



Bank Marketing

Group 2

Adarsh Salapaka

Evelyn Lee

Nien-Thing Chiang

Vidhi Gandhi



Overview



- We have a dataset related with direct marketing campaign of a Portuguese banking institution that was based on phone calls. The purpose of their campaign was to market term deposits towards different customers.
- 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.
- The data under study here is called Bank Marketing Dataset (BMD) and he was found in the Machine Learning Repository (UCI). The data is public available in the url <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>.
- In the BMD data we have 45211 observations, with 17 variables

Data Attributes



1 # Demographic data of customers

- age (numeric)
- job : type of job
(categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
- marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
- education (categorical: 'primary', 'secondary', 'unknown', 'missing')
- default: has credit in default? (categorical: 'no', 'yes', 'unknown')
- housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
- loan: has personal loan? (categorical: 'no', 'yes', 'unknown')

Data Attributes- continued



2 # Related with the last contact of the current campaign:

- contact: contact communication type (categorical: 'cellular','telephone')
- month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- day: last contact day of the week
- * • duration: last contact duration, in seconds (numeric).
- * • campaign: number of contacts performed during this campaign and for this client (numeric, includes the last contact)

3 # Related to previous Campaign:

- pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means the 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')

4 Output variable (desired target):

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

Business Goal



Problem:

Re-calling “wrong” customers

- High labor costs
- Harming customer relationship

Business Goal:

Improve marketing effectiveness by targeting the right customers in their next campaign.

Stakeholders:

- Bank Marketing Team
- Bank Service Employees
- Customers

Opportunities:

Gain revenues and lower costs by having more efficient marketing results

Challenges:

Very harmful for the mispredicted attributes

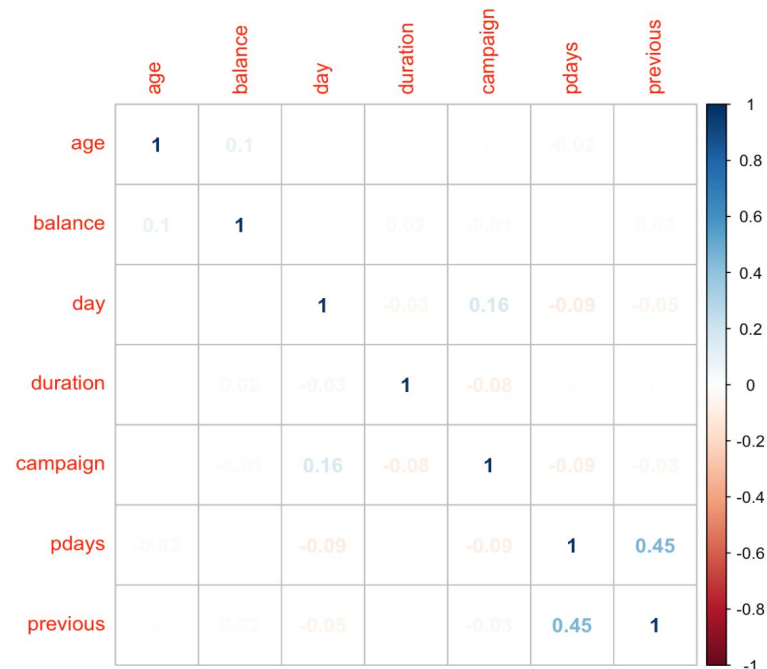
Summary Statistics & Plots

age	job	marital	education	default
Min. :18.00	blue-collar:9732	divorced: 5207	primary : 6851	no :44396
1st Qu.:33.00	management :9458	married :27214	secondary:23202	yes: 815
Median :39.00	technician :7597	single :12790	tertiary :13301	
Mean :40.94	admin. :5171		unknown : 1857	
3rd Qu.:48.00	services :4154			
Max. :95.00	retired :2264			
	(Other) :6835			

