Optimizing Efficiency for Bank Marketing Campaign

Era Kalaja, Kathleen Wang, Cather Zhang

The Dataset

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution.

The classification goal is to predict if the client will subscribe a term deposit (binary variable y).

Instances: 41188
Features: 20

- Demographic: age, job, education, marital
- Financial: credit default, personal loan, housing loan
- Previous Campaign information and outcome
- Social economic factors: employment variation rate, euribor rate, consumer price and confidence index, number of employees in the market

Motivation

Business Problem:
By analyzing past campaign data, find out:
- whether a client will subscribe to a term deposit
- what type of client is most likely to do so

Goal:
Ensure efficiency in marketing campaign while optimizing the number of people who subscribe through an **effective** and **reliable** solution

Conjectures

Demographic

'Higher' level professions and education are likely to subscribe.

Debt

Individuals facing already existing debts are less likely to subscribe.
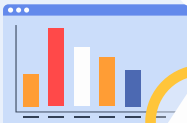
Previous Campaign

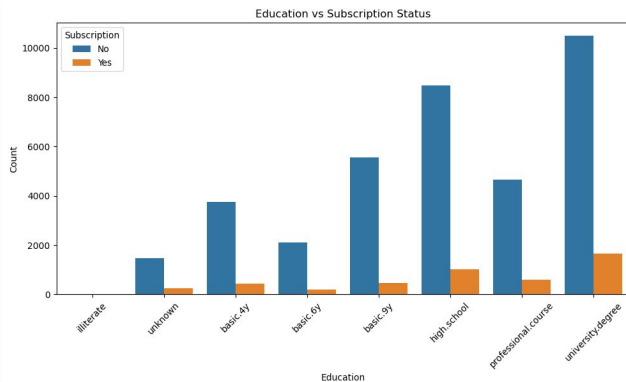Results from the previous marketing campaign can predict the results of this current campaign.
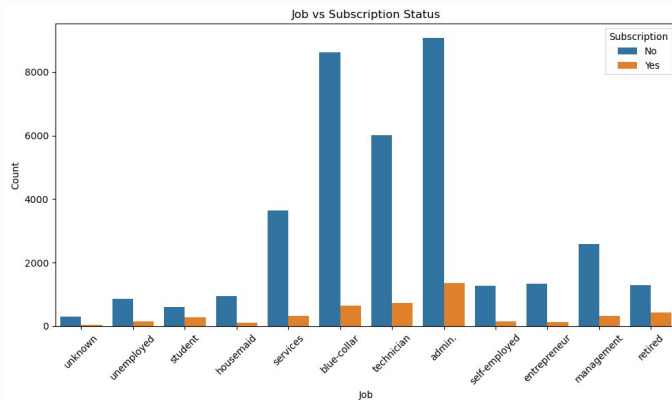
Social Economic Indicators

The behavior of the social and economic indicators impacts the campaign's success.

Conjecture 1: Demographic

Student: 31%

Retirement: 25%

Self defined 'order', therefore difficult to support.

Basic Avg.: 8.8%

High School +: 12%

Higher subscription rates in high school and up.

Conjecture 2: Debt
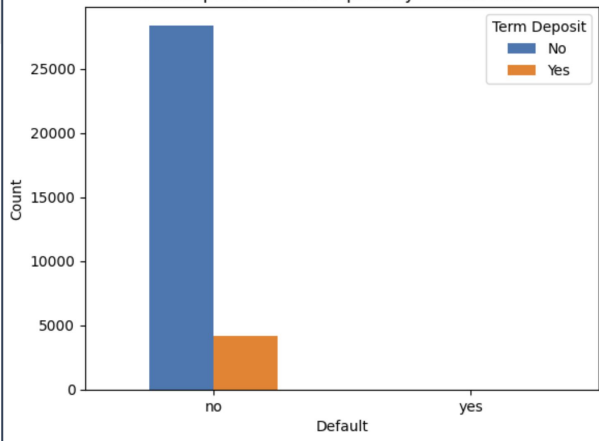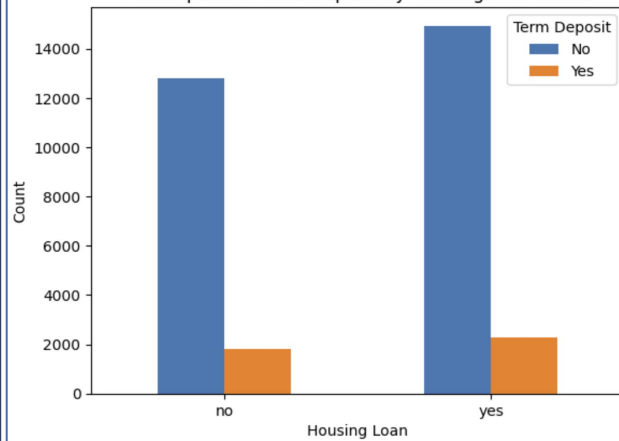

