

WHO IS OUR CUSTOMER?

Ha Minh Duc
Mai Le Hoang
Luu Xuan Nghia
Dinh Thi Ha Phuong

Group 11

Mar 2022





Agenda

1

Understanding problem
statement

2

EDA and Insights

3

Building model and
Evaluation

4

Conclusion and
Recommendation

Understanding problem statement





About problem

Situation

- Direct marketing campaigns of a banking institution.
- Based on phone calls (cellular, telephone)

Goals

- Predict if the client will subscribe (yes/no) a term deposit
- Finding implications and insights from data
- Recommendation for Marketing team



Original features of dataset

Customer data

- age
- job
- marital
- education
- default
- housing
- loan

Campaign data

- contact
- month
- day_of_week
- duration
- campaign
- pdays
- previous
- poutcome

Macro data

- emp.var.rate
- cons.price.idx
- cons.conf.idx
- euribor3m
- nr.employed



EDA and **Insights**

Preprocessing



✓
4639

Filter
→
duplicated data rows



✗
36537

Total customer: **41176**



Conversion Rate

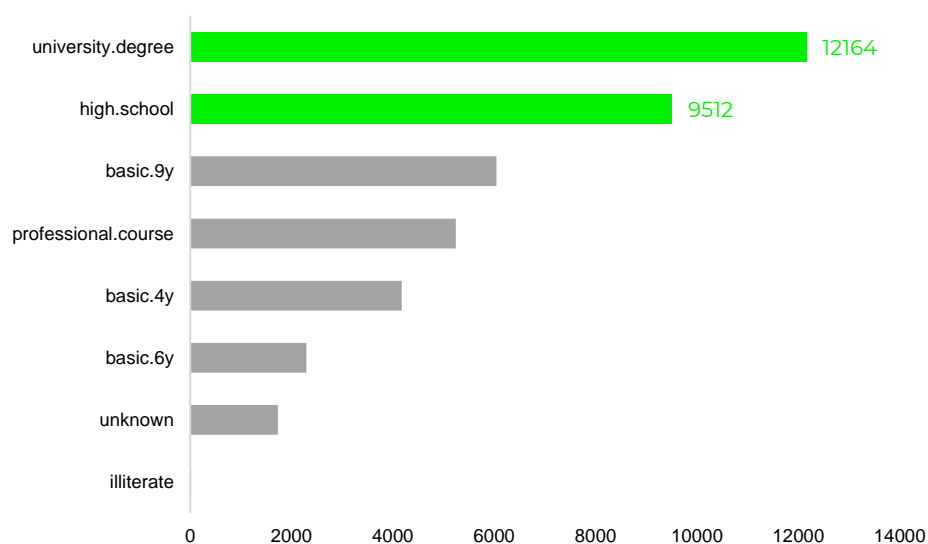
11.27%



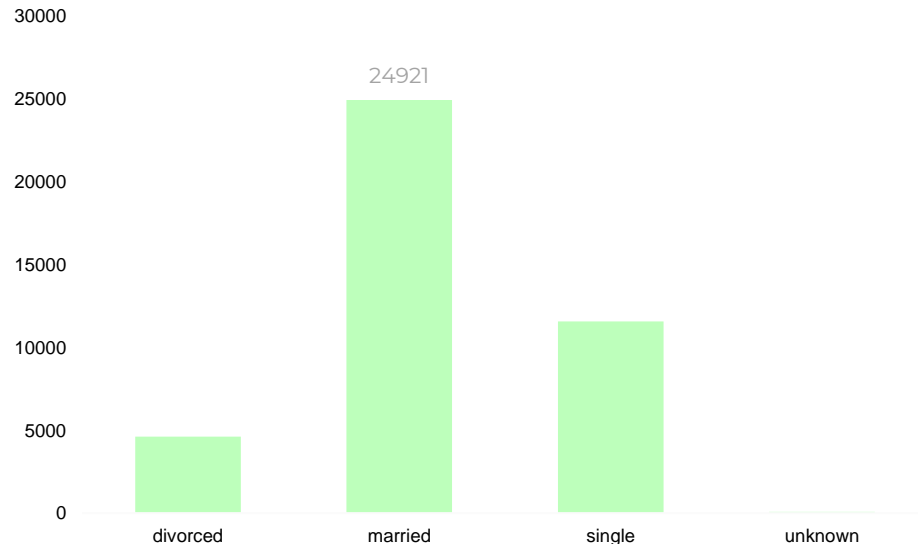
Overview (1)

40 is the average age of our customer

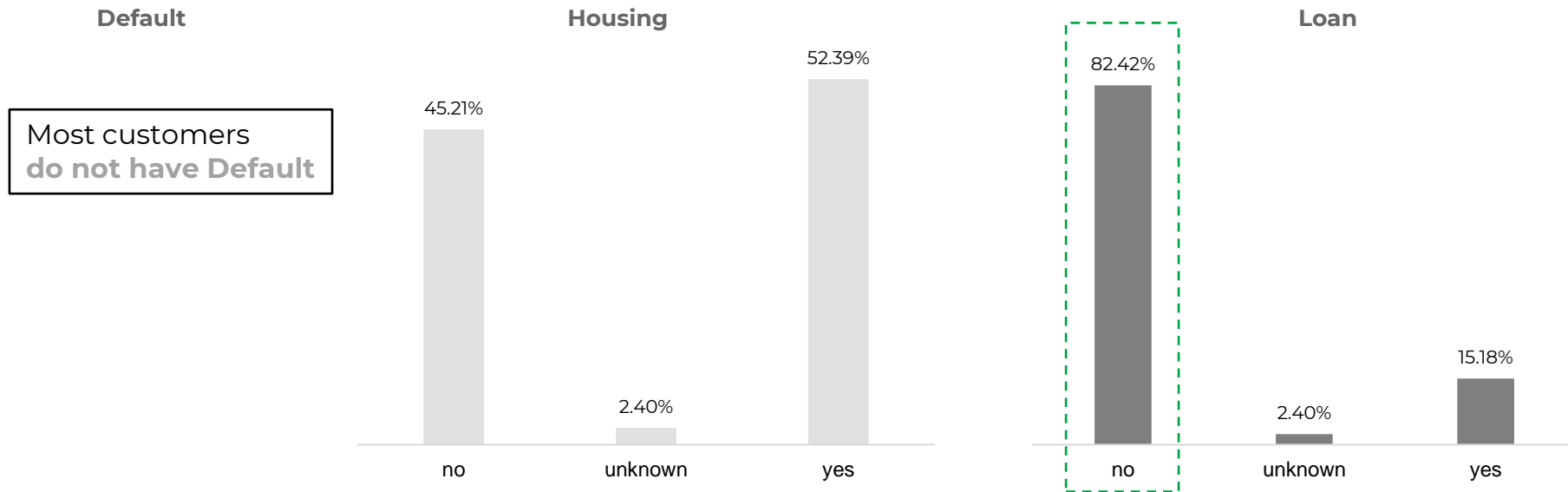
Education



Marital status



Overview (2)

