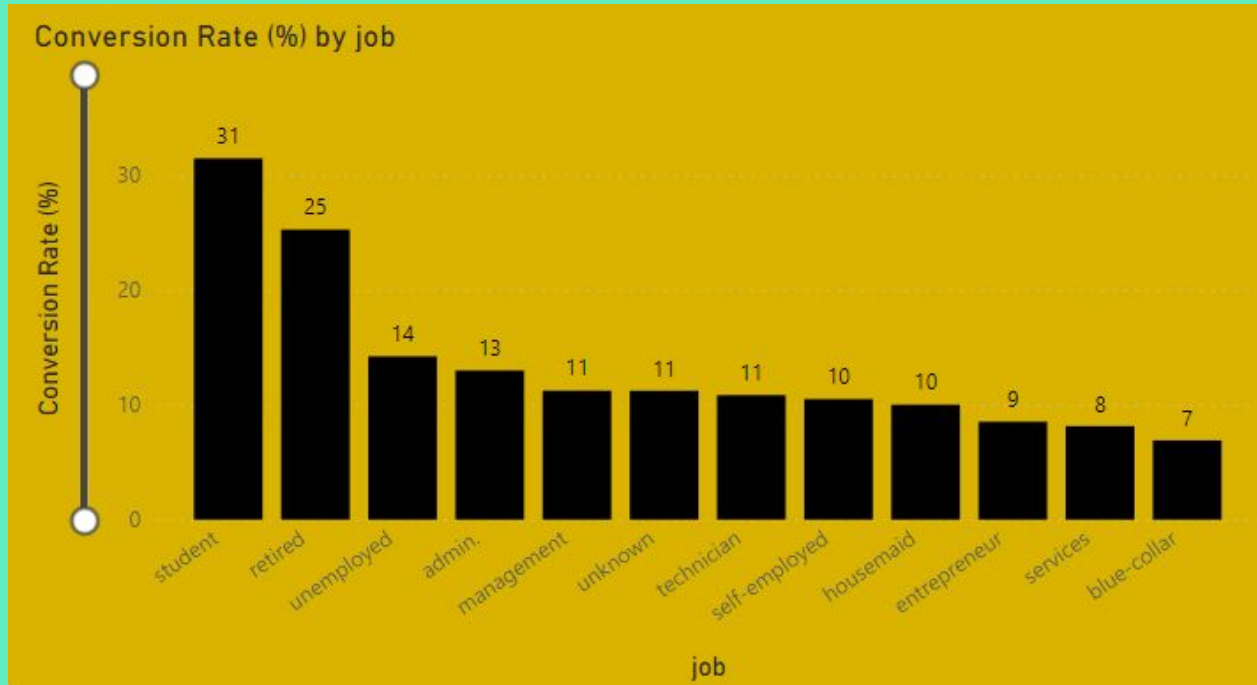
The number of customers agreed focuses heavily on the middle of the week, especially Thursday (1045 customers). At the same time, it gradually decreases towards the begin and the end of the week.

Kind of contact and conversion rate



We can see a clear effect when we use mobile with a conversion rate of 14.74%, nearly 3 times higher than using a phone (5.37%).

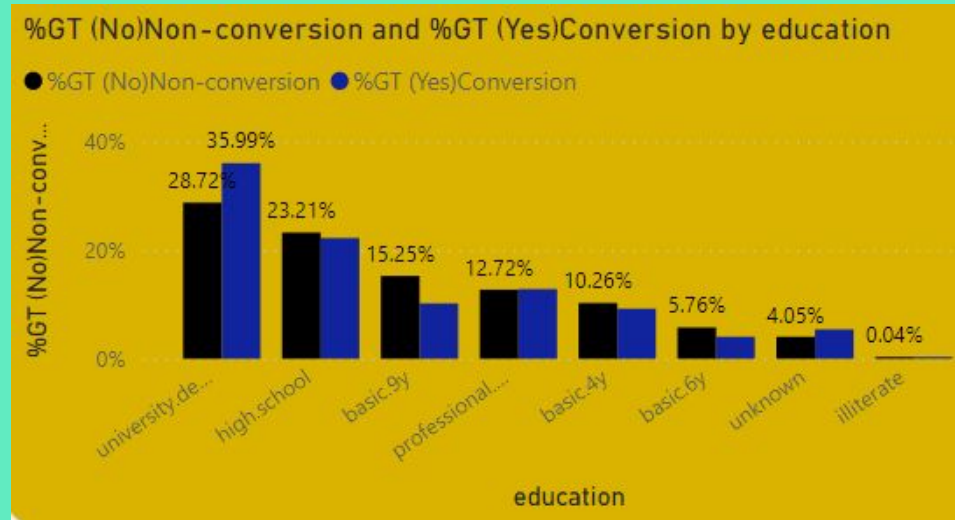
Job titles affection



- It can be seen that the percentage of students and retirees is higher than that of 32% and 25% respectively.
- Right after that is the unemployment group, with a higher rate than the rest of the occupational groups, 13%.

