BANK MARKETING CAMPAIGN

CASE STUDY PRESENTATION

BY

GOKULRAMESH DHANASEKAR


MARKETING
PLAN
advertising
research
finance
brand
sales
distribution
budget
action
performance
share
responsibilities
clients
strategy
training
analysis
results
incentives
promotion
cost
target
development
prices
benchmarks
channel
profit
expense
timescale
requirements
summary
management
sales
profit

About the Dataset:

- The data is related with direct marketing campaigns of a Portuguese banking institution.
- The marketing campaigns were based on phone calls.
- Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

ATTRIBUTES

Client data

- Age
- Job
- Marital status
- Education
- Housing loan
- Personal loan

Campaign data

- contact communication type
- last contact month of year
- last contact day of the week
- duration: last contact

ATTRIBUTES

Other campaign data

- Number of contacts performed
- Number of days that passed by after the client was last contacted
- Number of contacts performed before this campaign and for this client
- Poutcome

Social and Economic data

- Employment variation rate – quarterly
- Consumer price index - monthly
- Consumer confidence index – monthly
- Euribor 3 month rate - daily indicator
- Number of employees - quarterly indicator (numeric)

TASKS

To predict if the client will subscribe (yes/no) a term deposit (variable y).

- Exploratory Data Analysis
- Processing the Data
- Modelling
- Training and Testing the model
- Evaluation

