A DATA-DRIVEN APPROACH TO PREDICT TERM DEPOSIT SUBSCRIPTIONS

Team members:
Surin Singh
Fan Zhang
Maryam Zulfiqar
Mohammadreza Arabkhani

AGENDA

- Introduction
- Stakeholders
- Problem Statement
- Data Understanding
- Model Fitting
- Insights and Recommendations
- References

INTRODUCTION

Current Challenges with Direct Marketing
Frequency of contact
Keeping up-to-date information
Picking the right channels

What your institution is doing right?
Picking the right channels

What your marketing team is lacking
Keeping up-to-date information
Frequency of contact

STAKEHOLDERS

1) **Financial Institution: Bank**
   Primary stakeholder: interested in a product that will increase net revenue.

2) **Marketing Executives**
   Interested in carefully crafted marketing campaigns to meet bank targets.

3) **Customer Service Team**
   Will play an important part in activating and delivering the marketing campaign.

4) **Data Science Group**
   Benefits from all the insights and data to create meaningful and ready-to-implement solutions

PROBLEM STATEMENT

We aim to pinpoint the most important factors that affect a customer's decision to subscribe to a term deposit.

DATA UNDERSTANDING

- A Bank Marketing secondary dataset was obtained from the UCI Machine Learning Repository.
- It contains 41,188 entries and is related with direct marketing campaigns of a Portuguese banking institution based on phone calls.
- Often, more than one contact to the same client was required, in order to assess if the bank term deposit would be ('yes') or not ('no') subscribed.
- There are 21 attributes consisting of demographics, technological, competitive, social and economic and some other factors.

DATA UNDERSTANDING

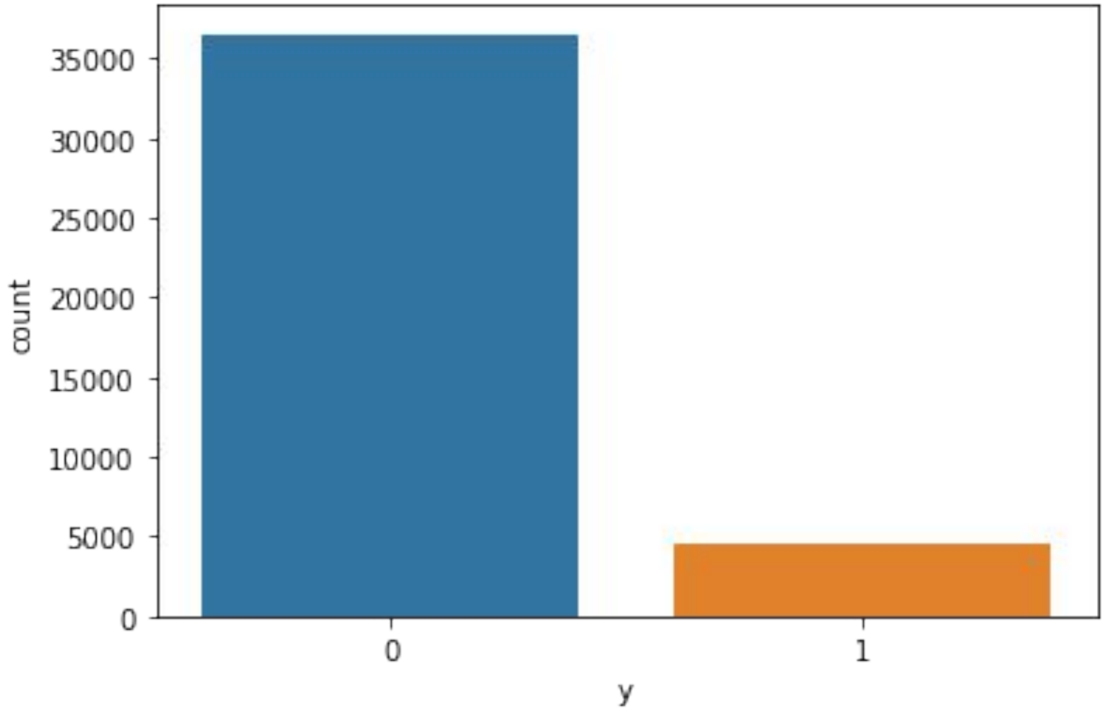
Input Variables:

1. age
2. job
3. marital
4. education
5. default
6. housing
7. loan
8. contact
9. month
10. day_of_week
11. duration
12. campaign
13. pdays
14. previous
15. poutcome
16. emp.price.idx
17. cons.price.idx
18. cons.conf.idx
19. euribor3m
20. nr.employed

Output Variable:

1. y

DATA UNDERSTANDING



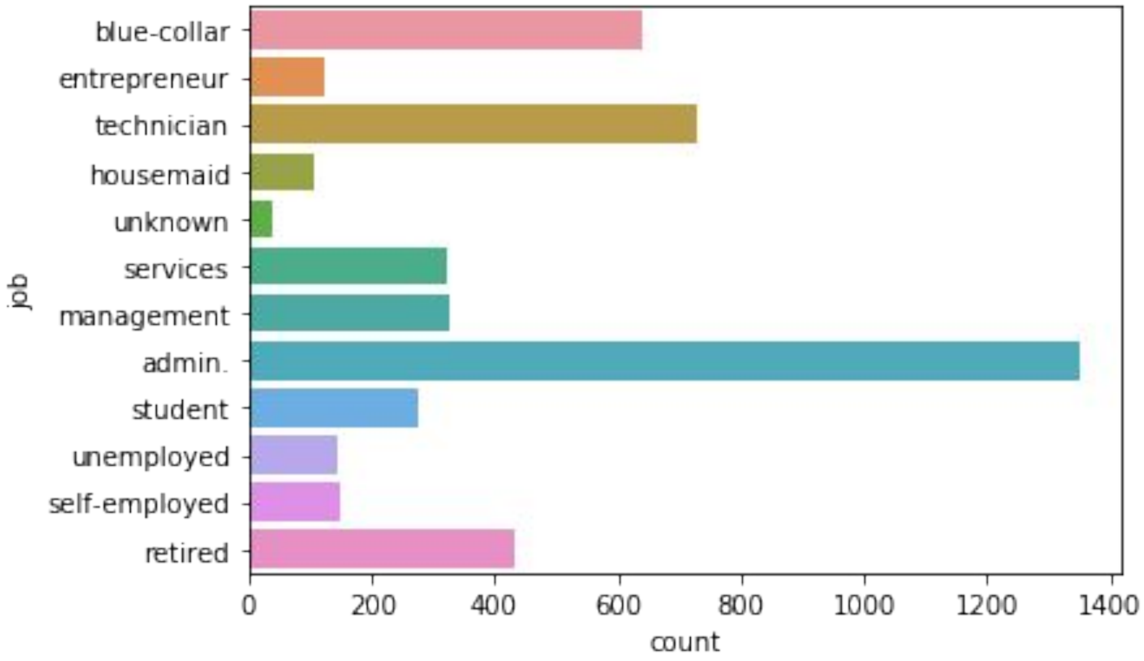


Figure 1: y categories, 0 & 1 for 'no' or 'yes' respectively

Figure 2: Count for the target variable 'y' when the outcome is 'yes' with job

DATA UNDERSTANDING

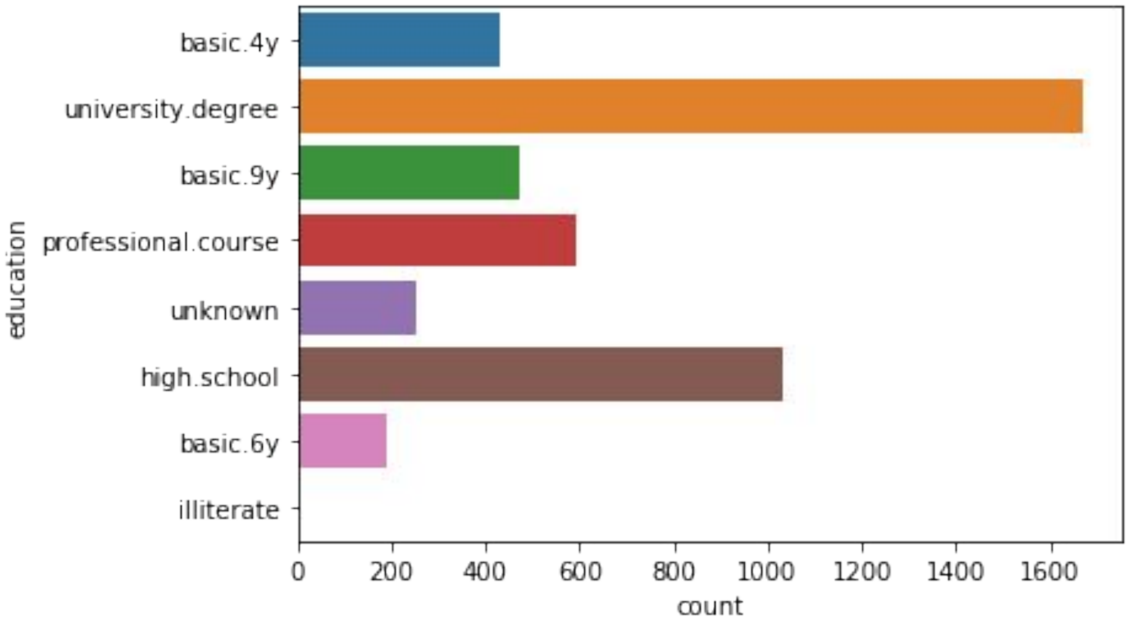


Figure 3: Count for the target variable 'y' when the outcome is 'yes' with education

