Numero Uno

Fajri Ardiansyah - Reza Falufi - Theresa Lalita Handaruputri - Widia Siska Yenti

Case Study:

Term Deposit Subscription

in a Portuguese bank

Olá! We are

Numero Uno

With our data-driven methods and analytics expertise, we deliver insightful business recommendations customized for your business needs

Members:

- Fajri Adriansyah
- Theresa Lalita
- Reza Falufi
- Widia Siska Yenti

NOVA
BANCO

We are currently providing our service to **Nova Banco**, a prominent Portuguese bank, of which products include:

Bank accounts

Savings and
term deposit

Loans

Personal, business,
and housing loans

Nova Banco sets to **increase term deposit** customer ...

Background

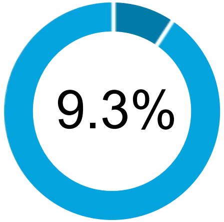
Nova Banco's customers **mostly** only have **saving accounts**, where customers can withdraw their funds anytime

Meanwhile **in term deposit**, Nova Banco can secure funds in certain time so that it can **invest in other financial products** with higher Rate of Return (RoR) or **lend the money out** to its other clients to get higher interest rate

Initial Strategy

Telemarketing campaign

However..



Initial campaign only attracts 9.3% new term deposit account

Why?

Random approach in choosing customers targeted for telemarketing campaign

We provide our solution for Nova Banco's marketing division to enhance campaign effectivity

How?

We create **machine learning model** to predict potential customer subscribing term deposit

Business Metrics

Conversion Rate

(no. of customers subscribing term deposit / no. of campaigns)