Exploratory Data Analysis

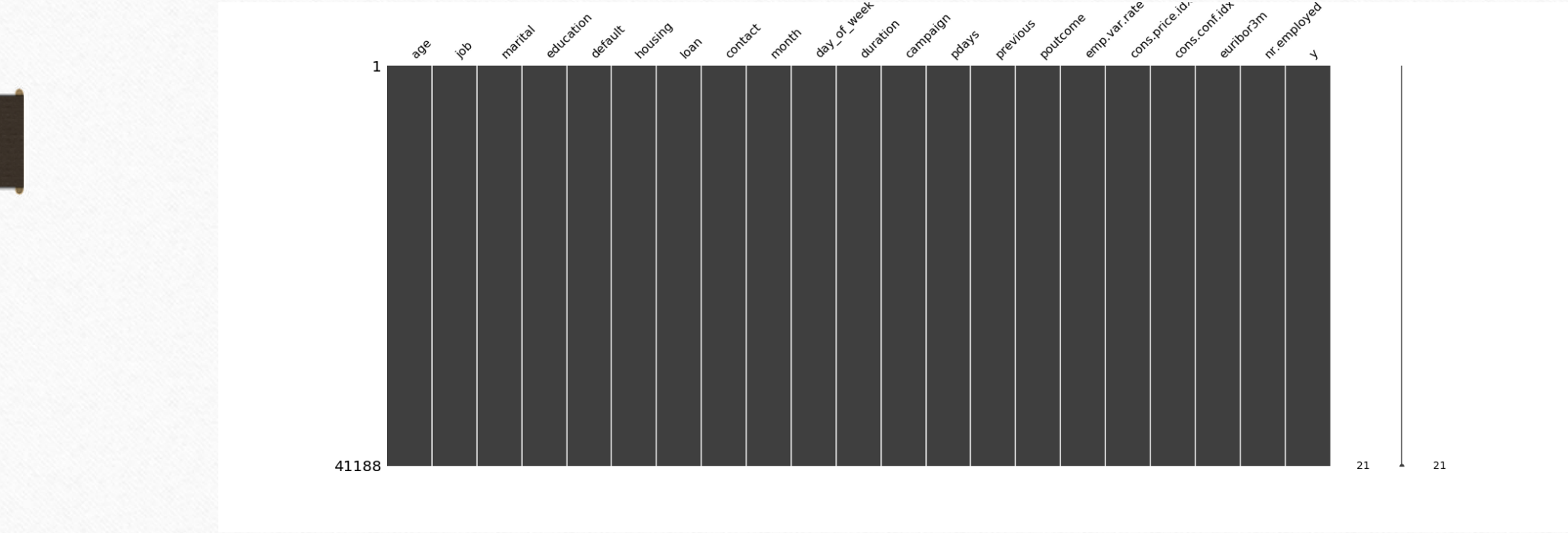
Detecting Missing values

Statistics

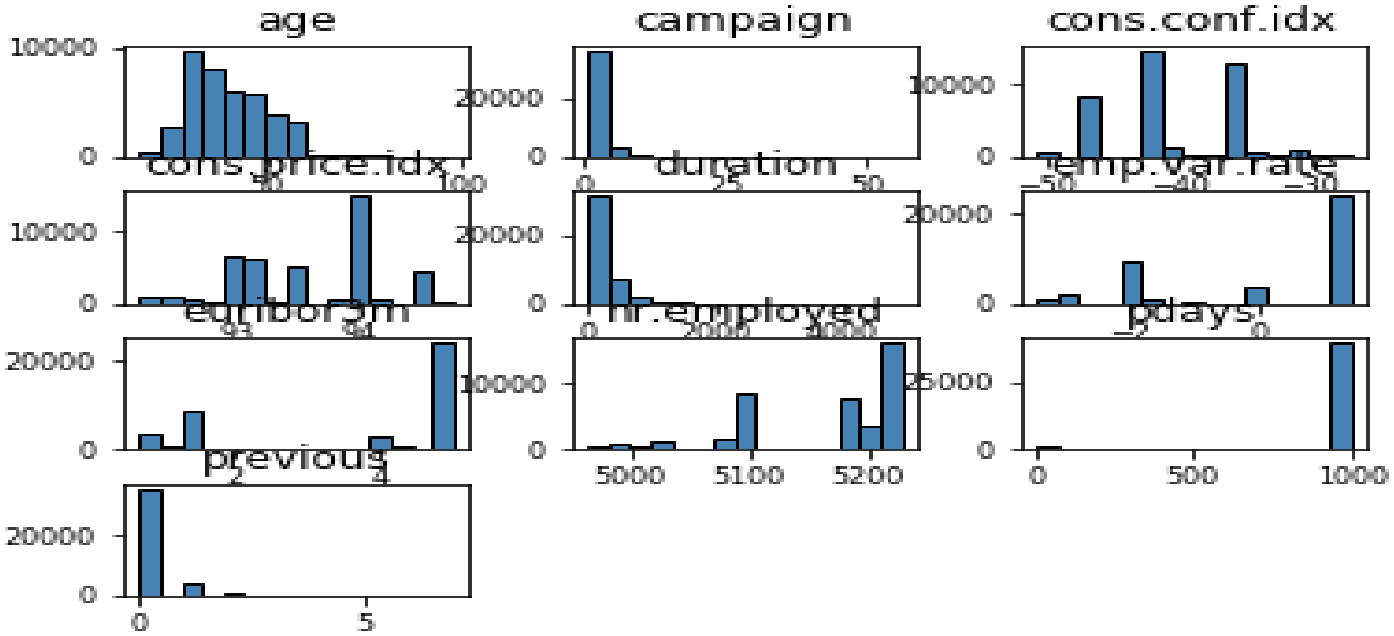