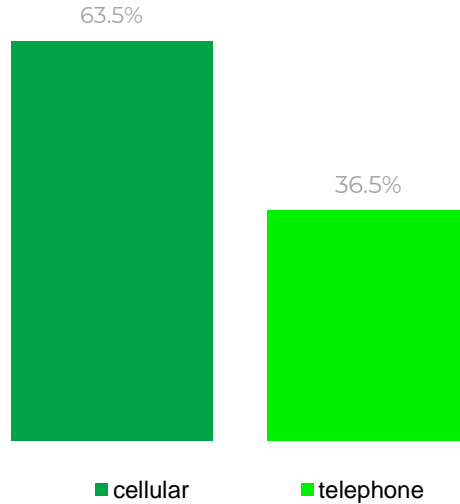


Overview (3)

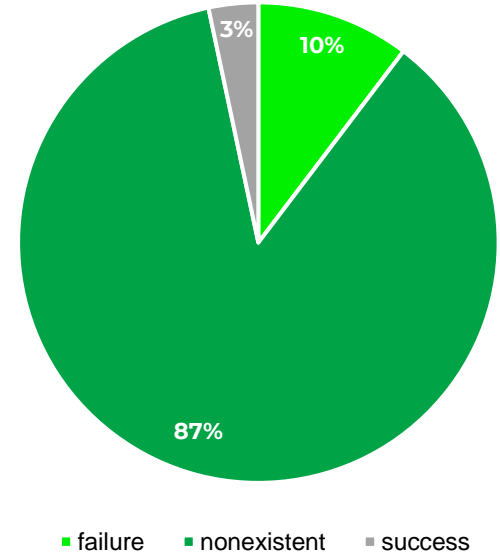
1 customer receives an average of **2.6** calls

Average time of calls is **4.3** minutes

Contact

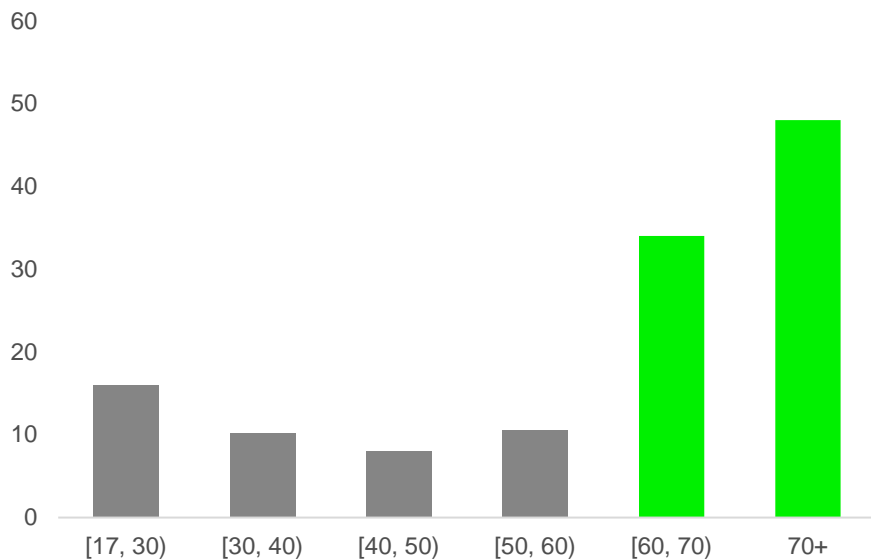


outcome



Does age affect conversion rate?

Conversion rate by Age groups (%)



Lots of noise in the older age group. The CR of 70+ age group fluctuates a lot (this age group data is very limited compared to other age groups)

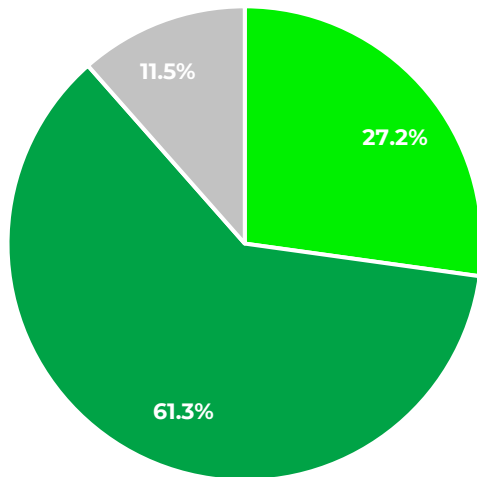
Conversion rate by Age (%)



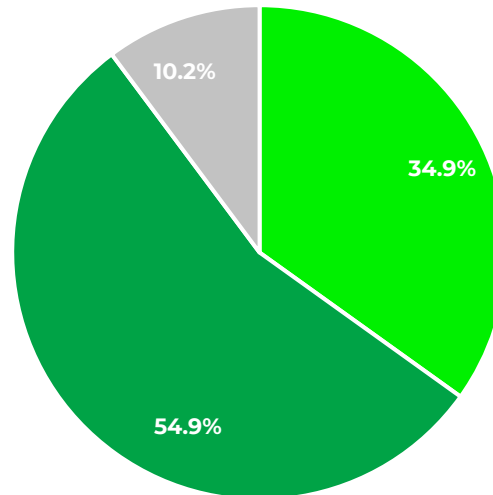
The CR of customers < 30 and > 60 years old is higher than other customer groups, especially customers > 60 years old
=> Age affects the CR

Does marital status affect conversion rates?

single | married | divorced



Non-conversion



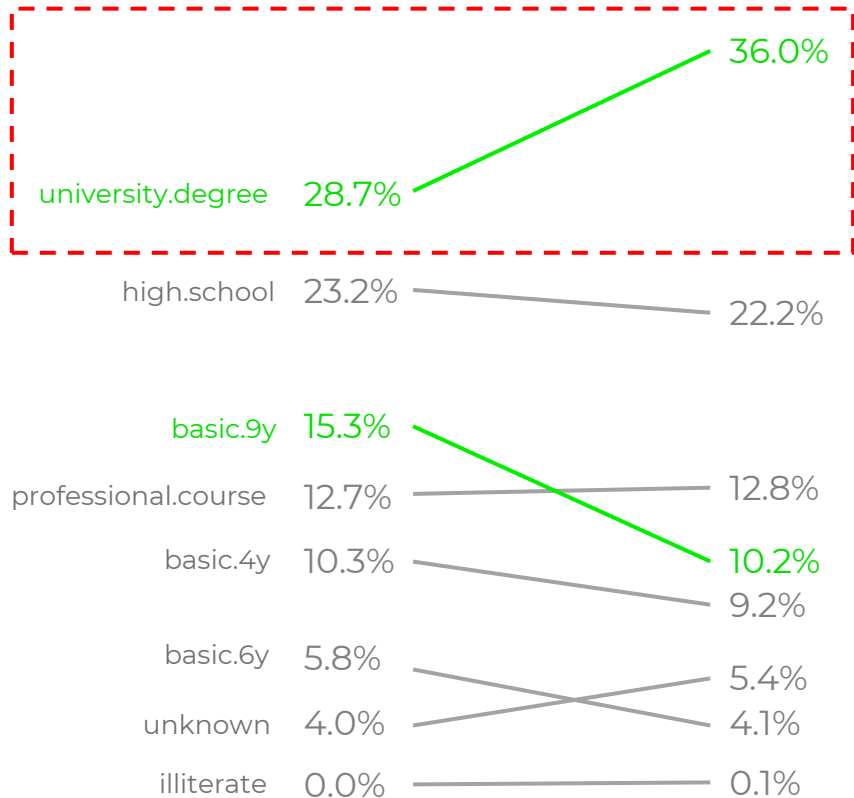
Conversion

In the conversion and non-conversion rates, married customers all account for the highest percentage, then single and divorced
=> Marital status does not affect the CR