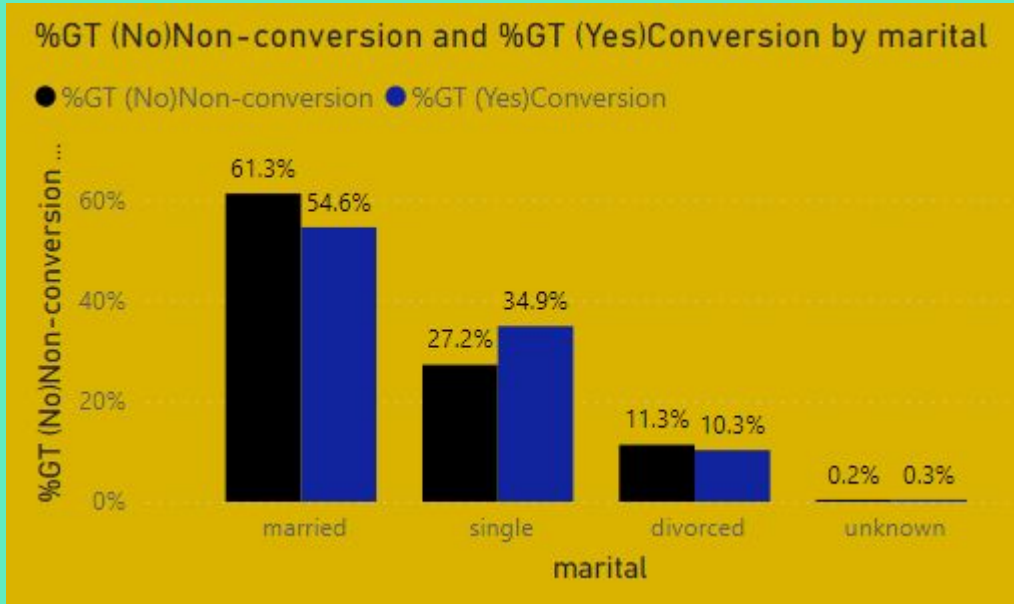
Education and conversion rate

- According to data from education, the university group has the largest percentage in both charts with a conversion rate of 35.99% and a non-conversion rate of 28.72%.



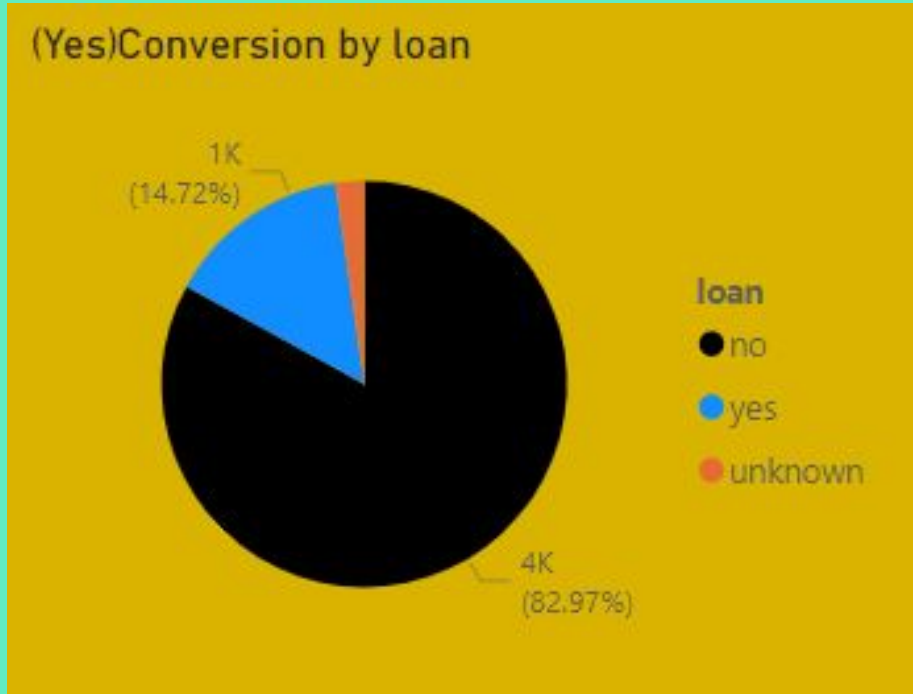
- This proves that the campaign can effectively reach the university group.
- So are the high school, basic 9y and professor groups.

Does marital affect conversion rate?



From the two charts, we can see that the single and married customer group accounted for the majority with 34.9%, 54.57% of conversion rate and 27%, 61.28% of non-conversion rate, respectively.

