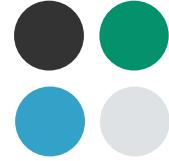


TERM DEPOSIT BANK MARKETING CAMPAIGN

FINAL PROJECT DATA SCIENCE
SCIPY



OUTLINE



Dataset & Case



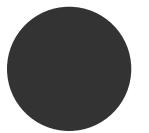
Scenario



Problem & Goals



Explanatory Data Analysis



Modelling



Simulation



Conclusion

DATASET & CASE

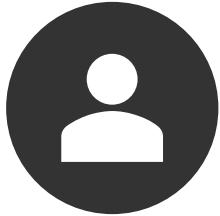




kaggleTM

<https://www.kaggle.com/volodymyrgavrysh/bank-marketing-campaigns-dataset>

FEATURES DESCRIPTION



Bank Customer Profile

<u>Age</u>	: (numerical continuous)
<u>Job</u>	: type of job (categorical: "admin.", "blue-collar", "entrepreneur", ..., "unknown")
<u>Marital</u>	: marital status (categorical: "divorced", "married", "single", "unknown")
<u>Education</u>	: (categorical: "illiterate", ..., "university.degree", "unknown")
<u>Default</u>	: has credit in default? (categorical: "no", "yes", "unknown")
<u>Housing</u>	: has housing loan? (categorical: "no", "yes", "unknown")
<u>Loan</u>	: has personal loan? (categorical: "no", "yes", "unknown")

FEATURES DESCRIPTION



Contact of the Current Campaign

- Contact** : contact communication type (categorical: "cellular", "telephone")
- Month** : last contact month of year (categorical: "mar", ..., "nov", "dec")
- Day of week** : last contact day of the week (categorical: "mon", ..., "fri")
- Duration** : last contact duration, in seconds (numeric).
 - Important note: this attribute highly affects the output target (e.g., if duration=0 then y="no"). , Can not get this feature before the campaign

FEATURES DESCRIPTION



Other Attributes

Campaign : number of contacts performed during this campaign and for this client (numeric, includes last contact)

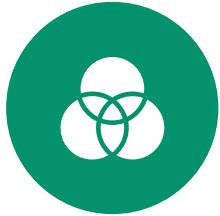
Important note : Can not get this feature before the campaign

Pdays : number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

Previous : number of contacts performed before this campaign and for this client (numeric)

Poutcome : outcome of the previous marketing campaign (categorical: "failure","nonexistent","success")

FEATURES DESCRIPTION



Social & Economics Index

- [Emp.var.rate](#) : employment variation rate - quarterly indicator (numeric)
- [Cons.price.idx](#) : consumer price index - monthly indicator (numeric)
- [Cons.conf.idx](#) : consumer confidence index - monthly indicator (numeric)
- [Euribor3m](#) : euribor 3 month rate - daily indicator (numeric)
- [Nr.employed](#) : number of employees - quarterly indicator (numeric)

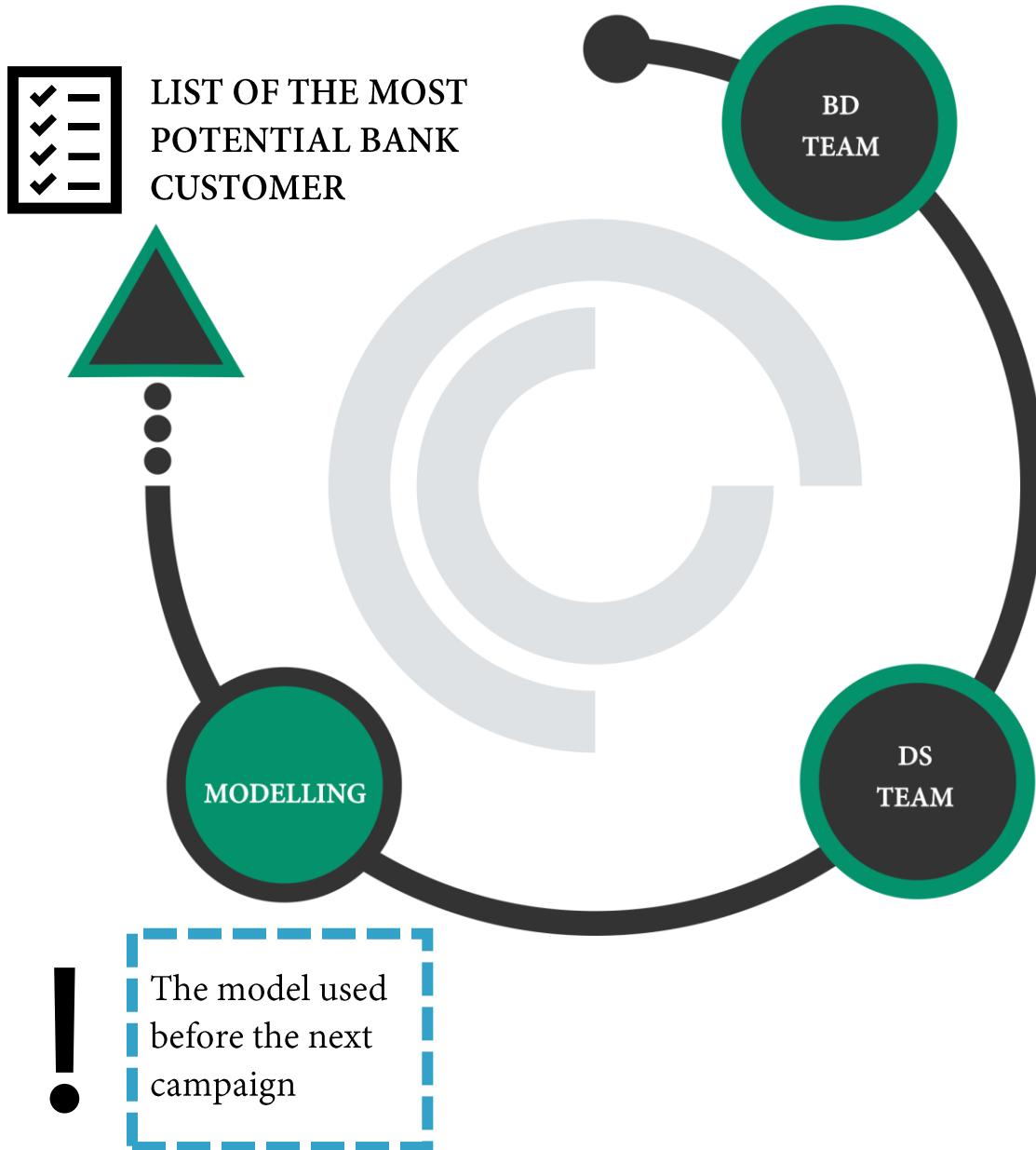


Target

- y : has the client subscribed a term deposit? ("yes","no")

SKENARIO



**Problem**

- Last campaign was not good enough compared to the **COST**

Requirement

- Needed improvement to reduce the marketing **COST**

- Receive evaluation data from last campaign
- Many fail to sell the term deposit (88%)

Solution

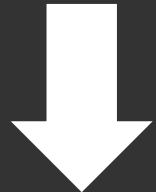
- Find the right person that will subscribed term deposit based on the previous data with **Machine learning**

PROBLEMS & GOALS

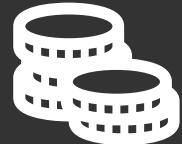




PROBLEMS



Subscribed deposit LOW
(11.27 %)



Cost TOO BIG



GOALS

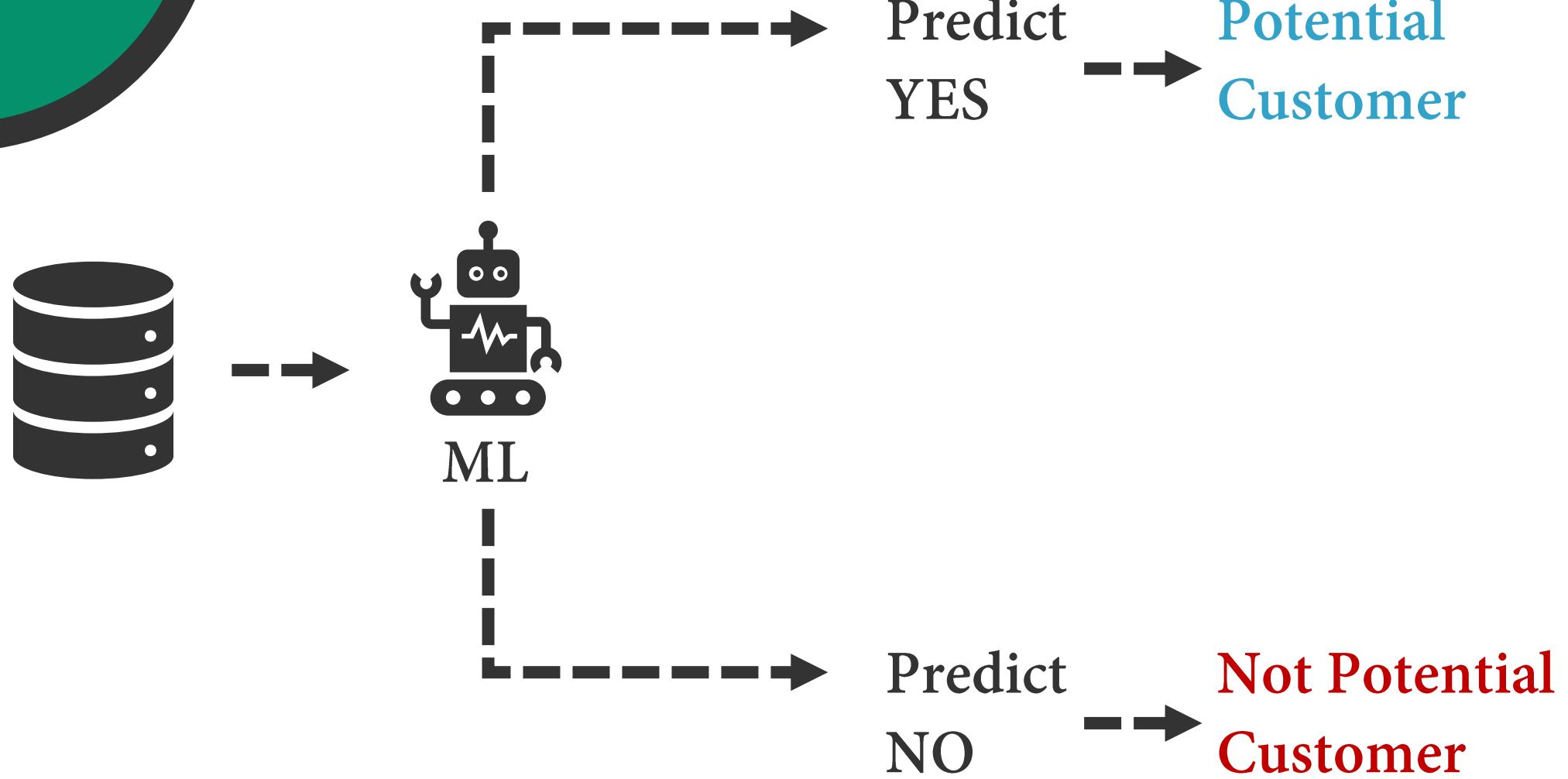


Predict Potential Bank Customer with Machine learning (ML) Method

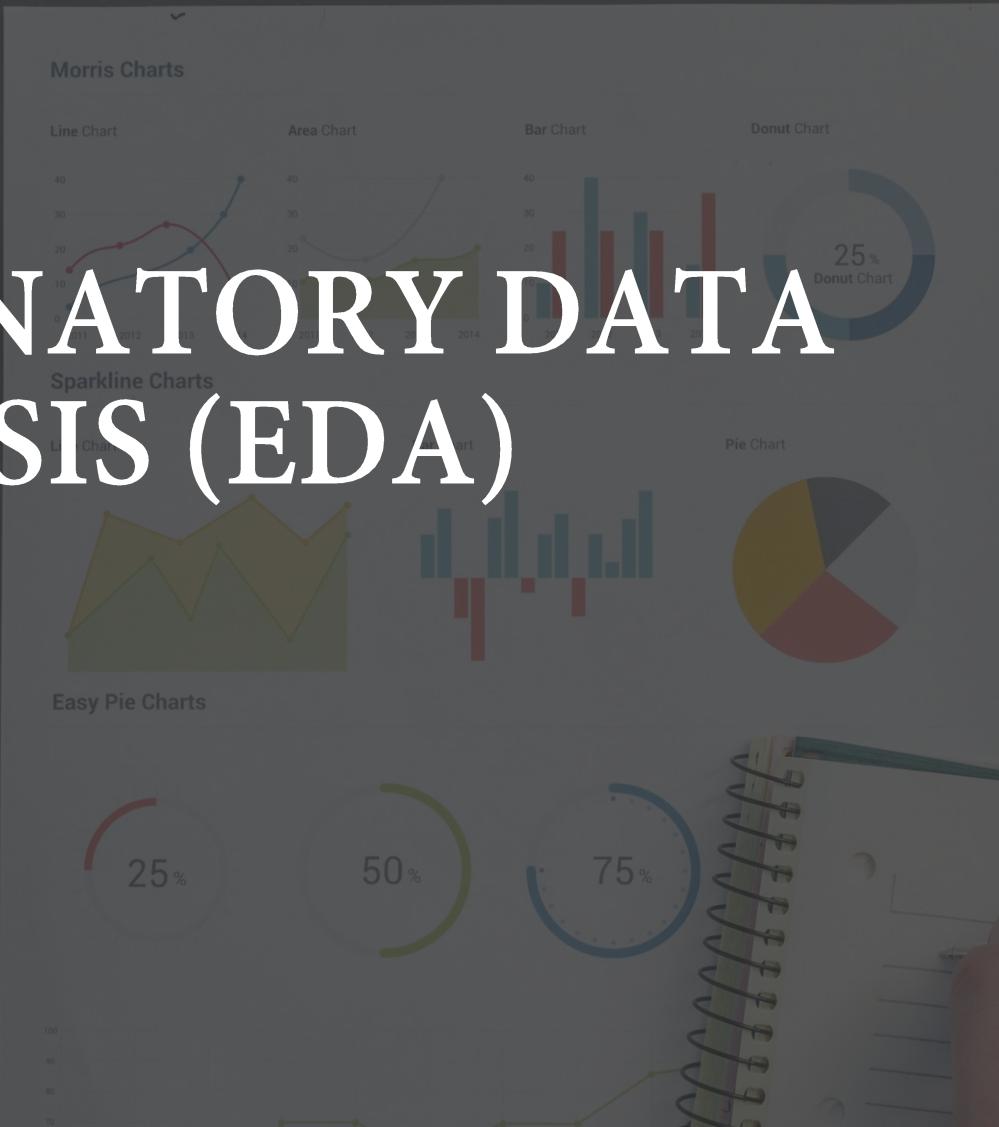


Give Cost Reduction
Simulation after using
ML

FLOW CHART



EXPLANATORY DATA ANALYSIS (EDA)



1 Uni variate

For every columns in Dataset

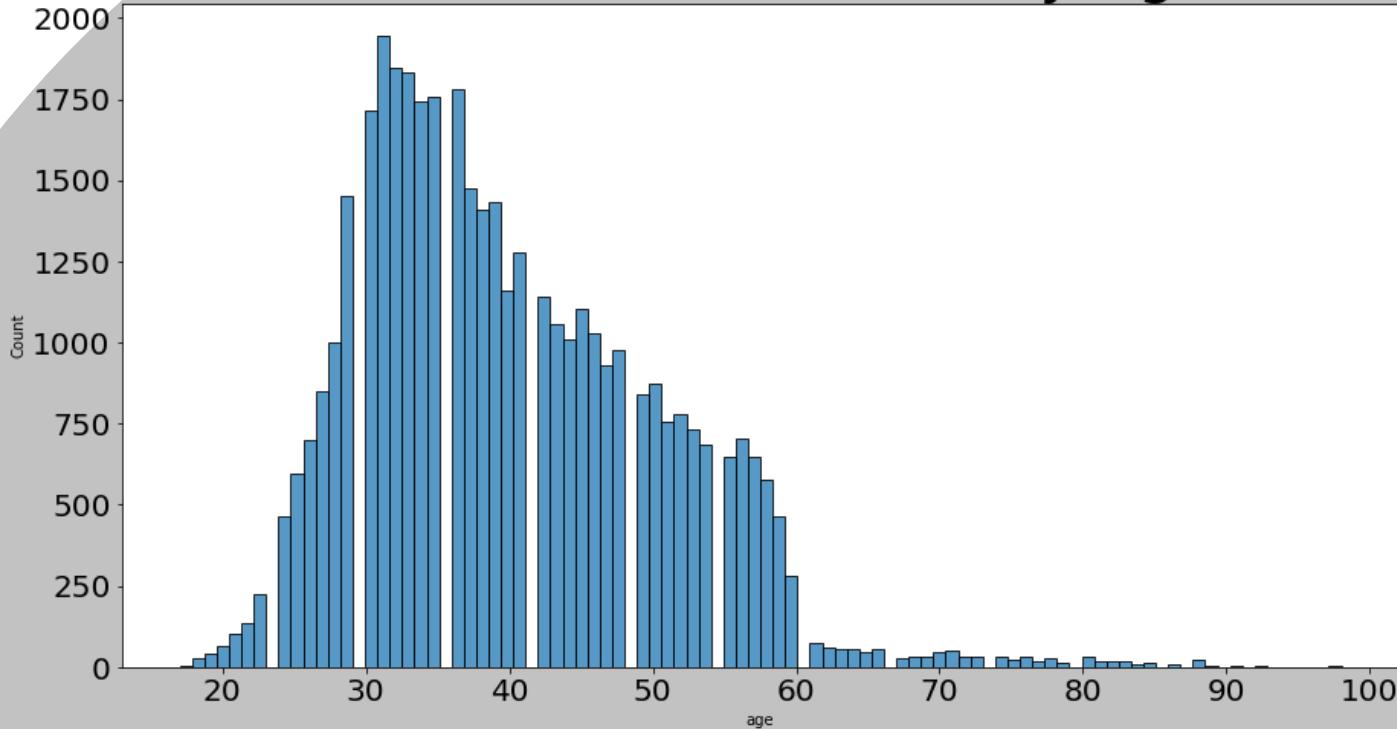
For every columns against
Target column (y)

Multi variate 2

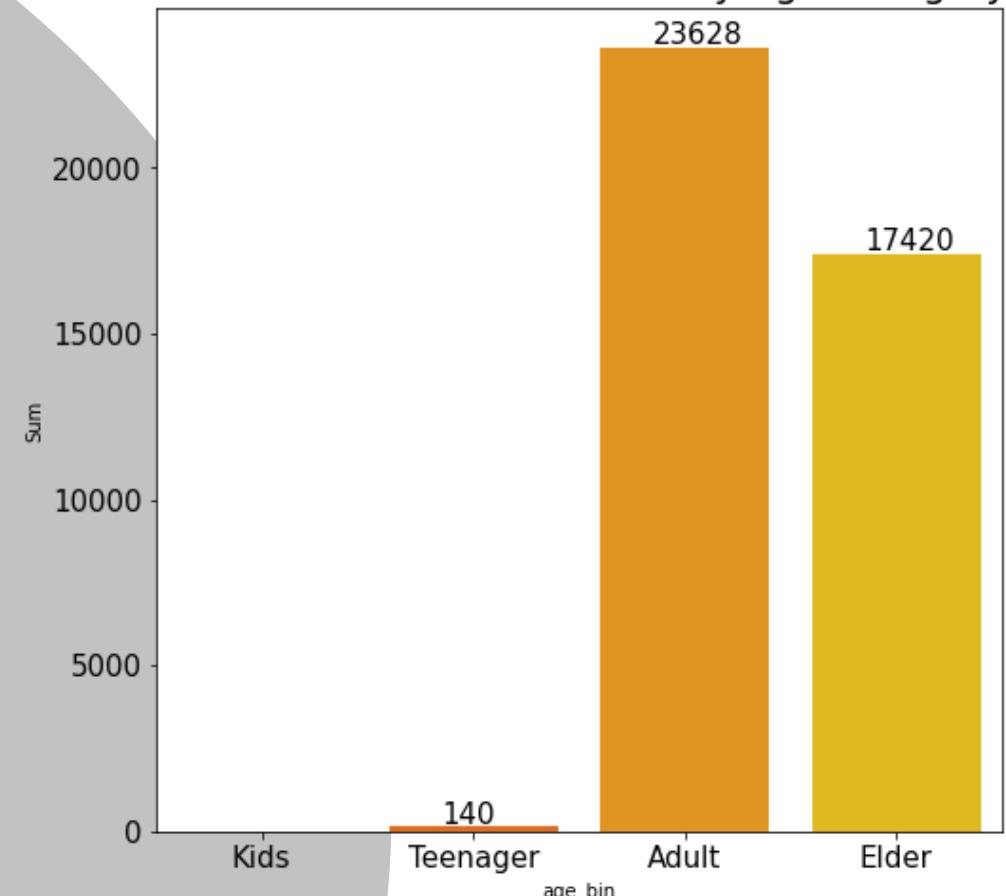
UNI

Customer Profile

Customer Distribution by Age



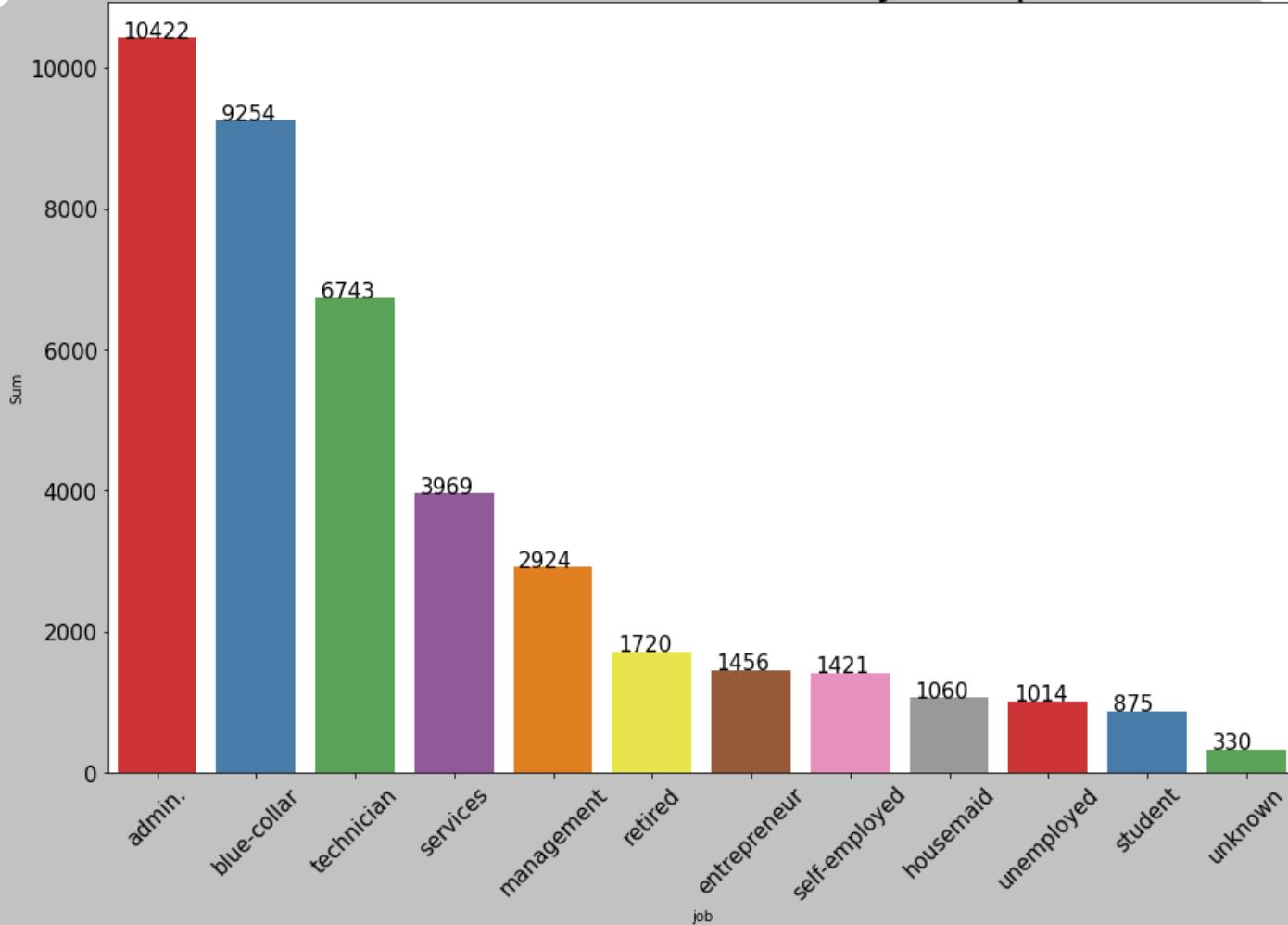
Number of Bank Customers by Age Category