Visualization

Detecting Outliers

Missing values



Distribution



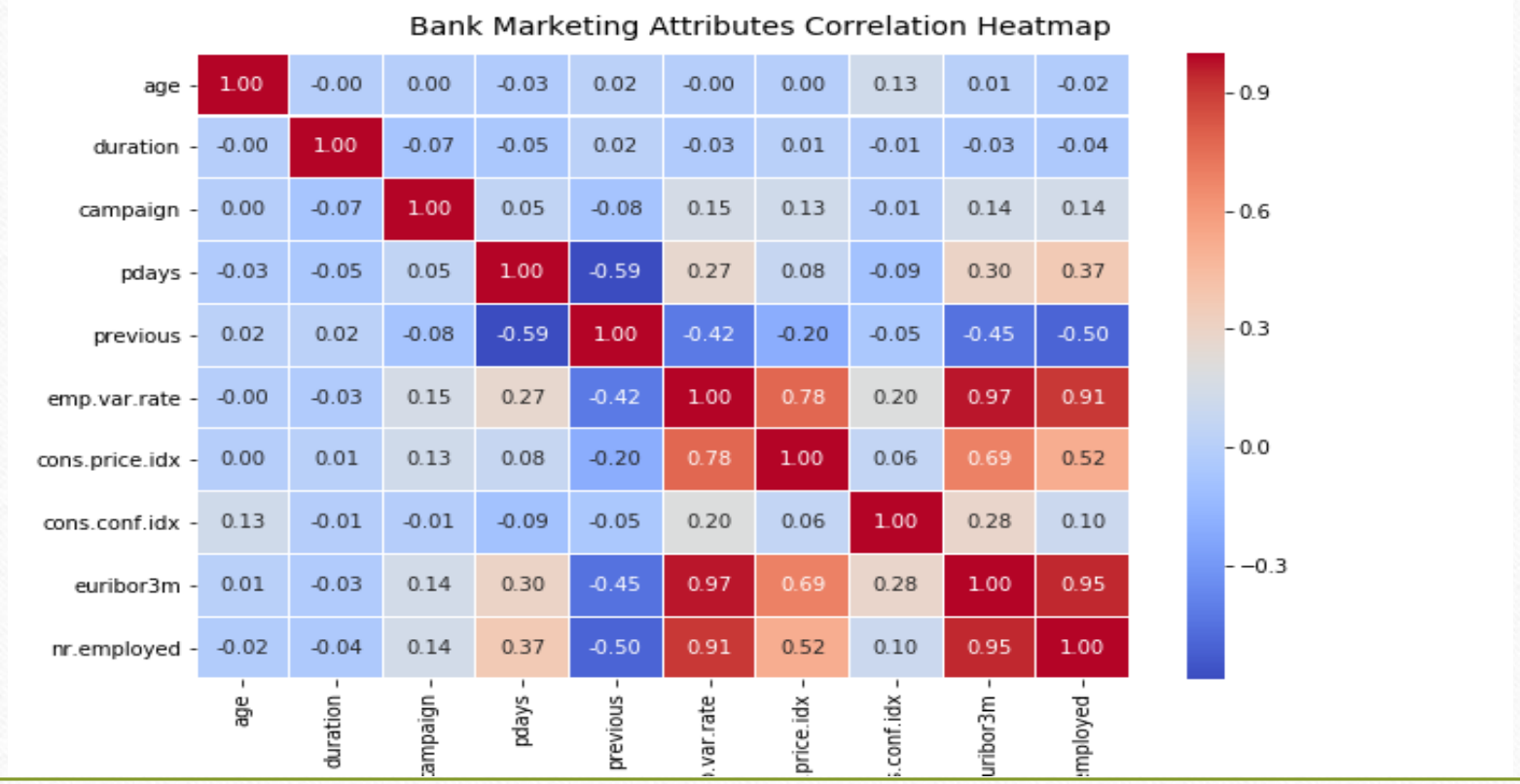
Statistics(numerical variable)