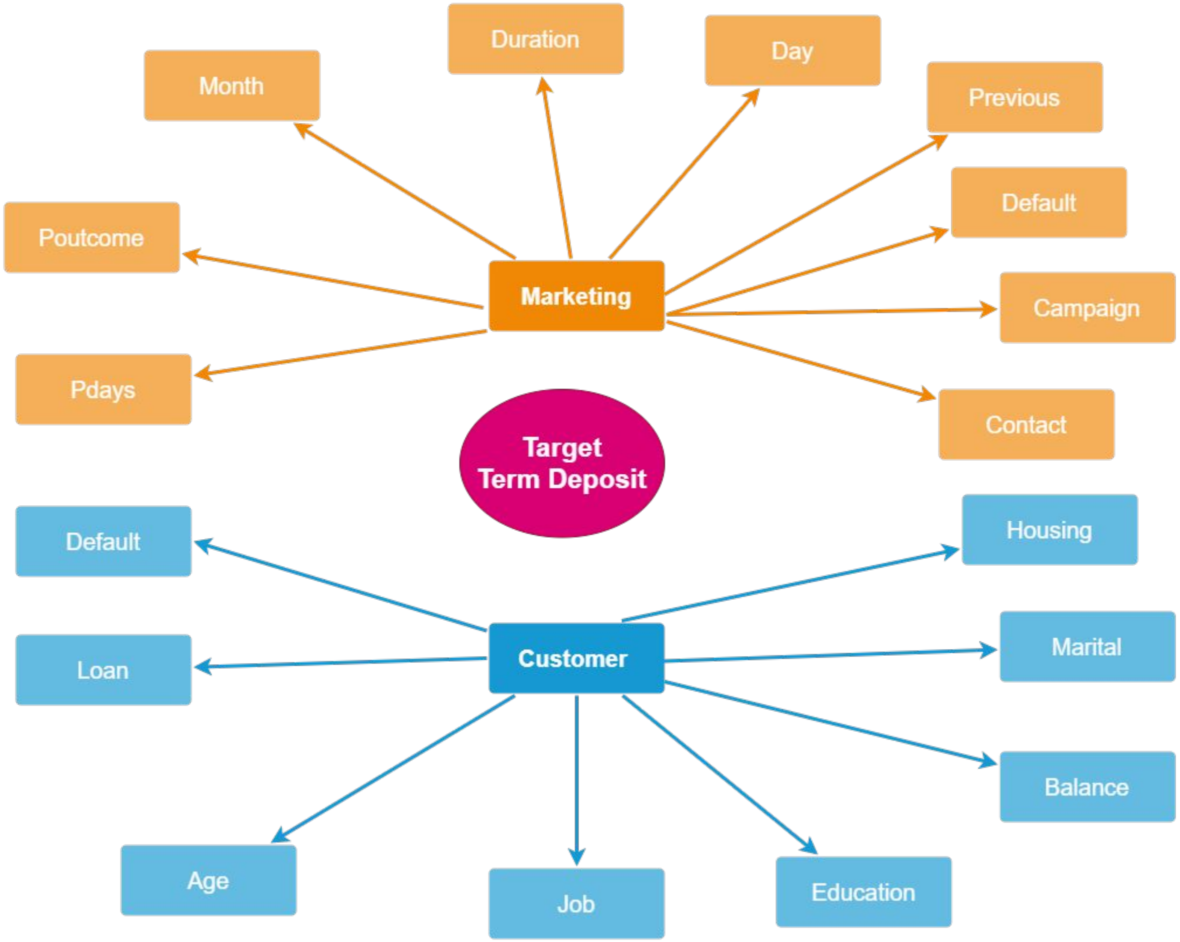
Expected result

Increase in the number of clients who subscribe to term deposits

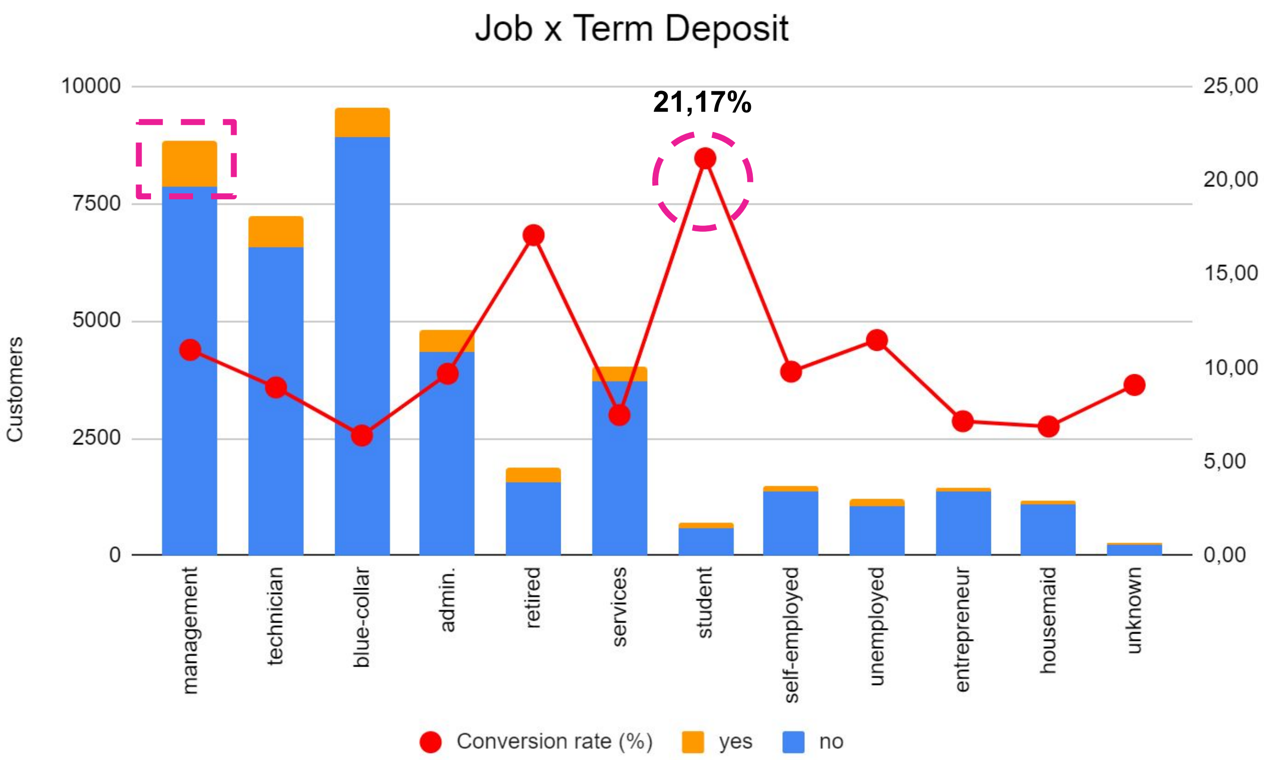
up to 2 times

Data Explora

(16 attributes)



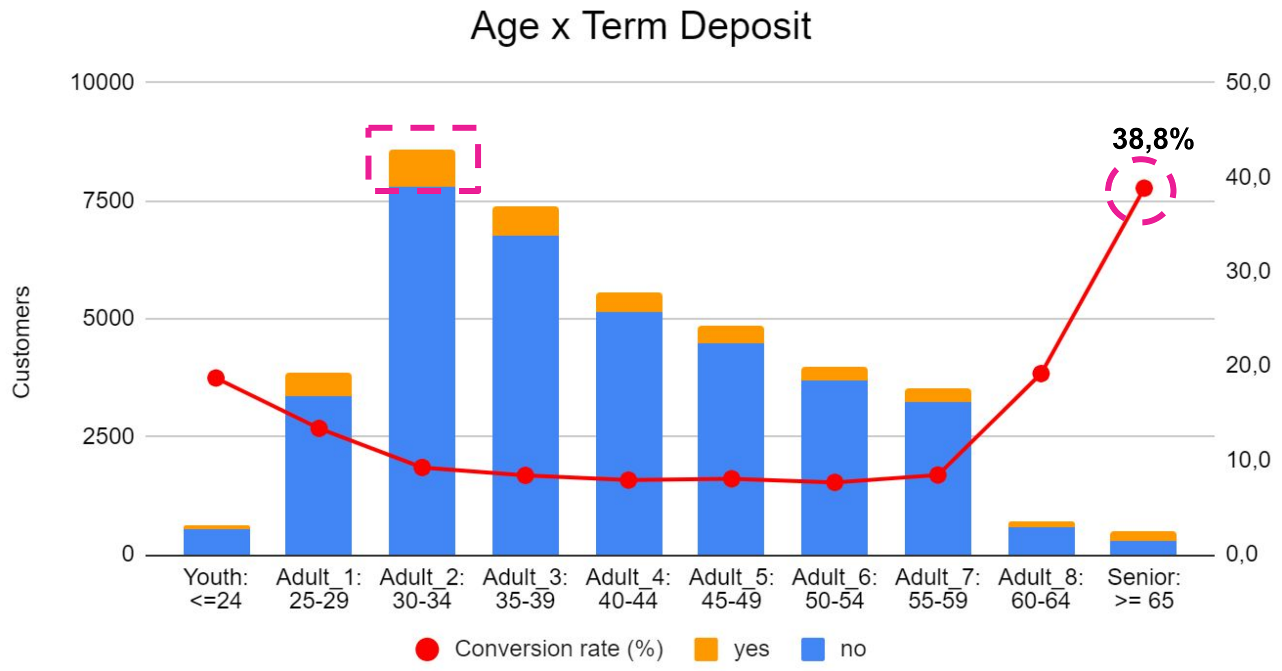
Data Exploration



While student has the highest conversion rate ...