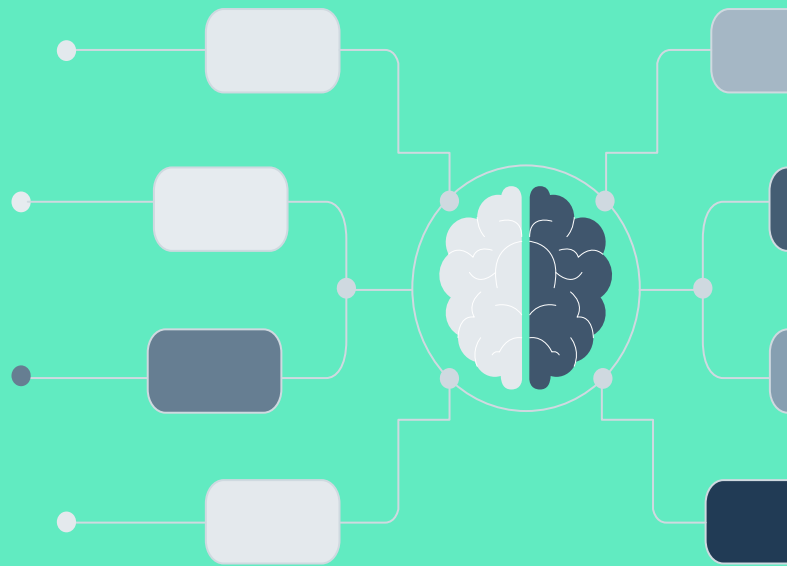
Is conversion rate higher in the group having loan?



Most of the customers who agree are the ones with no debt(82.97%).

3

MODEL



Fillna

| | Features | Nbr_Null | Pct_Null |
|---|-----------|----------|----------|
| 0 | age | 0 | 0.00 |
| 1 | job | 330 | 0.80 |
| 2 | marital | 80 | 0.19 |
| 3 | education | 1730 | 4.20 |
| 4 | default | 8596 | 20.88 |
| 5 | housing | 990 | 2.40 |
| 6 | loan | 990 | 2.40 |

1. Job: fillna by mode value

2. marital : groupby age

With person under 33 years old =>

single, over 33 years old => marital

6. education : fillna by job_rank

job_rank = {

 'unemployed': 0, 'student' : 1, 'retired' : 1,

 'housemaid': 1,

 'blue-collar' : 2, 'admin.' : 2, 'services' : 2,

 'technician' : 2,

 'management': 3, 'self-employed' : 3,

 'entrepreneur' : 3

}

With job_rank == 0 => education = 'illiterate'

With job_rank == 1 => education = 'basic.6y'

With job_rank == 2 => education = 'high.school'

Else => education = 'university.degree'

Fillna

3.housing : fillna by one hot encoding (job_rank, marital, age_group)

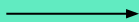
| job_rank_0 | job_rank_1 | job_rank_2 | job_rank_3 | marital_divorced | marital_married | marital_single | age_group_middle_age | age_group_old | age_group_young | ghepcode_housing |
|------------|------------|------------|------------|------------------|-----------------|----------------|----------------------|---------------|-----------------|------------------|
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0010100001 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1000100001 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1000100100 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1000100100 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1000100001 |

4. loan : fillna by onehot encoding (job_rank, housing, age_group, marital)

5. default : fillna by onehot encoding ('job_rank', 'housing', 'loan', 'age_group', 'marital')

Grouping column

| duration |
|------------|
| <120 |
| (120, 140) |
| else |



| duration_group |
|----------------|
| less_2mins |
| 2mins_4mins |
| more_4mins |

| campaign |
|----------|
| >=3 |
| 2 |
| 1 |



| campaign_group |
|----------------|
| more_than_two |
| two |
| one |

| previous |
|----------|
| 0 |
| 1 |
| 2 |
| >=3 |



| previous_group |
|-----------------|
| zero |
| one |
| two |
| more_than_three |

| pdays |
|-------|
| 999 |
| |



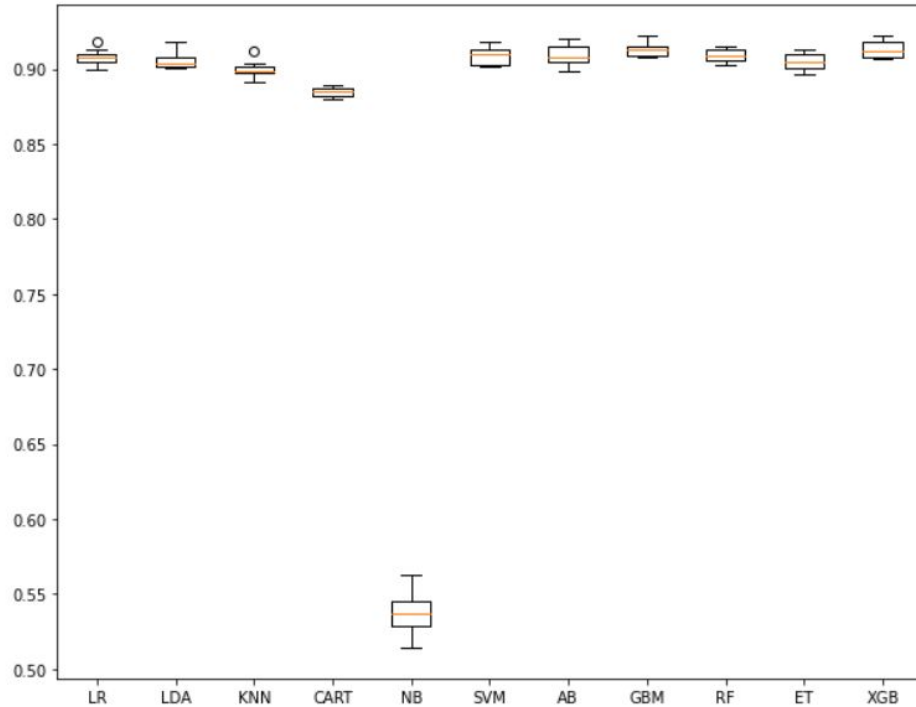
| pdays_group |
|-------------|
| more |
| less |