- Most customers were 25 to 60 years old
- There was little to none customers in kids and teenager group

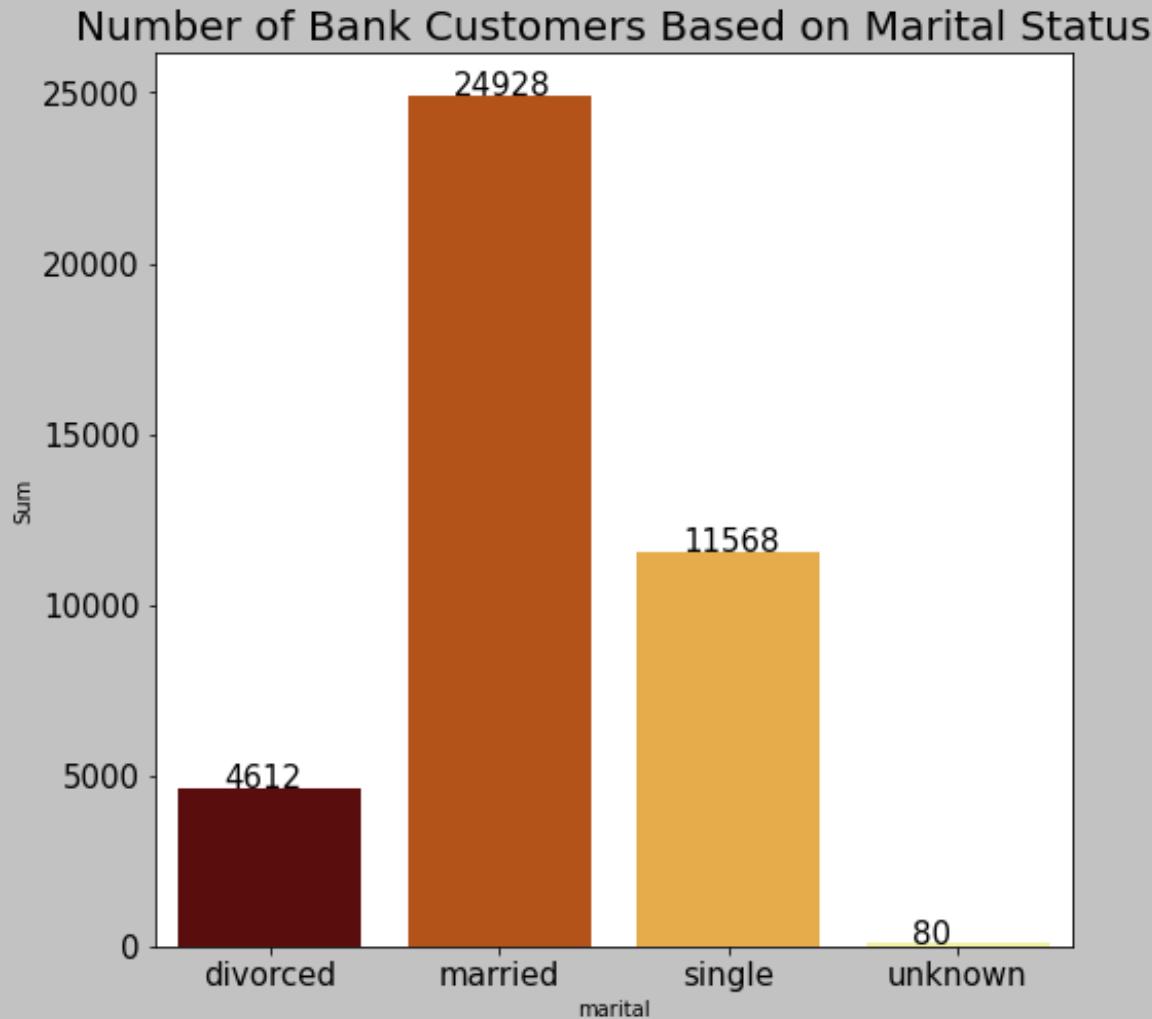
Customer Profile

Number of bank customers by occupation



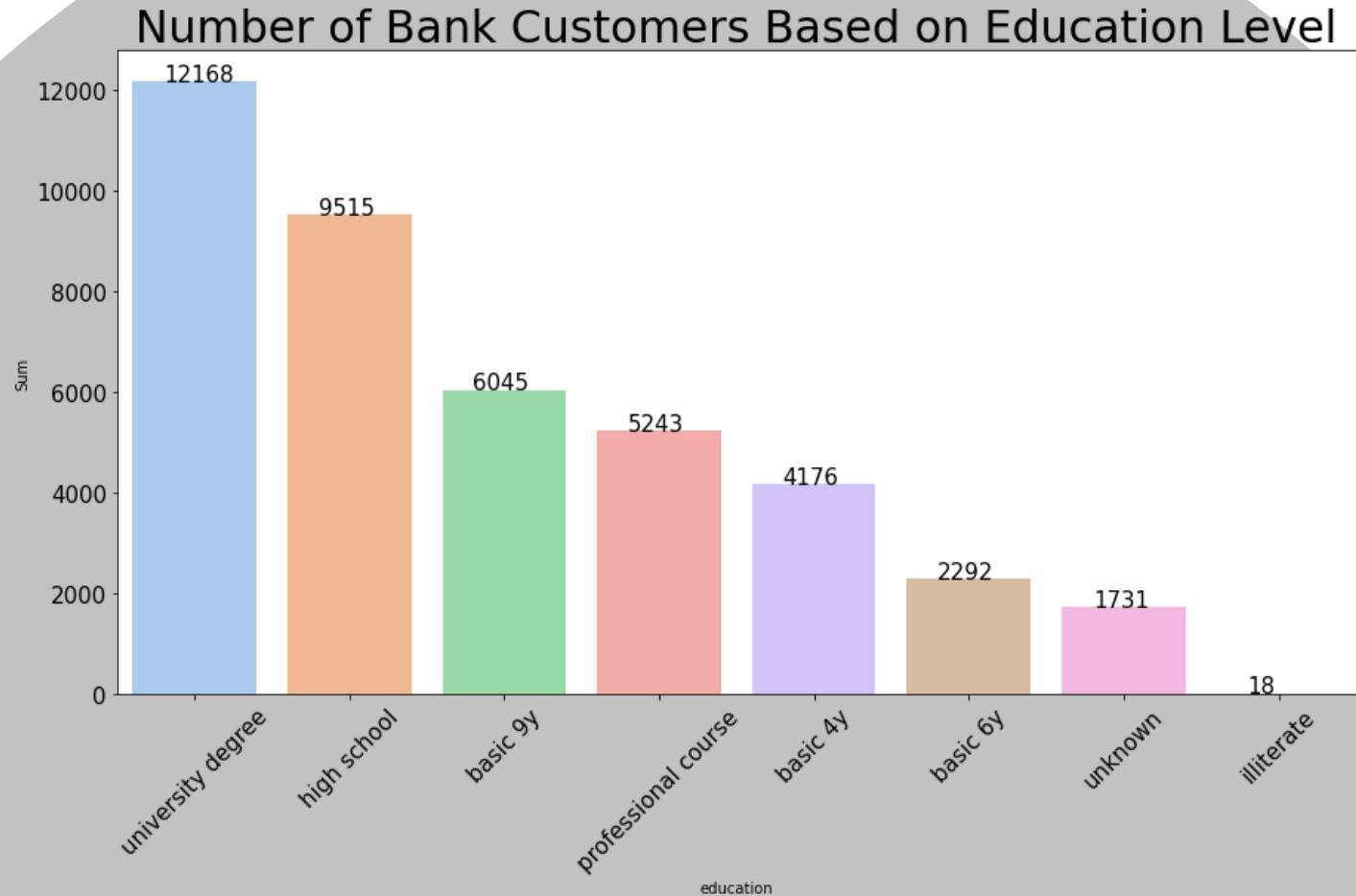
- Most customers were working as admin, blue-collar, and technician

Customer Profile



- Most customers were married

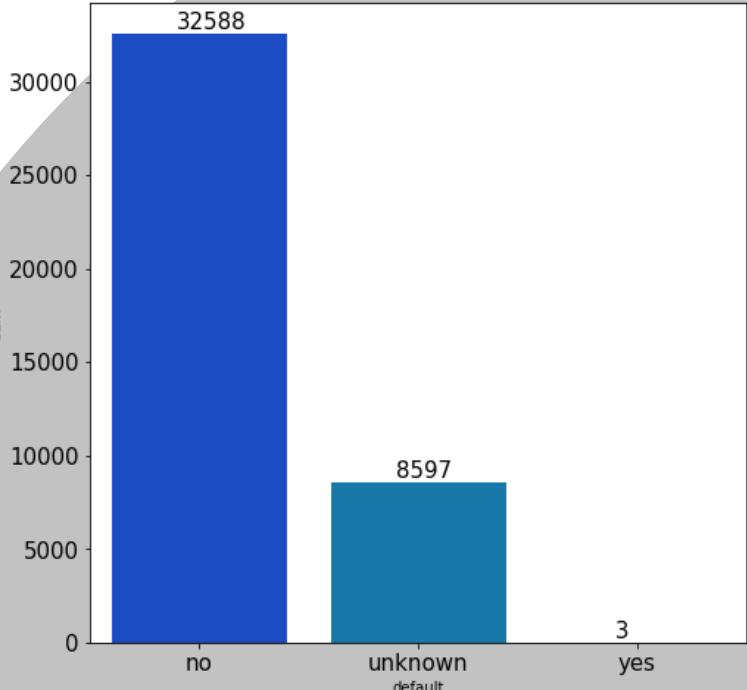
Customer Profile



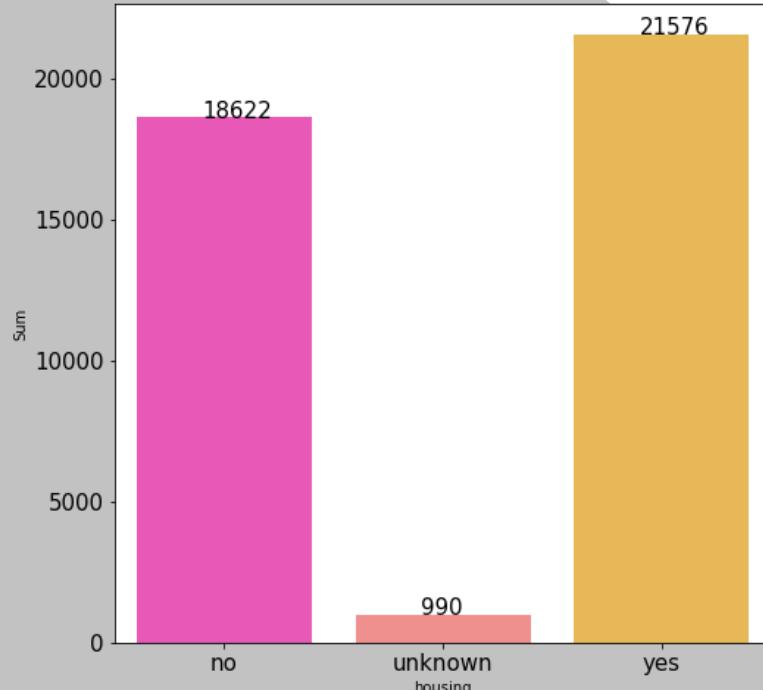
- Most customers have university degree and high school education

Credit Status

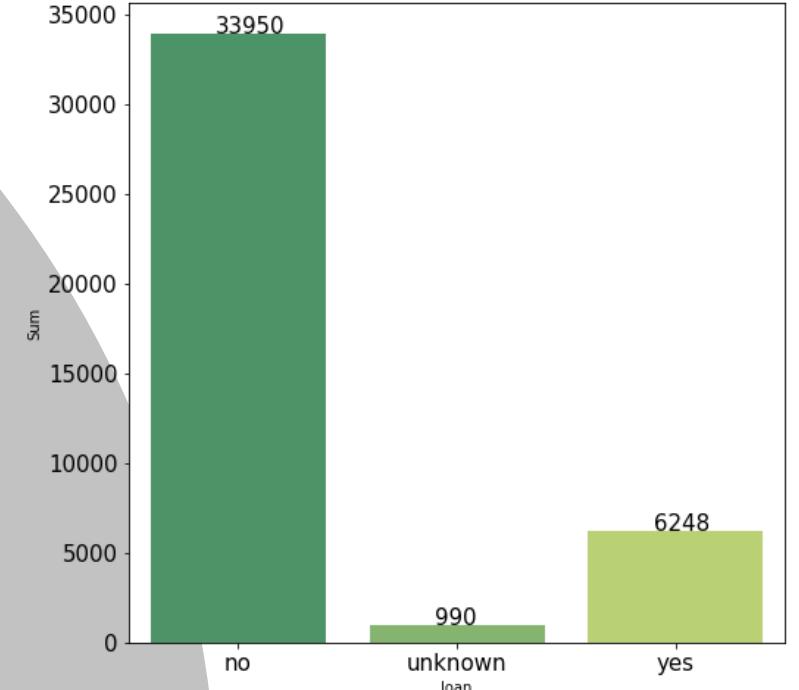
Number of Bank Customers Based on Default Status on Credit Cards



Number of Bank Customers Based on House Loan Status



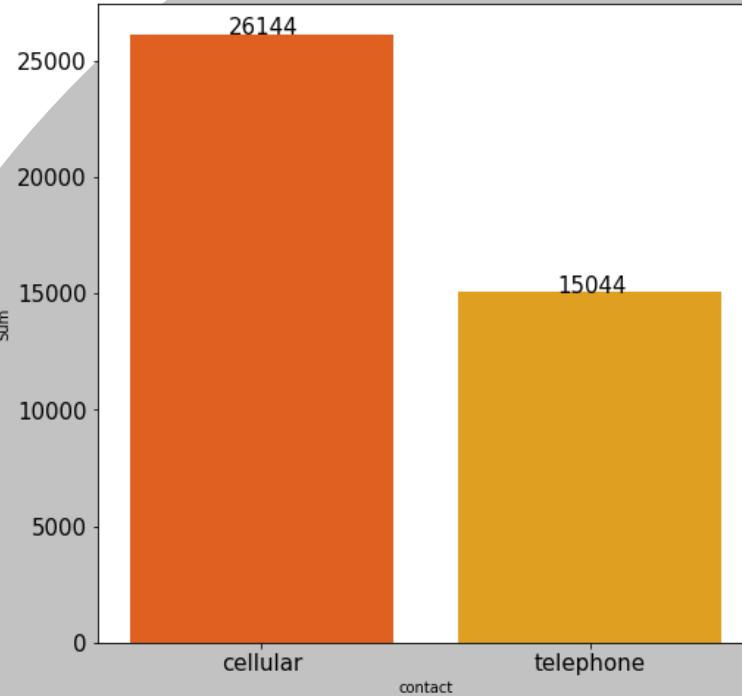
Number of Bank Customers Based on Personal Loan Status



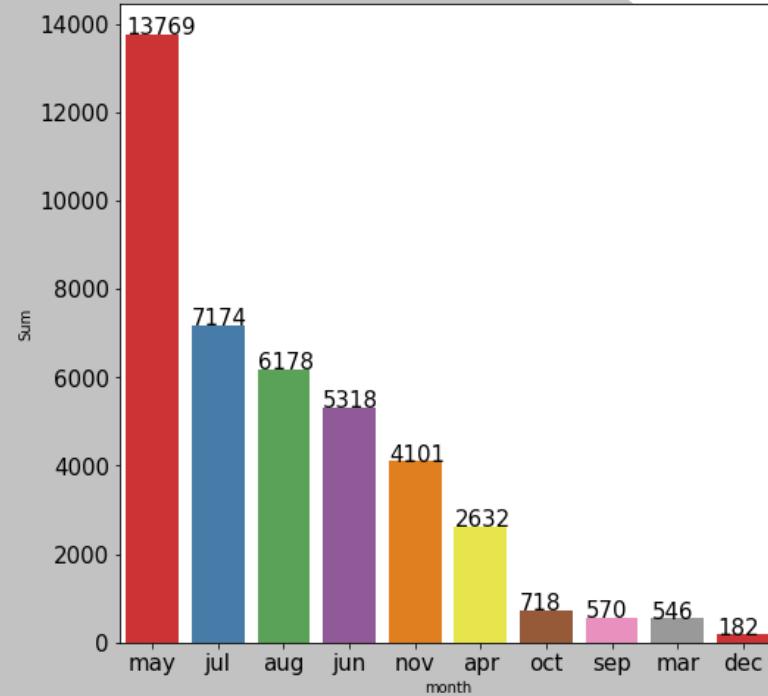
- Most customers were not credit card default and has no personal loan
- More than half of the customers have housing loan

Contact Communication Type

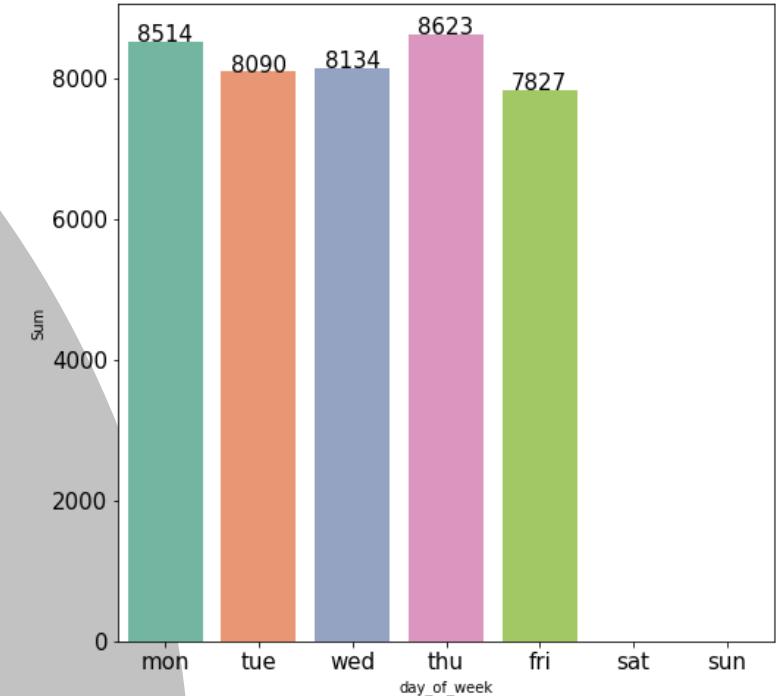
Number of Bank Customers Based on Communication Methods



Number of Bank Customers by Month Contacted

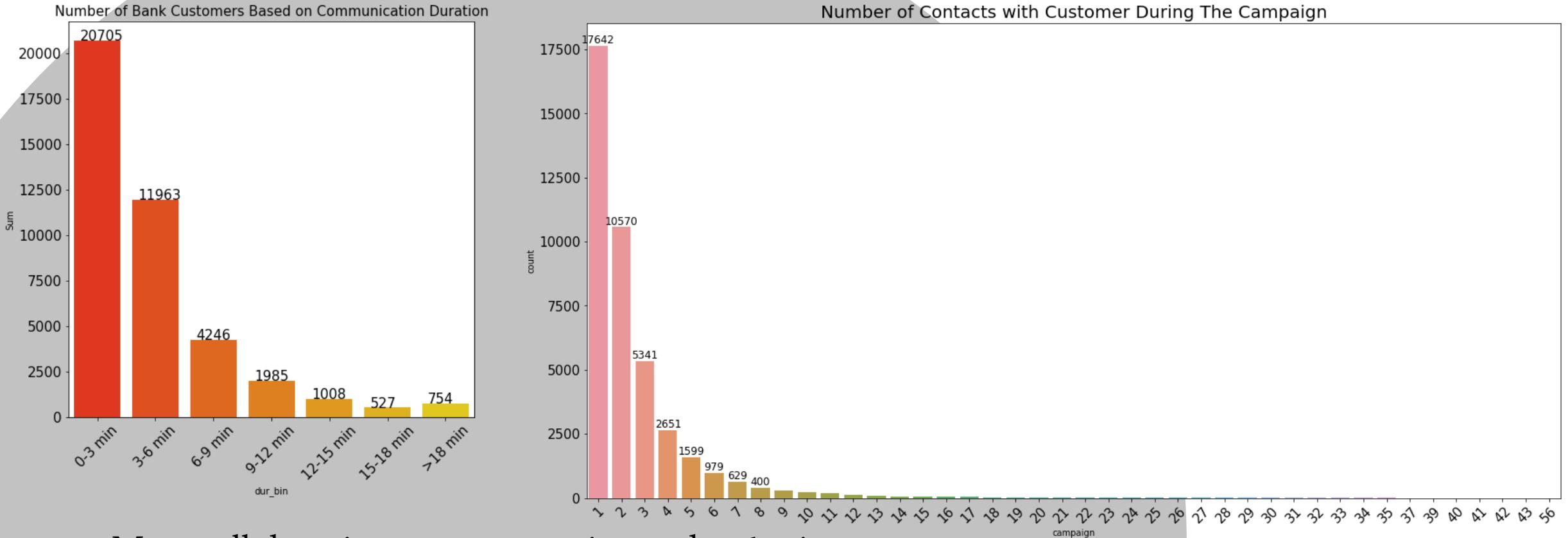


Number of Bank Customers by Weekday Contacted



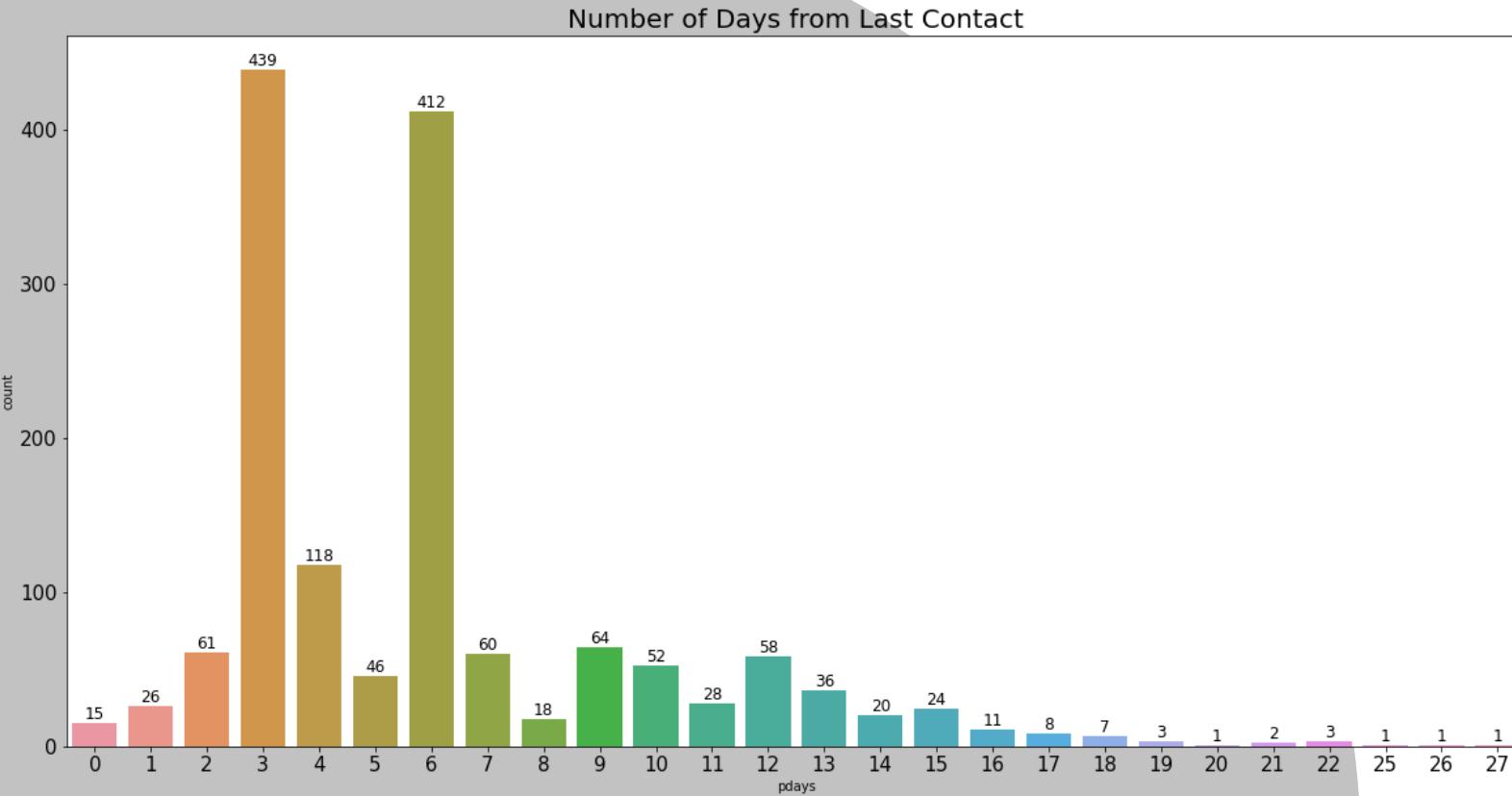
- Most customers were contacted through their mobile phone
- Most customers were contacted in May and all of them were contacted at weekdays