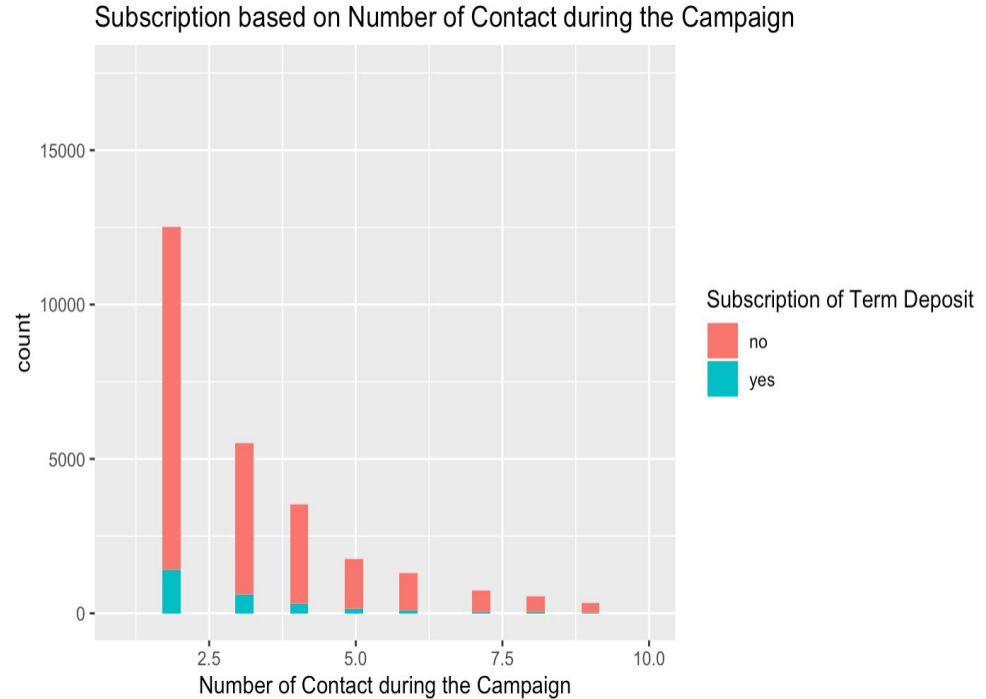
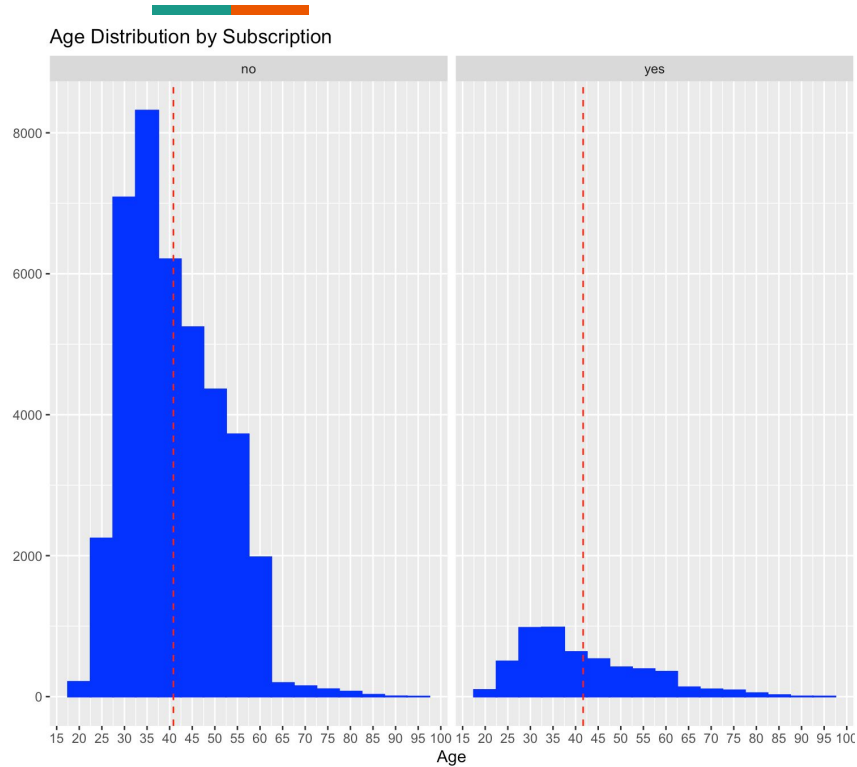
balance	housing	loan	contact	day	month
Min. : -8019	no :20081	no :37967	cellular :29285	Min. : 1.00	may :13766
1st Qu.: 72	yes:25130	yes: 7244	telephone: 2906	1st Qu.: 8.00	jul : 6895
Median : 448			unknown :13020	Median :16.00	aug : 6247
Mean : 1362				Mean :15.81	jun : 5341
3rd Qu.: 1428				3rd Qu.:21.00	nov : 3970
Max. :102127				Max. :31.00	apr : 2932
				(Other): 6060	