Management is the job that has the highest term deposit subscription, followed by technician, blue-collar, and admin

Data Exploration

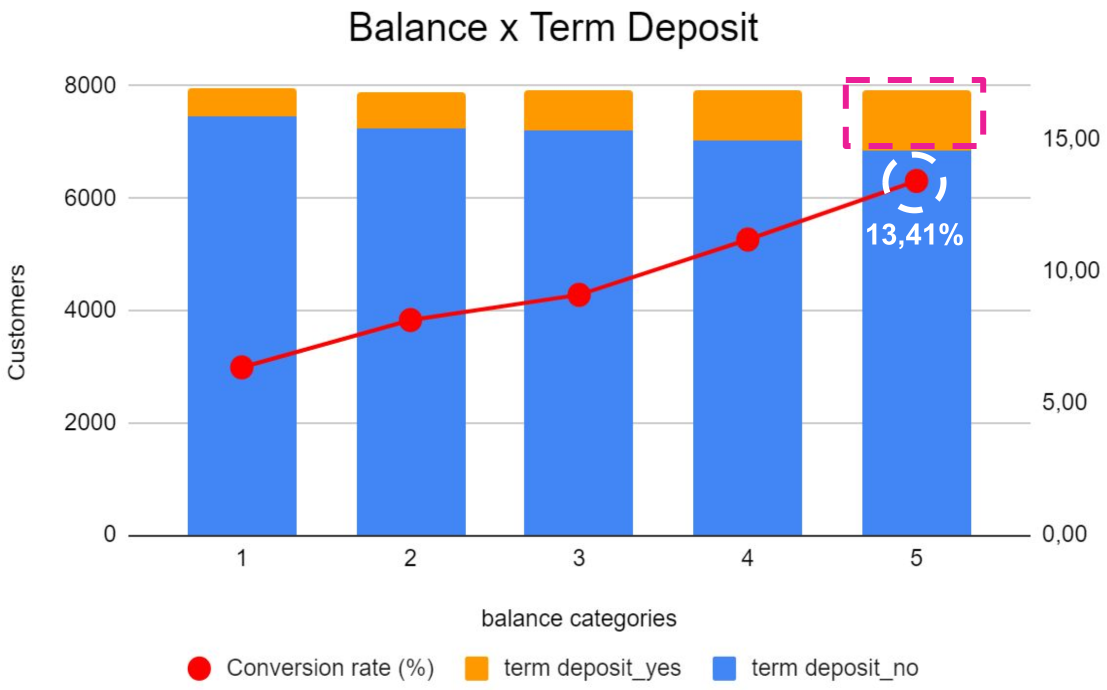


Conversion rate tends to:

- decrease from youths to early adults,
- stabilize during mid-to-late adults
- increase in seniors

Customers who subscribe term deposit are dominated by adults with age range of 30-34 years old

Data Exploration



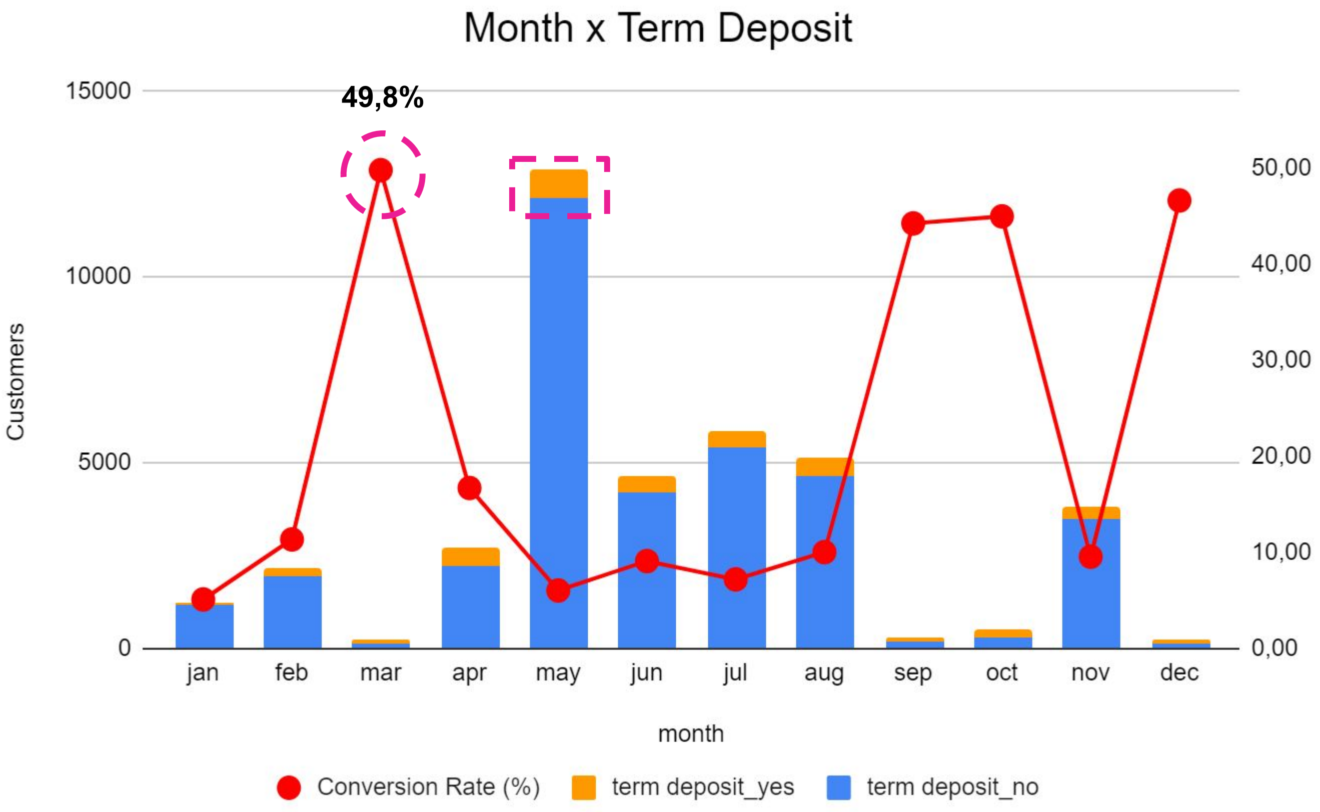
Number of customers are evenly distribution amongst Balance categories