Number of Contacts and Duration



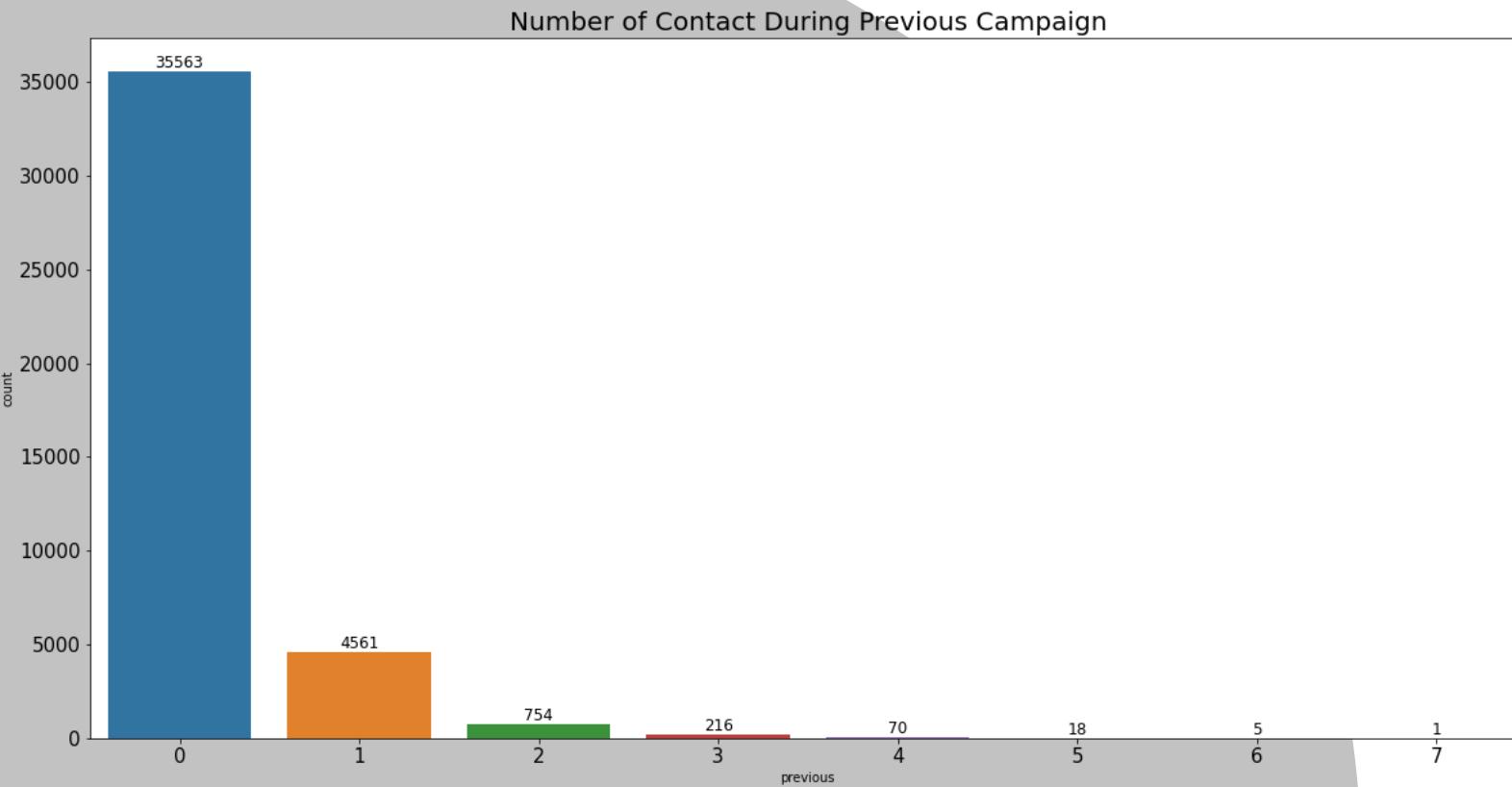
- Most call durations were 0-3 mins and 3-6 mins
- Most customers only contacted between 1 to 3 times during the campaign

Days from last contact



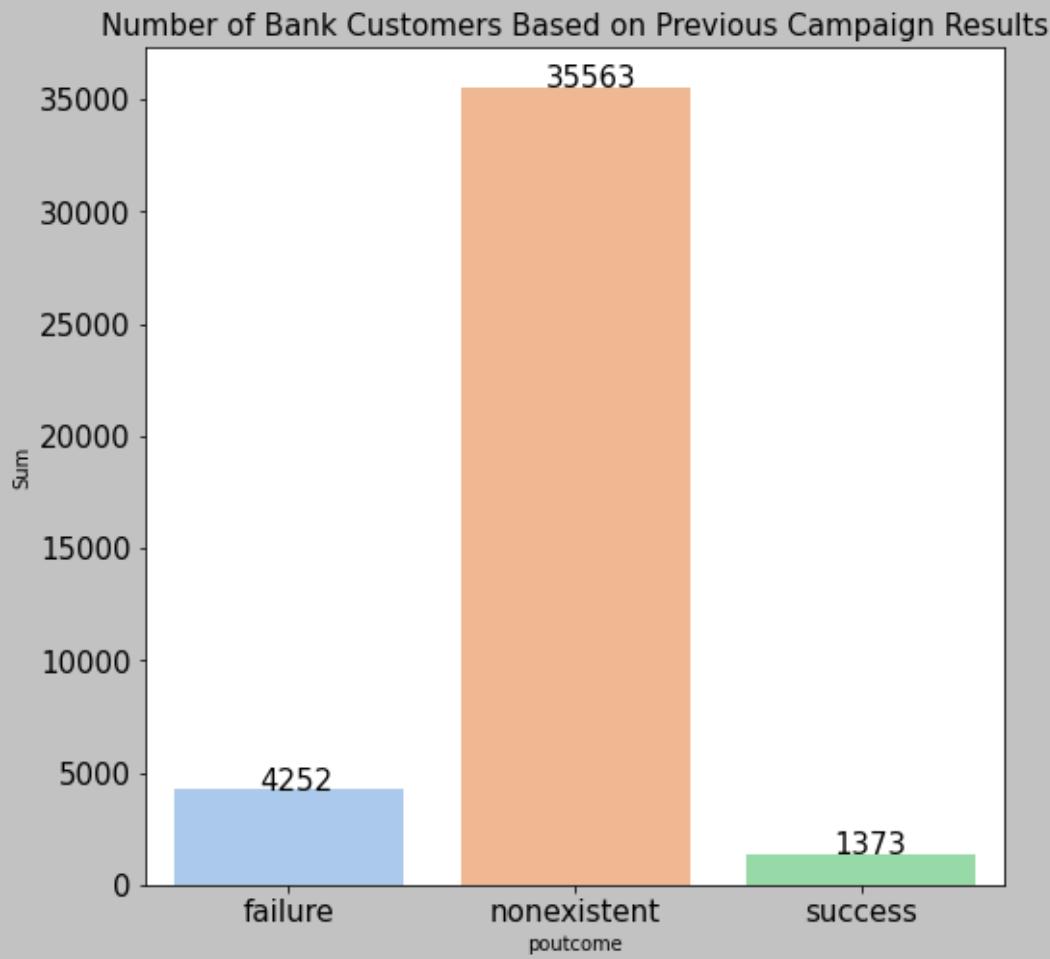
- The customers were contacted again in 12 days and most of them were contacted after three or six days

Previous Campaign Contacts



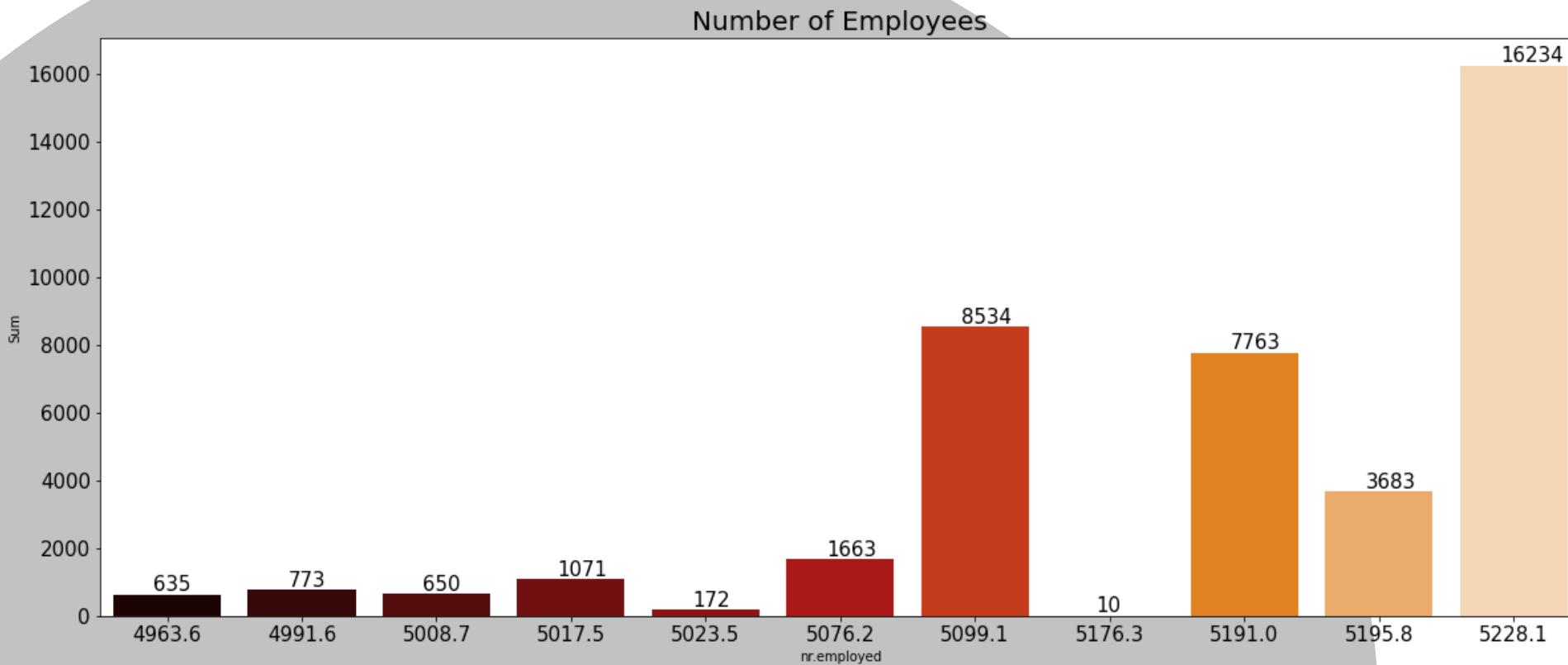
- Most customers only contacted between 0 to 1 times during the previous campaign

Previous Campaign Results



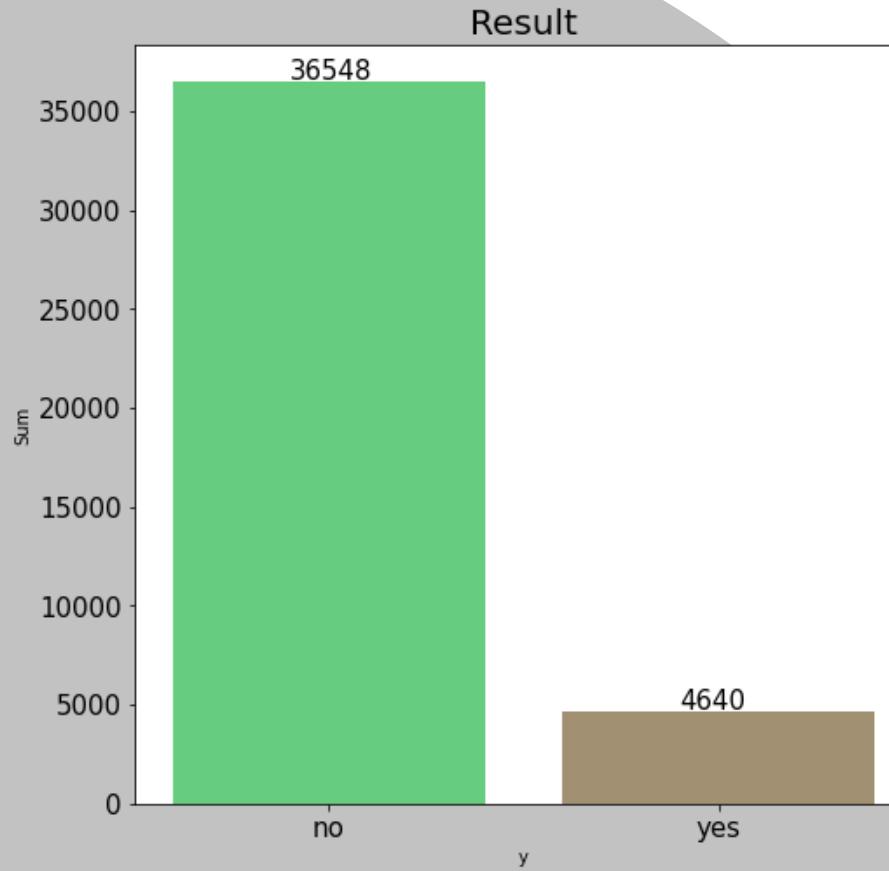
- Most customers were not included in previous campaign and have nonexistent results

Number of Employees



- Most call customers were contacted when number of employees were high

Deposit Subscribed



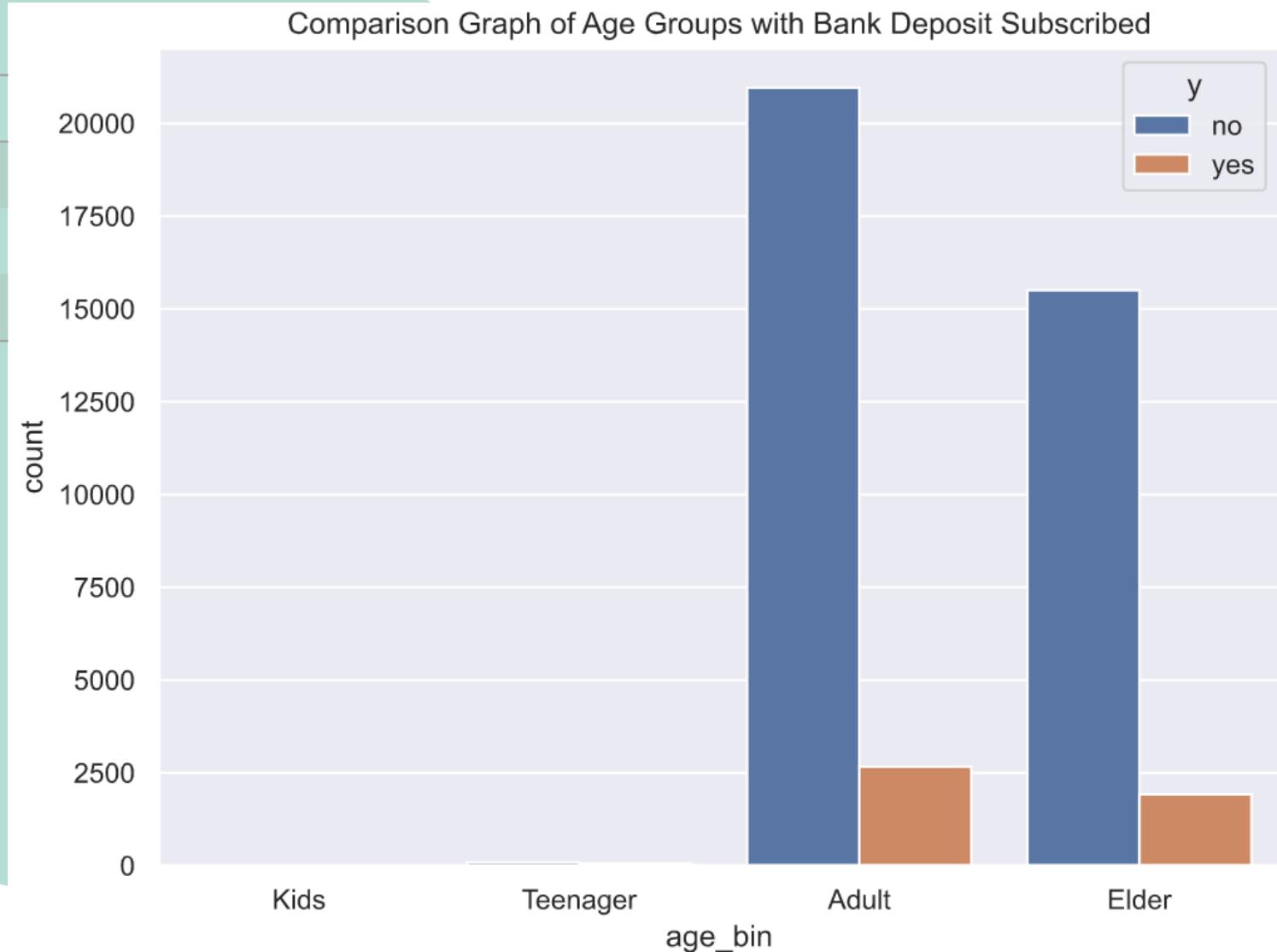
- Most customers didn't subscribe to term deposit offered

MULTI

Age and Deposit Subscribed

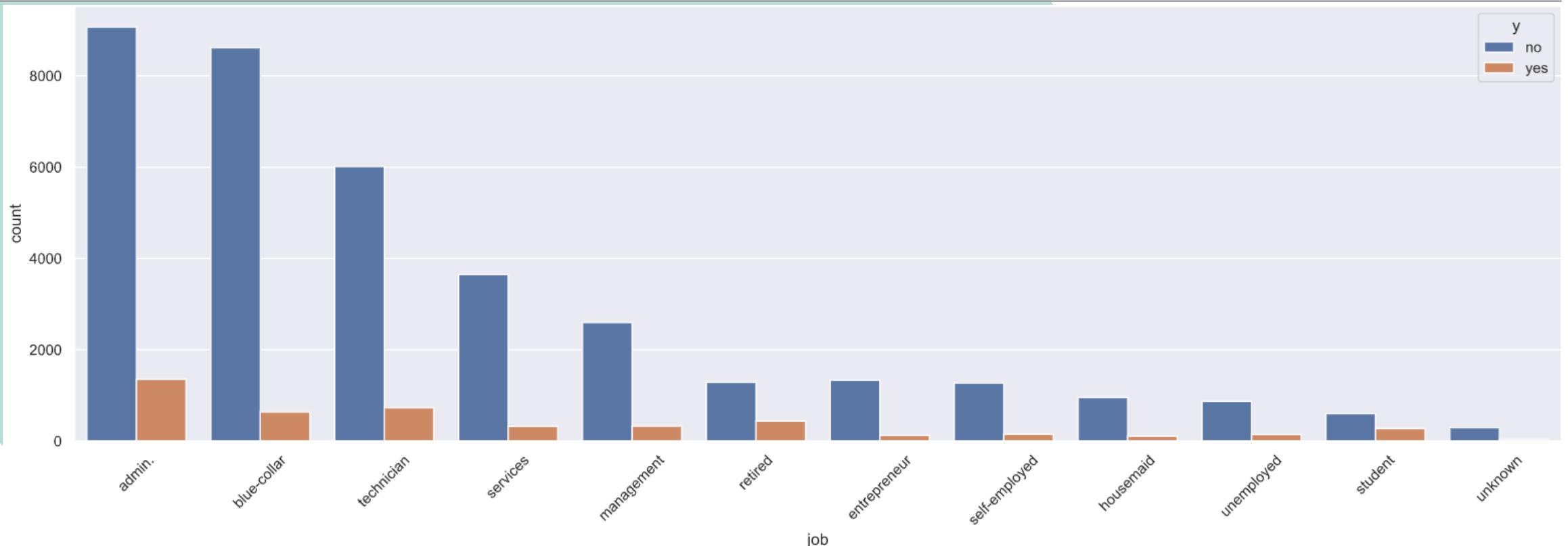
Age_bin	no	yes
Teenager	59.29	40.71
Adult	88.73	11.27
Elder	88.98	11.02

- The adult group dominates deposit subscribed in banks



Job and Deposit Subscribed

job	no	yes	job	no	yes	job	no	yes	job	no	yes
Student	68.57	31.43	Admin	87.03	12.97	Technician	89.17	10.83	Entrepreneur	91.48	8.52
Retired	74.77	25.23	Management	88.78	11.22	Self-employed	89.51	10.49	Services	91.86	8.14
Unemployed	85.80	14.20	Unknown	88.79	11.21	Housemaid	90.00	10.00	Blue-collar	93.11	6.89