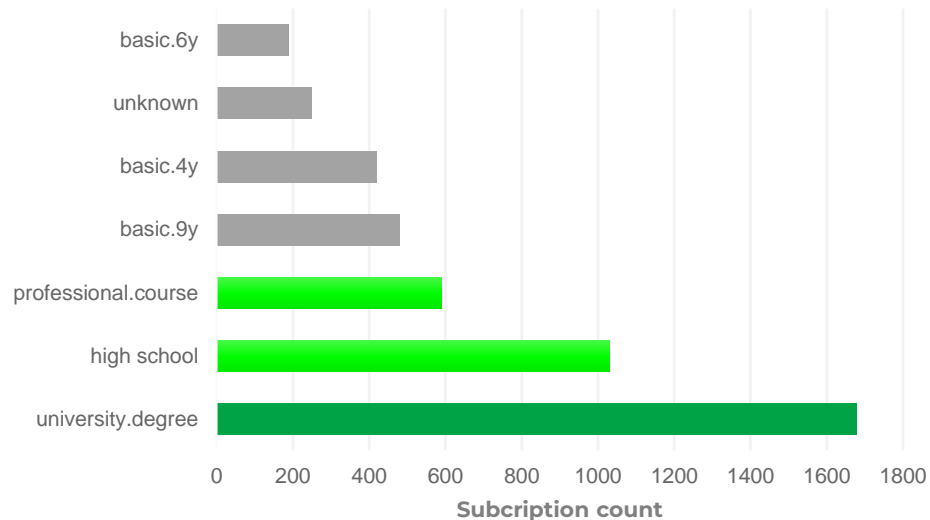
Does education affect conversion rate?

Non-conversion

Conversion



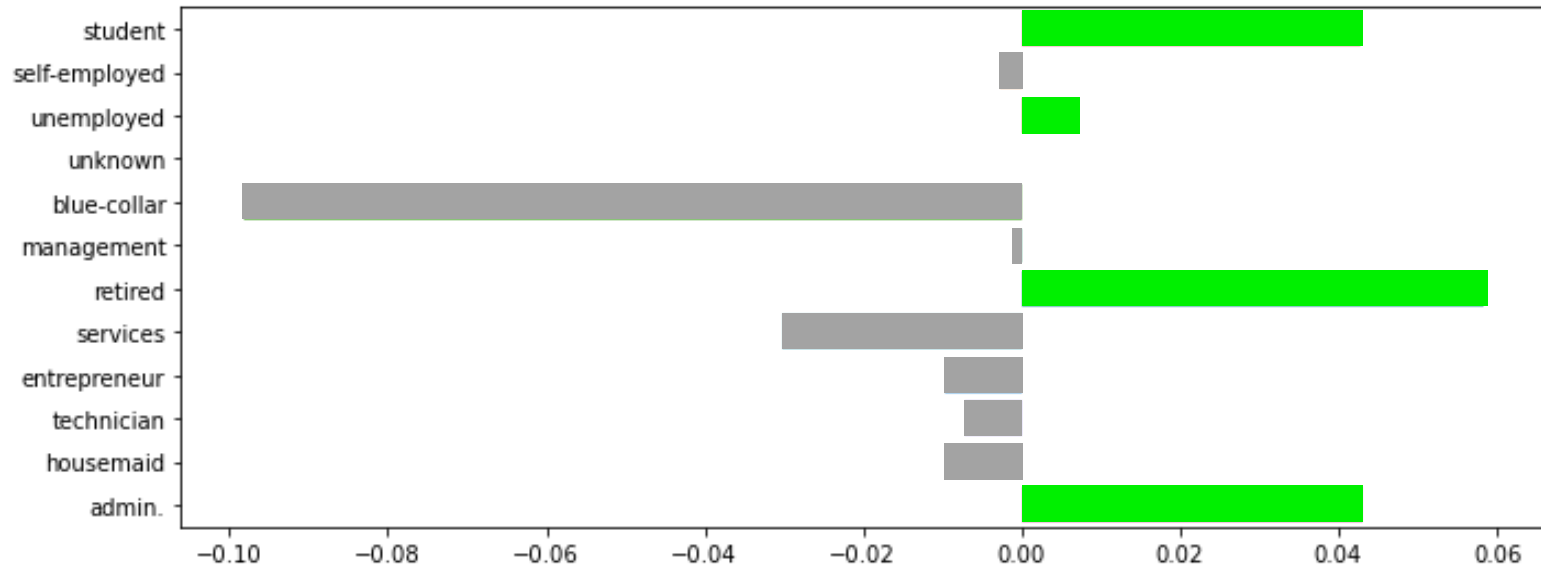
Education division vs. Supcription



- Customers with **university** and **high school** degrees both account for a high percentage of converting and non-converting
- In the group of customers who agree to supcript, the group with a **university degree** accounts for a high percentage, followed by high.school, professional.course

=> People with **high education** seem to have a **high CR**.