| y |
|-----|
| yes |
| no |



| y |
|---|
| 1 |
| 0 |

Model: XGBoost



Accuracy in training set : 0.9174262121774908

Accuracy in validation set: 0.9130645944633318

```
[[5317 144]
```

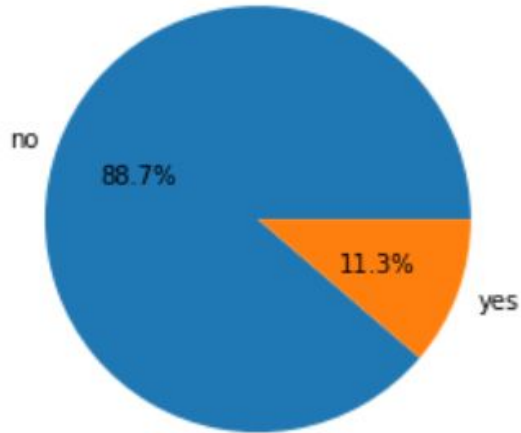
```
[ 393 323]]
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.93 | 0.97 | 0.95 | 5461 |
| 1 | 0.69 | 0.45 | 0.55 | 716 |
| accuracy | | | 0.91 | 6177 |
| macro avg | 0.81 | 0.71 | 0.75 | 6177 |
| weighted avg | 0.90 | 0.91 | 0.90 | 6177 |

Precision, recall of class 1 are pretty low
Problem: imbalance data

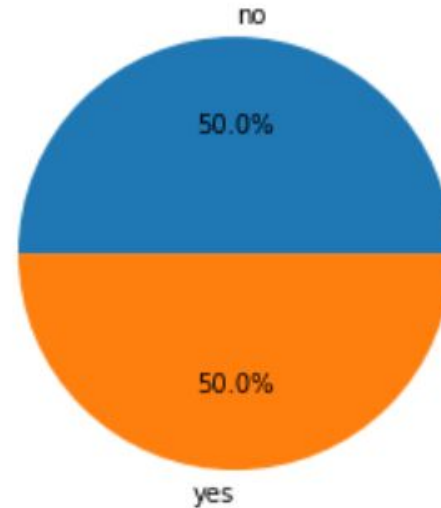
Raw Data

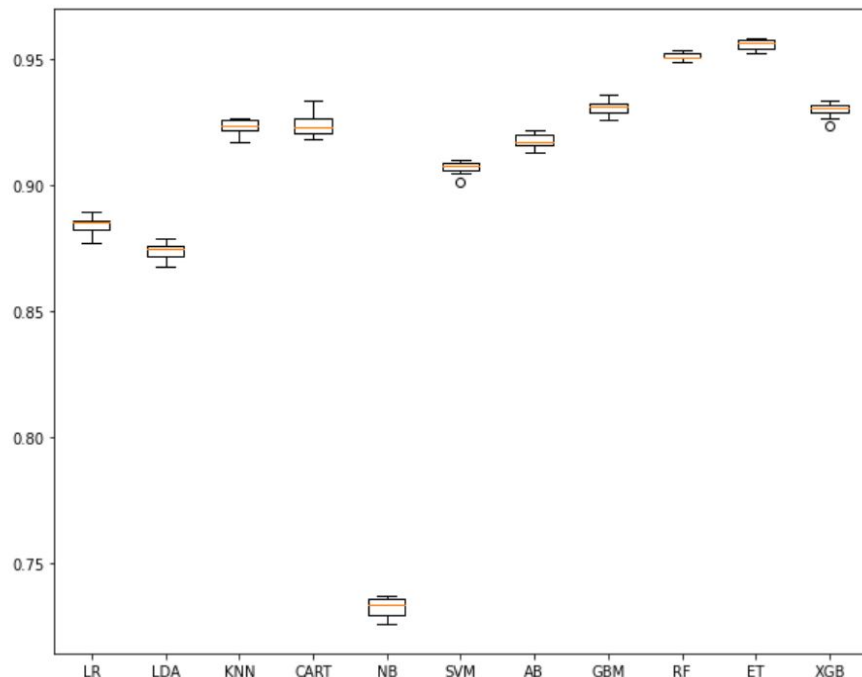
```
no      36537  
yes      4639  
Name: y, dtype: int64
```



Overcoming Class Imbalance using SMOTE Techniques