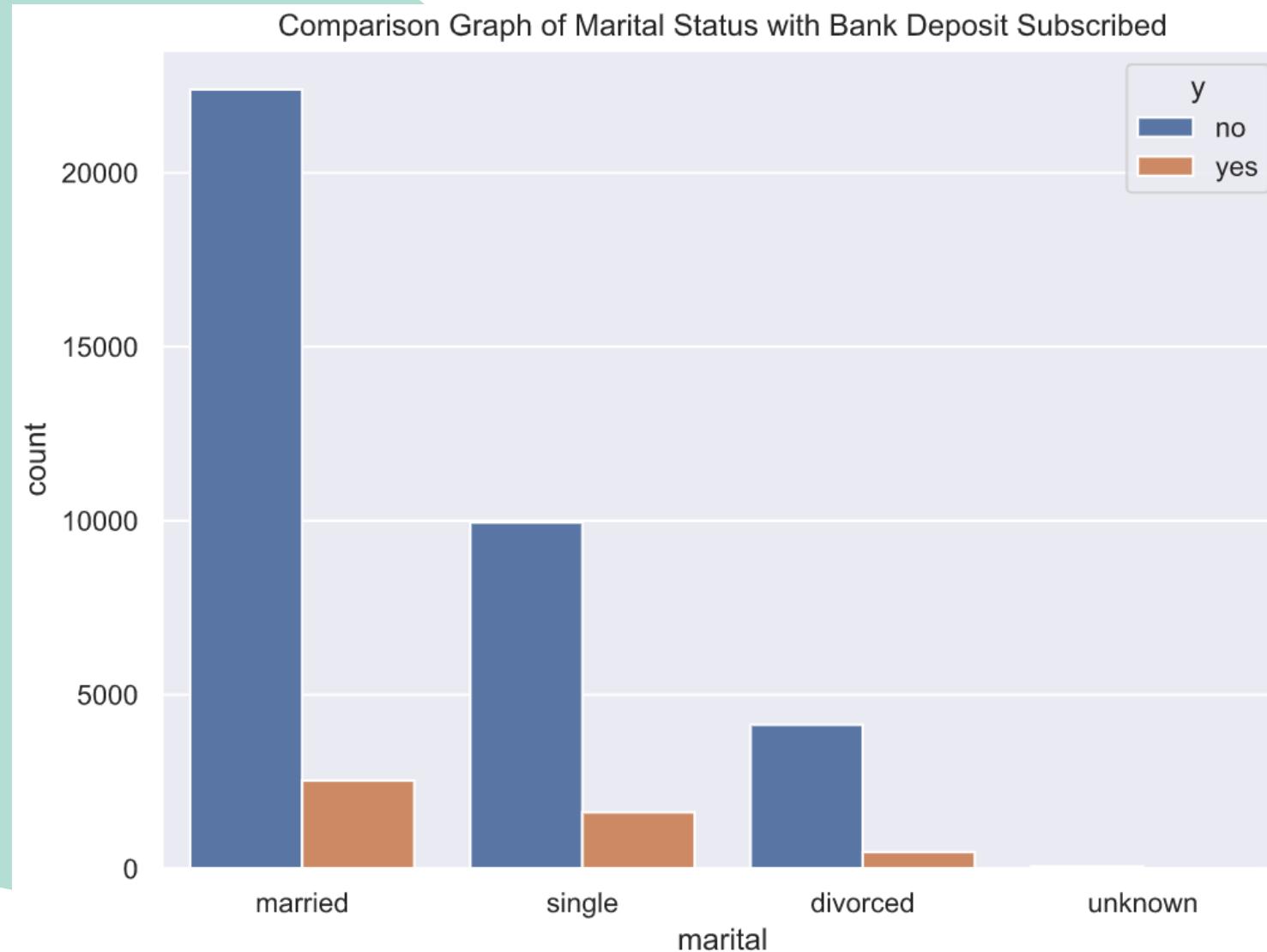


- The admin job-group dominates deposit subscribed in banks by quantity
- The student job-group has highest subscribe percentage

Marital and Deposit Subscribed

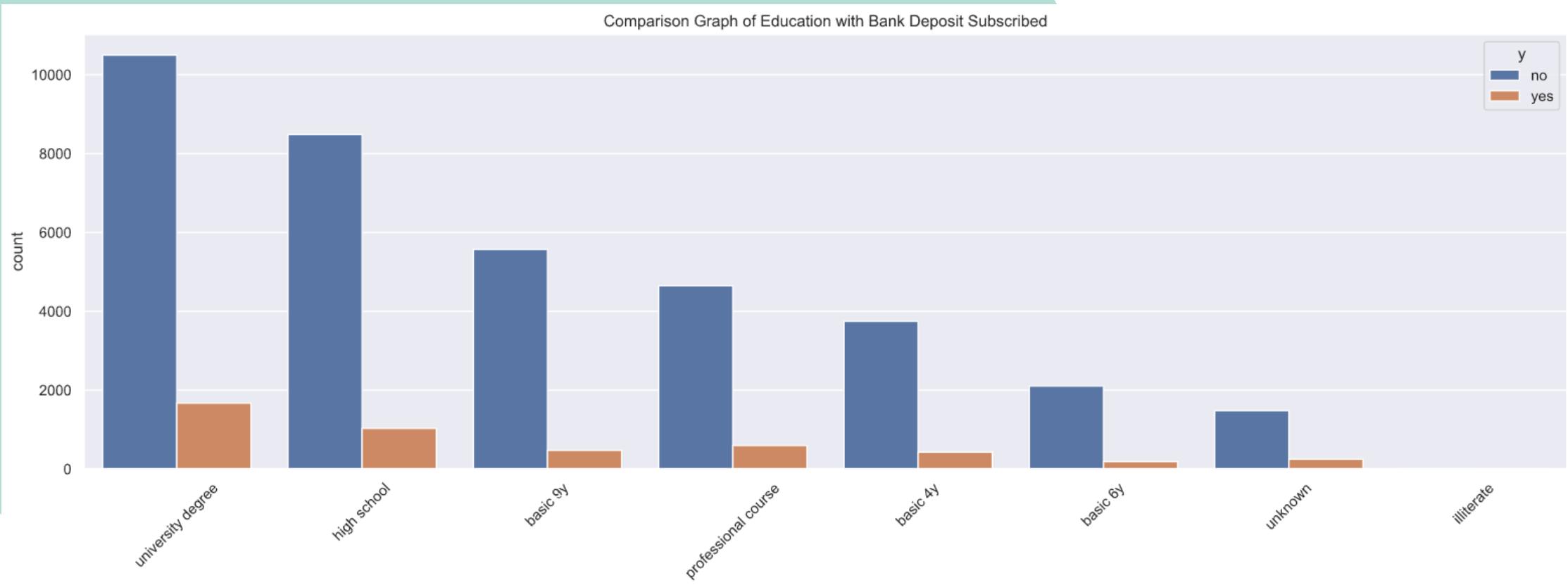
Marital status	no	yes
Divorced	89.68	10.32
Married	89.84	10.16
Single	86.00	14.00
unknown	85.00	15.00

- The marital status as married dominates deposit subscribed in banks
- Unknown marital status has the highest subscribe percentage



Education and Deposit Subscribed

Education	No	yes	education	no	yes	education	no	yes	education	no	yes
Basic 4 year	89.75	10.25	Basic 9 year	92.18	7.82	Illiterate	77.78	22.22	University degree	86.28	13.72
Basic 6 year	91.80	8.20	High school	89.16	10.84	Professional course	88.65	11.35	unknown	85.50	14.50

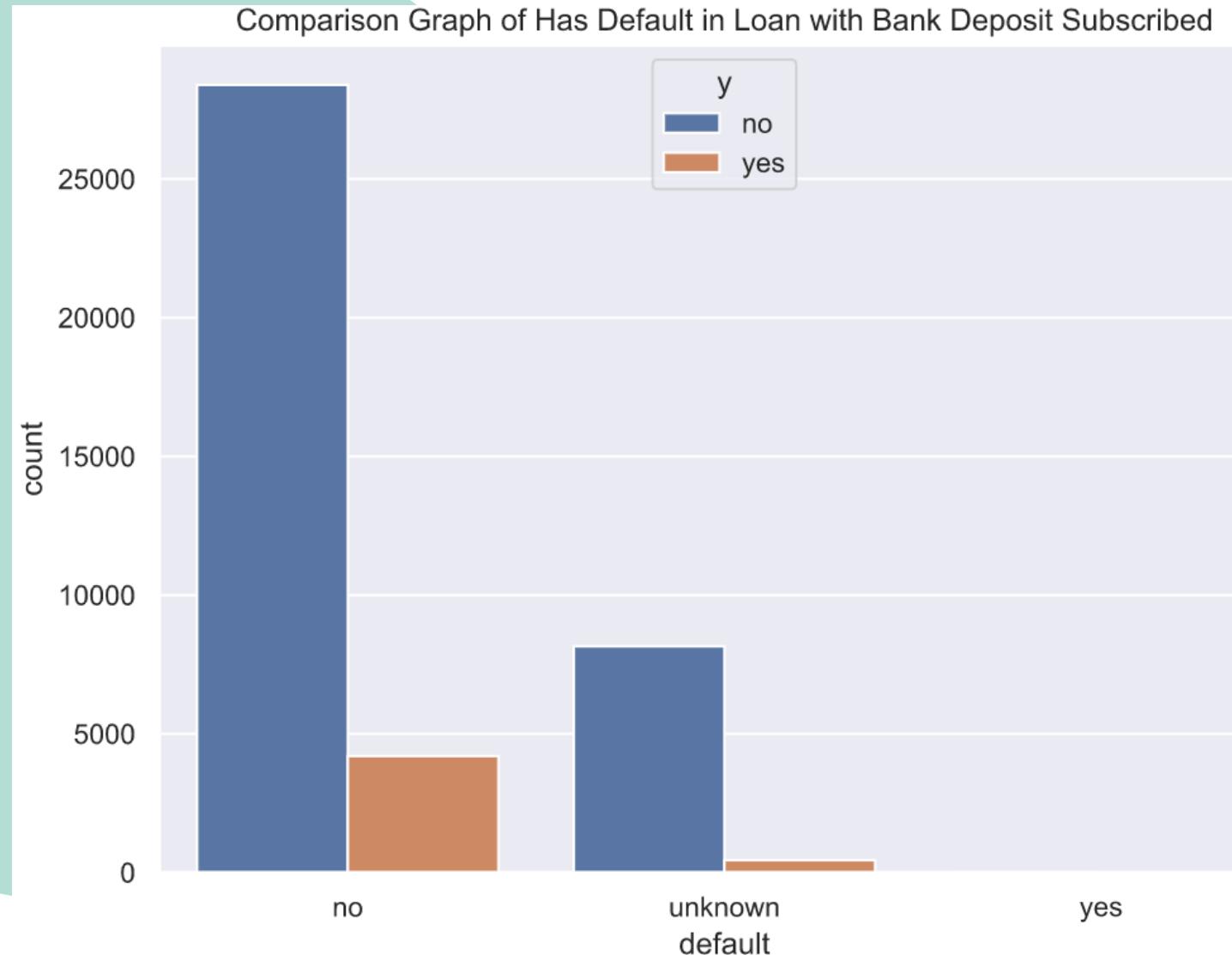


- The education level as university degree dominates deposit subscribed in banks
- Illiterate has the highest subscribe percentage but has the smallest quantity

Has Default in Loan and Deposit Subscribed

Default	no	yes
No	87.12	12.88
Unknown	94.85	5.15
Yes	100.00	0.00

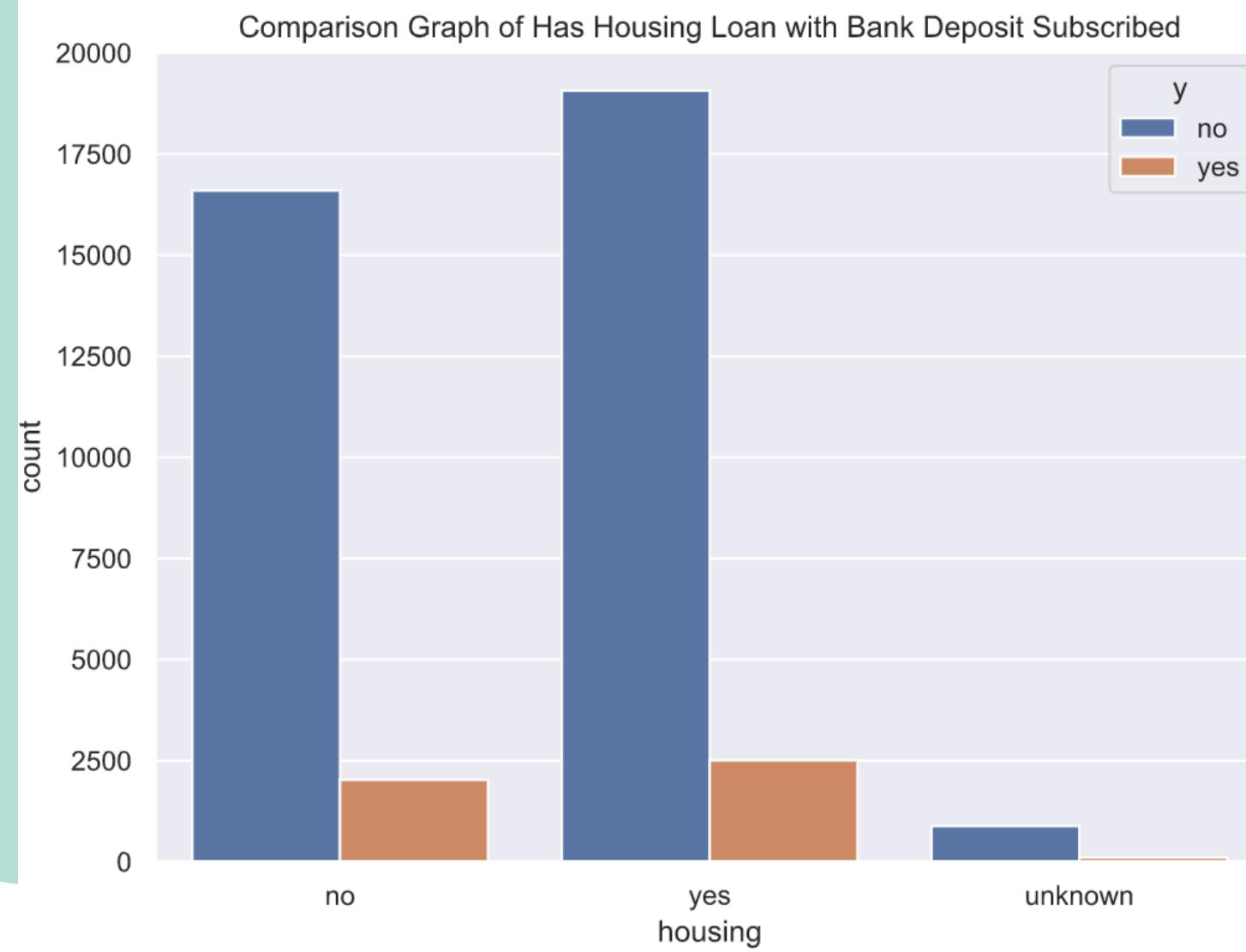
- Group which has no default loan dominates deposit subscribed in banks by quantity and percentage



Has housing loan and deposit subscribed

Housing	no	yes
No	89.12	10.88
Unknown	89.19	10.81
Yes	88.38	11.62

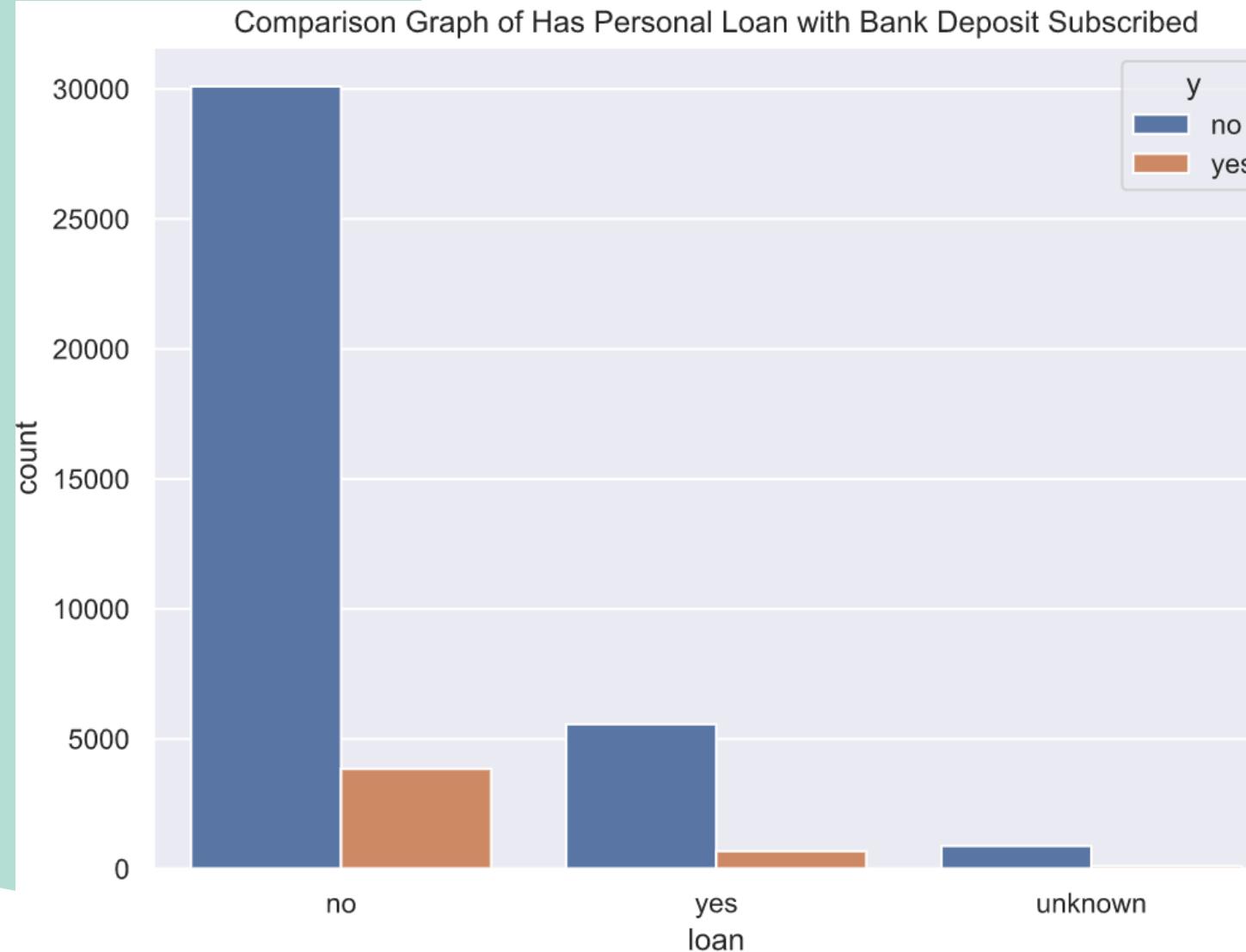
- Group of clients which has housing loan dominates deposit subscribed in banks by quantity and percentage



Has personal loan and deposit subscribed

Loan	no	yes
No	88.66	11.34
Unknown	89.19	10.81
Yes	89.07	10.93

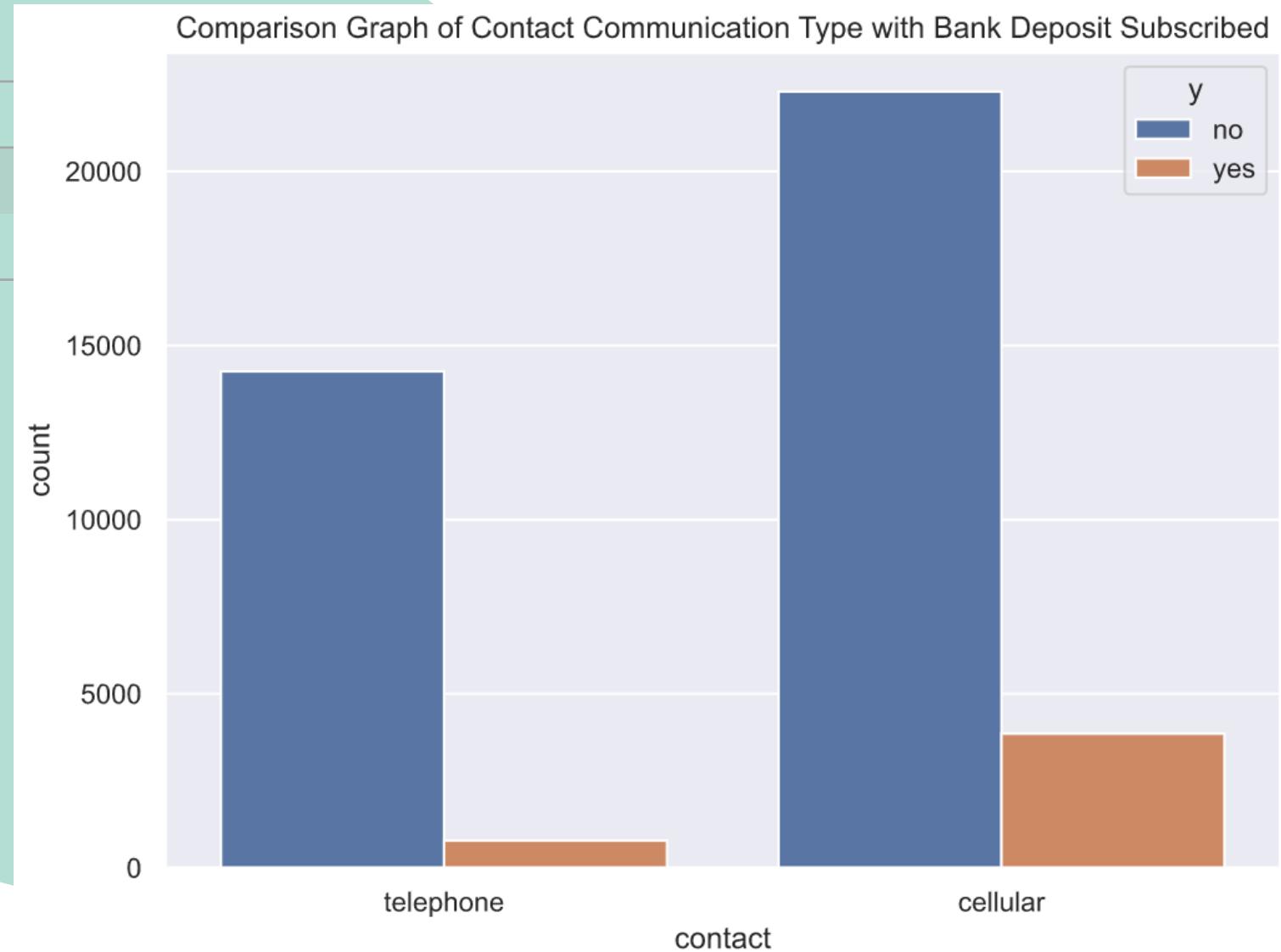
- Group which has no personal loan dominates deposit subscribed in banks by quantity and percentage



Communication type and deposit subscribed

Contact	no	yes
Cellular	85.26	14.74
Telephone	94.77	5.23

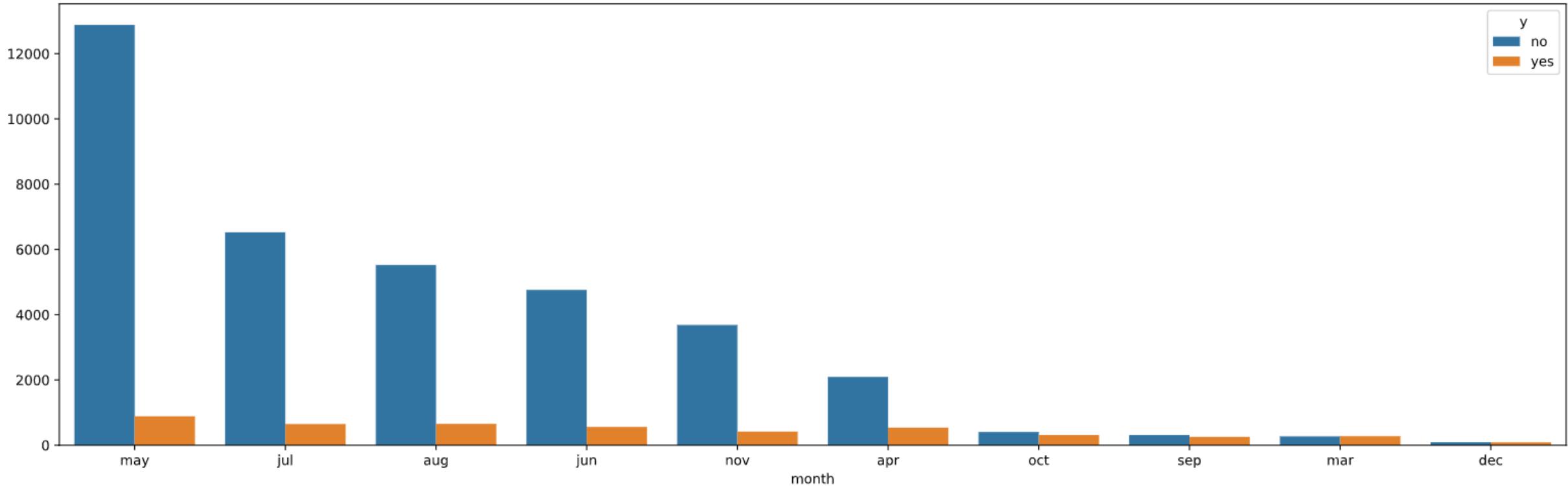
- Group which contacted by cellular dominates deposit subscribed in banks by quantity and percentage