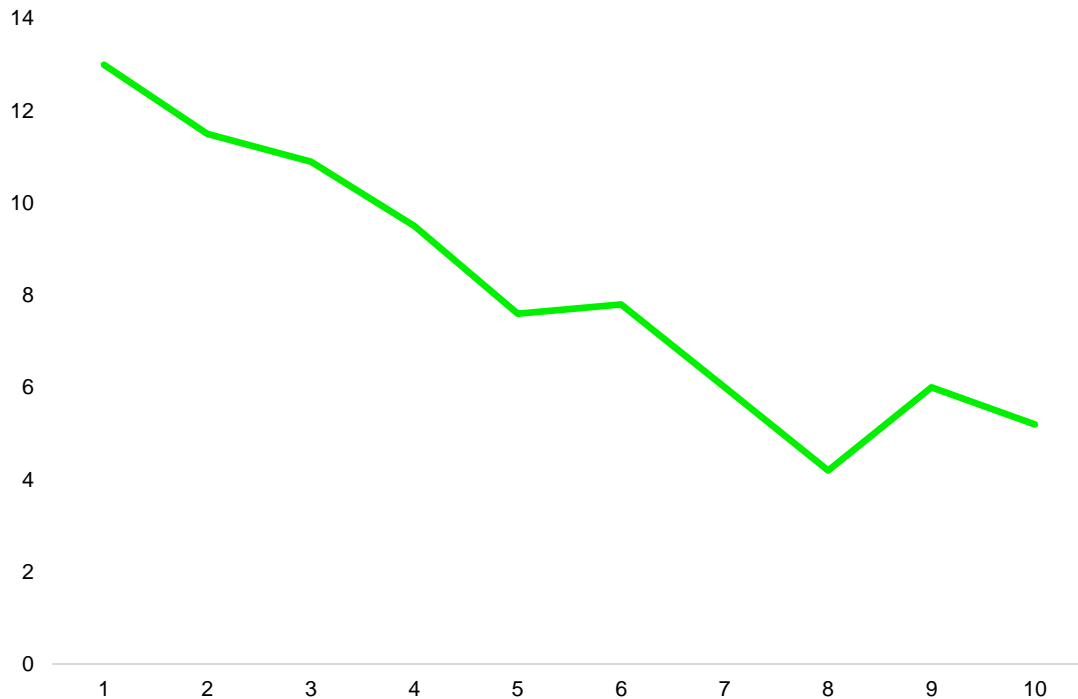
Does job affect conversion rate?



- The **blue-collar** occupation group has an exceptionally high Non-conversion rate when compared to the conversion rate. In addition, the **student** and **retired** groups have a high conversion rate compared to the non-conversion rate
- Retired groups (over 60 years old) tend to agree to open a deposit account to make money from interest

Does the number of contacts in previous campaigns affect the conversion rate?

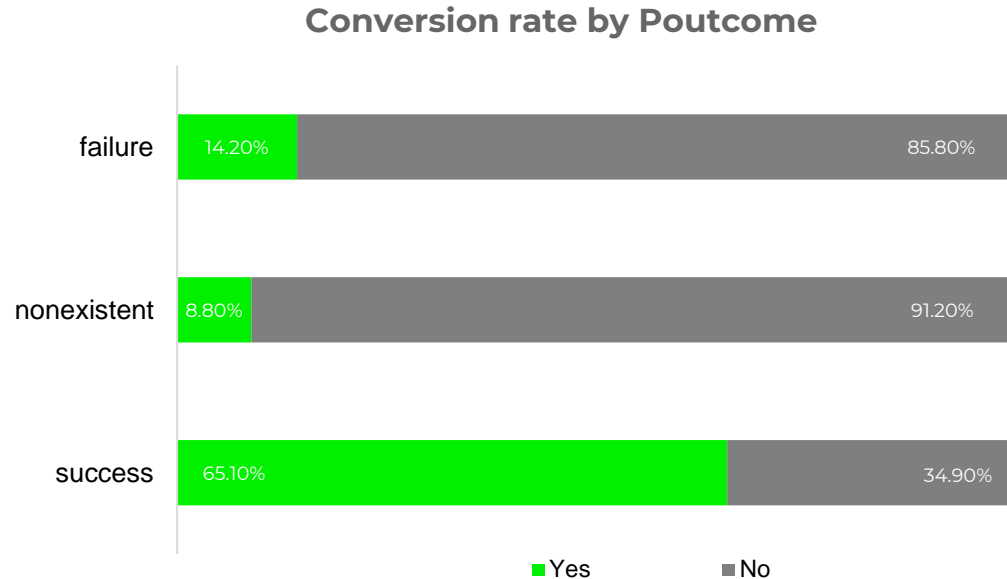
Conversion rate by number of Contacts (%)



The number of contacts of the previous campaign is **inversely proportional** to the CR. The higher the number of calls, the lower the CR.

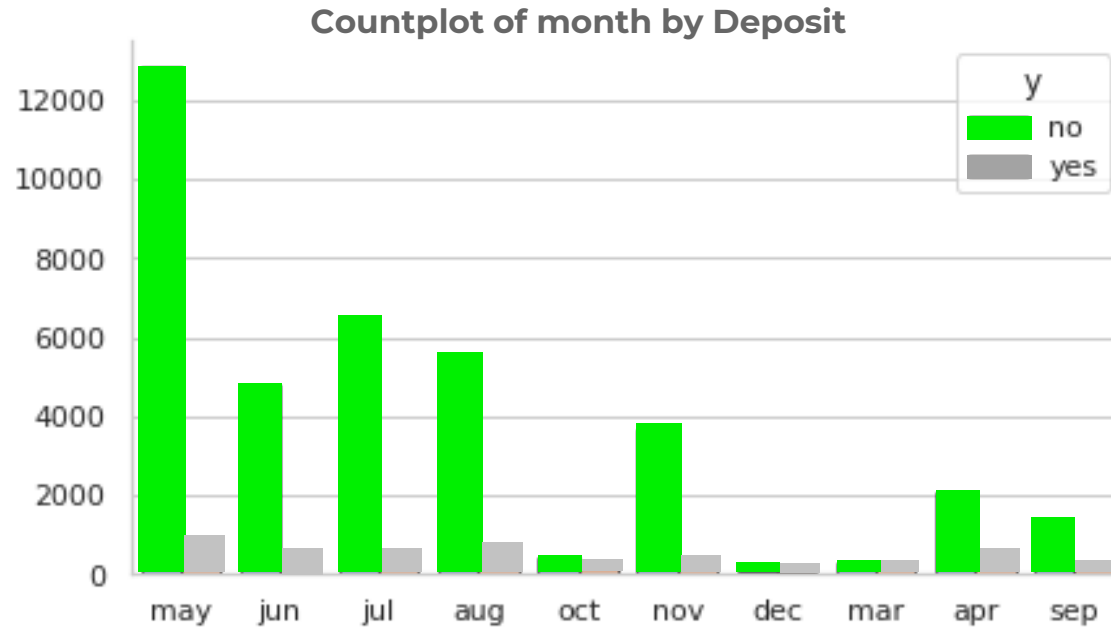
Should not call more than **3** calls for the same customer to save time and effort

Customers from previous campaign



For customers who have successfully agreed to register for savings in the previous campaign, 65.1% continue to agree to register for savings. Existing customers are potential customers to approach marketing in the following campaigns due to the trust and results of the previous registration
=> Boost marketing with a group of previously registered old customers