	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
count	41188.0	41188.0	41188.0	41188.0	41188.0
mean	0.0818855006319146	93.575664368312 63	-40.502600271913 99	3.6212908128580 68	5167.0359109418 44
std	1.5709597405172309	0.5788400489541813	4.628197856174375	1.7344474048511707	72.251527668265 27
min	-3.4	92.201000000000 01	-50.8	0.634	4963.6
25%	-1.8	93.075	-42.7	1.344	5099.1
50%	1.1	93.749	-41.8	4.857	5191.0
75%	1.4	93.994	-36.4	4.961	5228.1
max	1.4	94.767000000000 01	-26.9	5.045	5228.1

Statistics(numerical variables)

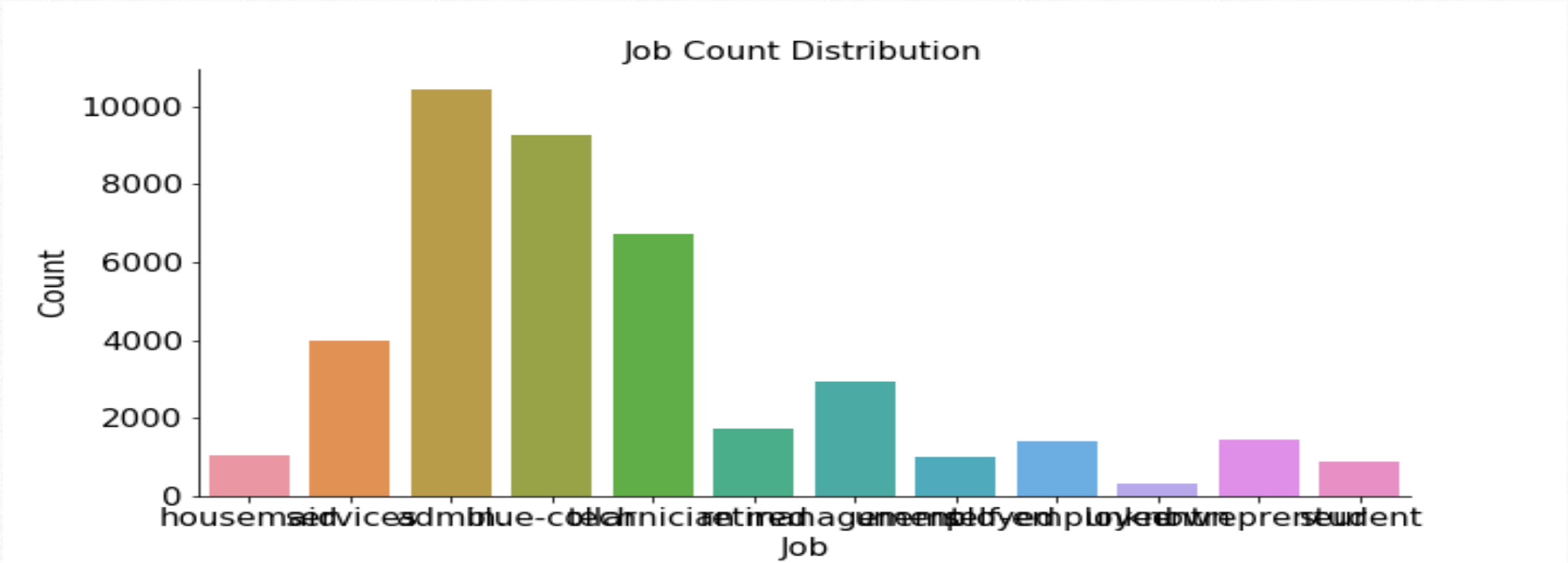
	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
count	41188.0	41188.0	41188.0	41188.0	41188.0
mean	0.08188550063191461	93.57566436831263	-40.50260027191399	3.621290812858068	5167.035910941844
std	1.5709597405172309	0.5788400489541813	4.628197856174375	1.7344474048511707	72.25152766826527
min	-3.4	92.20100000000001	-50.8	0.634	4963.6
25%	-1.8	93.075	-42.7	1.344	5099.1
50%	1.1	93.749	-41.8	4.857	5191.0
75%	1.4	93.994	-36.4	4.961	5228.1
max	1.4	94.76700000000001	-26.9	5.045	5228.1

