

Analysis of Bank Marketing Campaign

Problem Statement:

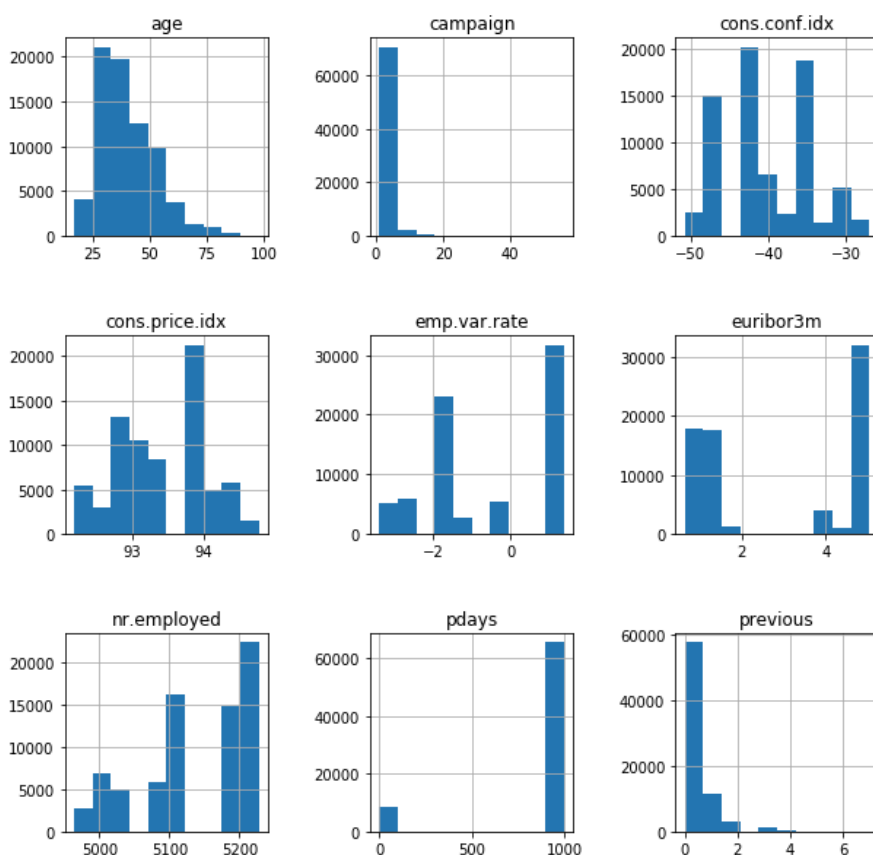
Data gathered and uploaded in the UCI Machine Learning Repository by the Portuguese Banking institution to produce an effective prediction model to make analysis of customer acquisition in the marketing campaign.

Data:

Data contains customer information like the age, job, marital status, education, loans etc. Together the data has 10 categorical variables, 9 Numerical variables and one output variable determining the customer interest in the campaign.

EDA:

Numerical Variables:



```
In [9]: idx_numeric=[0,10,11,12,14,15,16,17,18]
scaler = MinMaxScaler()
data[data.columns[idx_numeric]] = scaler.fit_transform(data[data.columns[idx_numeric]])

/Users/IAMNaveed/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/data.py:323: DataConversionWarning: Data with input dtype int64, float64 were all converted to float64 by MinMaxScaler.
return self.partial_fit(X, y)
```

```
In [10]: data.head()
```