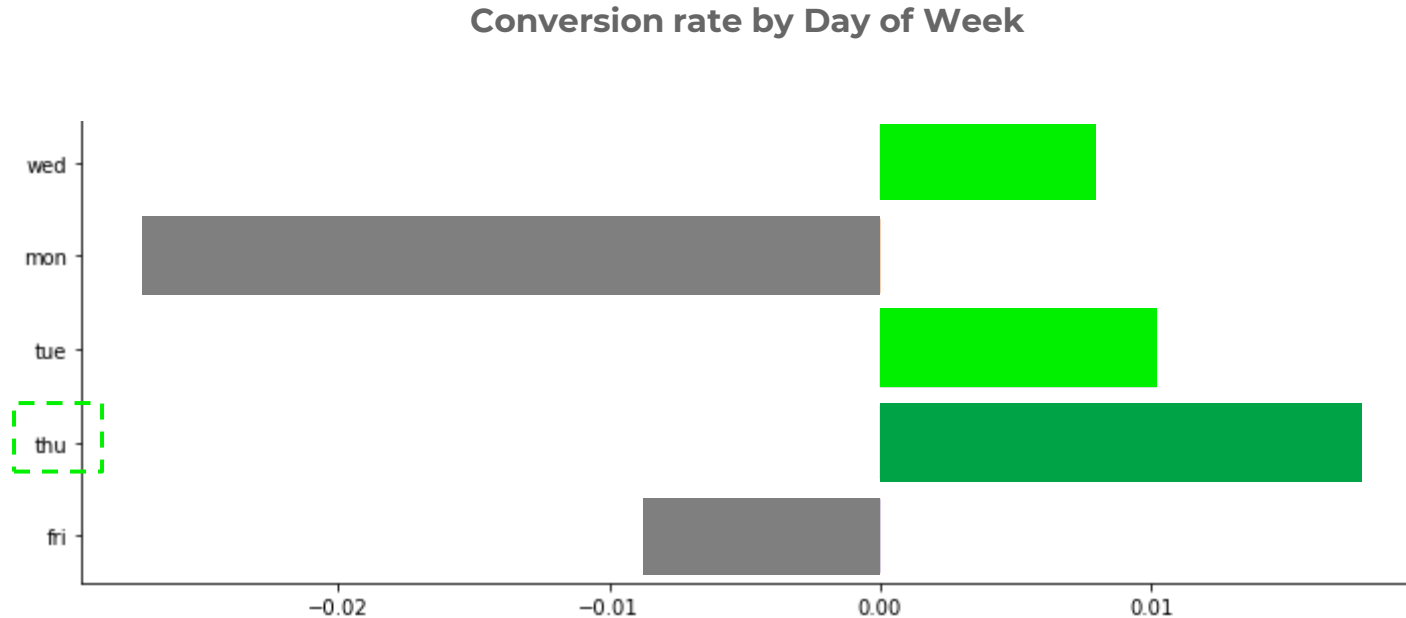
When should we contact customers?



May is the month with the highest rejection rate, followed by June, July, and August. Months with low rejection rate are September, October, December

=> Seasonality: Customers tend to open deposit accounts in **autumn** and **winter**.

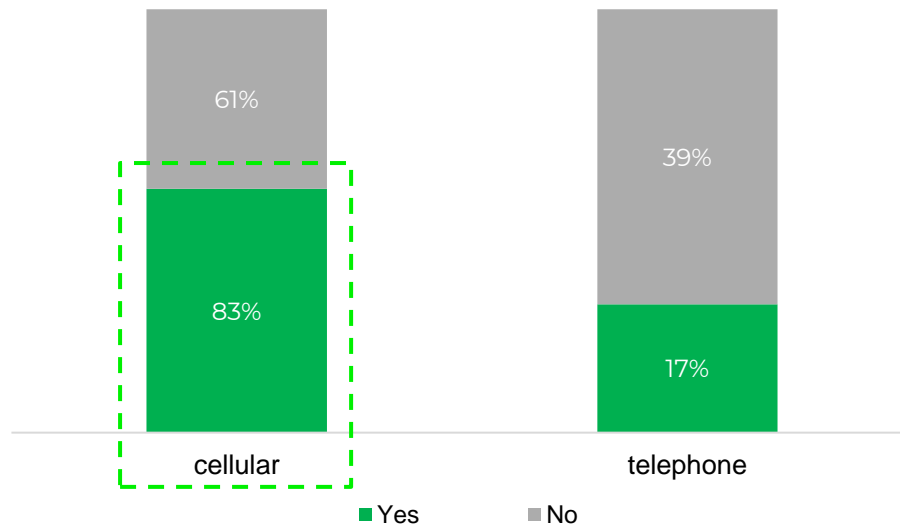
In a week, what day should we contact the customers?



Fridays and Mondays people have negatively responded (say no) as compared to other week days. We should **avoid** these 2 days