```
0      36537  
1      36537  
Name: y, dtype: int64
```





Best model

```
Accuracy in training set : 1.0
Accuracy in validation set: 0.9575094081423196
[[6838 470]
 [ 151 7156]]
```

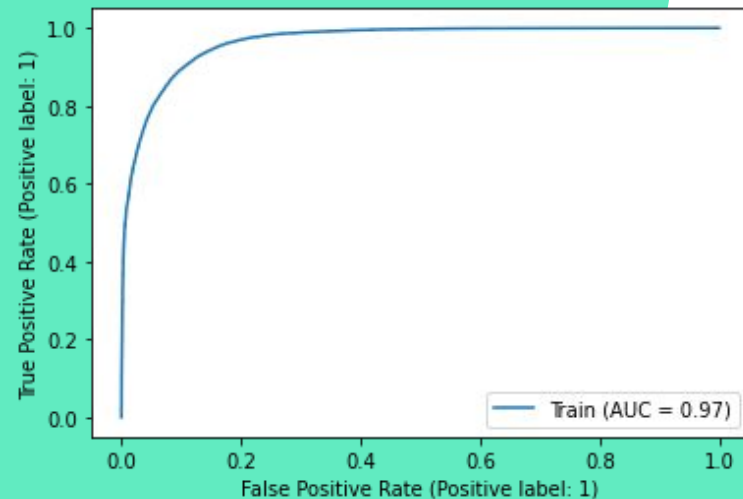
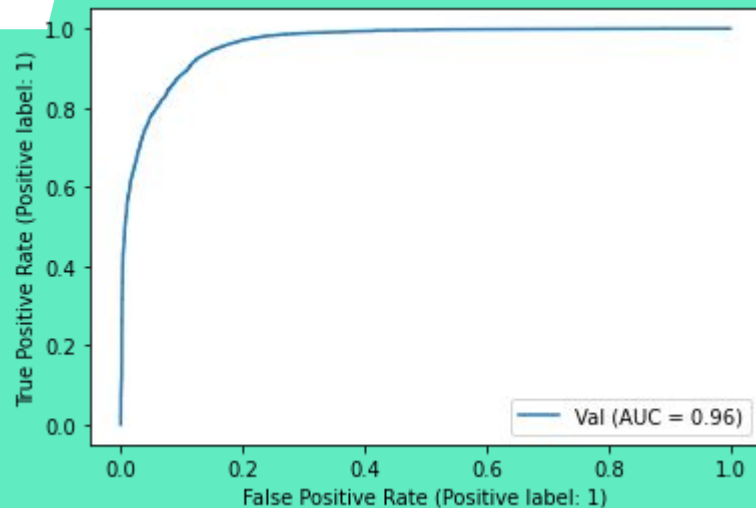
| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.98 | 0.94 | 0.96 | 7308 |
| 1 | 0.94 | 0.98 | 0.96 | 7307 |
| accuracy | | | 0.96 | 14615 |
| macro avg | 0.96 | 0.96 | 0.96 | 14615 |
| weighted avg | 0.96 | 0.96 | 0.96 | 14615 |

Model: XGBoost

```
Accuracy in training set : 0.9271113087805128
Accuracy in validation set: 0.9278823126924393
[[6539 769]
 [ 285 7022]]
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.96 | 0.89 | 0.93 | 7308 |
| 1 | 0.90 | 0.96 | 0.93 | 7307 |
| accuracy | | | 0.93 | 14615 |
| macro avg | 0.93 | 0.93 | 0.93 | 14615 |
| weighted avg | 0.93 | 0.93 | 0.93 | 14615 |