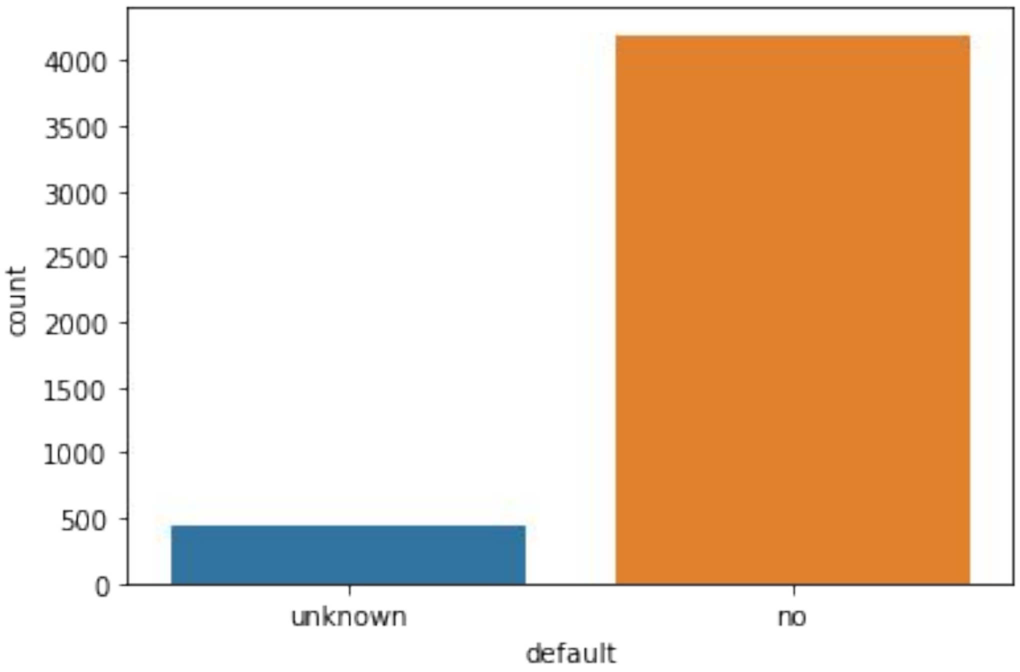
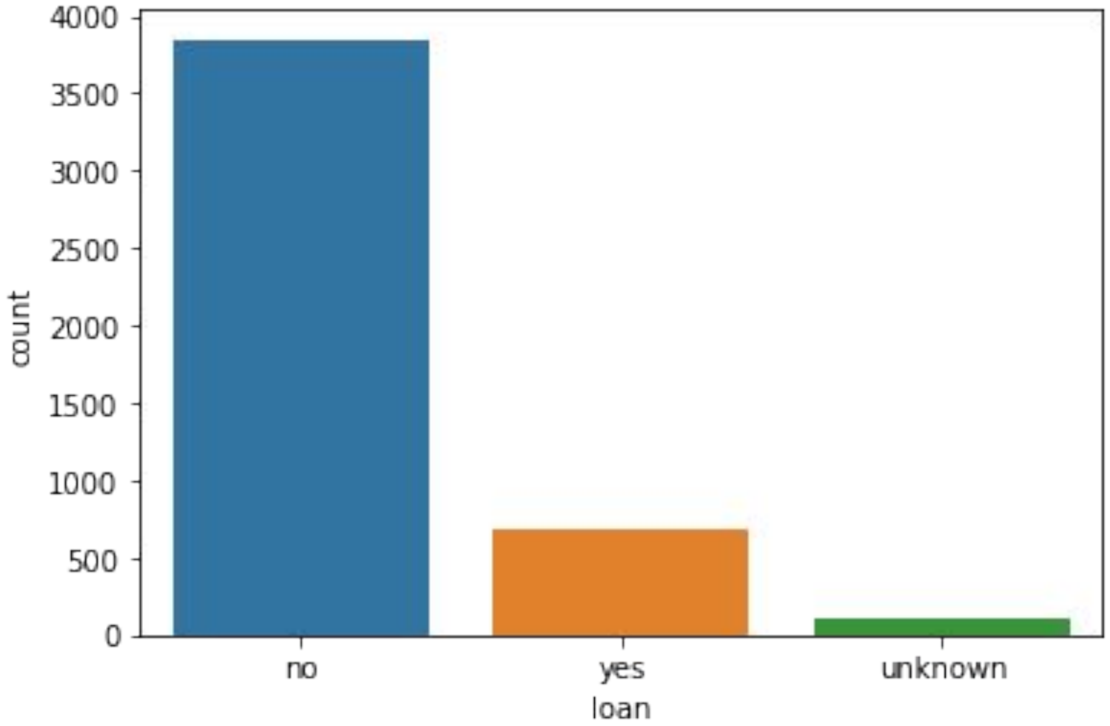