Last month contacted and deposit subscribed

month	no	yes												
Mar	49.45	50.55	May	93.57	6.43	Jul	90.95	9.05	Sep	55.09	44.91	Nov	89.86	10.14
Apr	79.52	20.48	Jun	89.49	10.51	Aug	89.40	10.60	Oct	56.13	43.87	Dec	51.10	48.90

Comparison Graph of Month Contacted with Bank Deposit Subscribed

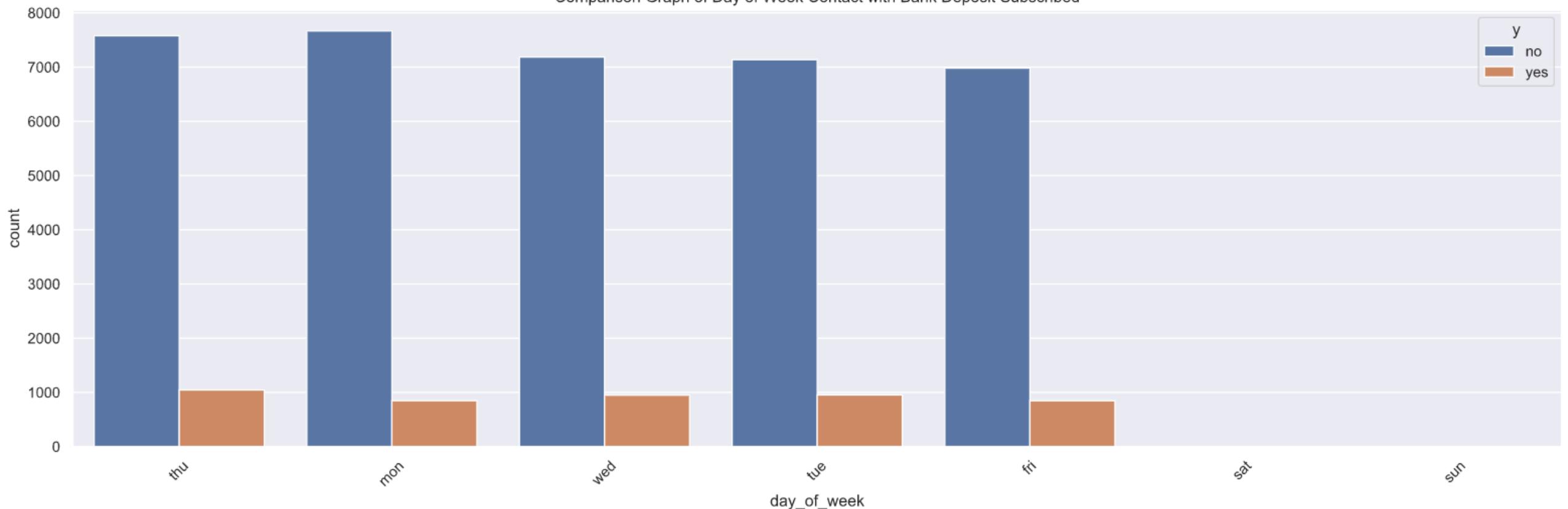


- Most customer was contacted in May and even though May has the lowest subscribe percentage, the quantity of deposit is the highest
- Mar has the highest subscribe percentage but is the second lowest by quantity

Day of Week contacted and Deposit Subscribed

Day of week	no	yes	Day of week	no	yes	Day of week	no	yes	Day of week	no	yes	Day of week	no	yes
mon	90.05	9.95	tue	88.22	11.78	wed	88.33	11.67	thu	87.88	12.12	fri	89.19	10.81

Comparison Graph of Day of Week Contact with Bank Deposit Subscribed

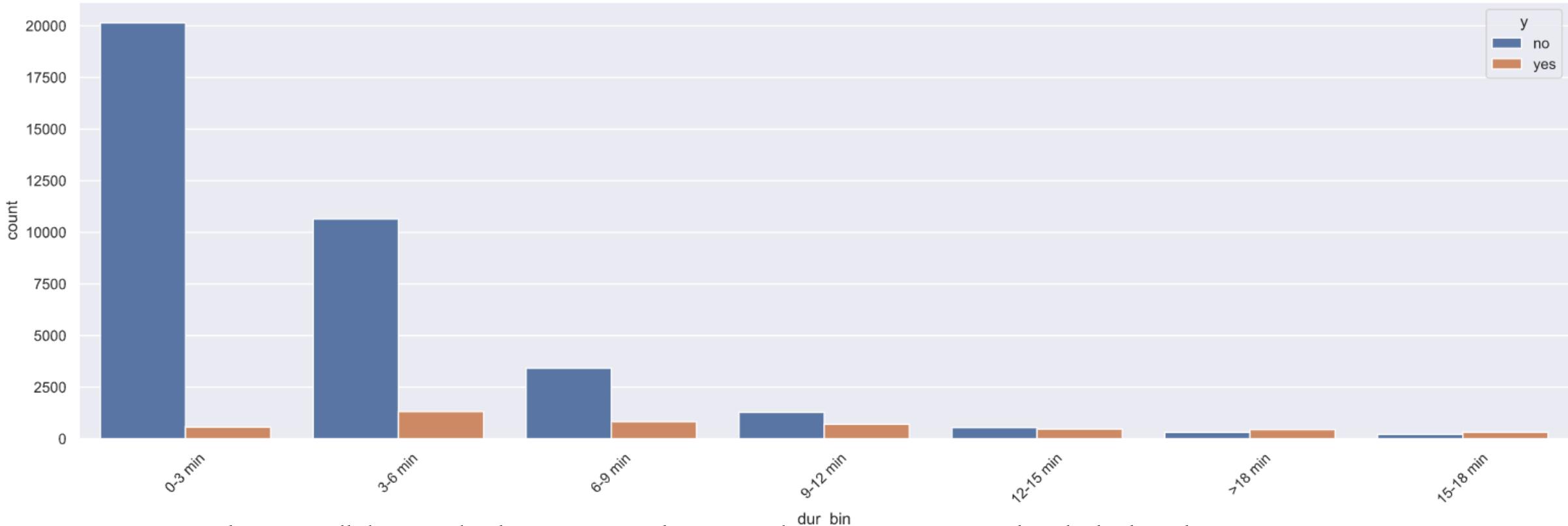


- Last day of week communication between client and campaign team was dominated in Thursday

Duration Contacted and Deposit Subscribed

Duration	no	yes	Duration	no	yes	Duration	no	yes	Day of week	no	yes
0 – 3 minutes	97.28	2.72	6 – 9 minutes	80.52	19.48	12 – 15 minutes	53.57	46.43	> 18 min	41.11	58.89
3 – 6 minutes	88.98	11.02	9 – 12 minutes	64.58	35.42	15 – 18 minutes	40.04	59.96			

Comparison Graph of Last Contact Duration with Bank Deposit Subscribed

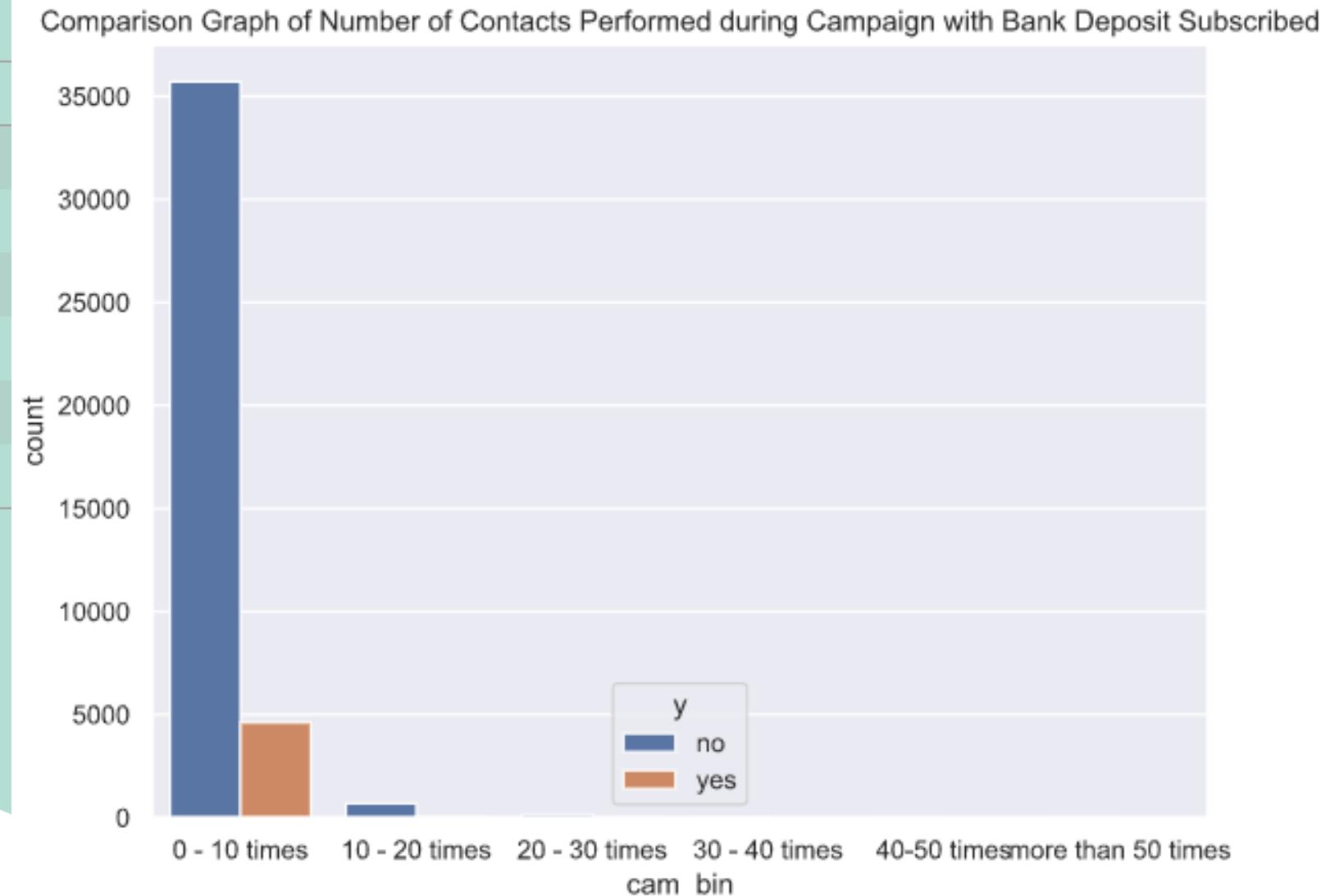


- Last duration calls between bank customers and campaign by range 3 - 6 minutes has the highest deposit quantity
- There is a trend that shows the longer the duration, the greater the chance for the customer to deposit

Campaign and deposit subscribed

Cam_bin	no	yes
0 – 10 times	88.56	11.44
10 – 20 times	96.35	3.65
20 – 30 times	99.19	0.81
30 – 40 times	100.00	0.00
40 – 50 times	100.00	0.00
More than 50 times	100.00	0.00

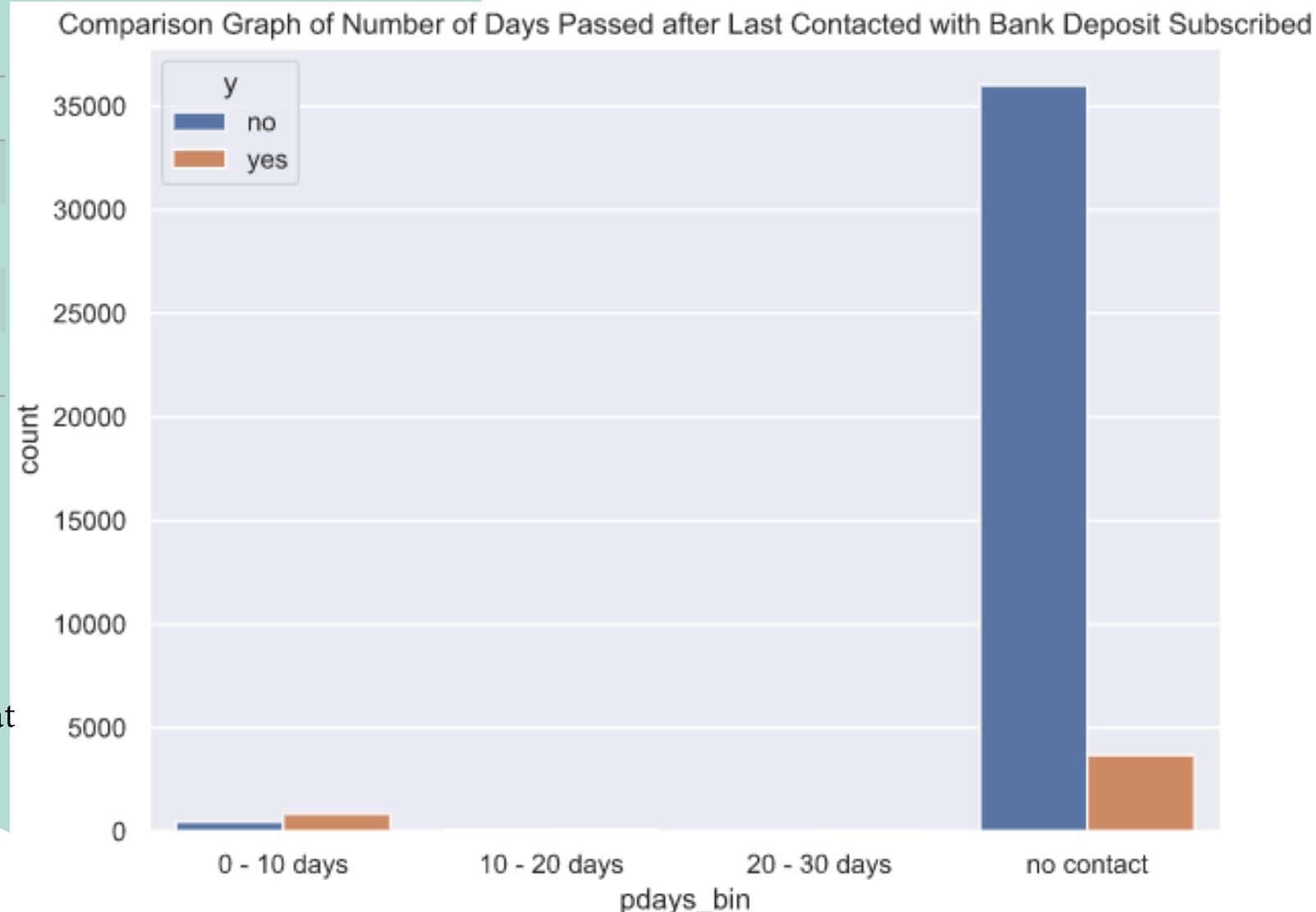
- Number of calls between bank customers and campaign team was dominated by range 0 - 10 times
- If the customer already contacted more than 10 times, the probability of deposit become slim to none



Previous Days and Deposit Subscribed

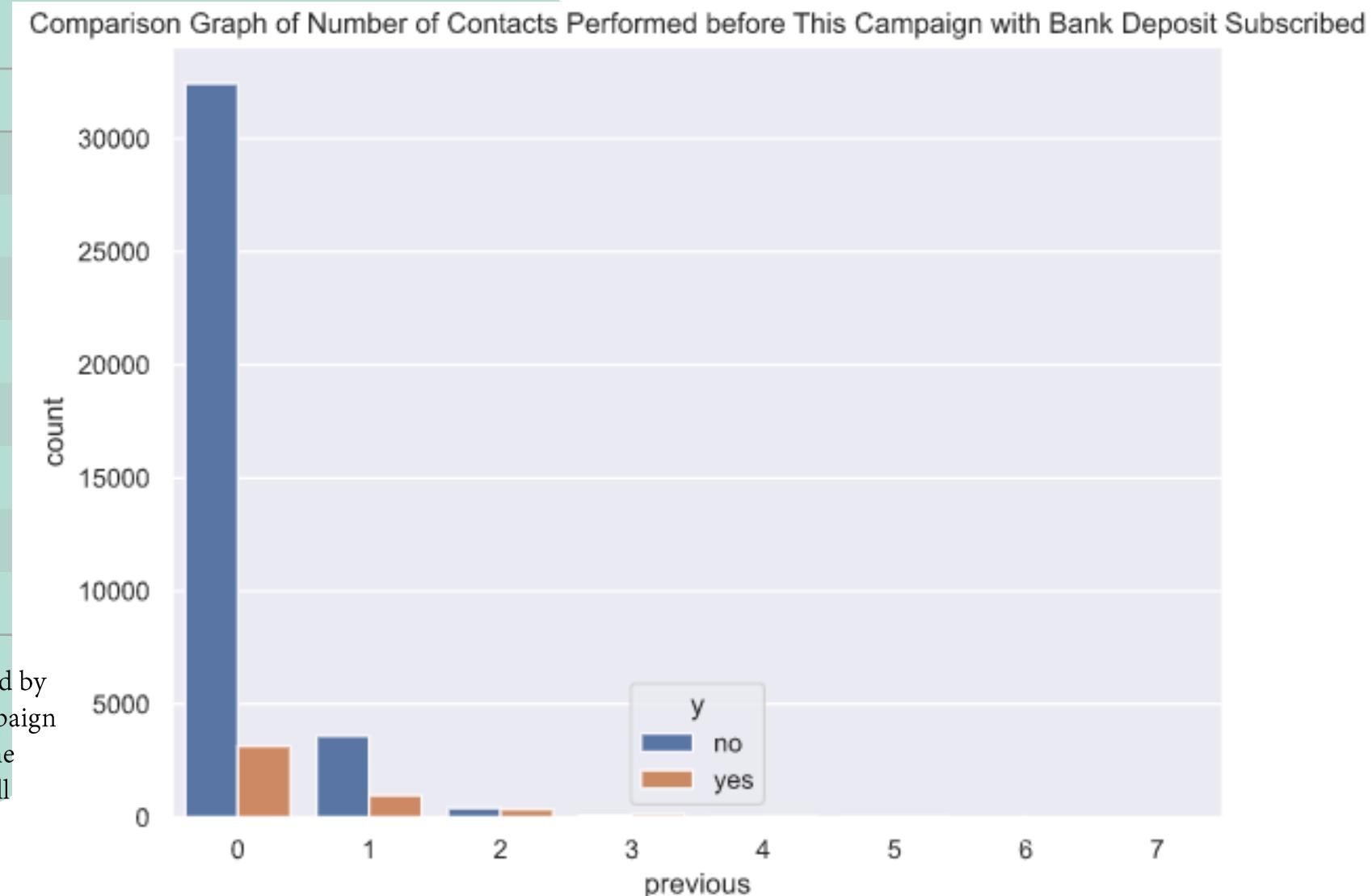
Pdays_bin	no	yes
0 – 10 days	35.11	64.89
10 – 20 days	44.39	55.61
20 – 30 days	12.50	87.50
No contact	90.74	9.26

- Clients who choose to subscribe deposits are dominated by who were not contacted from the previous campaign
- If the customer has been contacted in the previous campaign, there is higher chance that they will deposit



Number of previous contact and deposit subscribed

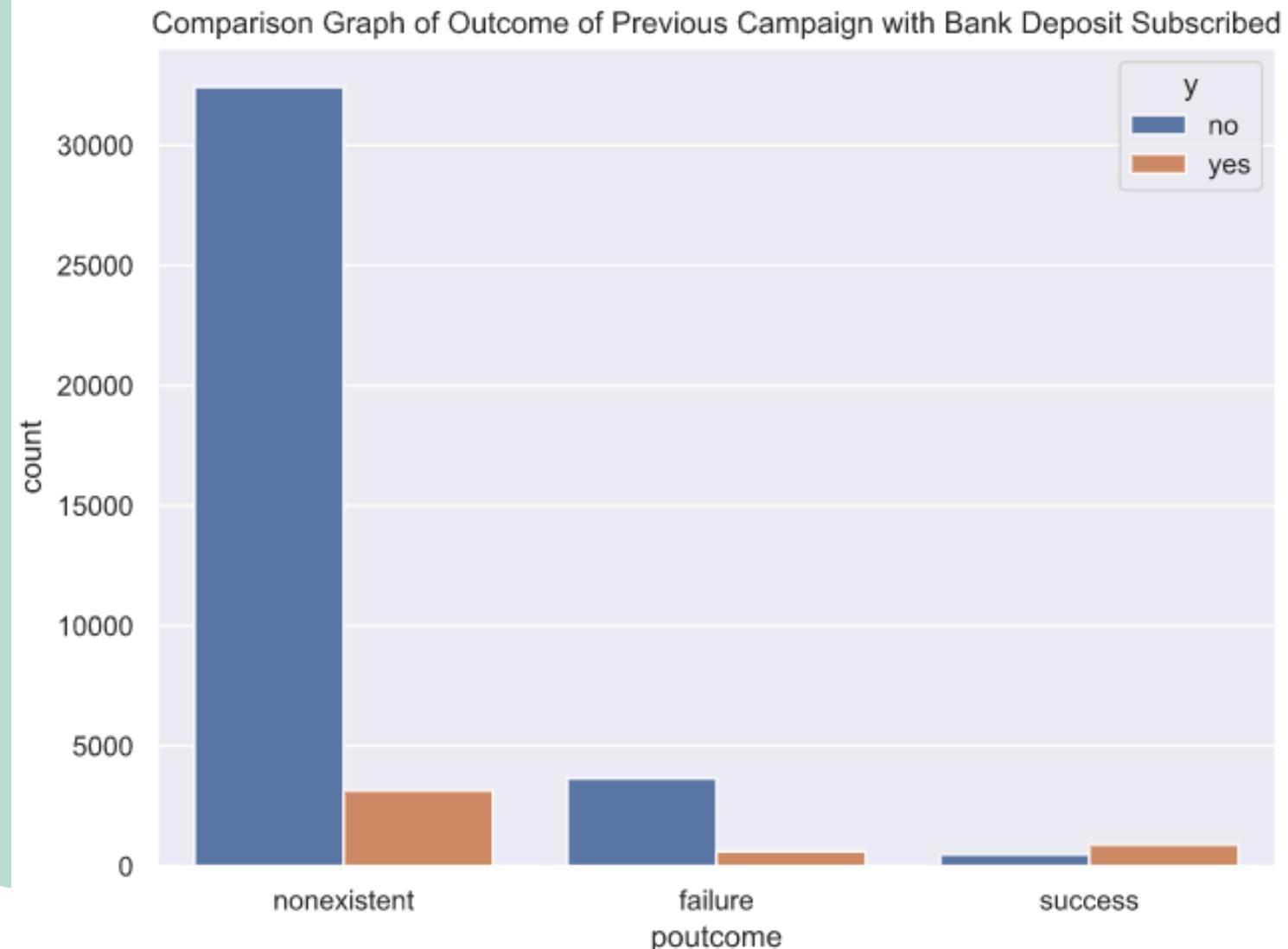
Previous contact	no	yes
0	91.17	8.83
1	78.80	21.20
2	53.58	46.42
3	40.74	59.26
4	45.71	54.29
5	27.78	72.22
6	40.00	60.00
7	100.00	0.00



- Clients who choose to subscribe deposits are dominated by clients who were not contacted from the previous campaign
- If the customer was contacted between 2 - 7 times in the previous campaign, there is higher chance that they will subscribe