Correlation Matrix

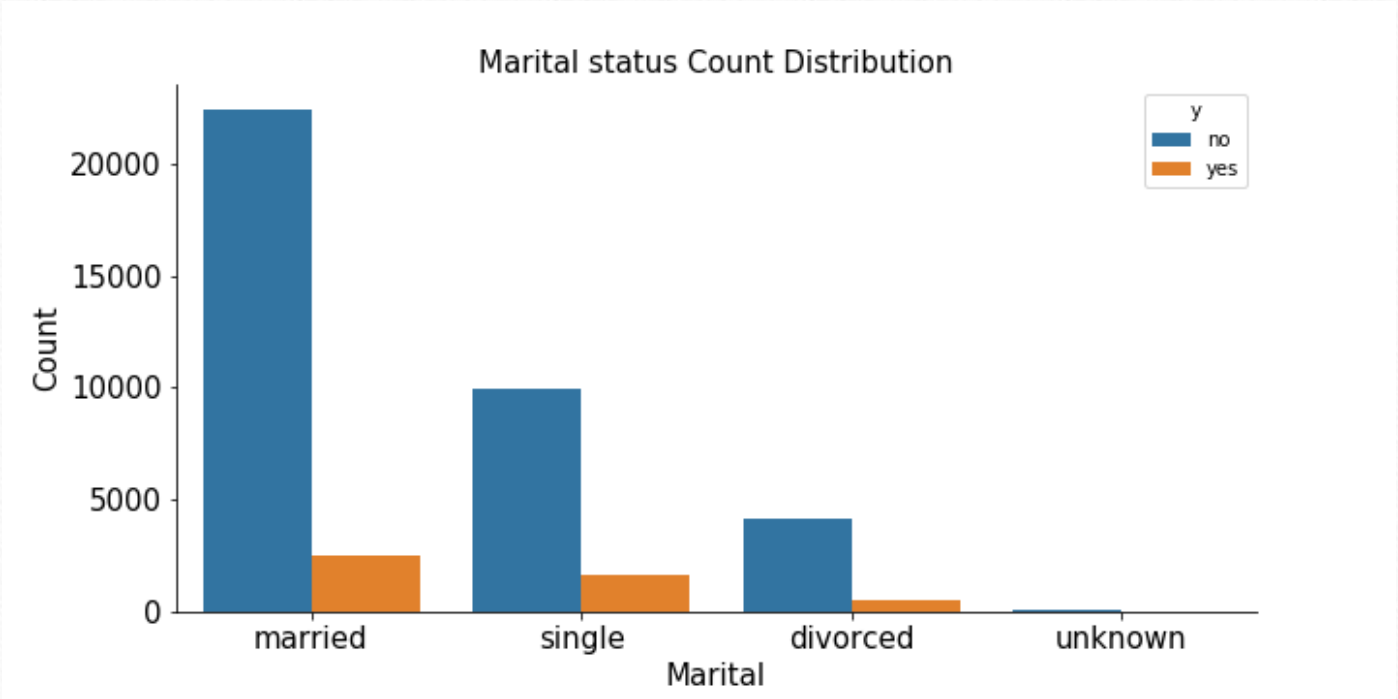


Categorical variables

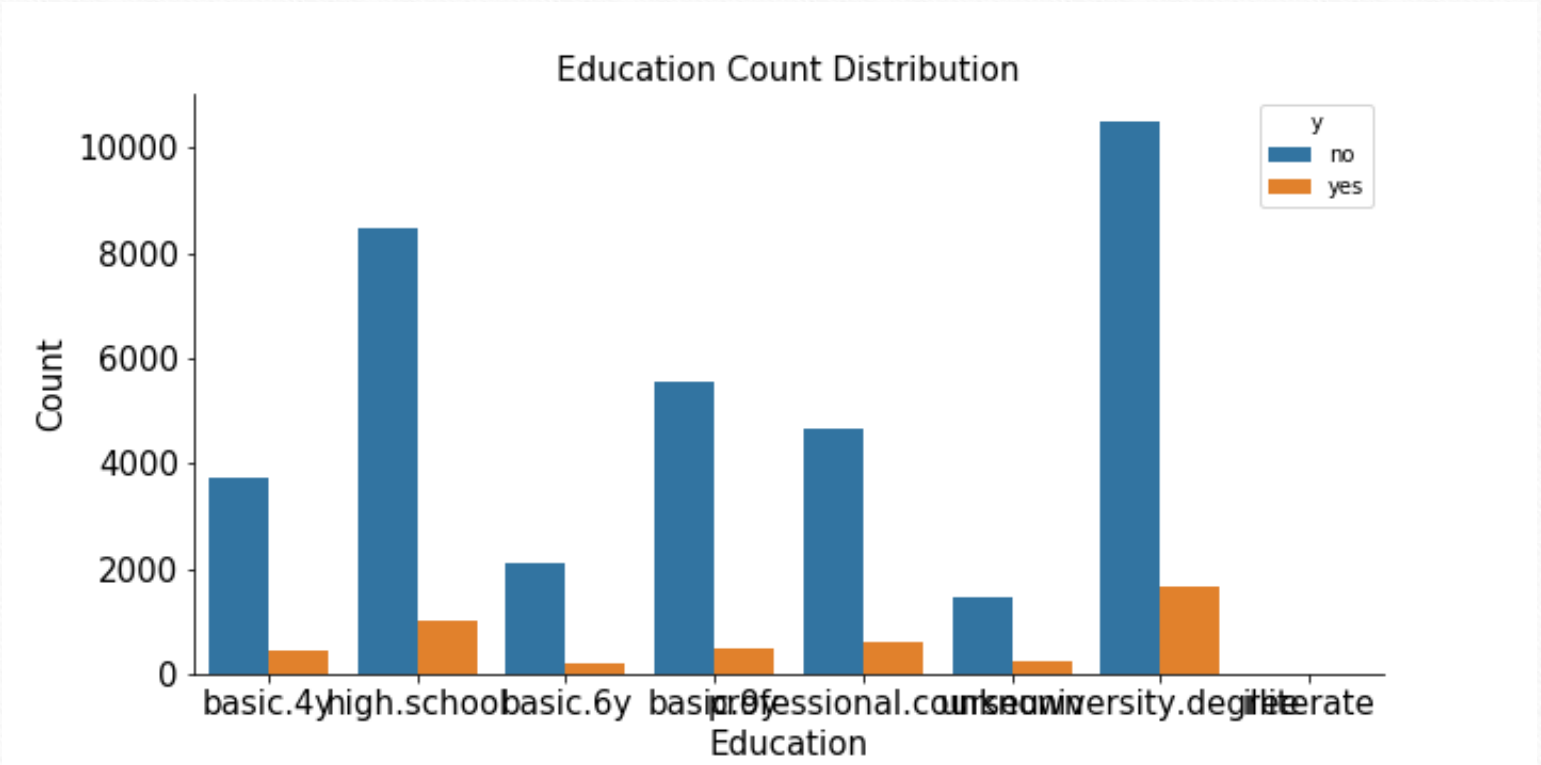
Job count distribution



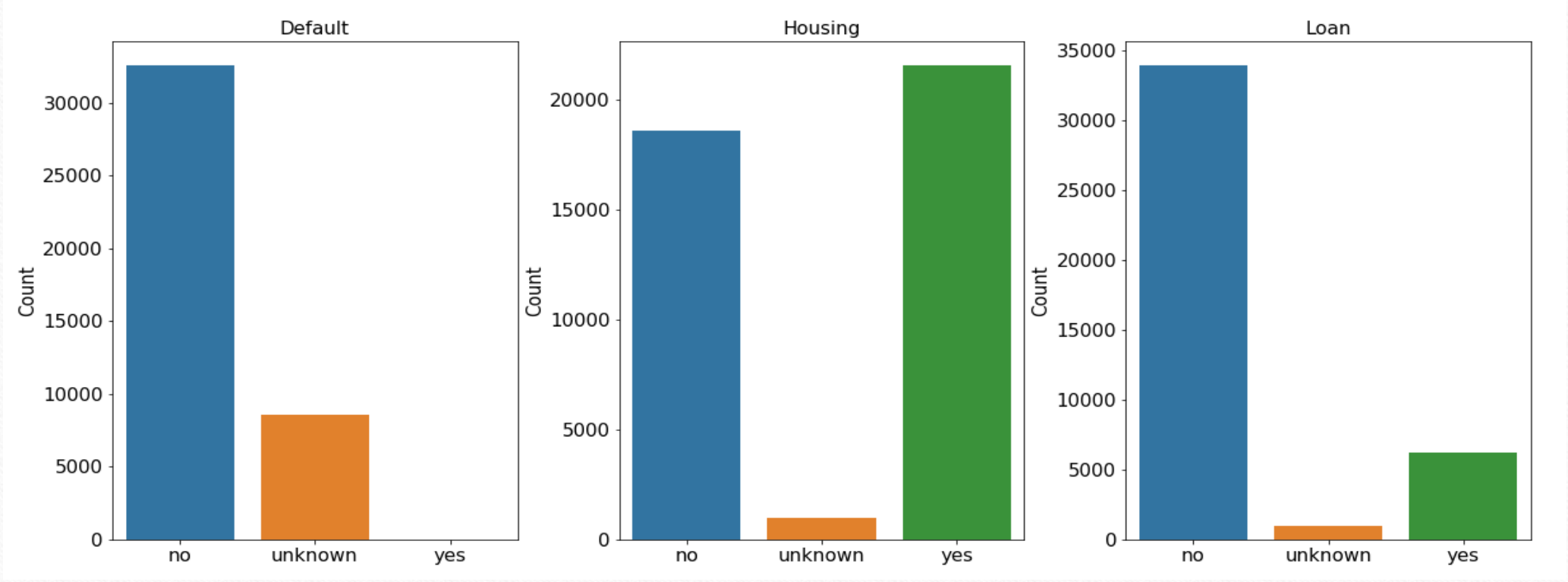
Marital status vs Deposit