Customers with the highest range of balance values have the **higher term deposit subscription and the highest conversion rate**

Note on Balance categories (based on percentile):

1. €9 or below
2. €9.1 - €260.8
3. €260.81 - €680
4. €681 - €1,815
5. €,1815 and above

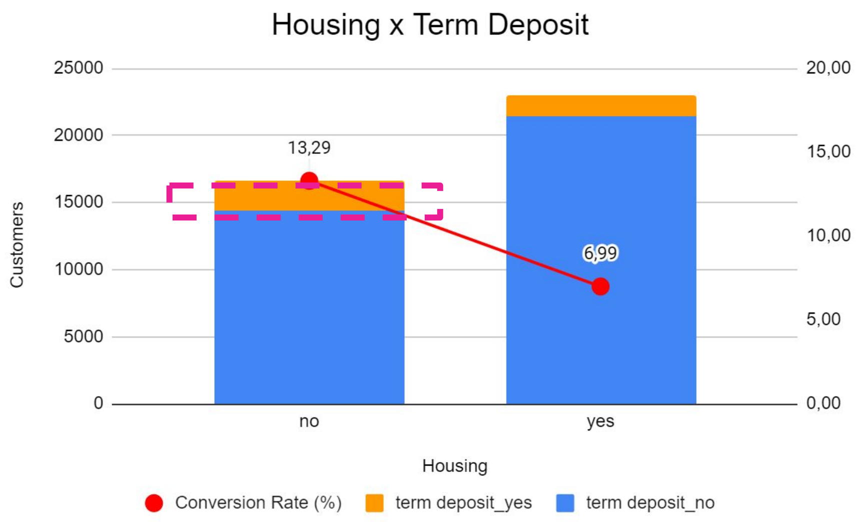
Data Exploration

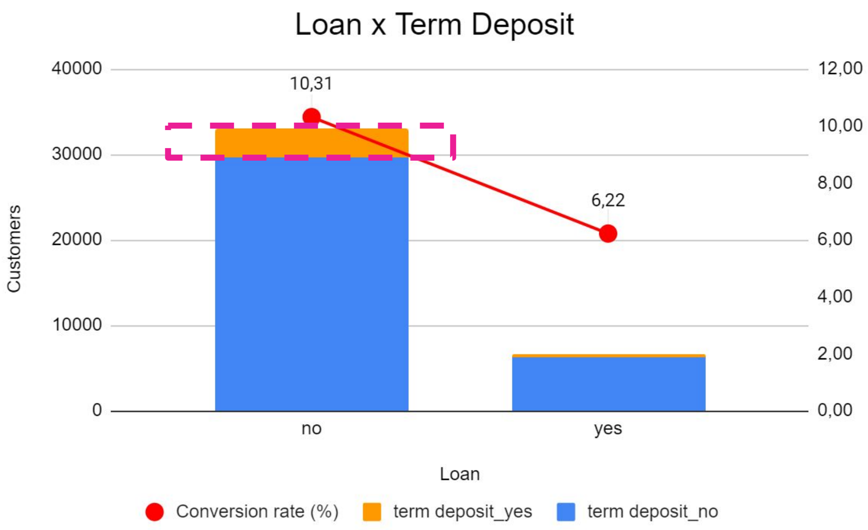


While the highest conversion rate occurs in March ...