duration	campaign	pdays	previous	poutcome
Min. : 0.0	Min. : 1.000	Min. : -1.0	Min. : 0.0000	failure: 4901
1st Qu.: 103.0	1st Qu.: 1.000	1st Qu.: -1.0	1st Qu.: 0.0000	other : 1840
Median : 180.0	Median : 2.000	Median : -1.0	Median : 0.0000	success: 1511
Mean : 258.2	Mean : 2.764	Mean : 40.2	Mean : 0.5803	unknown:36959
3rd Qu.: 319.0	3rd Qu.: 3.000	3rd Qu.: -1.0	3rd Qu.: 0.0000	
Max. :4918.0	Max. :63.000	Max. :871.0	Max. :275.0000	

y
no :39922
yes: 5289



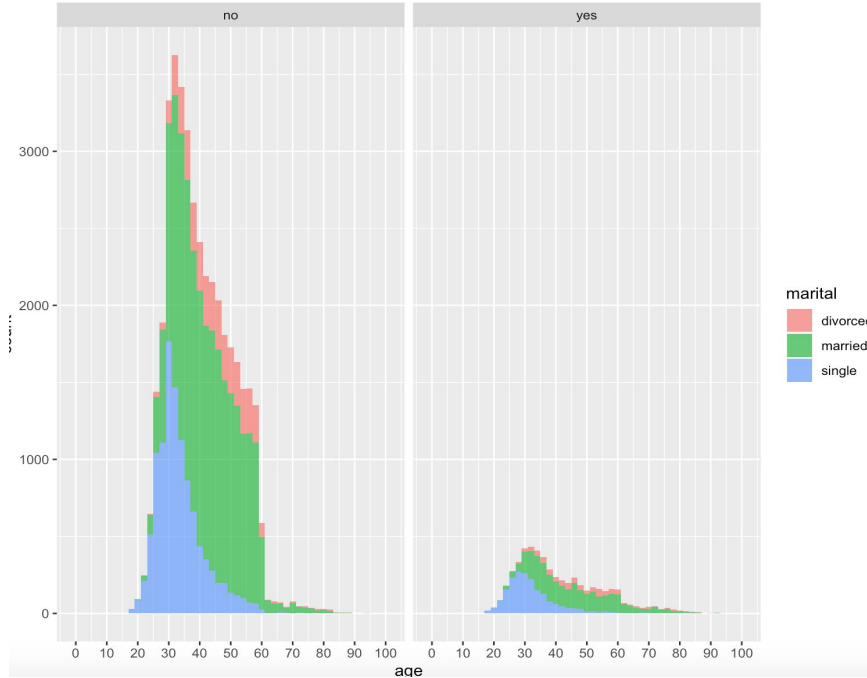
Exploratory Data Analysis



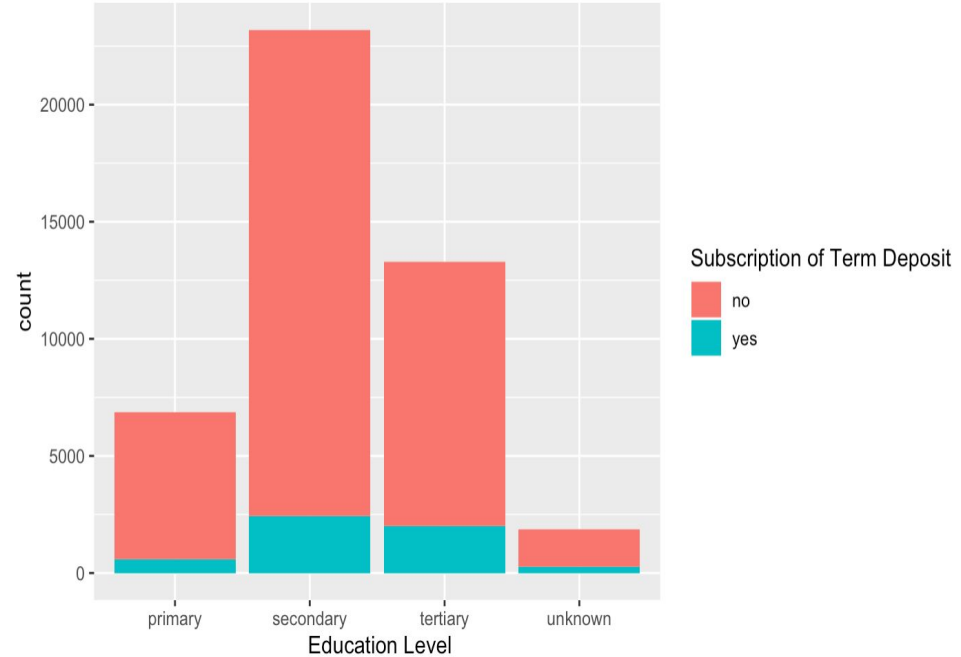
Exploratory Data Analysis



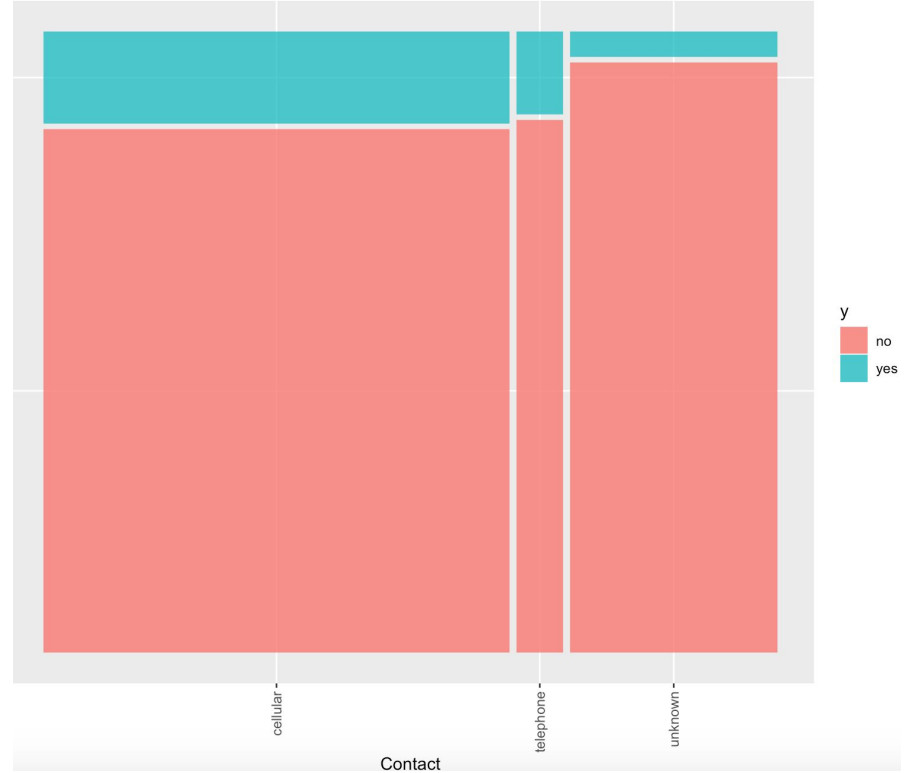
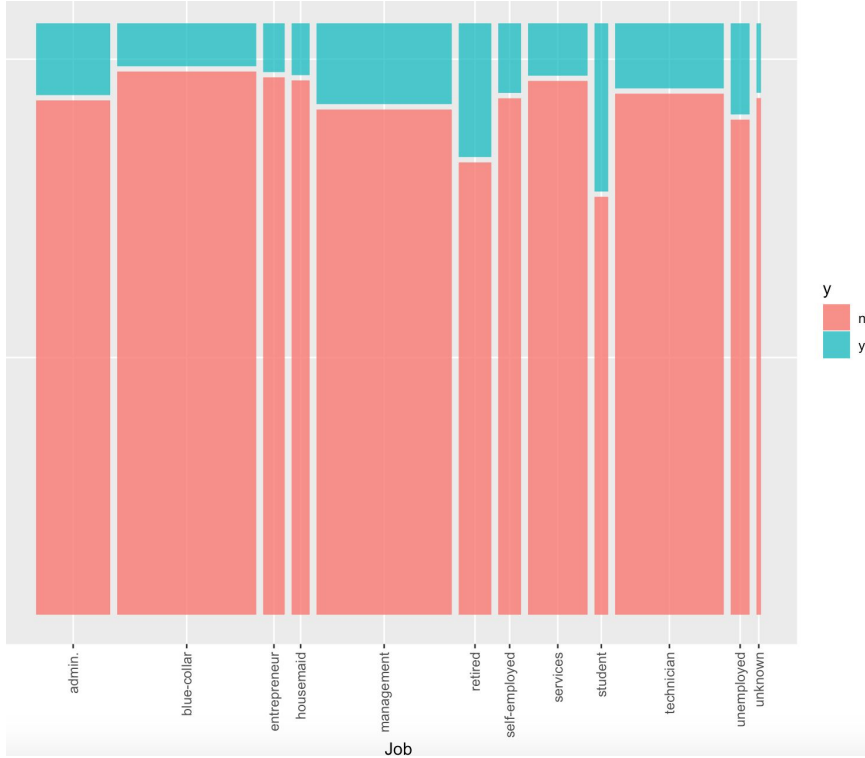
Age Distribution by Marital Status