Education Vs Deposit

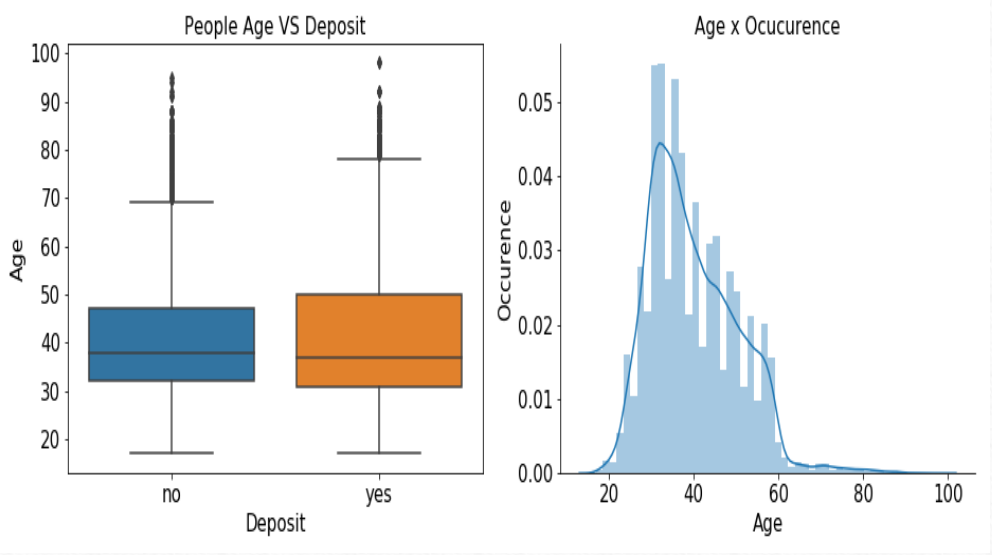


Default, Housing, Personal Loan

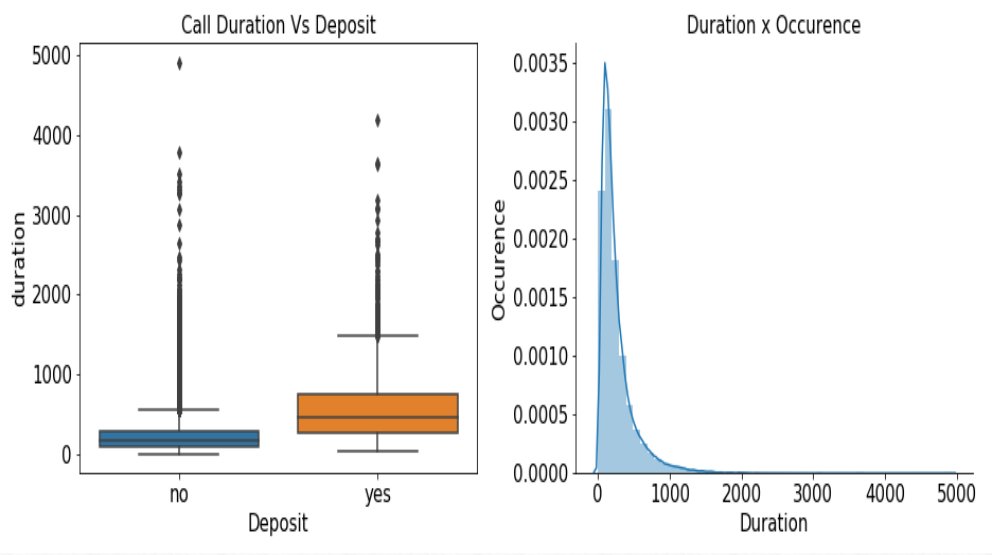


Outliers

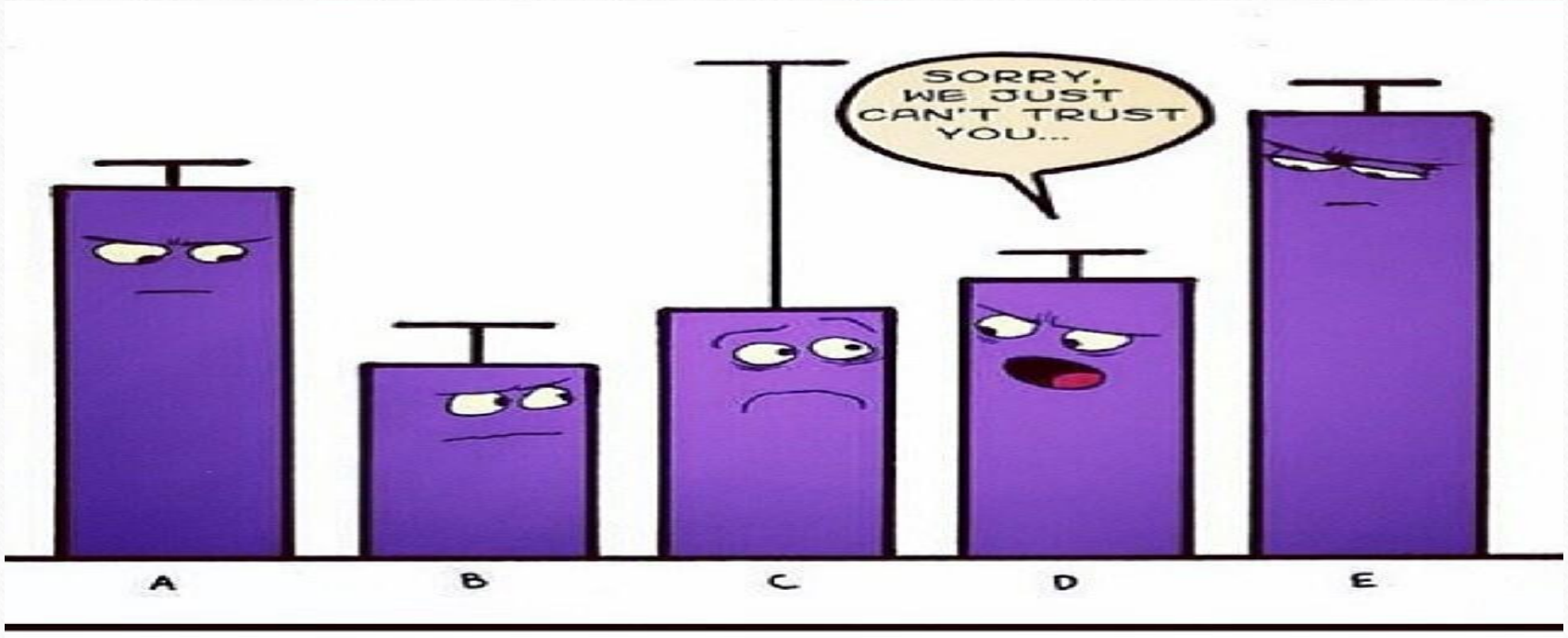
Age



Duration