Previous outcome and deposit subscribed

Poutcome	no	yes
Failure	85.77	14.23
Nonexistent	91.17	8.83
Success	34.89	65.11

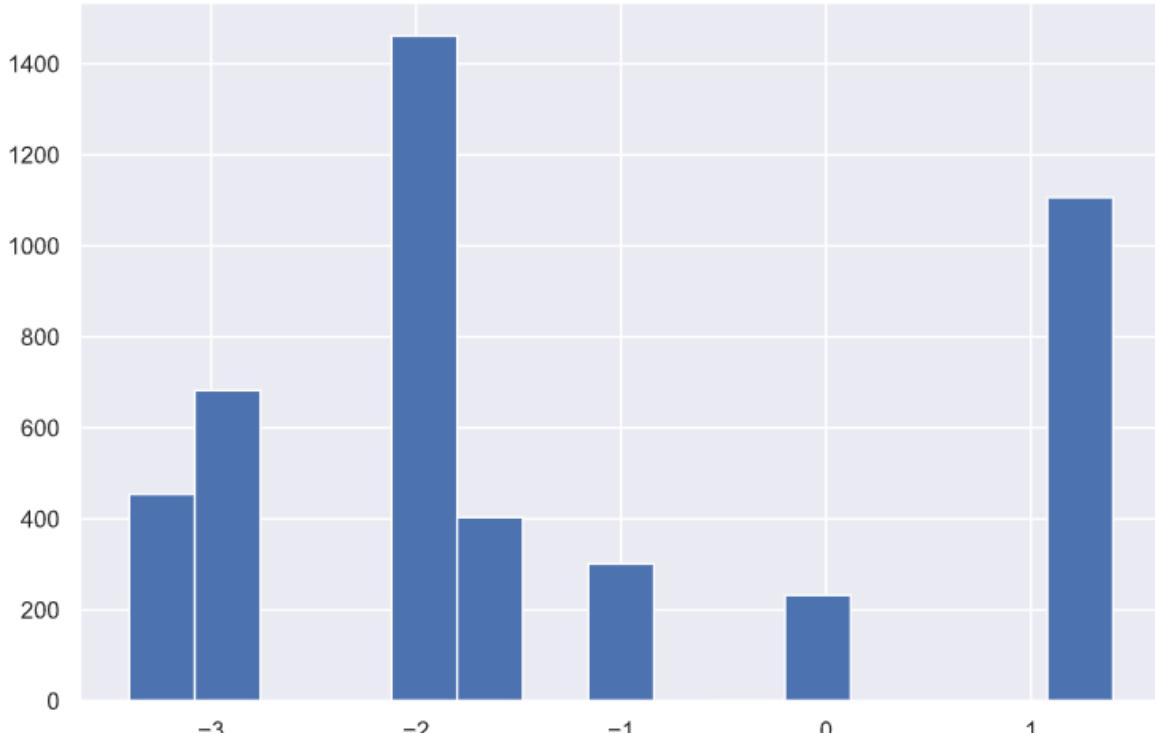


- Clients who choose to subscribe deposits are dominated by clients whose previous outcome of campaign were nonexistent
- Customer who subscribe at the previous campaign highest subscribe percentage in the current campaign

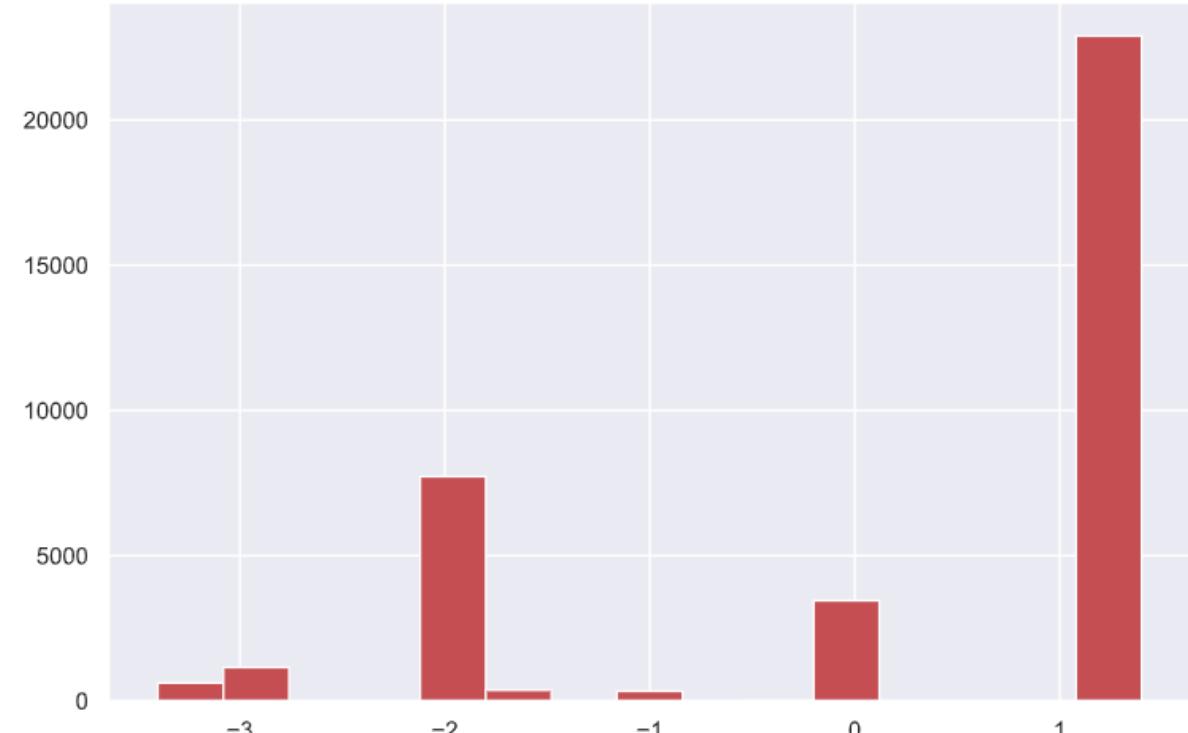
Employee variation rate and deposit subscribed

Emp var rate	no	yes	Emp var rate	no	yes									
-3.4	56.71	42.39	-2.9	64.28	35.72	-1.7	47.87	52.13	-0.2	90.00	10.00	1.1	96.91	3.09
-3.0	48.48	51.16	-1.8	84.09	15.91	-1.1	52.60	47.40	-0.1	93.70	6.30	1.4	94.67	5.33

Comparison Graph of Employee Variation Rate and Bank Deposit Subscribed as yes



Comparison Graph of Employee Variation Rate and Bank Deposit Subscribed as no

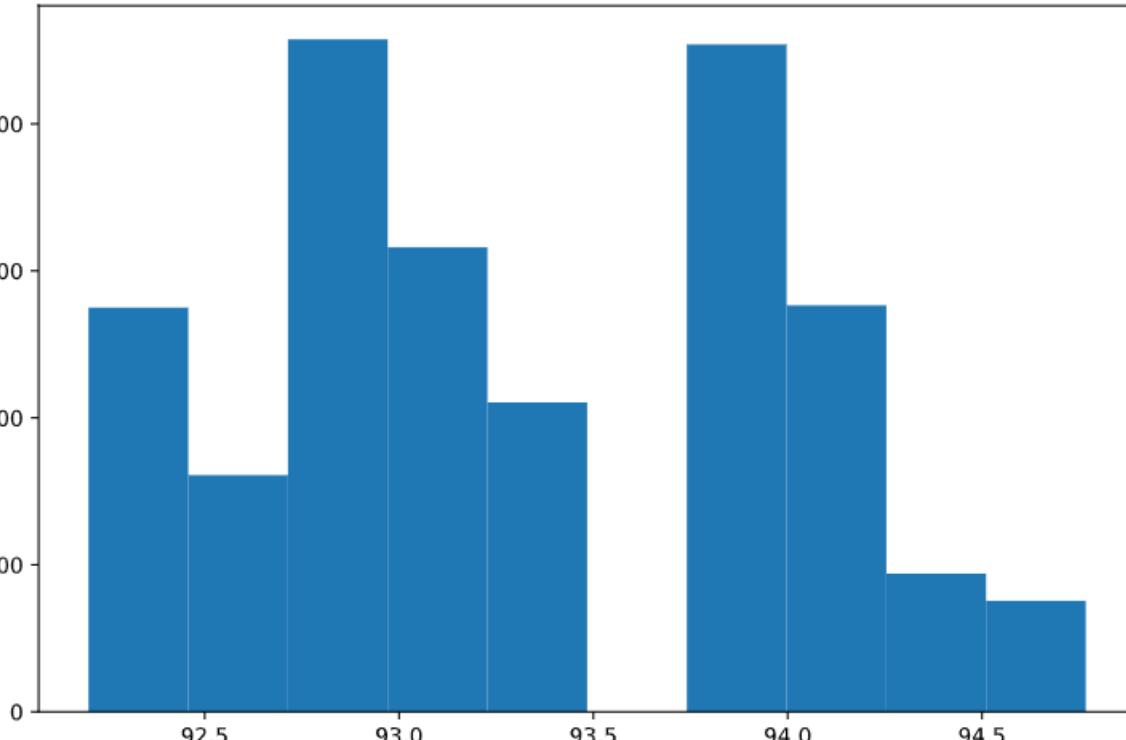


- Clients who choose to subscribe deposits are dominated by clients whose employee variation rate as -1.8
- There is a trend that shows if the employee variation rate become better (less negative or more positive), the subscribe percentage become lower

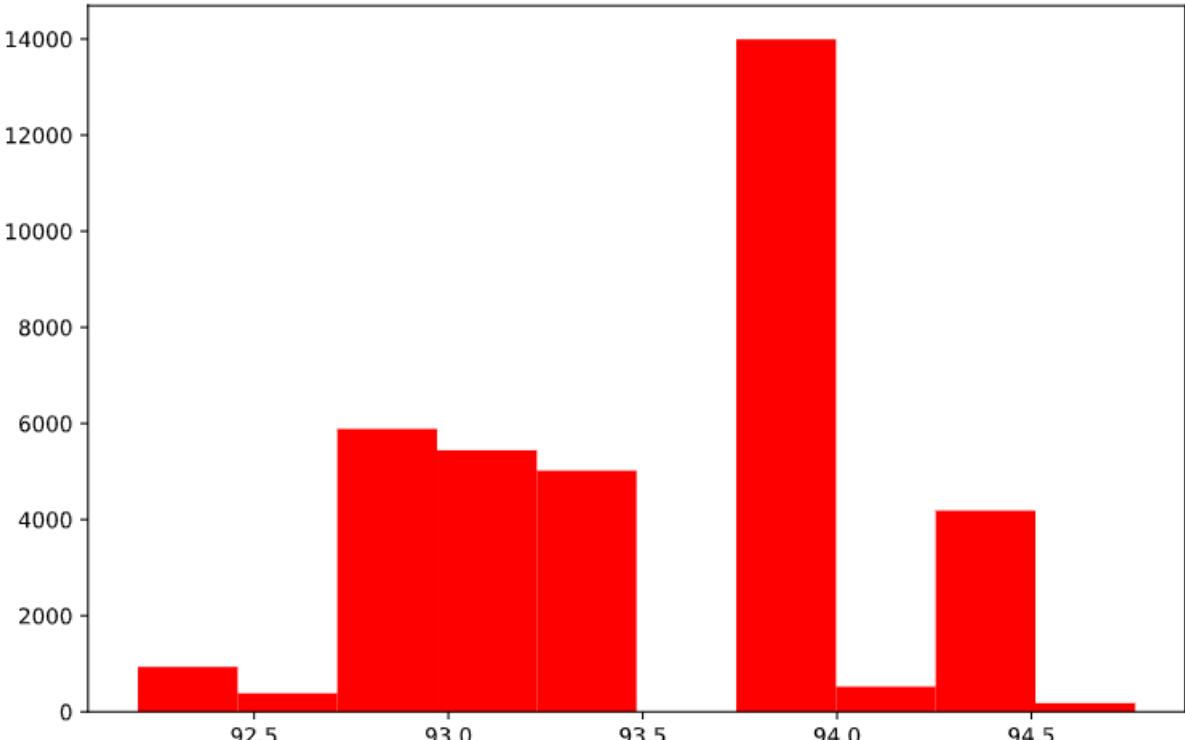
Consumer price index and deposit subscribed

Cons price idx	no	yes	Cons price idx	no	yes												
92.201	65.71	34.29	92.713	48.84	51.16	93.075	82.02	17.98	93.798	37.31	62.69	94.055	53.28	46.27	94.767		
92.379	60.30	39.70	92.756	90.00	10.00	93.200	94.75	5.25	93.876	42.45	57.55	94.199	50.50	49.50			
92.431	59.73	40.27	92.843	55.32	44.68	93.369	43.18	56.82	93.918	93.91	6.09	94.215	43.41	56.59			
92.469	62.92	37.08	92.893	90.96	9.04	93.444	94.76	5.24	93.994	96.91	6.09	94.465	95.70	4.30			
92.649	52.94	47.06	92.963	63.08	36.92	93.749	44.25	55.75	94.027	48.50	51.50	94.601	54.41	45.59			

Comparison Graph of Consumer Price Index with Bank Deposit Subscribed as yes



Comparison Graph of Consumer Price Index with Bank Deposit Subscribed as no

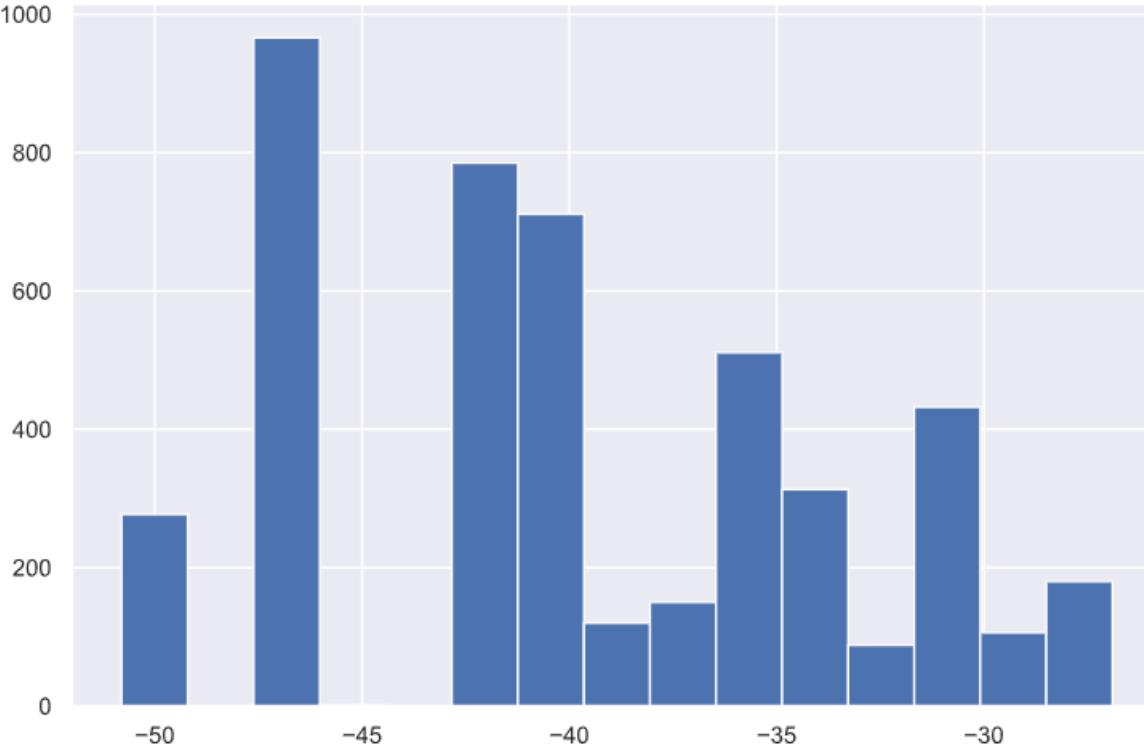


- Client who choose to subscribe are dominated by clients whose consumer price index was 92.893

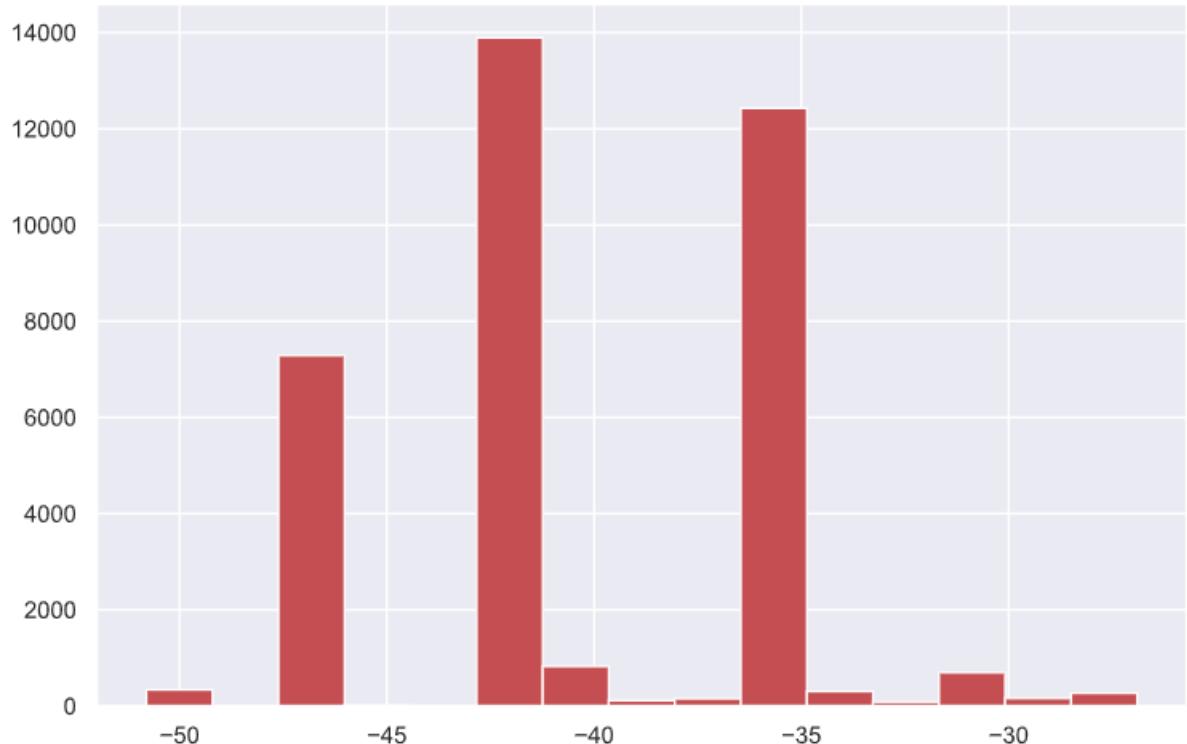
Consumer confidence index and deposit subscribed

Cons confidx	no	yes															
-50.8	54.69	54.31	-45.9	90.00	10.00	-40.4	37.31	62.69	-37.5	50.50	49.50	-33.6	62.92	37.08	-26.9	59.73	40.27
-50.0	55.32	44.68	-42.7	93.91	6.09	-40.3	43.41	56.59	-36.4	96.91	3.09	-33.0	48.84	51.16			
-49.5	54.41	45.59	-42.0	94.75	5.25	-40.0	42.45	57.55	-36.1	94.76	5.24	-31.4	65.71	34.29			
-47.1	82.02	17.98	-41.8	95.70	4.03	-39.8	53.28	46.72	-34.8	43.18	56.82	-30.1	52.94	47.06			
-46.2	90.96	9.04	-40.8	63.08	36.92	-38.3	48.50	51.50	-34.6	44.25	55.75	-29.8	60.30	39.70			

Comparison Graph of Consumer Confidence Index with Bank Deposit Subscribed as yes



Comparison Graph of Consumer Confidence Index with Bank Deposit Subscribed as no

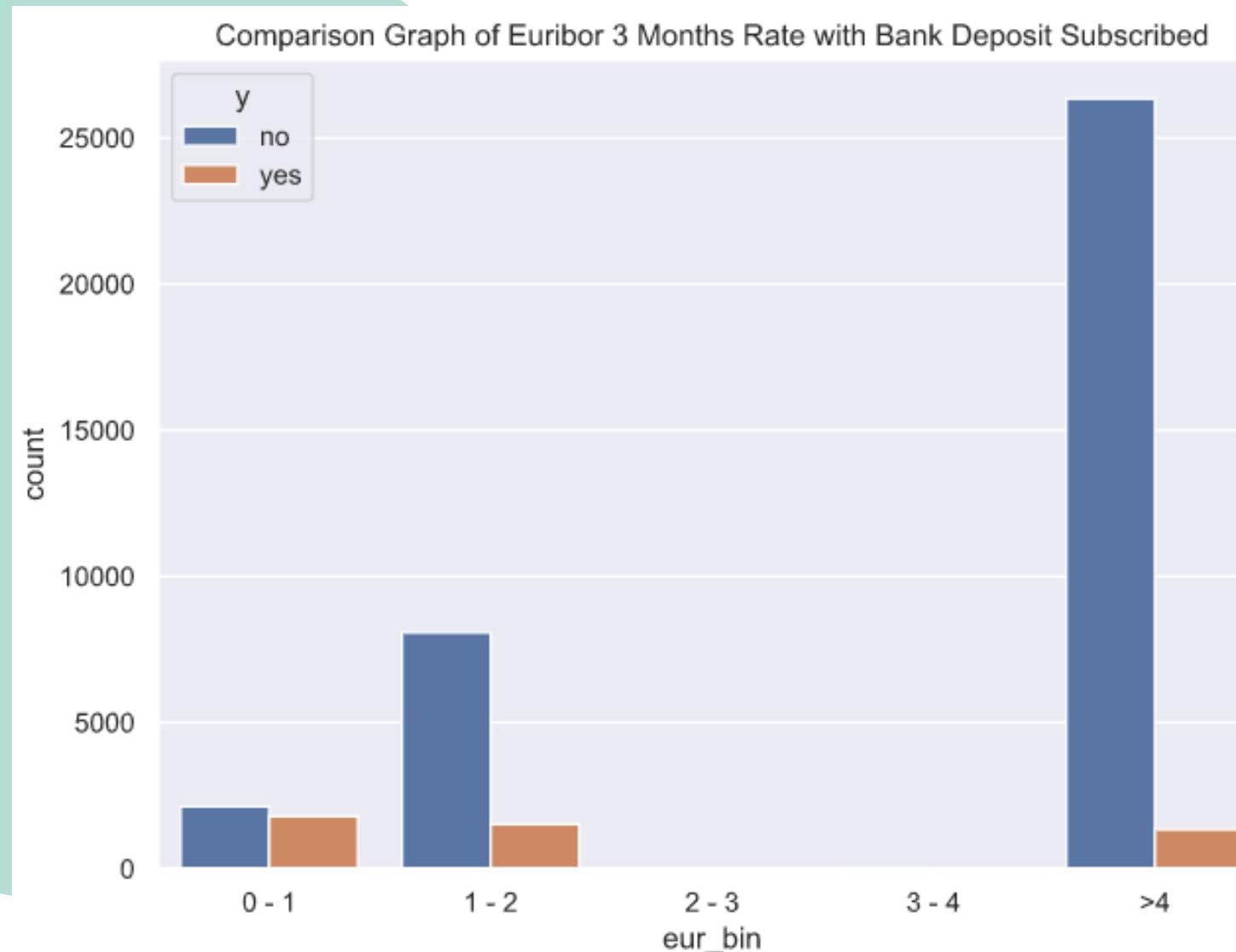


- Clients who choose to subscribe deposits are dominated by clients whose consumer confidence index as -46.2

Euribor 3 months rate and deposit subscribed

Eur_bin	no	yes
0 - 1	54.30	45.70
1 - 2	84.20	15.80
3 - 4	92.86	7.14
> 4	95.17	4.83