Logistic regression model

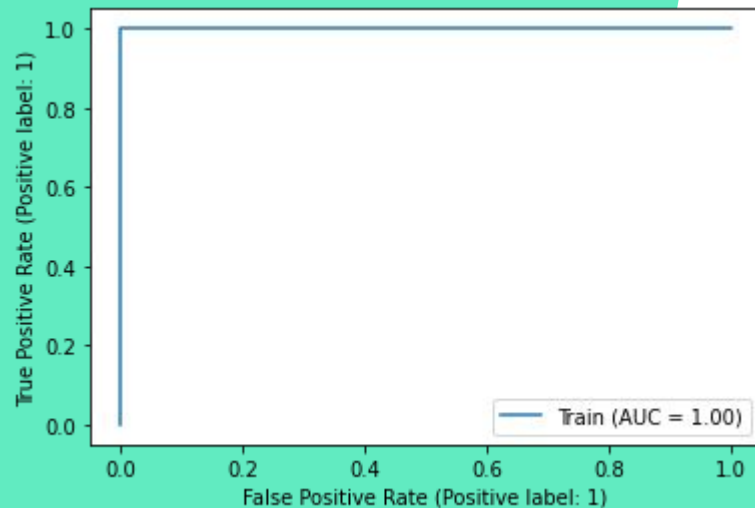
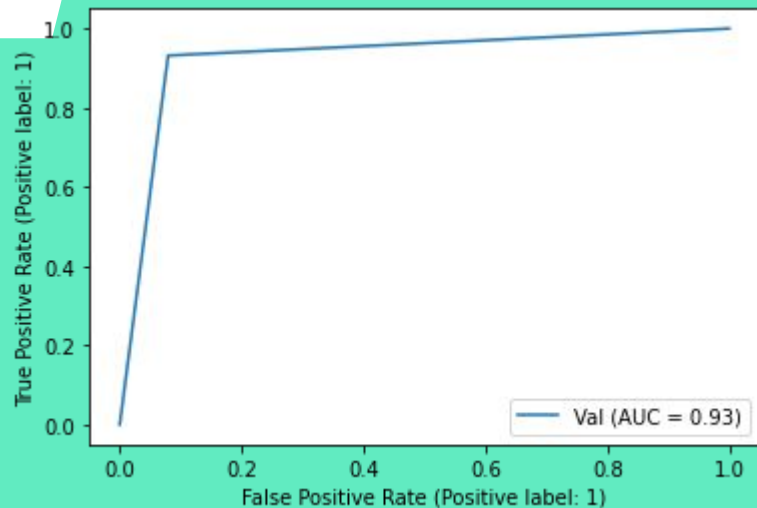


```
Accuracy in training set : 0.8966969481508968
Accuracy in validation set: 0.8927126847271913
      precision    recall  f1-score   support

     0       0.89       0.90       0.89       7308
     1       0.90       0.89       0.89       7307

   accuracy                0.89       14615
  macro avg       0.89       0.89       0.89       14615
 weighted avg     0.89       0.89       0.89       14615
```


Decision Tree

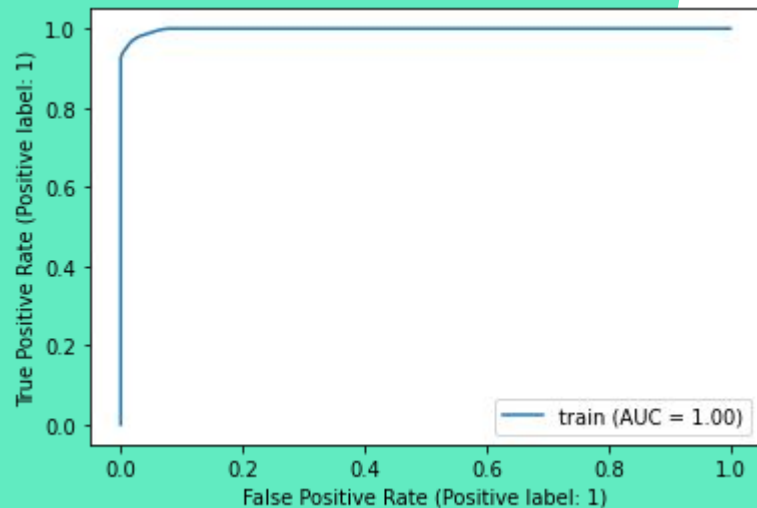
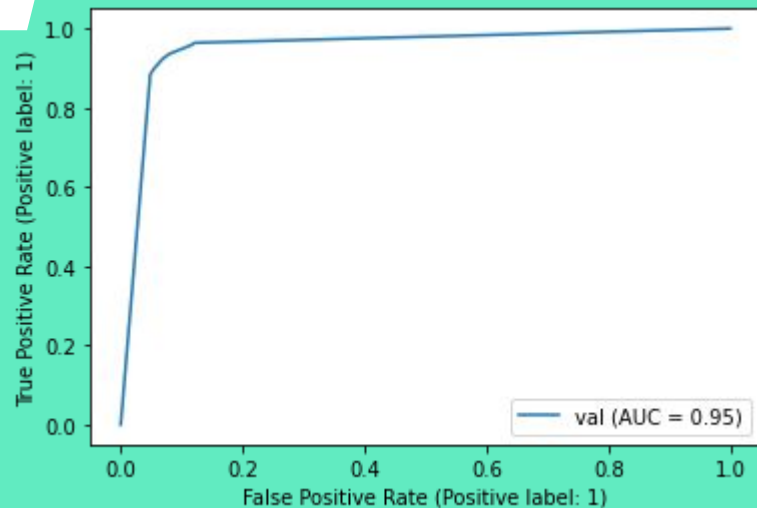


```
Accuracy in training set : 1.0
Accuracy in validation set: 0.9262405477678504
      precision    recall  f1-score   support

     0       0.93      0.92      0.93       7308
     1       0.92      0.93      0.93       7307

   accuracy          0.93          14615
  macro avg       0.93      0.93      0.93      14615
 weighted avg     0.93      0.93      0.93      14615
```