Boxplot and outliers



Pre-processing Data

Outlier processing

Label Encoding

Scaling

Modelling

Model-1

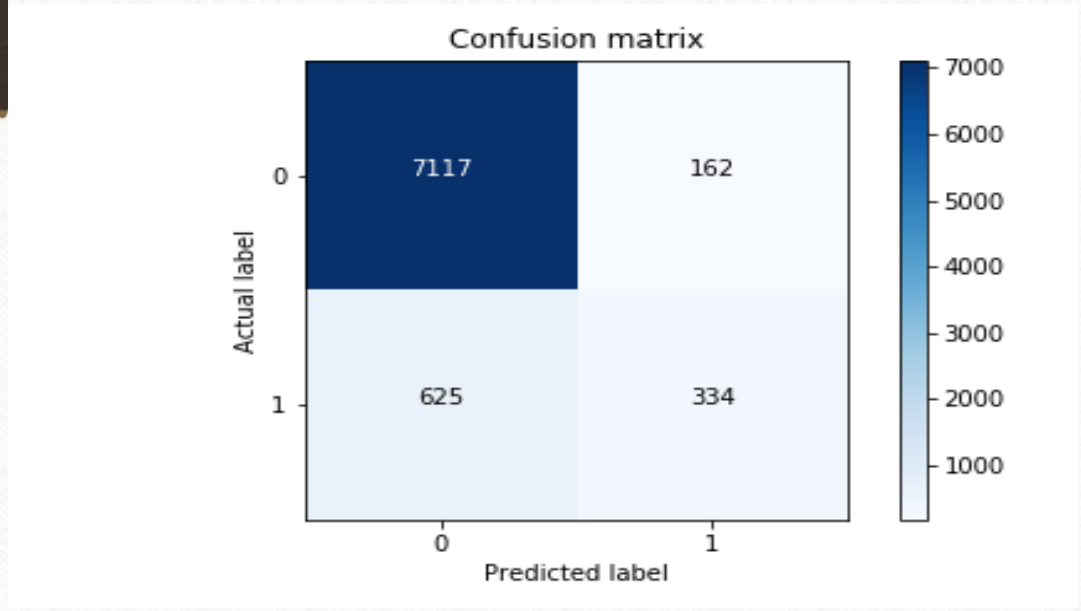
- Unscaled data
- 20 variables
- Algorithm:
  - Logistic regression