Figure 4: Count for the target variable 'y' when the outcome is 'yes' with default

DATA UNDERSTANDING



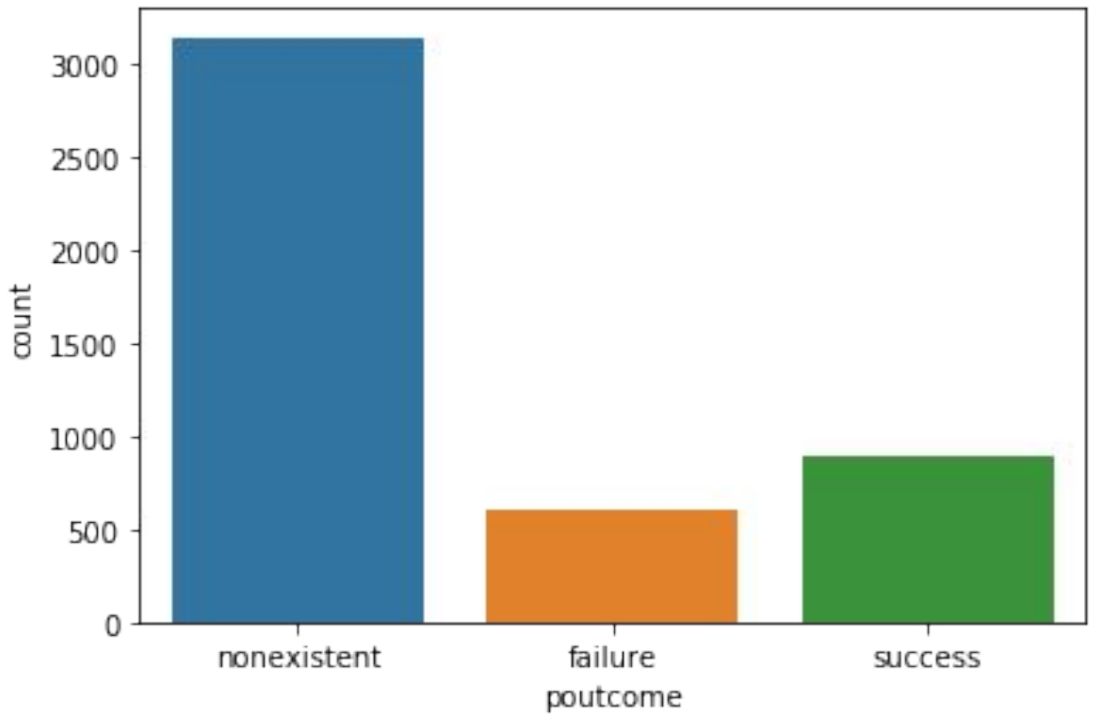


Figure 5: Count for the target variable 'y' when the outcome is 'yes' with loan

Figure 6: Count for the target variable 'y' when the outcome is 'yes' with poutcome