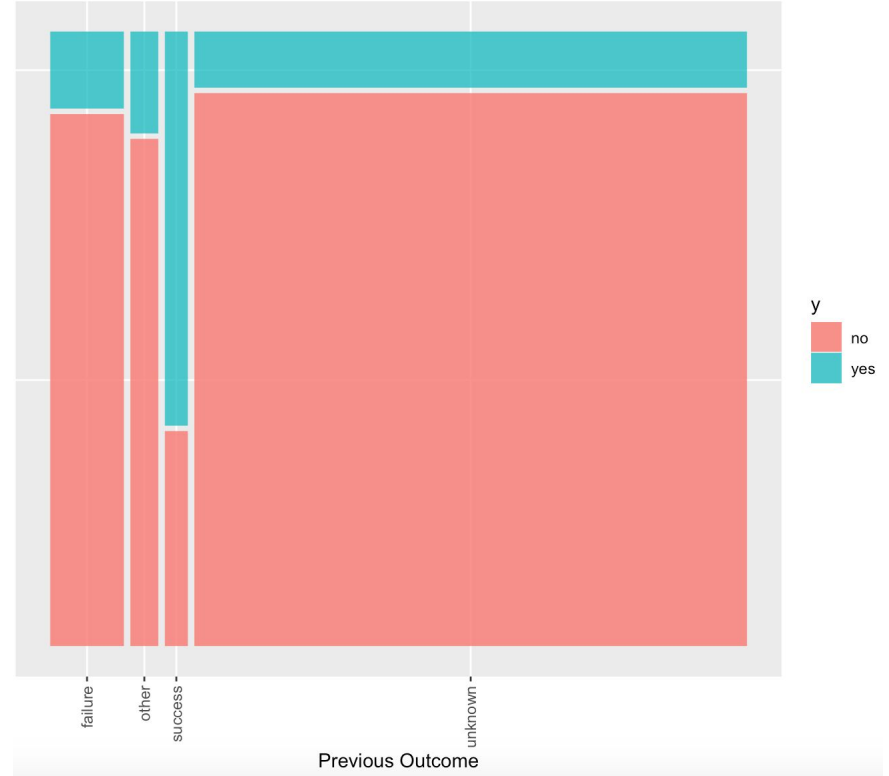
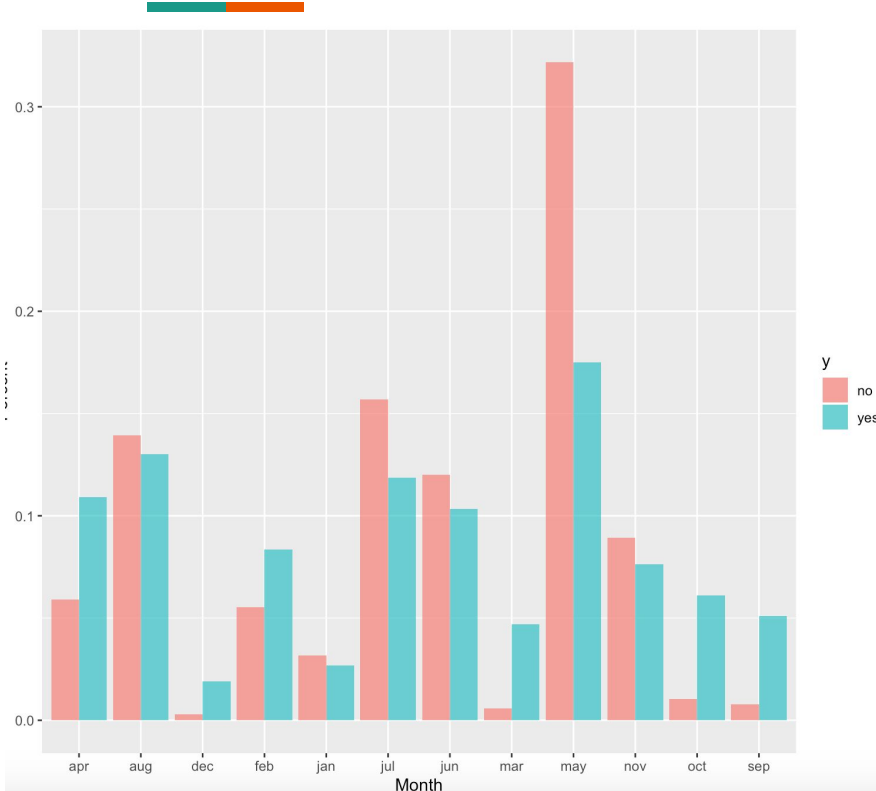
Term Deposit Subscription based on Education Level



Exploratory Data Analysis



Exploratory Data Analysis



Summary of key points



- Higher response among students (31.4%) and retired people (25.2%).
- Majority of people over 60 years old subscribed to a term deposit, which is a lot in comparison with younger individuals or young adults (aged lower than 30) and for the remaining observations said yes (aged between 30 and 60)).
- Singles slightly more like to say “yes” than divorced or married customers.
- A positive correlation exists between the number of years of education and the odds of subscribing to a term deposit. Users with Secondary and tertiary level of education have most number of term deposits as compared to primary education level.