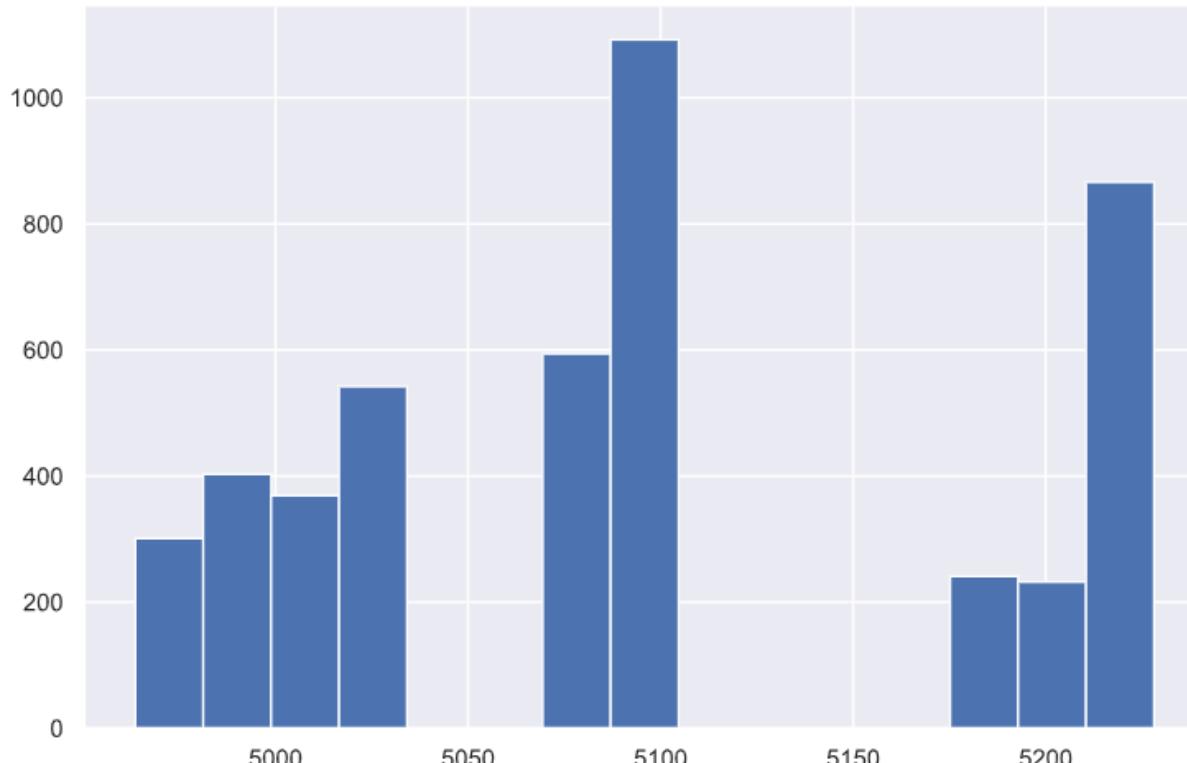
- Clients who choose to subscribe deposits are dominated by clients whose euribor 3 months rate in range 0 - 1 by quantity
- If euribor become higher then subscribe percentage become lower



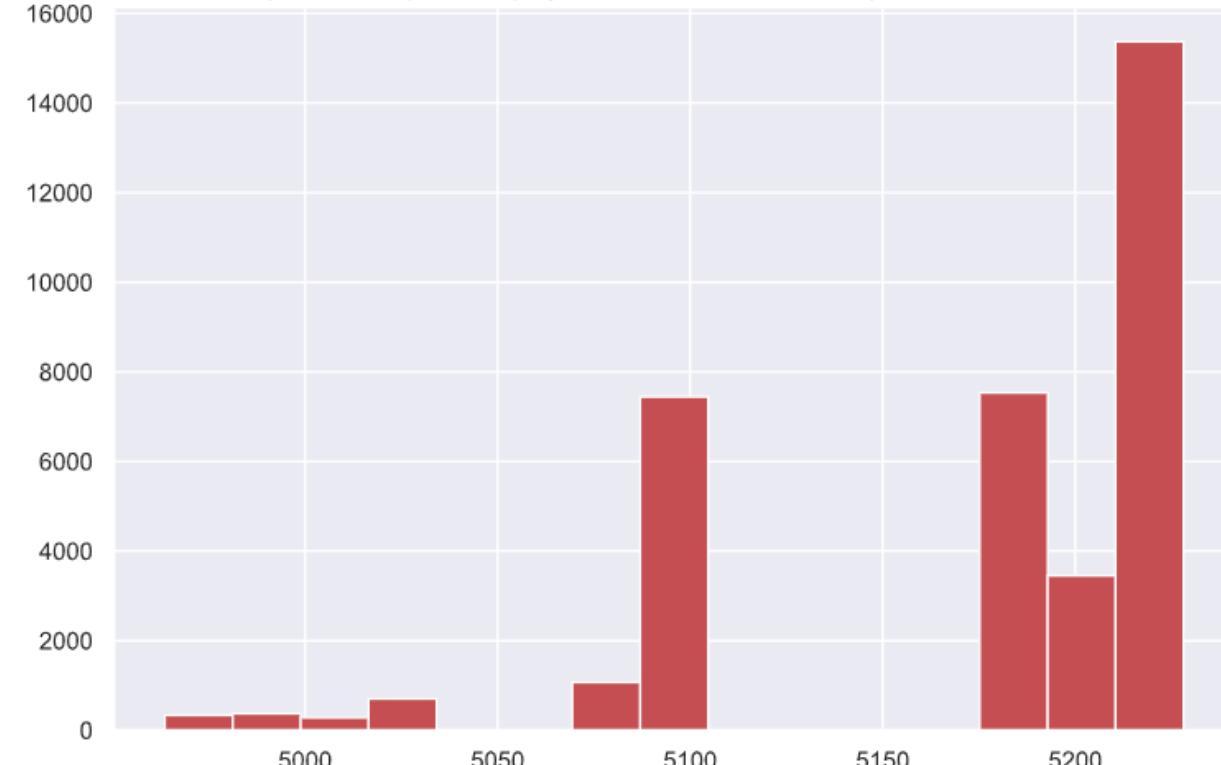
Number of employees and deposit subscribed

Nr employed	no	yes	Nr employed	no	yes	Nr employed	no	yes	Nr employed	no	yes
4963.6	52.60	47.40	5017.5	57.61	42.39	5099.1	87.20	12.80	5195.8	93.70	6.30
4991.6	47.87	52.13	5023.5	48.84	51.16	5176.3	90.00	10.00	5228.1	94.67	5.33
5008.7	43.23	56.77	5076.2	64.28	35.72	5191.0	96.91	3.09			

Comparison Graph of Employees Number with Bank Deposit Subscribed as yes



Comparison Graph of Employees Number with Bank Deposit Subscribed as no



- Clients who choose to subscribe deposits are dominated by number of employee at 5228 workers

MACHINE LEARNING / MODELLING

```
padding: 4px 6px;  
text-align: left;  
  
  &:hover {  
    color: $c-link-hover;  
  }  
  
&.selected {  
  background-color: $c-action;  
  color: white;  
}  
  
.amount {  
  float: right;  
  font-weight: bold;  
}
```

OVERVIEW

The model used before running the campaign

Using features that can be obtained before the campaign starts

DATA PREPARATION

DROP COLUMNS

Column
'duration' &
'campaign'

NULL VALUES

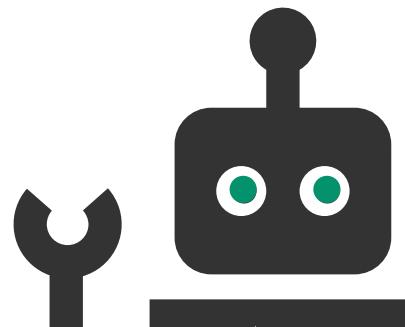
CHECK the
null values in
all columns

CHANGE VALUES

In target
column into
numeric
values

EN CODING

The
categorical
columns



ENCODING

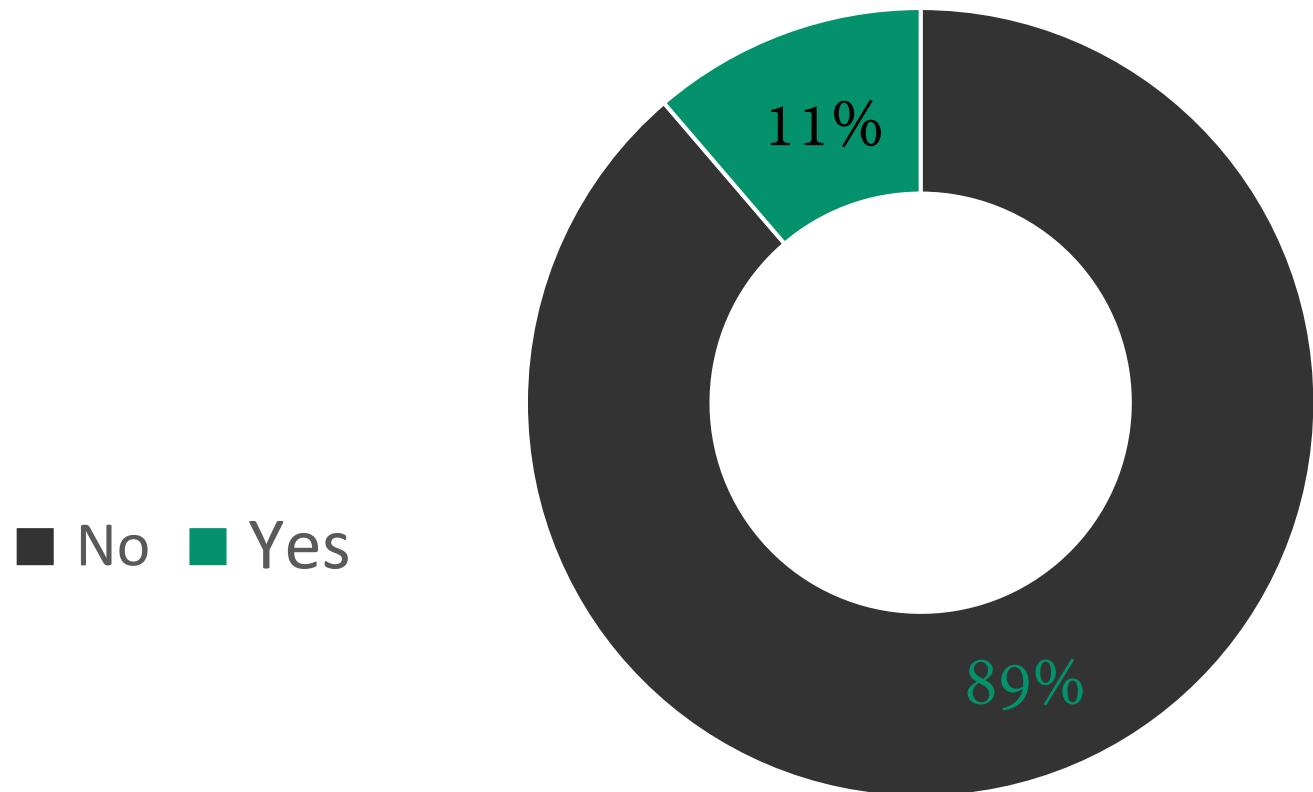
LABEL ENCODING

A	A
X	1
Y	2
Z	3

ONE HOT ENCODING

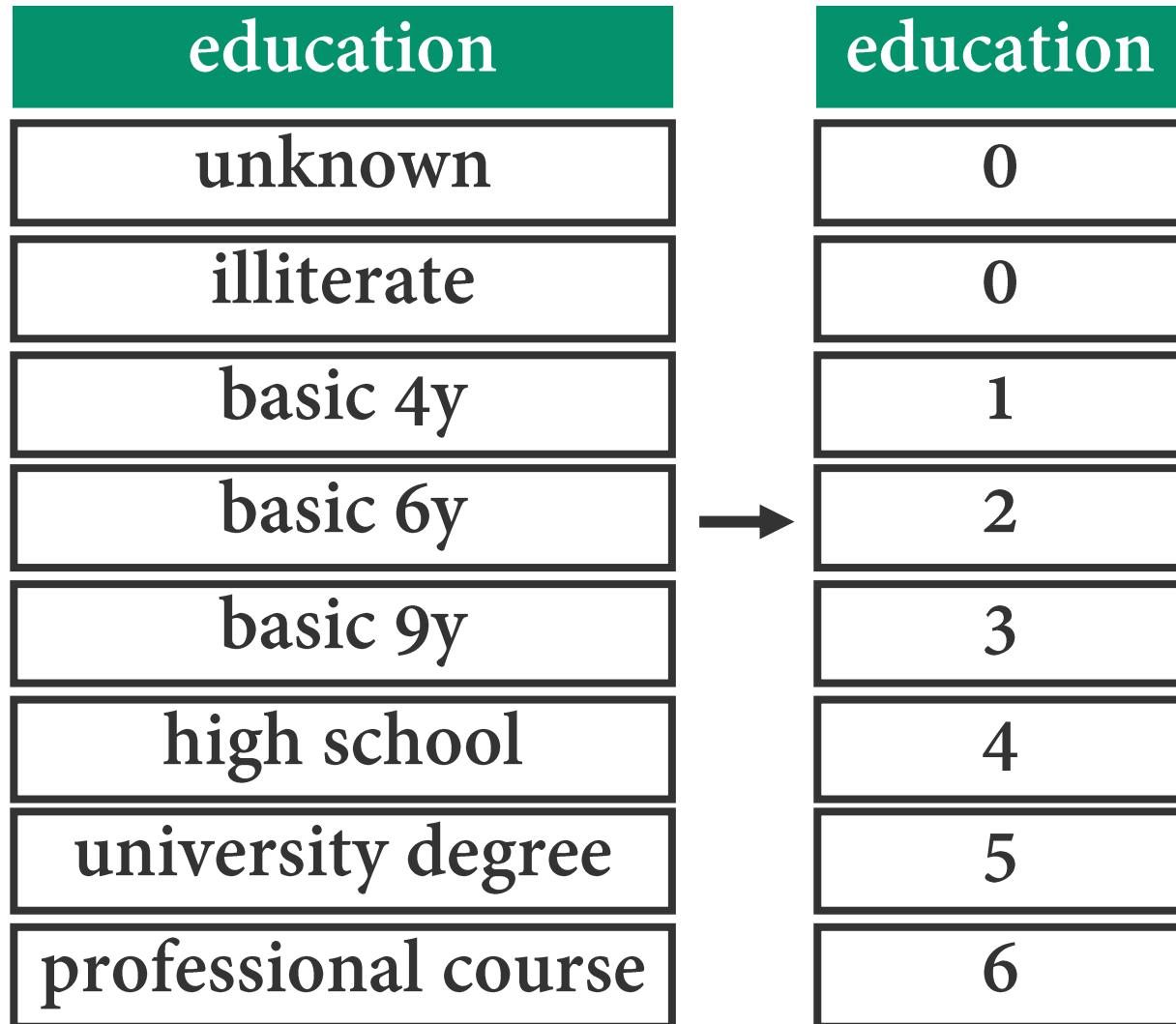
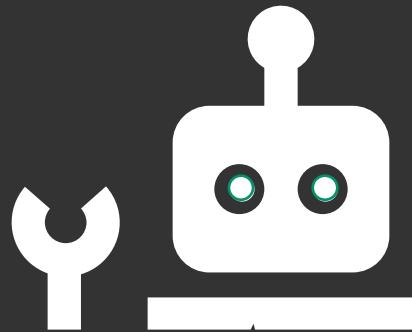
A	X	Y	Z
X	1	0	0
Y	0	1	0
Z	0	0	1

Subscribed Deposit

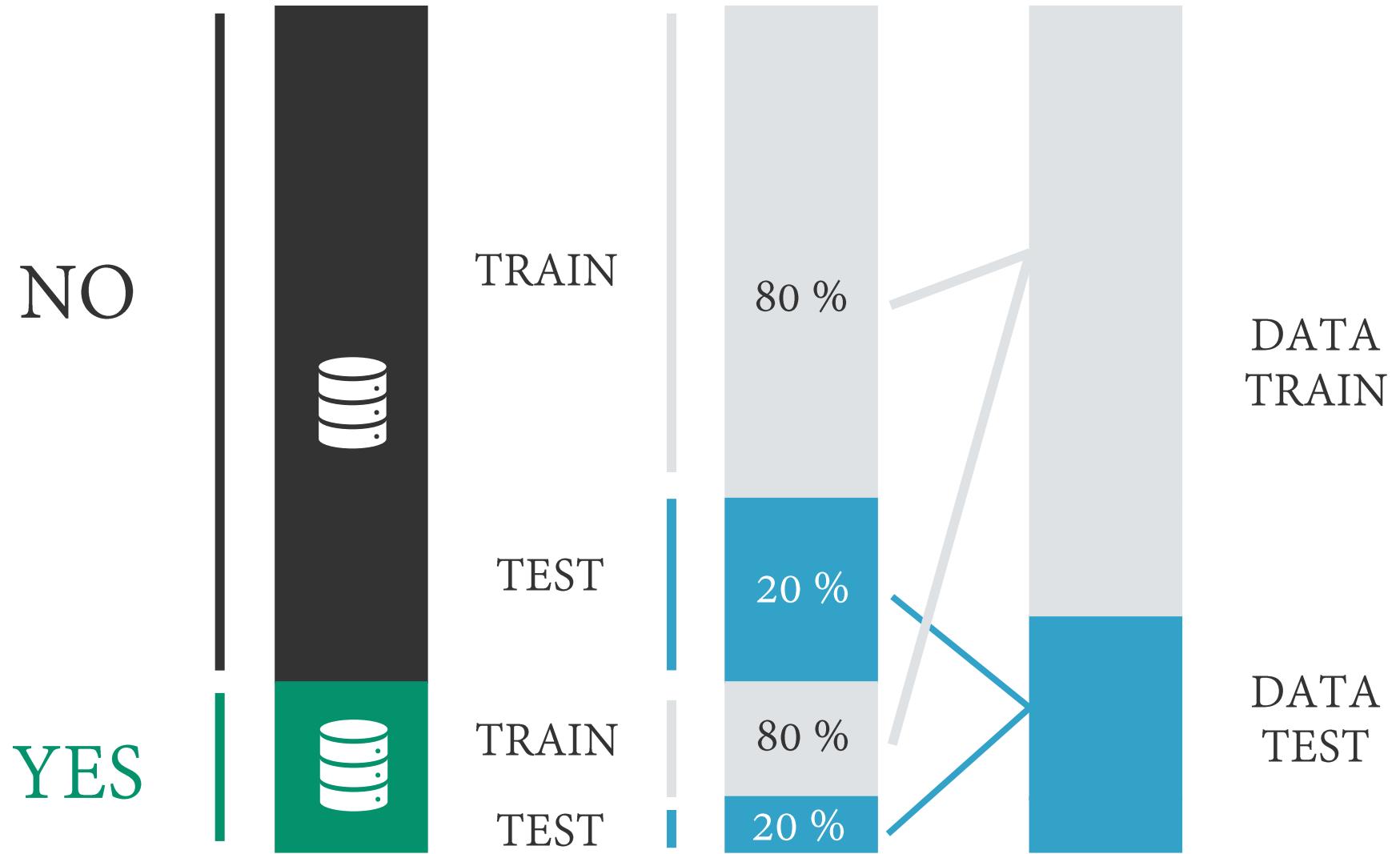


Data
Imbalance

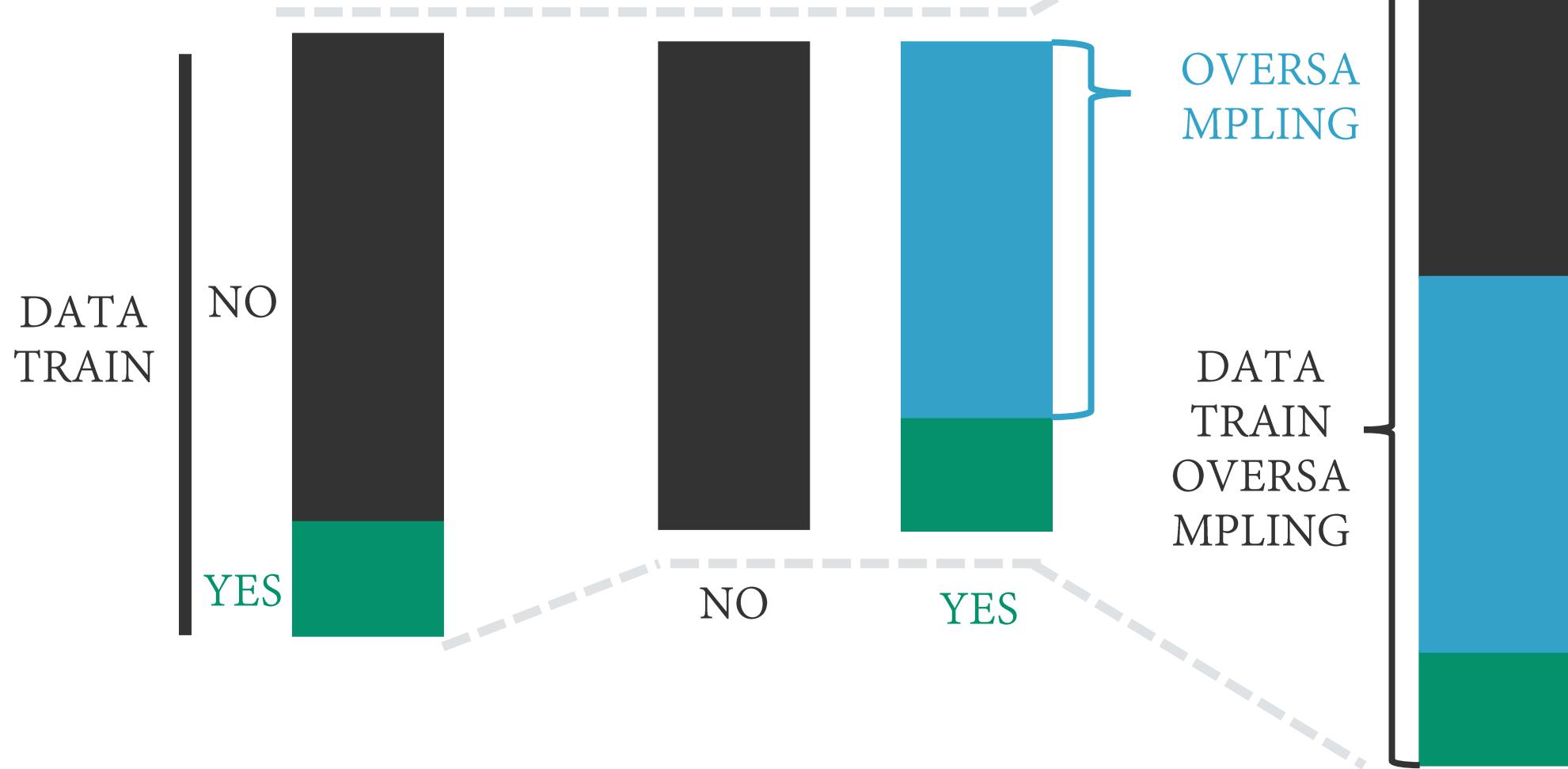
Label Encoding

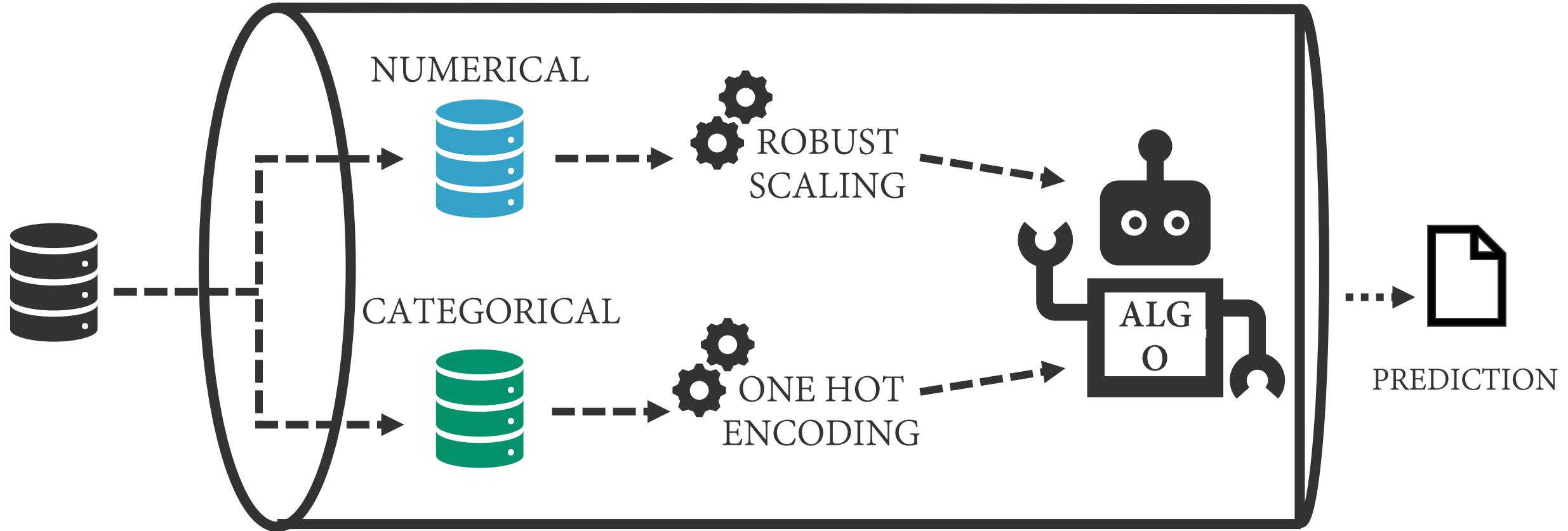


Data Splitting



Oversampling





NUMERICAL COLUMNS

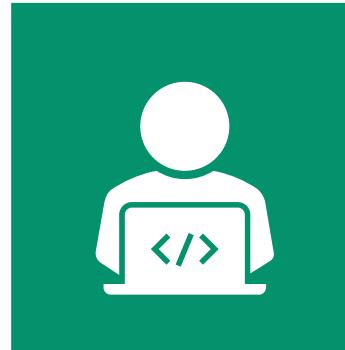
- Age
- Education
- Pdays
- Previous
- Cons.price.idx
- Cons.conf.idx
- Euribor3m
- Nr.employed
- Emp.var.rate