The highest term deposit subscription occurs in May

Data Exploration



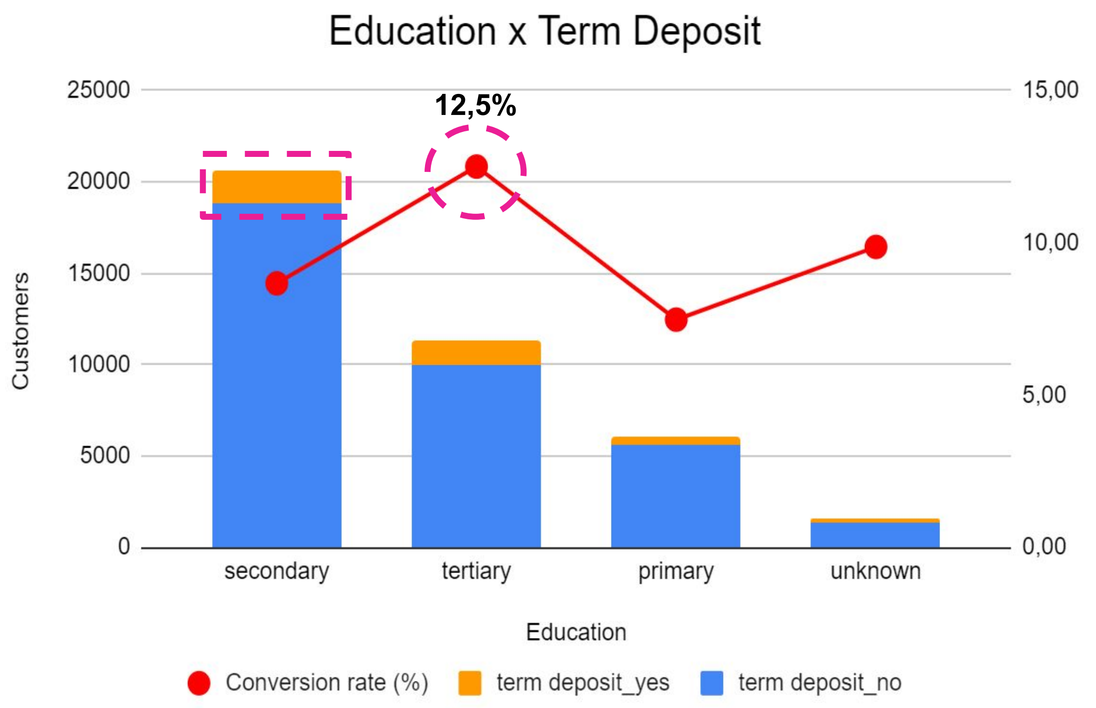


Housing loan and Personal loan have similar data exploration outcome

:

Customers without housing loan and personal loan have **higher term deposit subscription** and **higher conversion rate**

Data Exploration



While customer with tertiary education have highest conversion rate...

Customers with secondary education dominate term deposit subscription

Note on Education stages in Portugal:

- **Primary:**
  primary school - junior high school
- **Secondary:**
  senior/vocational high school
- **Tertiary:**
  undergraduate and postgraduate education

Data Preprocessing

Missing data

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42639 entries, 0 to 42638
Data columns (total 17 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   age           42639 non-null  int64 
 1   job           42639 non-null  object
 2   marital       42639 non-null  object
 3   education     42639 non-null  object
 4   default       42639 non-null  object
 5   balance       42639 non-null  int64 
 6   housing       42639 non-null  object
 7   loan          42639 non-null  object
 8   contact       42639 non-null  object
 9   day           42639 non-null  int64 
 10  month         42639 non-null  object
 11  duration      42639 non-null  int64 
 12  campaign      42639 non-null  int64 
 13  pdays         42639 non-null  int64 
 14  previous      42639 non-null  int64 
 15  poutcome      42639 non-null  object
 16  term_deposit  42639 non-null  object
dtypes: int64(7), object(10)
memory usage: 5.5+ MB
```

- Dataset has 42639 rows and 17 columns in total
- There are no null values

Duplicate data