Tuning Decision Tree



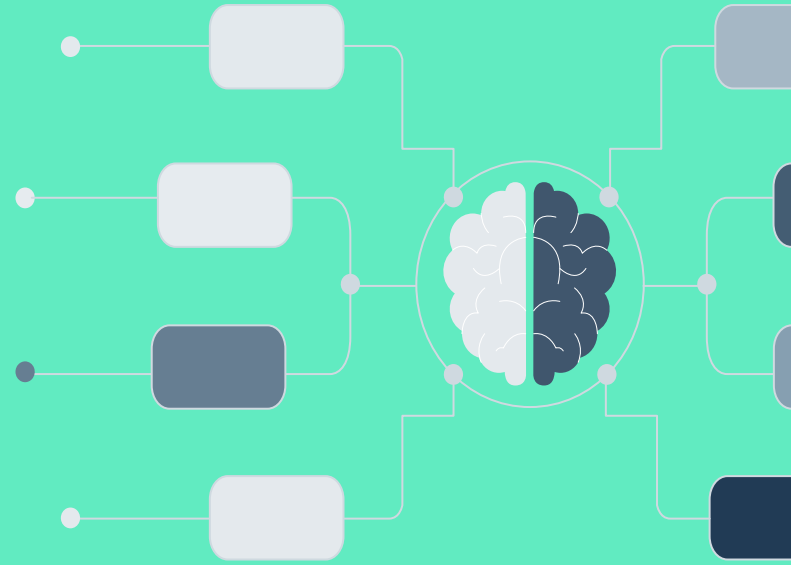
```
Accuracy in training set : 0.975470053636527
Accuracy in validation set: 0.9281559756788988
      precision    recall  f1-score   support

     0       0.93      0.93      0.93     7308
     1       0.93      0.93      0.93     7307

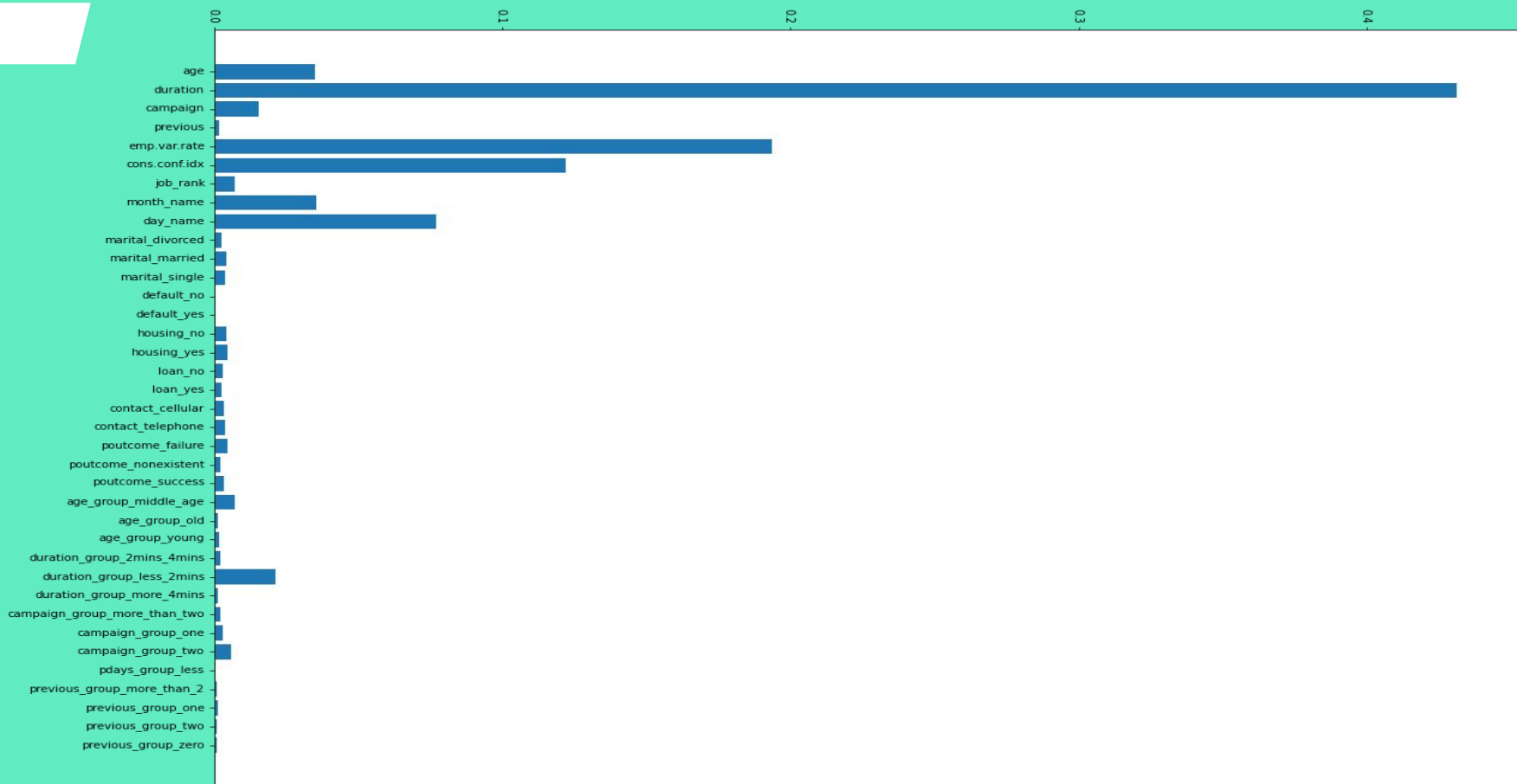
   accuracy          0.93     14615
  macro avg       0.93      0.93      0.93     14615
 weighted avg       0.93      0.93      0.93     14615
```