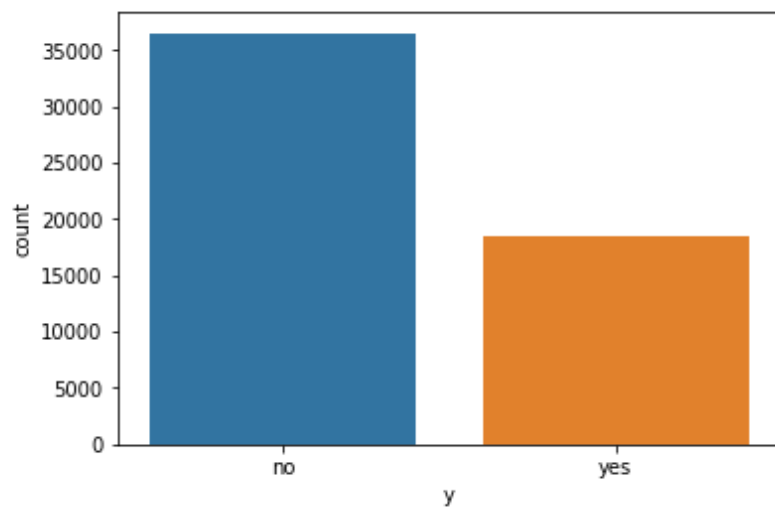
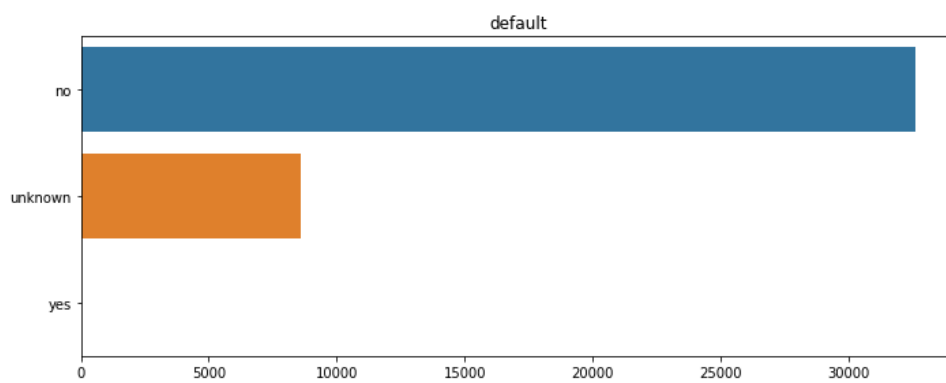
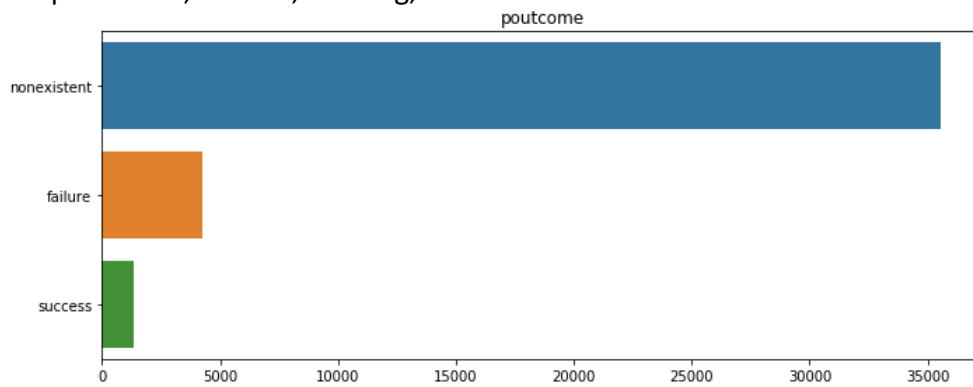
```
Out[10]:
```

using	loan	contact	month	day_of_week	...	pdays	previous	poutcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y	pdays2
no	no	telephone	may	mon	...	1.0	0.0	nonexistent	0.9375	0.698753	0.60251	0.957379	0.859735	no	0
no	no	telephone	may	mon	...	1.0	0.0	nonexistent	0.9375	0.698753	0.60251	0.957379	0.859735	no	0
yes	no	telephone	may	mon	...	1.0	0.0	nonexistent	0.9375	0.698753	0.60251	0.957379	0.859735	no	0
no	no	telephone	may	mon	...	1.0	0.0	nonexistent	0.9375	0.698753	0.60251	0.957379	0.859735	no	0
no	yes	telephone	may	mon	...	1.0	0.0	nonexistent	0.9375	0.698753	0.60251	0.957379	0.859735	no	0

Categorical:

Nominal: Age, Campaign, education, contact, month, day_of_week.

Ordinal: poutcome, default, housing, loan.



```

In [11]: data['poutcome'] = data['poutcome'].map({'failure': -1, 'nonexistent': 0, 'success': 1})
data['default'] = data['default'].map({'yes': -1, 'unknown': 0, 'no': 1})
data['housing'] = data['housing'].map({'yes': -1, 'unknown': 0, 'no': 1})
data['loan'] = data['loan'].map({'yes': -1, 'unknown': 0, 'no': 1})

In [12]: # One hot encoding of nominal variables
nominal = ['job', 'marital', 'education', 'contact', 'month', 'day_of_week']
dataProcessed = pd.get_dummies(data, columns=nominal)
dataProcessed['y'] = dataProcessed['y'].map({'yes': 1, 'no': 0})
dataProcessed.head()

Out[12]:

```

idx	...	month_mar	month_may	month_nov	month_oct	month_sep	day_of_week_fri	day_of_week_mon	day_of_week_thu	day_of_week_tue	day_of_week_wed
'53	...	0	1	0	0	0	0	1	0	0	0
'53	...	0	1	0	0	0	0	1	0	0	0
'53	...	0	1	0	0	0	0	1	0	0	0
'53	...	0	1	0	0	0	0	1	0	0	0
'53	...	0	1	0	0	0	0	1	0	0	0

Expectation:

For the above processed data, I would like to try and execute different classification problems.