DATA UNDERSTANDING

	age	job	marital	education	default	loan	contact	duration	campaign	previous	poutcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y
age	1	0.001	-0.4	-0.1	0.2	-0.007	0.007	-0.0008	0.005	0.02	0.02	-0.0001	0.001	0.1	0.01	-0.02	0.03
job	0.001	1	0.03	0.1	-0.03	-0.01	-0.02	-0.006	-0.007	0.02	0.01	-0.009	-0.02	0.05	-0.008	-0.02	0.03
marital	-0.4	0.03	1	0.1	-0.08	0.006	-0.05	0.01	-0.007	0.04	0.002	-0.08	-0.06	-0.03	-0.09	-0.09	0.05
education	-0.1	0.1	0.1	1	-0.2	0.006	-0.1	-0.02	0.0004	0.04	0.02	-0.04	-0.08	0.08	-0.04	-0.04	0.06
default	0.2	-0.03	-0.08	-0.2	1	-0.004	0.1	-0.01	0.03	-0.1	0.02	0.2	0.2	0.03	0.2	0.2	-0.1
loan	-0.007	-0.01	0.006	0.006	-0.004	1	-0.009	-0.001	0.005	-0.001	-0.001	0.002	-0.003	-0.01	0.0001	0.004	-0.005
contact	0.007	-0.02	-0.05	-0.1	0.1	-0.009	1	-0.03	0.08	-0.2	0.1	0.4	0.6	0.3	0.4	0.3	-0.1
duration	-0.0008	-0.006	0.01	-0.02	-0.01	-0.001	-0.03	1	-0.07	0.02	0.03	-0.03	0.005	-0.008	-0.03	-0.04	0.4
campaign	0.005	-0.007	-0.007	0.0004	0.03	0.005	0.08	-0.07	1	-0.08	0.03	0.2	0.1	-0.01	0.1	0.1	-0.07
previous	0.02	0.02	0.04	0.04	-0.1	-0.001	-0.2	0.02	-0.08	1	-0.3	-0.4	-0.2	-0.05	-0.5	-0.5	0.2
poutcome	0.02	0.01	0.002	0.02	0.02	-0.001	0.1	0.03	0.03	-0.3	1	0.2	0.2	0.2	0.2	0.1	0.1
emp.var.rate	-0.0001	-0.009	-0.08	-0.04	0.2	0.002	0.4	-0.03	0.2	-0.4	0.2	1	0.8	0.2	1	0.9	-0.3
cons.price.idx	0.001	-0.02	-0.06	-0.08	0.2	-0.003	0.6	0.005	0.1	-0.2	0.2	0.8	1	0.06	0.7	0.5	-0.1
cons.conf.idx	0.1	0.05	-0.03	0.08	0.03	-0.01	0.3	-0.008	-0.01	-0.05	0.2	0.2	0.06	1	0.3	0.1	0.05
euribor3m	0.01	-0.008	-0.09	-0.04	0.2	0.0001	0.4	-0.03	0.1	-0.5	0.2	1	0.7	0.3	1	0.9	-0.3
nr.employed	-0.02	-0.02	-0.09	-0.04	0.2	0.004	0.3	-0.04	0.1	-0.5	0.1	0.9	0.5	0.1	0.9	1	-0.4
y	0.03	0.03	0.05	0.06	-0.1	-0.005	-0.1	0.4	-0.07	0.2	0.1	-0.3	-0.1	0.05	-0.3	-0.4	1

Figure 7: Correlation of all the variables with the target variable

MODEL FITTING-LOGISTIC REGRESSION

- To predict if people will subscribe to a term deposit: no or yes
- The input variables are the independent variables and the the output is the dependent variable ‘y’ that we divide into binary class - ‘no’ or ‘yes’
- To understand model performance, we divided the dataset into a training set and a test set in a ratio 90:10.
- We have two classes 0 and 1 for ‘no’ and ‘yes’ respectively.
- We evaluate the model using model evaluation metrics of precision and recall.

MODEL FITTING-LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.93	0.98	0.95	3644
1	0.69	0.40	0.50	473
accuracy			0.91	4117
macro avg	0.81	0.69	0.73	4117
weighted avg	0.90	0.91	0.90	4117

Table 1: The final percentage (weighted average) of right prediction for if a client subscribed or did not subscribe is 90%.

MODEL FITTING-RANDOM FOREST CLASSIFIER

- Splits 90% of the dataset into our training set and the other 10% into test data.
- For greater accuracy random_state parameter was set to 0 and the n_estimators was equal to 200.

	precision	recall	f1-score	support
0	0.94	0.96	0.95	3665
1	0.63	0.52	0.57	452
accuracy			0.91	4117
macro avg	0.78	0.74	0.76	4117
weighted avg	0.91	0.91	0.91	4117

Table 2: The final percentage of right prediction for if a client subscribed or did not subscribe is 91%.

INSIGHTS AND RECOMMENDATIONS

- To get people to subscribe for a term deposit, look for individuals with an administration or technician job with a high level of education.

- They shouldn’t have any credit in default or a personal loan and are new clients or their last campaign was a successful one.

- It would be difficult to reduce phone calls since our analyses show that people have a higher chance of subscribing if the call lasts longer and contacts are made more often.

- Random Forest Classifier enabled us to predict to a high level of probability (91%), if a person will subscribe to a term deposit based on demographics, social, and economic factors.

REFERENCES

- Moro, S., Cortez,P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Retrieved from http://media.salford-systems.com/video/tutorial/2015/targeted_marketing.pdf