```
1  df.duplicated().sum()
```

```
0
```

- There are no duplicate values

Feature Engineering

Feature Drop

- **Duration:**
  Irrelevant because effectiveness of duration is only known when campaign finish.
- **Default:**
  Irrelevant because it has extreme imbalance between those with default credit (2%) and those without (98%)
- **Previous**
  Irrelevant because it has the redundancy with other feature (pdays)

Reformatting Numerical attributes into Categorical attributes

- Age
- Balance
- Day
- Campaign
- Pdays

Fixing invalid values

- Pdays
- Campaign

Feature Engineering

Final Features (13)

Age

Month

Education

Day of Week
(Mon-Sun)

Pdays
(number of days that passed by after the client was last contacted from a previous campaign)

Loan

Contact

Housing

Poutcome
(outcome of the previous marketing campaign)

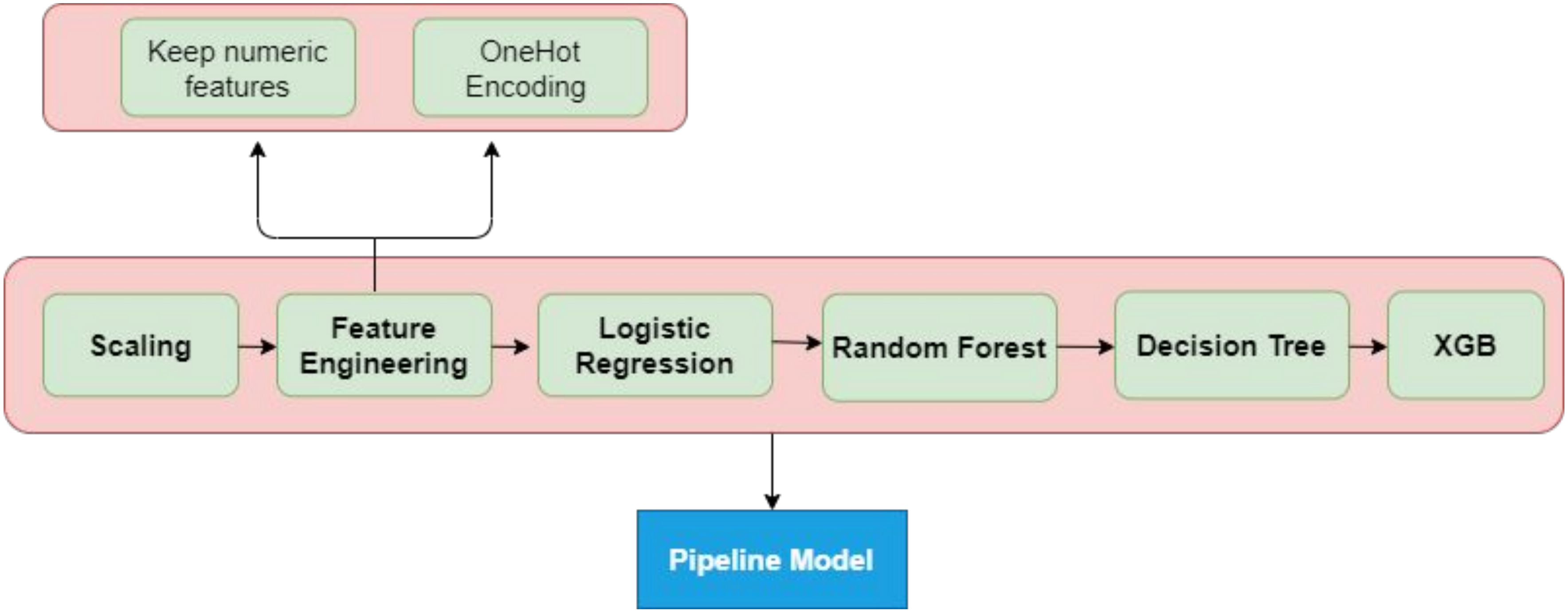
Balance

Jobs

Marital

Campaign

Modeling


Keep numeric features
OneHot Encoding
Scaling
Feature Engineering
Logistic Regression
Random Forest
Decision Tree
XGB
Pipeline Model

Model Result

Test size : 0.3

Model	Accuracy	Precision	Recall	AUC
Logistic Regression	72%	20%	63%	68%
Decision Tree	67%	16%	59%	64%
Random Forest	77%	23%	59%	69%
XGboost	91%	53%	13%	56%

Logistic regression model has the highest recall value than any others model

Model Result

Logistic Regression
Class Weight : **0** :0.095, **1**: 0.905
Solver : Liblinear
Scoring : Recall
CV : 5

Our model is reliable for both train and test dataset (not over fitting)

Train

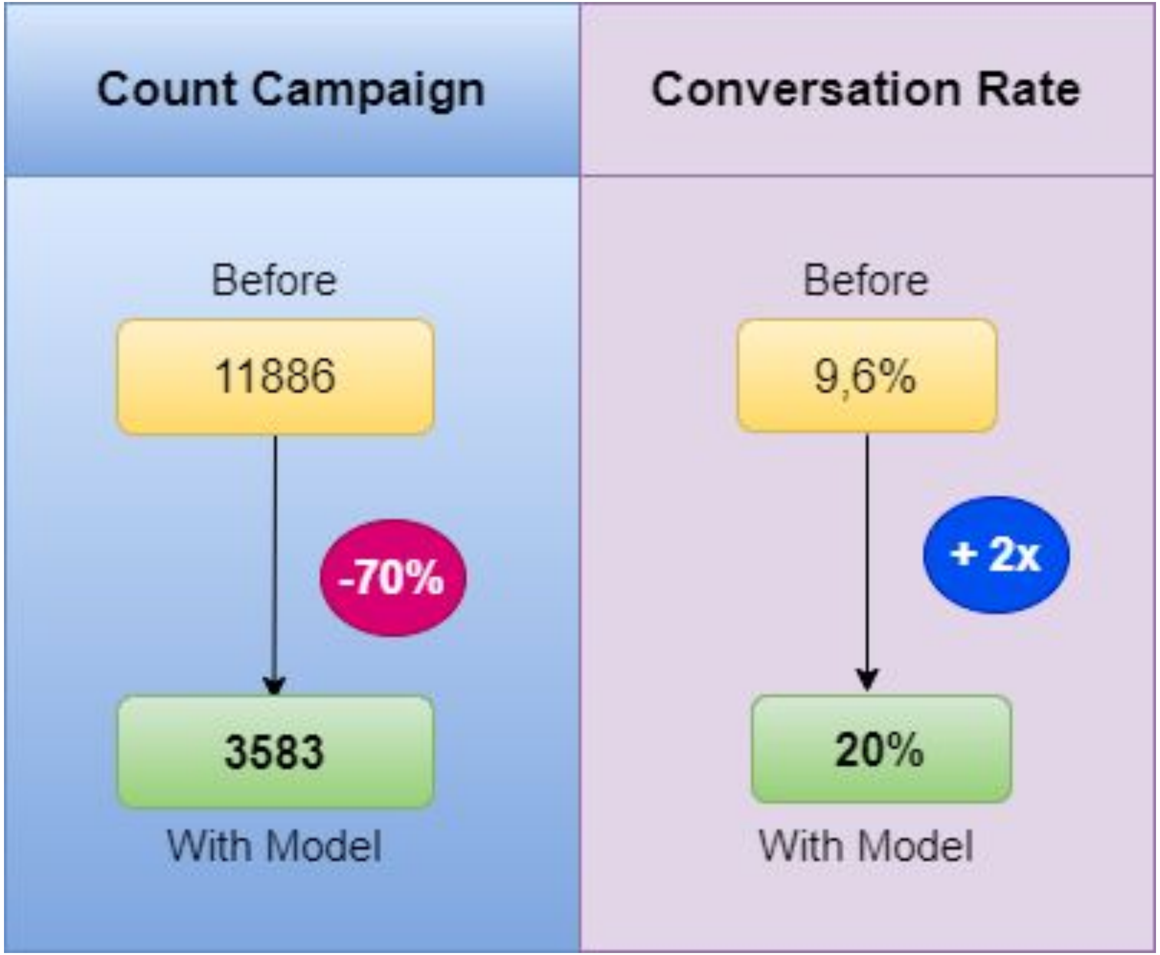
	precision	recall	f1-score	support
0	0.95	0.73	0.83	25061
1	0.20	0.63	0.30	2671
accuracy			0.72	27732
macro avg	0.57	0.68	0.56	27732
weighted avg	0.88	0.72	0.77	27732

Test

	precision	recall	f1-score	support
0	0.95	0.73	0.83	10741
1	0.20	0.63	0.30	1145
accuracy			0.72	11886
macro avg	0.57	0.68	0.57	11886
weighted avg	0.88	0.72	0.78	11886