Summary of key points- continued



- Cellular responders subscribed to a term deposit more than telephone responders did.
- Most of the calls were in May but there is higher conversion in March, September, October, and in December.
- The highest spike occurs during May, but it has the worst ratio of subscribers over persons contacted. Every month with a very low frequency of contact (March, September, October and December) shows very good results
- Majority of people who already subscribed to a term deposit after a previous contact have accepted to do it again. Even if they were denied before, they're still more enthusiastic to accept it than people who haven't been contacted before. So even if the previous campaign was a failure, recontacting people seems important.

Analysis Approach / Methodology



1. Clean Data

- *Get rid of columns "pdays" and "day"*
- *Transform categorical variables into dummy variables.*

2. Split data into **Training and Test Dataset**. Then apply **SMOTE** on training dataset.

3. Predict whether a certain customer will subscribe a term deposit or not using different **classification models**: Naive Bayes, Decision Tree, Random Forest, Binary Logit.

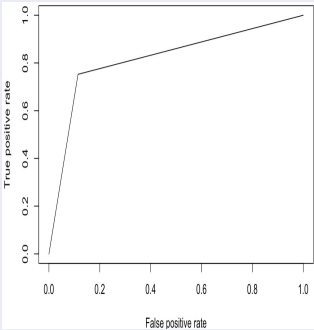
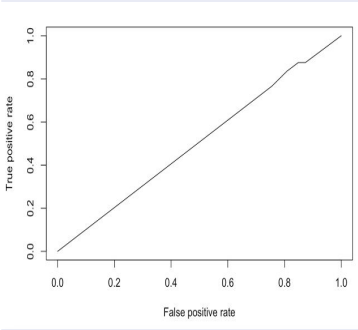
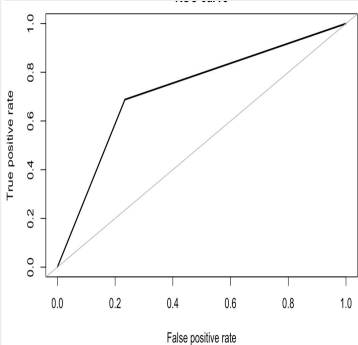
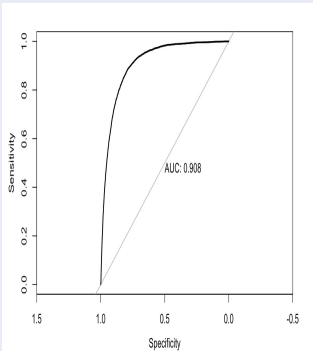
4. Compare **AUC ROC** to gauge the performance of all models.

5. Use **Random Forest** to gain insights on which factor are most important in determining subscription.

6. Perform **Logistic Regression** on the top attributes suggested by Random Forest.

7. Extract **managerial insights** from Logistic Regression Results.

Models AUC/ ROC Comparison

	Random Forest	Decision Tree	Naïve Bayes	Binary Logit
Accuracy	86.99%	84.36%	76.39%	N/a
AUC	81.89%	83.41%	68.17%	90.3%
ROC Curve				