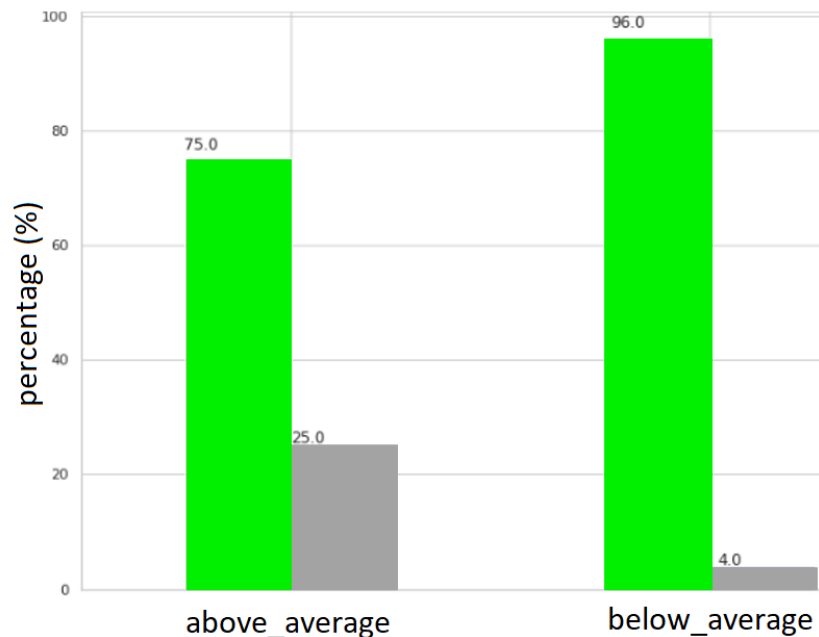
Customers from previous campaign

Conversion rate by contact

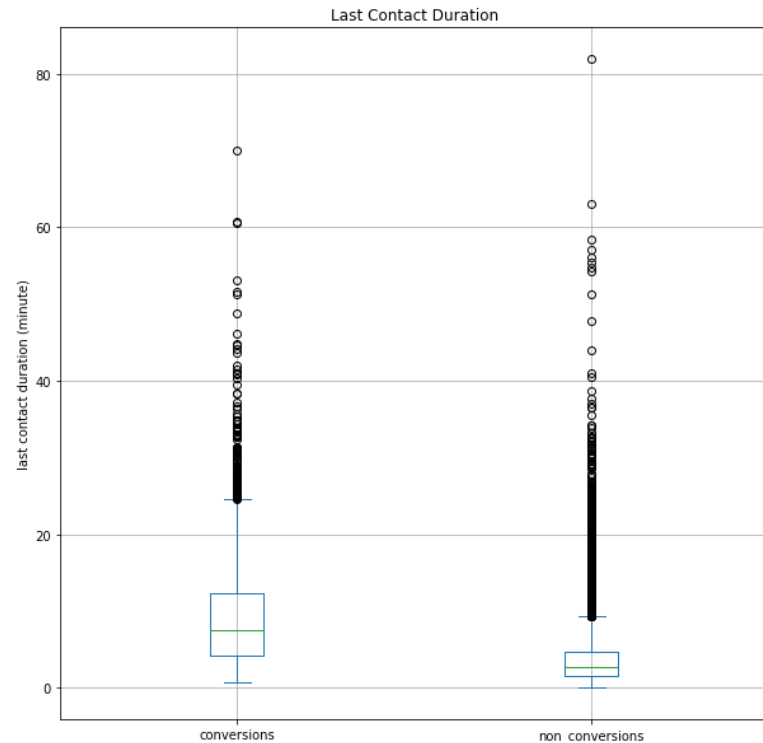


Conversion rate in the group of customers using cellular is nearly 3 times higher than that of the group using telephone. Customers using cellular are more likely to sign up for savings accounts than customers using telephone => Focus on making marketing calls to customers **using cellular**

Does duration call affect conversions?



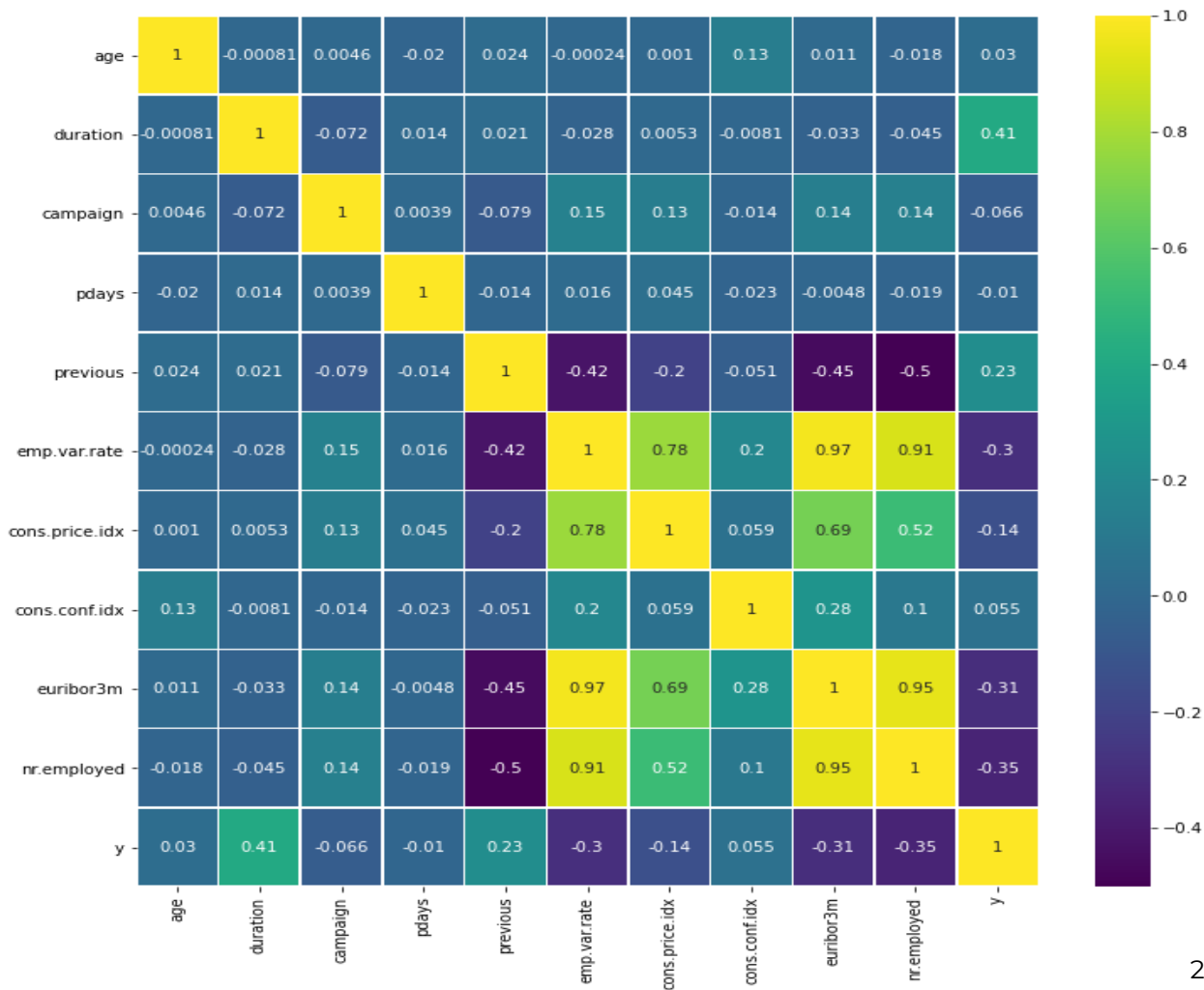
- The CR of customers with **above-average duration call** are much **higher** than those of below-average duration call.
- Customers who agree to deposit usually have a call lasting 7-12 *minutes* while customers who do not agree usually have a call lasting less than 7 *minutes*.



Expand customer duration calls (ask questions, share suitable information,...)

Heatmap

The **macro** features (especially *nr.employed*, *emp.var.rate*, *euribor3m*) are strongly correlated with each other, and strongly correlated with the **target variable**.



Handling missing value (categorical)

“Unknown”

imputed by using other independent variables to **infer** the value of the missing variable