```
Train score: 0.7196379633636233
Test score:0.7226148409893993
```

Impact



Based on data test

Model Insight - Job conversion rate

Management

month	conversion rate	yes	no	total	yes_ percentage	yes_cum_ percentage
may	6,8	137	1877	2014	14,7	15
aug	10,1	165	1468	1633	17,7	32
jul	6,6	75	1059	1134	8,1	41
nov	9,1	97	968	1065	10,4	51
jun	13,2	104	686	790	11,2	62
apr	23,1	114	380	494	12,2	74
feb	14,0	63	388	451	6,8	81
jan	2,8	7	246	253	0,8	82
oct	47,2	59	66	125	6,3	88
sep	45,9	39	46	85	4,2	92
mar	58,0	47	34	81	5	97
dec	48,0	24	26	50	2,6	100

Admin

month	conversion rate	yes	no	total	yes_ percentage	yes_cum_ percentage
may	7,0	116	1534	1650	26,1	26,1
apr	14,8	55	317	372	12,4	38,5
jun	9,9	51	464	515	11,5	50,0
jul	5,8	41	663	704	9,2	59,2
aug	13,9	40	247	287	9	68,2
nov	9,0	36	366	402	8,1	76,3
feb	12,1	31	225	256	7	83,3
oct	33,8	23	45	68	5,2	88,5
mar	41,5	17	24	41	3,8	92,3
sep	50,0	17	17	34	3,8	96,1
dec	52,2	12	11	23	2,7	98,8
may	7,0	116	1534	1650	26,1	26,1

Model Insight - Job conversion rate

Technician

month	conversion rate	yes	no	total	yes_ percentage	yes_cum_ percentage
aug	7,6	114	1387	1501	18,2	18,2
may	5,8	112	1803	1915	17,9	36,1
apr	19,6	76	312	388	12,2	48,3
jun	10,6	62	521	583	9,9	58,2
jul	6,7	61	856	917	9,8	68,0
nov	9,4	58	557	615	9,3	77,3
feb	13,0	45	301	346	7,2	84,5
oct	58,6	41	29	70	6,6	91,1
mar	47,2	17	19	36	2,7	93,8
sep	53,1	17	15	32	2,7	96,5
dec	36,7	11	19	30	1,8	98,3
jan	4,9	11	213	224	1,8	100,0

Blue-collar

month	conversion rate	yes	no	total	yes_ percentage	yes_cum_ percentage
may	5,1	211	3923	4134	36	36,0
jul	7,8	95	1130	1225	16,2	52,2
jun	6,4	70	1028	1098	11,9	64,1
apr	9,0	66	668	734	11,3	75,4
nov	7,3	43	546	589	7,3	82,7
aug	8,5	41	444	485	7	89,7
feb	5,4	19	333	352	3,2	92,9
oct	39,0	16	25	41	2,7	95,6
sep	83,3	10	2	12	1,7	97,3
jan	3,4	6	170	176	1	98,3
dec	45,5	5	6	11	0,9	99,2
mar	40,0	4	6	10	0,7	99,9

Conclusion and Recommendation

Conclusion

The impact of machine learning utilization will **increase campaign efficiency** and **conversion rate**.

x2

The new term deposit account will be doubled (conversion rate from 9.6% to 20%) while **number of campaign is -70% decreased** (from 11886 campaigns to 3583 campaigns).

Recommendation

Since our recall value left 37% customer not being campaigned, our suggestion are to carry out another campaign according to these recommendations:

1. To focus on doing telemarketing campaign on customers with these profiles:

Jobs: ***management, technician, blue collar,admin***

AGE

Age: ***30-34*** years old

2. We also suggest to give ***promos*** to customers in the above-mentioned profile in:

%

April, June, July, and August

Thank You

Appendix

Appendix 1 - EDA Insights

EDA Insights

Education:
Proporsi pendidikan masyarakat di Portugal memang banyak di level SMA/SMK

Balance:
Sesuai *common sense*, di mana customer dengan balance besar memiliki *spare fund* yang besar pula, sehingga bisa dimasukkan ke term deposit

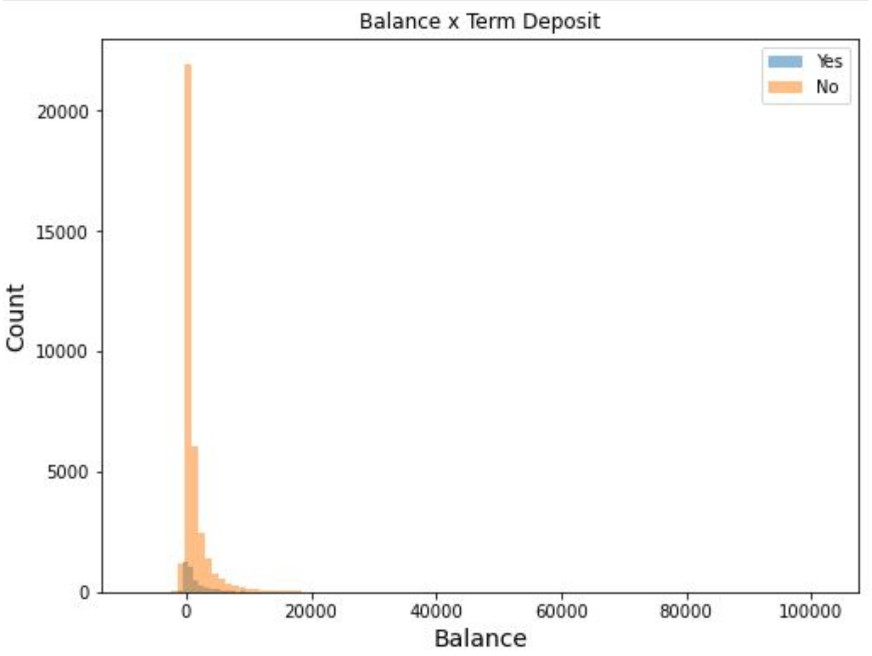
Housing loan dan personal loan:
Sesuai *common sense*, di mana customer tanpa pinjaman memiliki *spare fund* yang besar, sehingga bisa dimasukkan ke term deposit

Reformatting Numerical attributes into Categorical attributes

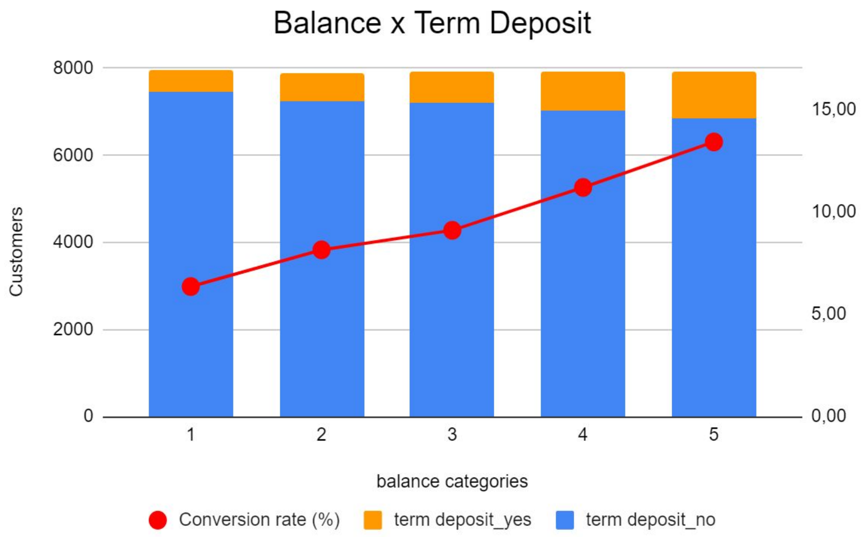
- **Age**
  Age grouping based on **WHO age range standardization** (https://www.who.int/healthinfo/paper31.pdf)
- **Balance**
  Balance grouping based on **percentile** (20%, 40%, 60%, 80%)
- **Day**
  Day grouping is based on **weekly grouping**
- **Campaign**
  Campaign grouping is based on **frequency division**, according to **consultation result with Nova Banco's marketing department**
- **Pdays**
  Pdays grouping is based on **monthly grouping (per 30 days)**

Appendix 2 - Balance comparison

Balance tanpa grouping



Balance dengan grouping (categorical)



Appendix 3 - Term Deposit Subscription Percentage

term_deposit	no	yes	total_data	conversion_rate	%conversion
job					
management	7244	931	8175	0.113884	0.243973
technician	6032	625	6657	0.093886	0.163784
blue-collar	8281	586	8867	0.066088	0.153564
admin.	4053	445	4498	0.098933	0.116614
retired	1480	315	1795	0.175487	0.082547
services	3465	293	3758	0.077967	0.076782
student	536	149	685	0.217518	0.039046
self-employed	1244	141	1385	0.101805	0.036950
unemployed	1000	134	1134	0.118166	0.035115
entrepreneur	1243	98	1341	0.073080	0.025681
housemaid	1016	76	1092	0.069597	0.019916
unknown	208	23	231	0.099567	0.006027

Appendix 4 - Modeling history

Percobaan 1

- Tanpa Pipeline
- Best model : Random Forest