CATEGORICAL COLUMNS

- Job
- Marital
- Default
- Housing
- Loan
- Contact
- Month
- Day_of_week
- Poutcome

A L G O

R I T H M

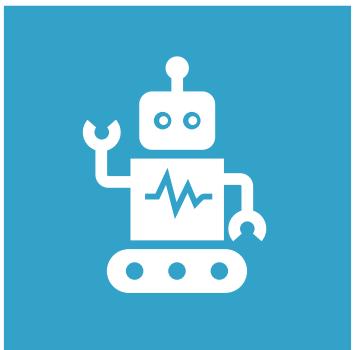


LOGISTIC
REGRESSION

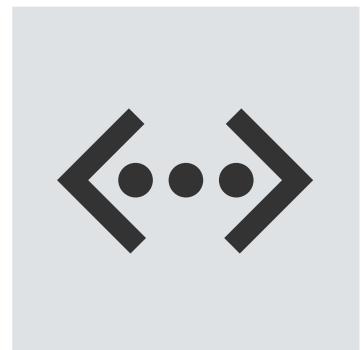
SVM

DECISION
TREE

KNN



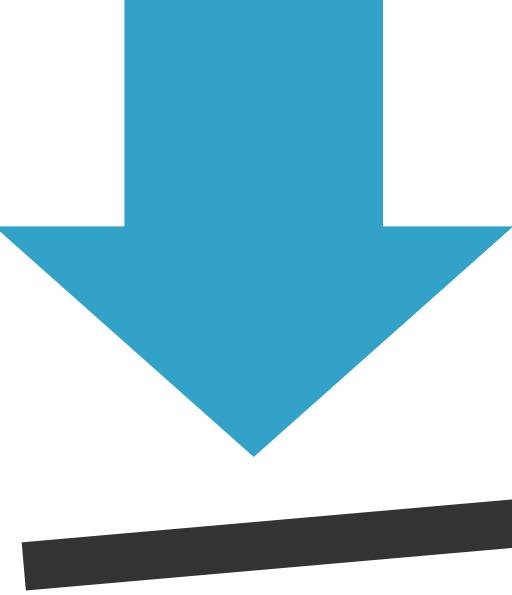
RANDOM
FOREST



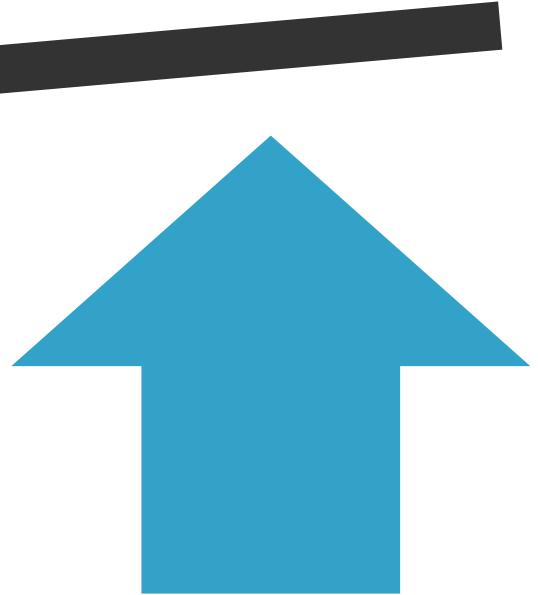
XGBOOST

“Precision is more
IMPORTANT”

“Predict 1 (buy deposit) but in
actual 0 (not buying) have
more risk and wasting
more cost”



Precision



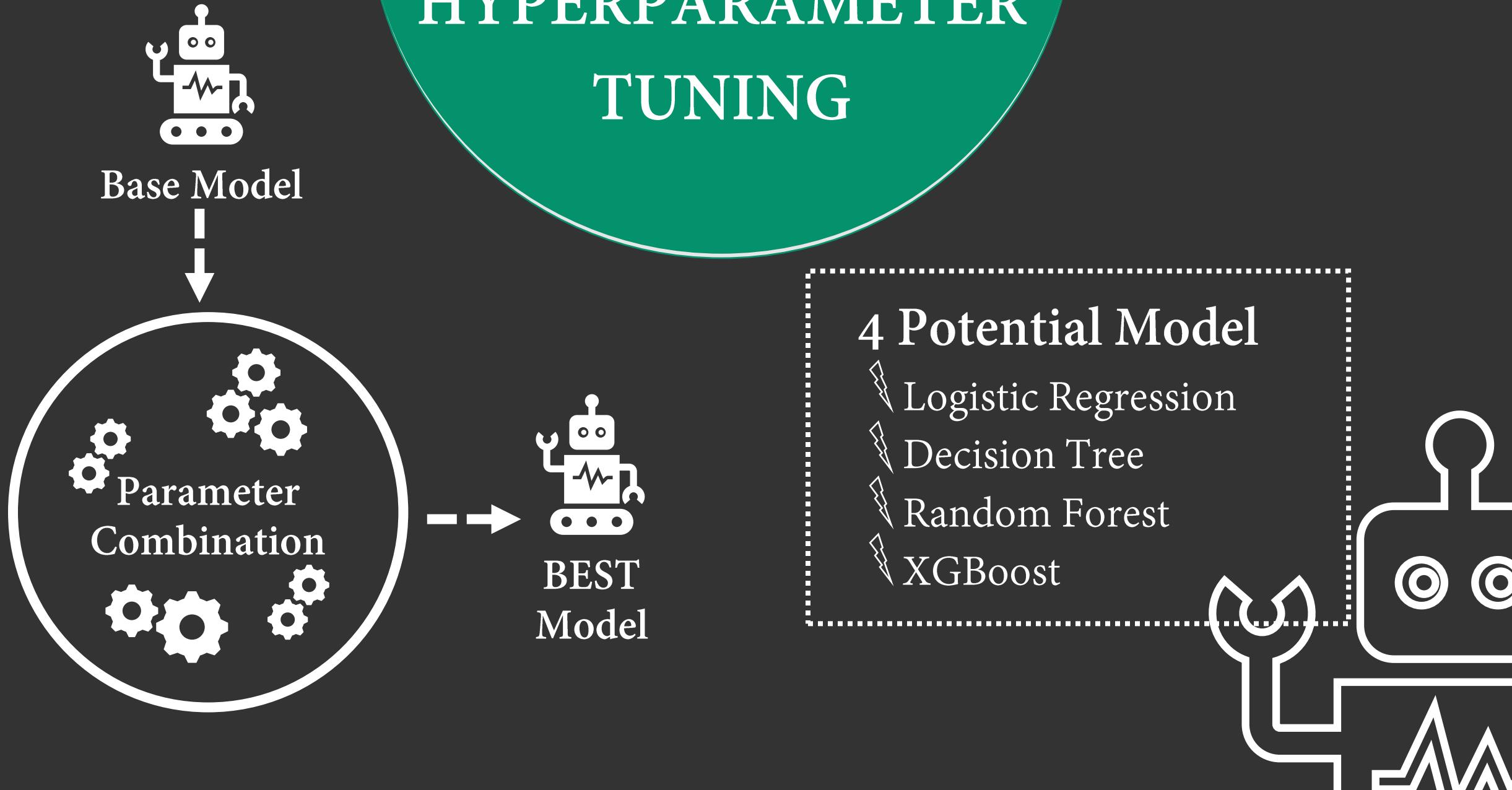
False
Positive

EVALUATION MATRIX

BASE MODEL

	Logistic Regression	KNN	SVM	Decision Tree	Random Forest	XGB
Precision	0.356	0.272	0.093	0.306	0.447	0.388
Recall	0.645	0.572	0.798	0.357	0.412	0.595

HYPERPARAMETER TUNING



5 MODEL WITH HIGH PRECISION

	Precision	Recall
★ Logistic Reg. HPT 2	0.647	0.202
Random Forest Base	0.447	0.412
Random Forest HPT 1	0.446	0.409
Random Forest HPT 3	0.433	0.394
Random Forest HPT 2	0.414	0.360

CONFUSION MATRIX FOR BEST MODEL

		Predict
		1 0
		1 0
Actual	1	187 741
	0	102 7208

False Positive : 102

False Negative : 741

PREDICTION USING ALL DATA

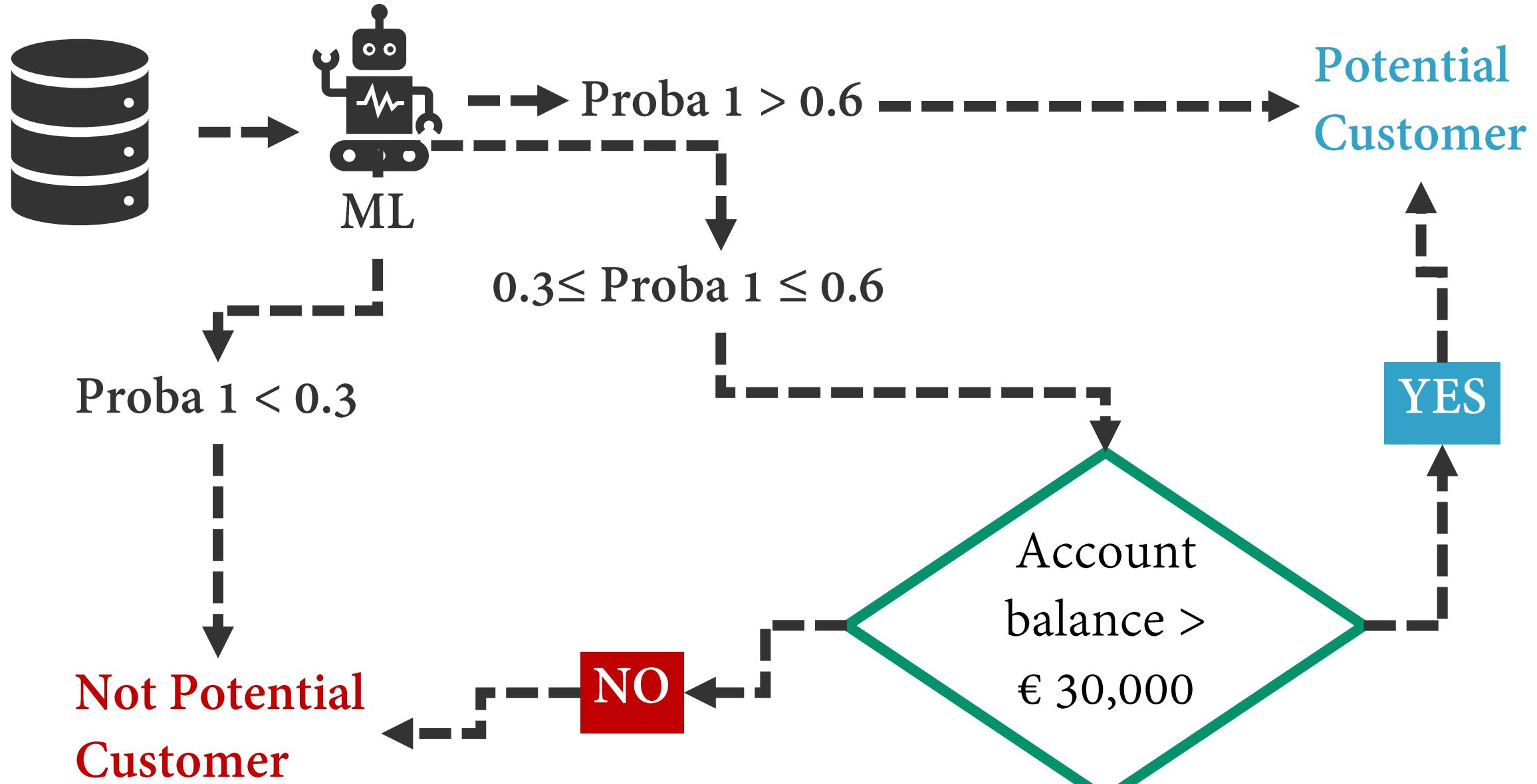
		Predict
		1 0
Actual	1	967 3673
	0	548 36000

False Positive : 548

False Negative : 3673 



Predict Probability with certain threshold
Considering Customer Bank Acc Balance



WHY ACCOUNT BALANCE ?

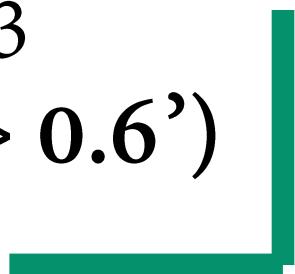
before someone decide to subs a deposit,
they sure will check their account balance
first to see if they have enough money or
not

WHY € 30,000?

Refers to the minimum number of
premium customer accounts balance in
Indonesia

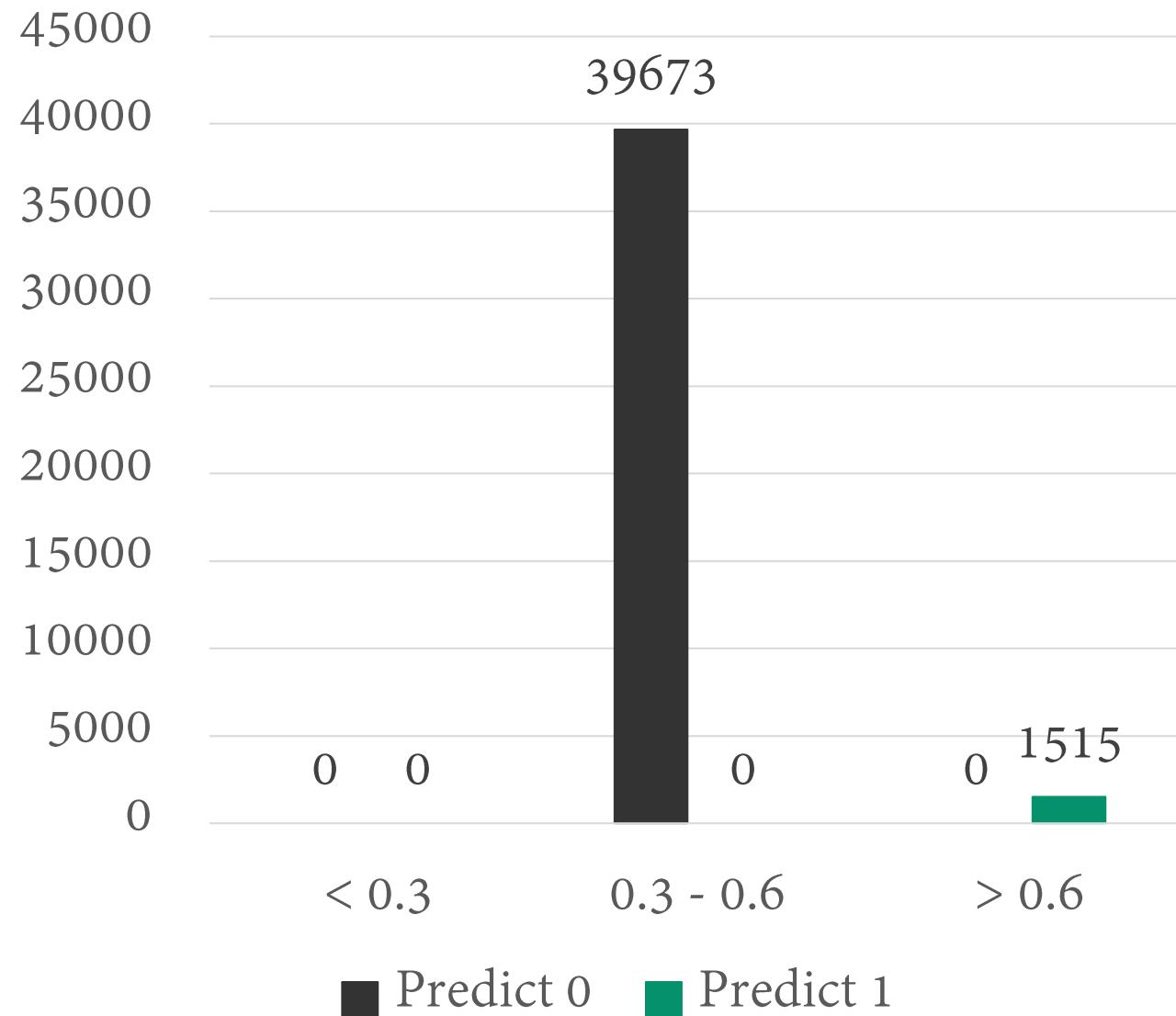
PREDICT PROBABILITY

 predict the **probability of subscribing the deposit** for every bank customer

 categorizing the probability into 3 groups ('< 0.3', '0.3 – 0.6' & '> 0.6')

Distibution of Prob. 1 for each y

Prediction



1. All Predict 0 in range proba
 $1 \ 0.3 - 0.6 \rightarrow \text{all FN's}$
Customer Account Balance should be checked
2. All Predict 1 in range proba
 $1 > 0.6$

COST SIMULATION



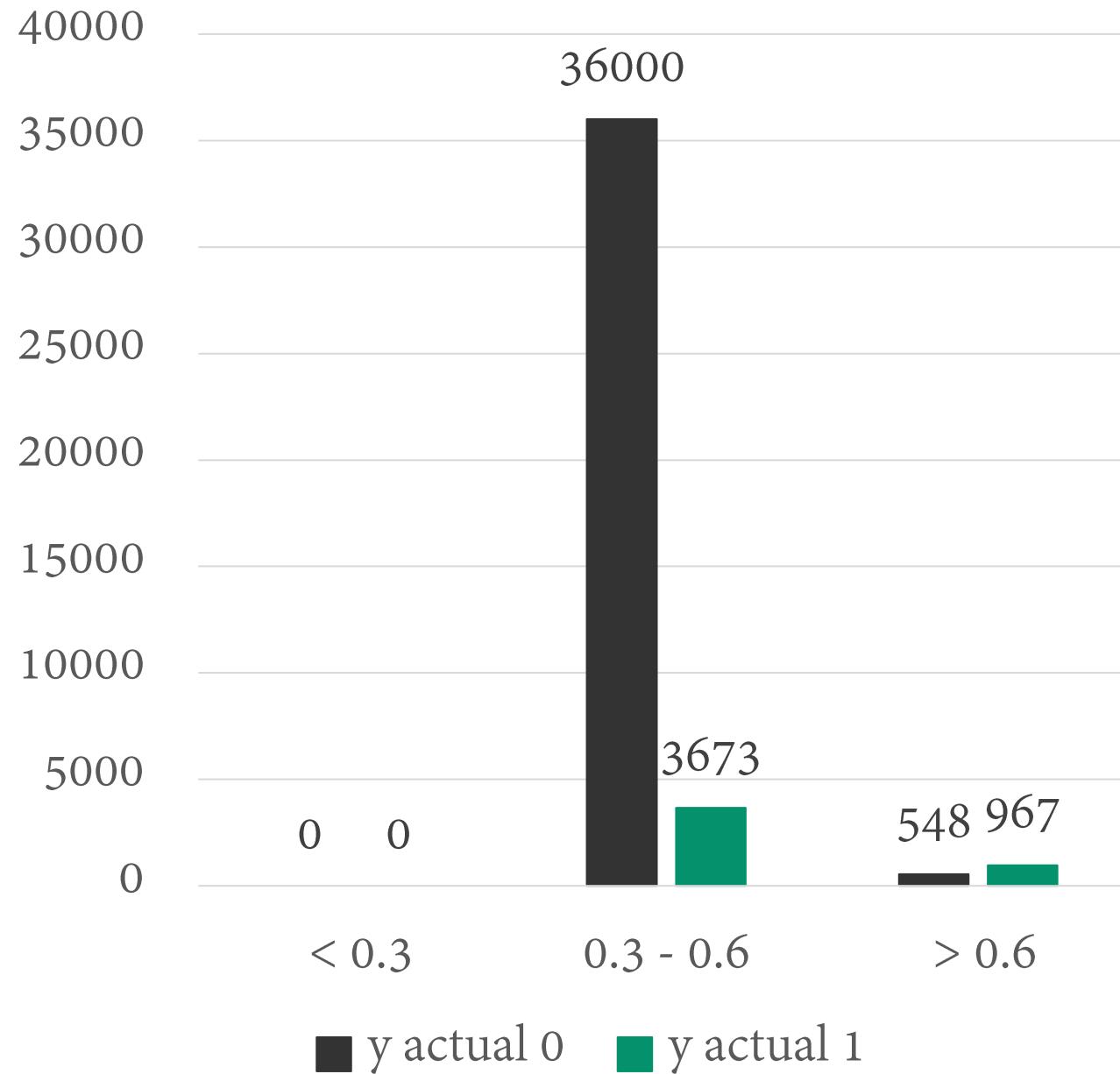
APPROACH AND ASSUMPTION

DURATION APPROACH

ASSUMPTION :

- Cost driven by two main things : Call Charges per Minutes, Salary
- The campaign duration is can be adjust
- Call charges: 0.3 euro per minutes
- Salary : 1600 euro/person/month
- 4 hours/240 minutes of call (effective time) that can handled by 1 employee per day
- 20 working days per month

Distribution of Prob. 1 for each y Actual

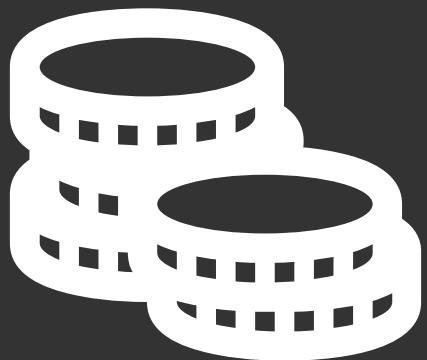


we assume for people that have **bank account more than 30,000 Euro** (from Proba 1 in range 0.3 - 0.6) is **people that subs the deposit for real ('y' = 1)** [The false negative] which is in this case there are **3673**

COST BEFORE AND AFTER

		Price	Quantity	Duration	Total (all months) (euro)
BEFORE	Call Charges	0.3 euro/minutes	1	177304 minutes (in 10 months)	53191.215
	Salary	1600 euro/month/emp	4 emp	10 months	64000.000
			TOTAL		117191.215
AFTER	Call Charges	0.3 euro/minutes	1	44977 minutes (in 10 months)	13493.094
	Salary	1600 euro/month/emp	3 emp	3 months	14400.000
			TOTAL		27893.094
				SAVING	89298.121

COST REDUCTION



76.2%



A close-up photograph of a snowy owl in flight. The owl's head is turned slightly to the right, revealing its large, bright orange eyes with dark pupils. Its white feathers are heavily mottled with dark brown and black spots, particularly on the wings and tail. The background is a solid, dark green.

CONCLUSION

CONCLUSION

1. Machine Learning is a solution that can predict potential customer who will buy the deposit.
2. After predicting the potential customer, we do cost simulation that can reduce the cost up to 76.2 % from initial cost.

REFERNCES

- <https://www.lonelyplanet.com/portugal/narratives/practical-information/directory/telephone#:~:text=Local%2C%20Regional%20%26%20National%20Calls&text=Local%20calls%20cost%20around%20%E2%82%AC,from%20anywhere%20in%20the%20country>.
- <http://www.salaryexplorer.com/salary-survey.php?loc=174&loctype=1&job=12702&jobtype=3>

THANK'S !