education

Inferring from *jobs*

job

Inferring from *age* and *education*

housing

mode

loan

mode

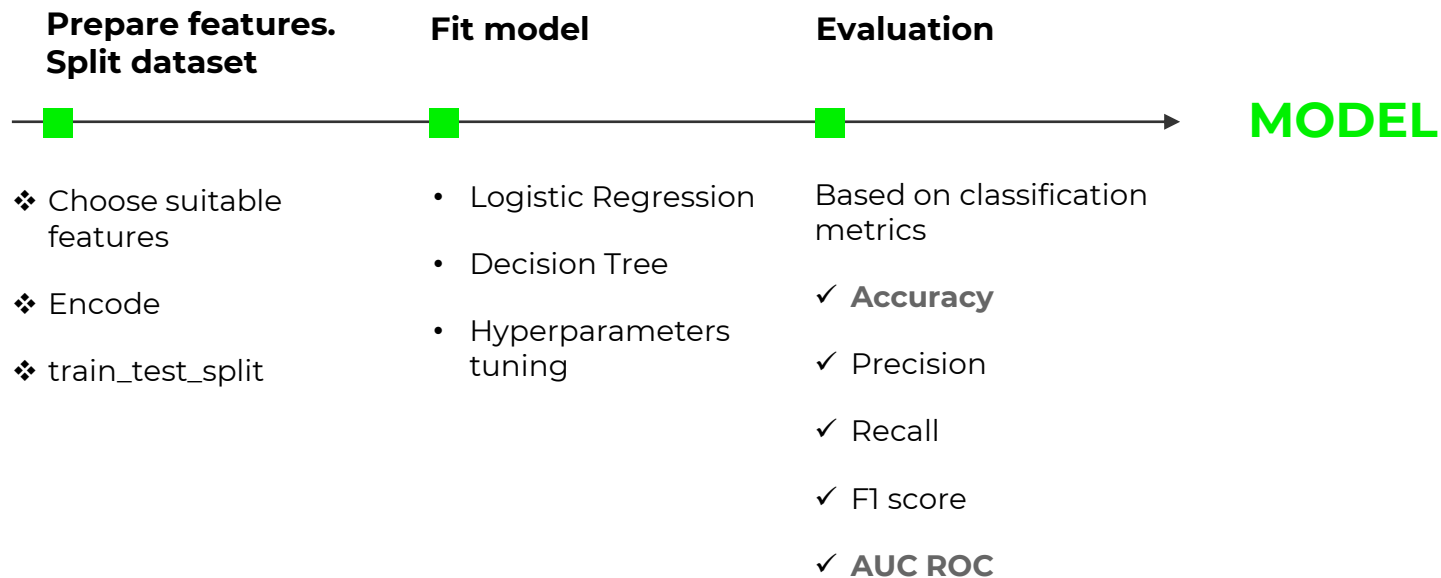
Imputation

<i>before</i>		<i>after</i>	
Education	Job	Education	Job
Unknown	Management	University.degree	
Unknown	Service	High.school	
unknown	Housemaid	Basic.4y	
Basic.4y	unknown		Blue-collar
Basic.6y	unknown		Blue-collar
Basic.9y	unknown		Blue-collar
Professional.course	unknown		technician
Age	Job	Job	
> 60	unknown	retired	

Building **model** and **Evaluation**



Process

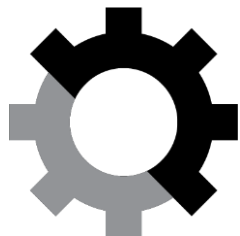


Prepare features. Split dataset



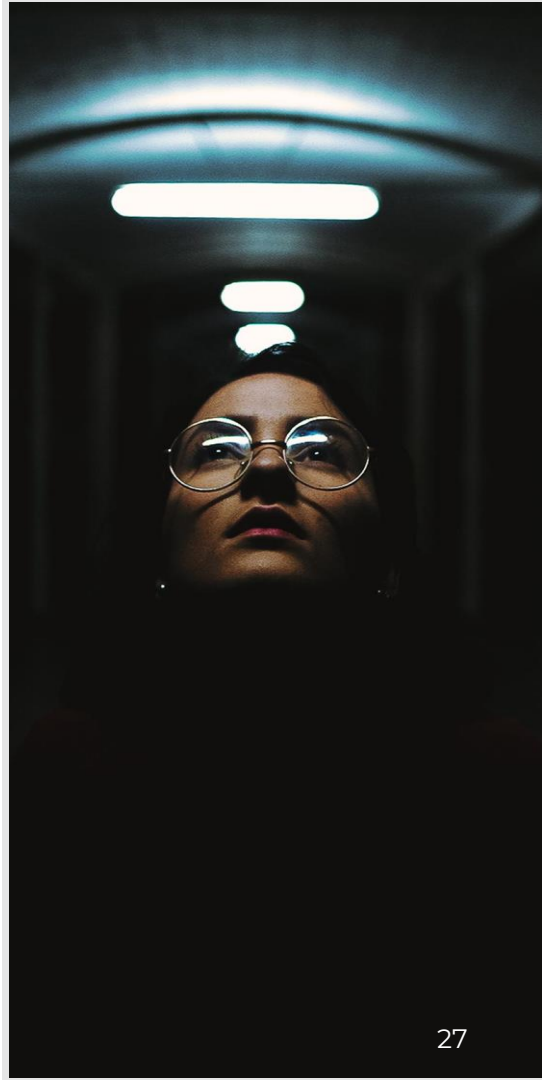
Prepare features

- Drop *duration*. This feature **highly affects** the output target (e.g., if $\text{duration}=0$ then $y=\text{'no'}$). Yet, the duration is not known before a call is performed. After the end of the call y is obviously known. This input should only be included for benchmark purposes and should be discarded if the intention is to have a *realistic model*.
- Label Encoder



Train_test_split

- test size = 0.2





Fit **model**



Logistic Regression.

Hyperparameters tuning

GridSearchCV



`{'C': 100, 'max_iter': 500, 'penalty': 'L2', 'solver': 'lbfgs'}`

Evaluation (1)



Metrics

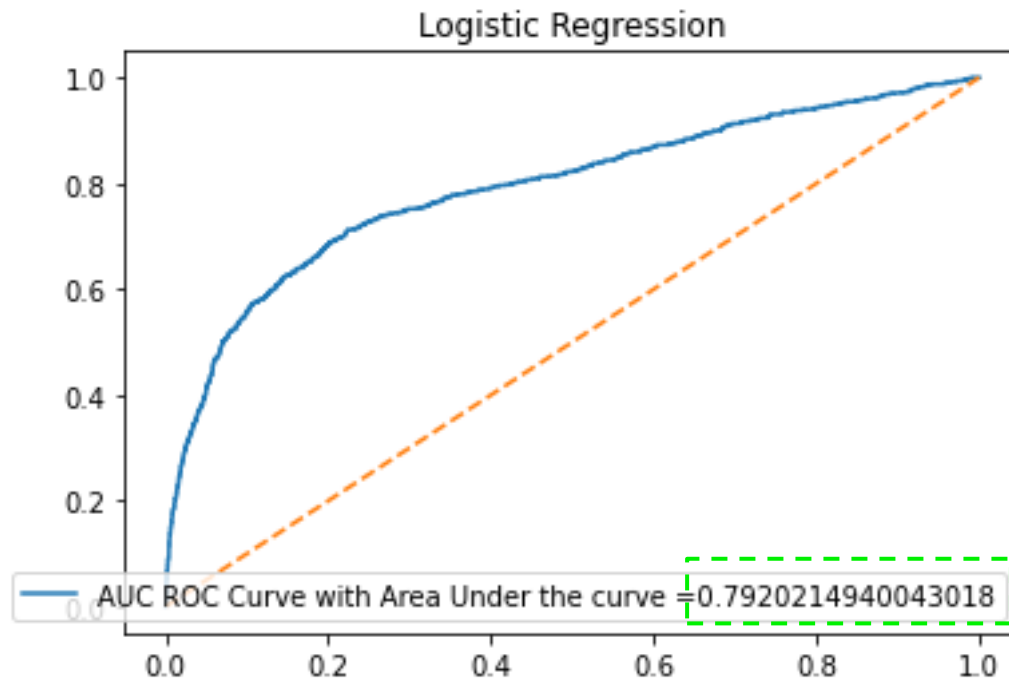
- Accuracy = 0.895580
- Precision = 0.711462
- Recall = 0.186143
- F1 Score = 0.295082



Confusion matrix

7196	73
787	180

Evaluation (2) – AUC ROC





Decision Tree.