Random Forest Factors Regression Results

Propensity Score

y <dbl>	mnfit <dbl>
0	0.08323575
1	0.37172669

2 rows

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.817e+00	3.312e-01	-5.485	4.13e-08	***
duration	6.521e-03	2.283e-04	28.561	< 2e-16	***
mar	2.340e+00	3.709e-01	6.309	2.80e-10	***
balance	6.534e-05	1.598e-05	4.090	4.32e-05	***
poutcome_success	3.744e+00	2.884e-01	12.983	< 2e-16	***
con_unknown	-2.355e+00	2.706e-01	-8.700	< 2e-16	***
oct	2.213e+00	3.509e-01	6.305	2.88e-10	***
previous	6.578e-02	2.878e-02	2.285	0.022298	*
campaign	-1.637e-01	2.973e-02	-5.508	3.62e-08	***
aug	-4.037e-01	1.699e-01	-2.376	0.017507	*
jun	6.568e-01	2.019e-01	3.253	0.001140	**
age_3060	-5.217e-01	1.495e-01	-3.491	0.000482	***
housing_yes	-8.095e-01	1.143e-01	-7.083	1.41e-12	***
jul	-5.823e-01	1.671e-01	-3.485	0.000492	***
sep	1.897e+00	4.342e-01	4.370	1.24e-05	***
con_cellular	-3.766e-01	2.159e-01	-1.744	0.081102	.
nov	-7.238e-01	1.880e-01	-3.849	0.000118	***
apr	2.631e-01	1.809e-01	1.455	0.145739	
marital_married	-3.735e-01	1.072e-01	-3.485	0.000492	***
loan_no	5.795e-01	1.658e-01	3.495	0.000474	***
tertiary	6.204e-01	1.549e-01	4.006	6.18e-05	***
age_60	2.780e-01	3.113e-01	0.893	0.371824	
secondary	4.155e-01	1.446e-01	2.874	0.004053	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5961.1 on 4299 degrees of freedom
Residual deviance: 2753.8 on 4277 degrees of freedom
AIC: 2799.8

Result Observations



- **Duration** is statistically significant, suggesting a strong association of the last contact duration and the chance of subscribing to a term deposit.
- The **months** (last month contacted) *march, october* and *september* are statistically significant and have positive coefficient, suggesting these months have most positive effect on clients' subscription. On the other hand, the **months** *august, june, july* and *november* are statistically significant and have negative coefficient, suggesting these months have most negative effect on clients' subscription.
- **Campaign** (signifying number contacts during the most recent campaign) is statistically significant. The increase in contacts actually decreases the chance of customers subscribing to a term deposit.
- **Balance** is statistically significant. The higher the customer's average yearly balance (in euros), the higher the chance of them subscribing to a term deposit.
- **Poutcome Success:** Those who had successfully subscribed to a term deposit as a result of the previous campaign are more likely to subscribe again.

** Although Unknown Contact type shows as statistically significant, it is not included in our observations because there isn't any actionable insights that can be gained from this factor.

Result Observations



Previous is statistically significant, suggesting a strong association between number of contacts performed before this campaign and for this client and the chance of subscribing to a term deposit by that client.

The ***age_3060*** is statistically significant, with a negative coefficient. This suggests that clients in this age group are less likely to subscribe to a term deposit.

Marital_married (signifying marital status as married) is statistically significant but with a negative coefficient. This suggests that clients who are married are less likely to subscribe to a term deposit.

Housing_yes (signifying has a house loan) is statistically significant but with a negative coefficient. This suggests that clients who have a house loan are less likely to subscribe to a term deposit.

Loan_no is statistically significant with a positive coefficient. This suggests that clients who do not have a loan are more likely to subscribe to a term deposit.

Findings and Managerial Suggestions



Target those who:

- Have had previously subscribed to a term deposit as a result of previous campaign efforts.
- Have a higher average yearly balance.
- Have no loan and are preferably not married.
- Have education level secondary and above.

Suggestions for Campaign strategy:

- Try to minimize the amount of times a customer is contacted.
- Focus more campaign efforts during March, September and October and avoid June, July, August and November.

Note:

Although duration of a customer call is not completely under the company's control, a longer duration suggests a customer has higher interest in subscribing for a term deposit. Also, recontacting someone again might increase the chance of a customer subscribing for a term deposit.

A hand is shown from the top, holding a US dollar coin and dropping it into the slot of a pink piggy bank. The piggy bank is a classic pig shape, made of glossy pink ceramic, with a small slot on its back for coins. The background is a plain, light gray.


Thank you!!

Questions?