Default
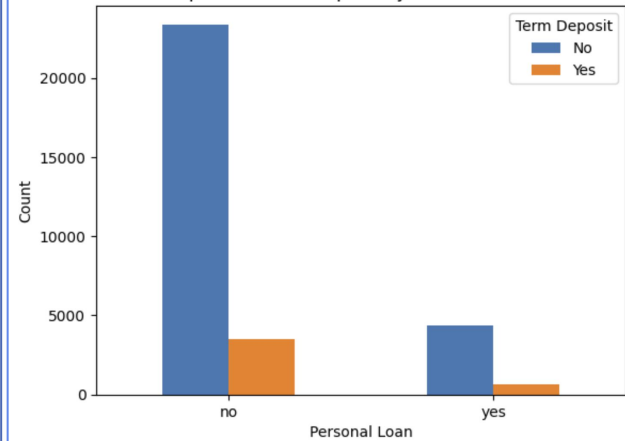
People in default are less likely to subscribe

Housing

People with a housing loan less likely to subscribe

Personal

People with a personal loan less likely to subscribe

Conjecture 3: Previous Campaign

poutcome ●────────────● **y**
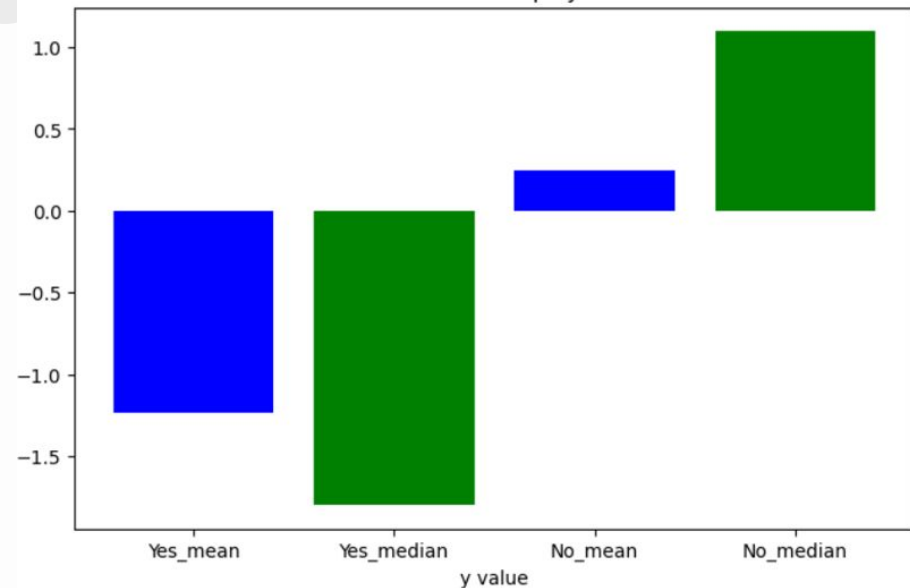
Previous campaign outcome

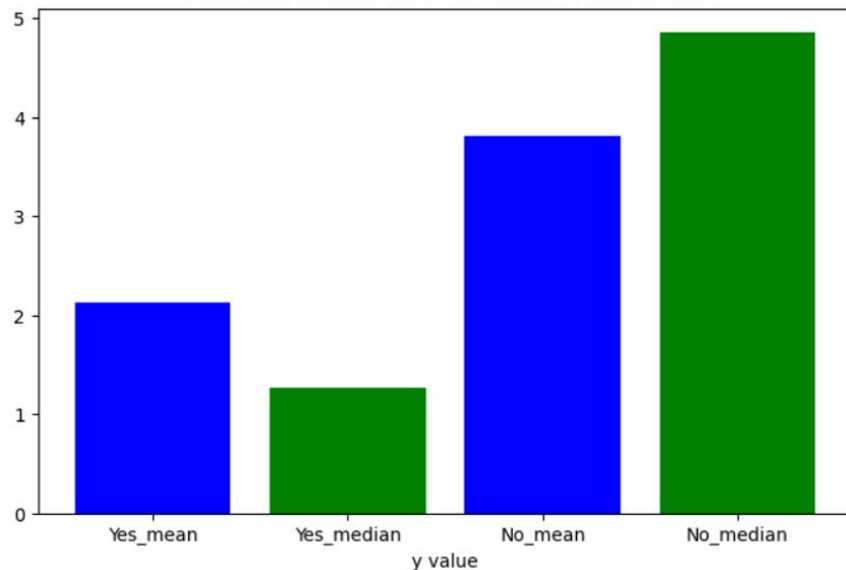Target variable, subscribed (y/n)?