4

CONCLUSION

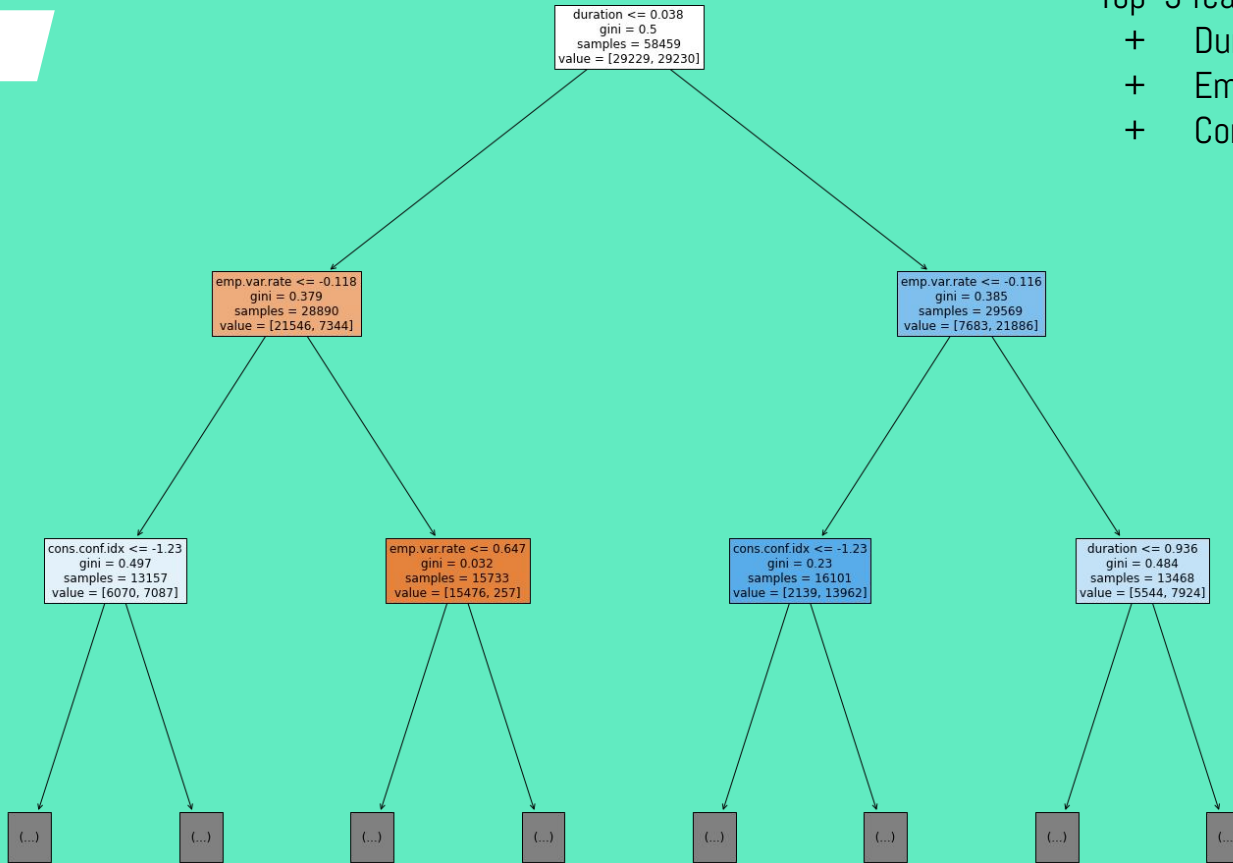


Decision Tree Features Importance



Top 3 features ảnh hưởng đến model :

- + Duration
- + Emp.var.rate
- + Cons.con.idx



Full tree with top 5 features importance model