Hyperparameter tuning

GridSearchCV



`{'criterion': 'entropy', 'max_depth': 5}`

Evaluation (1)



Metrics

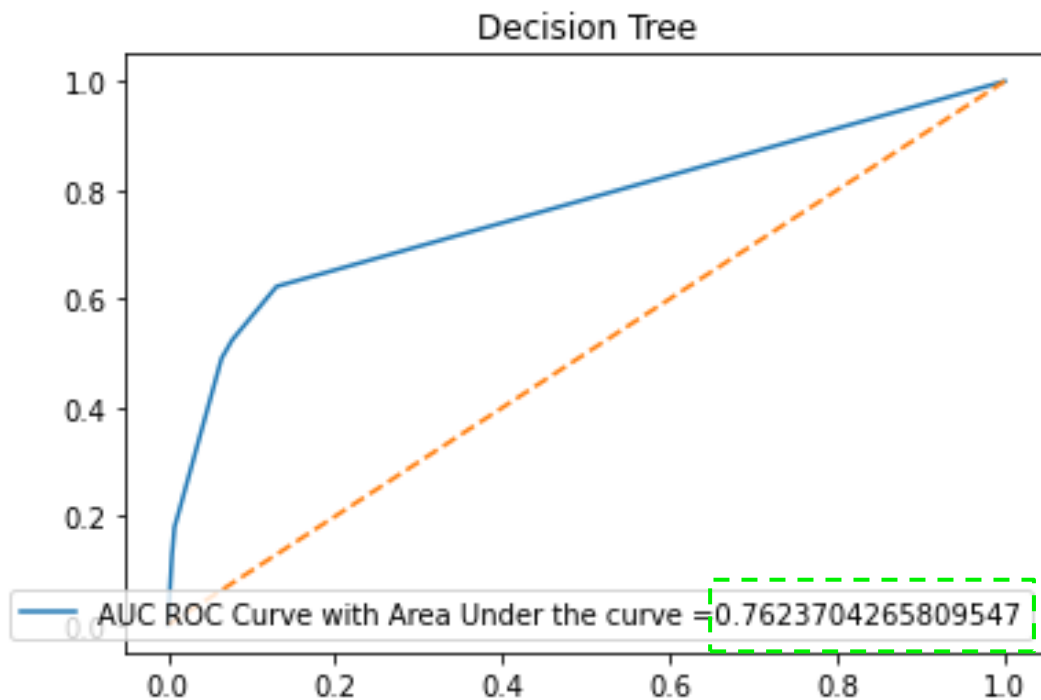
- Accuracy = 0.896795
- Precision = 0.753247
- Recall = 0.179938
- F1 Score = 0.280484



Confusion matrix

7212	57
793	174

Evaluation (2) – AUC ROC

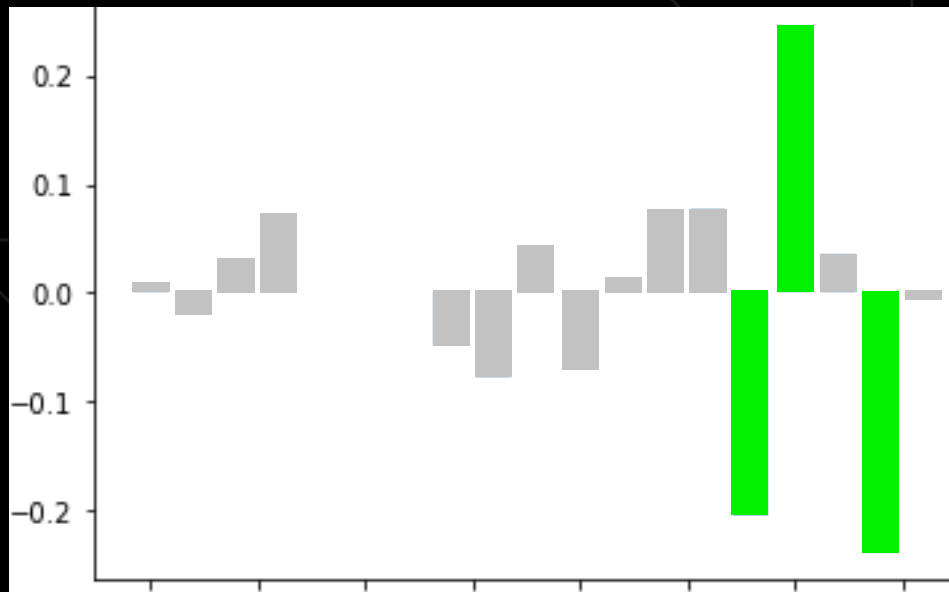


Comparison (after tuning hyperparameter)

Test set	<i>Logistic Regression</i>	▶▶▶	Accuracy ~ 90%
		▶▶▶	AUC ~ 0.79
	<i>Decision Tree</i>	▶▶▶	Accuracy ~ 90%
		▶▶▶	AUC ~ 0.76

We choose **Logistic Regression** because of its **AUC** metric.

Feature importance



Top 3 features that have the most influence on the target variable

▶▶▶ **cons.price.idx**

▶▶▶ **euribor3m**

▶▶▶ **emp.var.rate**

Conclusion and Recommendation



Conclusion

- 1** euribor3m, CPI, CCI, emp.var.rate are important macro variables
- 2** The target customer should be a group of people of relatively high age. And student (could be from university)
- 3** Try to engage customers and have longer calls
Prioritize using cellular as a means of communication.
- 4** Priority is given to customers who have contacted from previous campaigns (focus on using cellular)
- 5** Should not call more than **3** calls for the same customer to save time and effort.
- 6** The correlation of '*duration*' with the target variable showed that the higher the duration, the better it is more likely that the customer will sign up for a time deposit. A longer duration shows that the customer is interested in the product.
- 7** *Fridays* and *Mondays* people have Negatively responded as compared to other week days. We should avoid these 2 days
- 8** Seasonality: Customers tend to open deposit accounts in autumn and winter

Recommendation



Work with **economists** so that as soon as they have some signal of an improving economic situation, i.e. an increase in the CPI or CCI.

=> Expect more customers to sign up than term deposits and should **proactively contact** them before your bank's competitors do.



Target relatively **older** customers who are looking for **safe** and **profitable investment** options that must ensure conveying the peace of mind and steady income that the products offer.

This product provides a value proposition to these customers
=> Run marketing campaigns that convey these **suitable messages** for each target customer group



Prioritize customers they've reached before in **previous campaigns**. They have the ability to apply for a term deposit.

=> Should **reconnect** and **personalized marketing**



Thank you.