Correlation Coefficient: 0.19

Conjecture 4: Econ

Mean and median value of employment variation rate

Mean and median value of euribor 3 month rate

Employment variation rate, number of employees, and euribor rate have the highest correlation coefficient to "y", and the three factors are highly correlated with each other
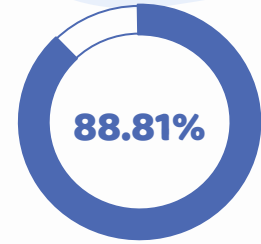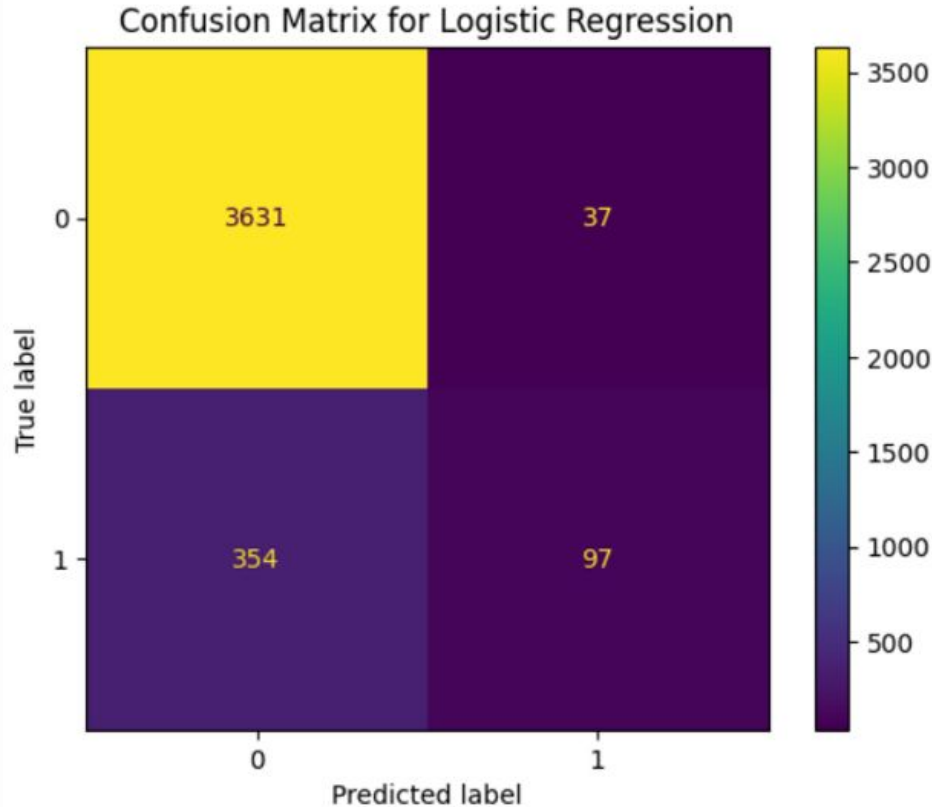
During periods of economic uncertainty or job market downturns, people tend to seek a safer and more secure way to invest

Based on the outcome of our conjectures, we developed classification models and tested for feature importance.

|	Feature	Coefficient	Abs_Coefficient
12	cons.price.idx	0.323182	0.323182
14	euribor3m	−0.181200	0.181200
11	emp.var.rate	−0.166905	0.166905
4	default	−0.082857	0.082857
2	marital	−0.078254	0.078254
3	education	0.057127	0.057127
7	campaign	−0.054267	0.054267
13	cons.conf.idx	0.025195	0.025195
9	previous	0.021180	0.021180
1	job	−0.019857	0.019857
6	loan	0.008175	0.008175
5	housing	−0.007399	0.007399
0	age	0.006030	0.006030
15	nr.employed	−0.005762	0.005762
10	poutcome	−0.005166	0.005166
8	pdays	−0.001808	0.001808

Model Accuracy

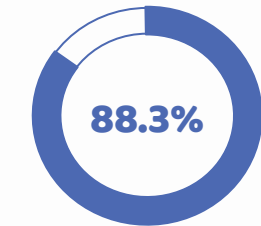Eg. Bank reach out to 37+97=134 candidates out of all, 97 of them will subscribe

-> **72.4%**

Confusion Matrix for Logistic Regression

88.81%
Random forest

90.51%
Logistic Regression
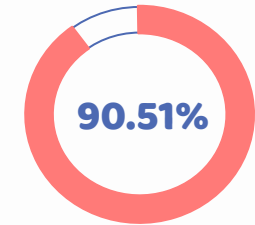
88.3%
KNN Classification

Thank you...
Questions?