```
Accuracy (Test Set): 0.50
Precision (Test Set): 0.11
Recall (Test Set): 0.62
F1-Score (Test Set): 0.18
AUC: 0.55
```

Percobaan 2

- Dengan Pipeline
- Best model : Logistic Regression

```
Accuracy (Test Set): 0.72
Precision (Test Set): 0.18
Recall (Test Set): 0.59
F1-Score (Test Set): 0.28
AUC: 0.66
```

Percobaan 2

- Dengan Pipeline
- Reformatting categorical
- Best model : Logistic Regression

```
Accuracy (Test Set): 0.72
Precision (Test Set): 0.20
Recall (Test Set): 0.63
F1-Score (Test Set): 0.30
AUC: 0.68
```

month	apr	yes	no	total
balance_cat				
5	20.559211	125	483	608
4	16.167665	108	560	668
3	14.219474	92	555	647
2	15.740741	68	364	432
1	16.616314	55	276	331

month	jun	yes	no	total
balance_cat				
4	10.731244	113	940	1053
5	9.499576	112	1067	1179
3	9.699769	84	782	866
2	9.388336	66	637	703
1	5.402161	45	788	833

month	jul	yes	no	total
balance_cat				
1	5.378832	93	1636	1729
2	6.989247	91	1211	1302
4	8.216433	82	916	998
3	8.051690	81	925	1006
5	8.849558	70	721	791

month	aug	yes	no	total
balance_cat				
5	13.971292	146	899	1045
4	12.143611	115	832	947
2	9.082814	102	1021	1123
3	8.682635	87	915	1002
1	6.169154	62	943	1005