Model-2

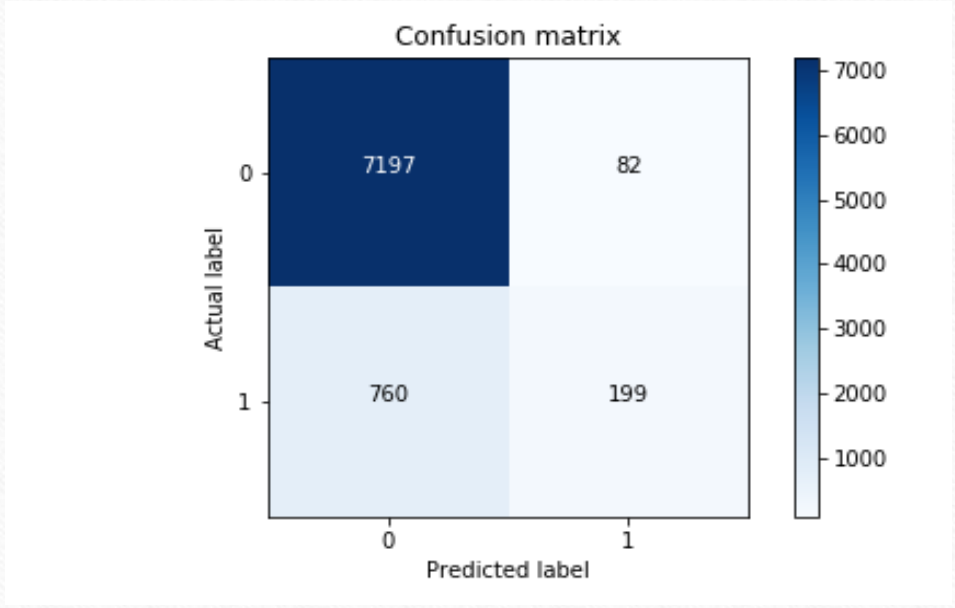
- Scaled
- PCA (comp=5) &Clustered(k=4)
- 6 variables
- Algorithms:
  - PCA, K-Means, Logistic Regression

Confusion matrix

Model-1



Model-2



Classification Metrics

Measure	Formula
Precision	$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
Recall / Sensitivity	$Recall/Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
Selectivity	$Selectivity = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}$
Accuracy	$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$
F-Measure	$F - Measure = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$

Model Evaluation

Accuracy

Precision

Recall

Specificity

ROC

AUC score

Model Evaluation

Model-1

	precision	recall	f1-score
0	0.92	0.98	0.95
1	0.67	0.35	0.46
Avg/Tot	0.89	0.90	0.89

Model-2

	precision	recall	f1-score
0	0.90	0.99	0.94
1	0.71	0.21	0.32
Avg/Tot	0.88	0.90	0.87

Other Evaluations

Model-1

- Accuracy:
  - 90.0
- Cross validation Score:
  - 0.90998
- Specificity:
  - 0.9777

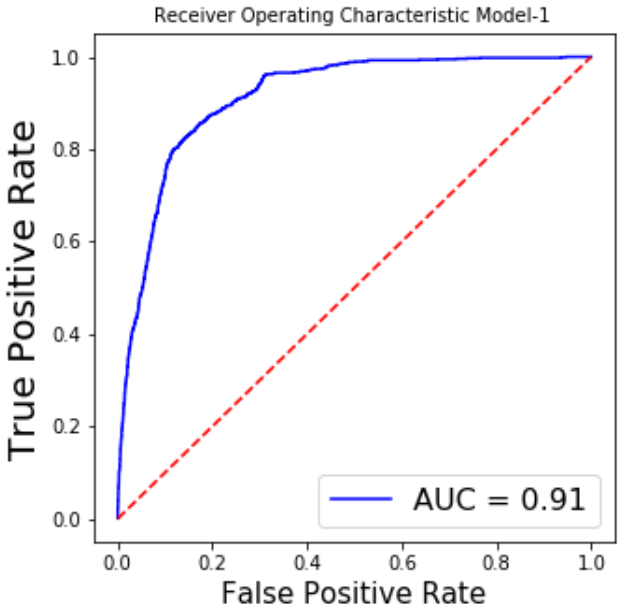
Model-2

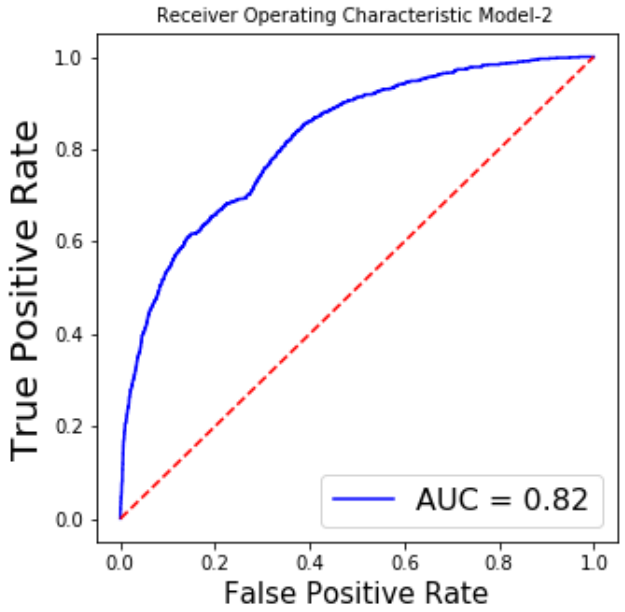
- Accuracy:
  - 90.0
- Cross Validation Score:
  - 0.898300
- Specificity
  - 0.9887

Insights

Both the models have almost same performance .

ROC(Model-1 Vs Model-2)AUC





Conclusion

Model – I is more simple and have a higher AUC score compared to Model-II

Hence the Model-I is more suitable for predicting the customers for this case.

Thank you!!!

Have a nice day!!!!!!