A regression model to predict the outcome of a tele-marketing campaign.

Logistic Regression

Assignment 2

Ankita Avadhani (A0178495J)

Bethapudi Keerthana (A0178493M)

Sabrish Gopalakrishnan (A0178314A)

Sanchit Mittal (A0178507W)

Tan Zhi Wen David (A0056418A)

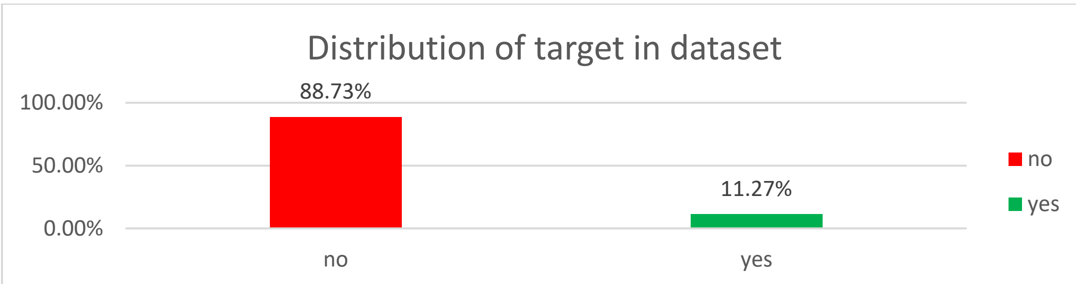
Contents

Business Scenario

A Portuguese Banking Institution is interested in selling a term deposit product to potential clients through tele-marketing. Based on the results from previous tele-marketing campaigns, the bank wants to assess if a given individual would subscribe to the term deposit. In order to do this, data from the previous direct marketing campaigns was collected for the period May 2008 – November 2010.

Objective

The objective is to analyze the given data and predict if a client would subscribe to Term Deposit product, using a Logistic regression model.



Data Preparation

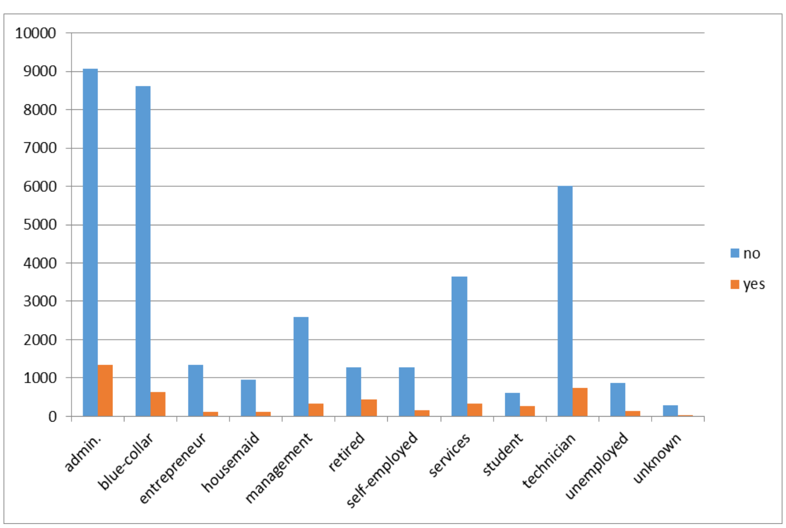
The observations with “unknown” values for job, marital, education, default, housing and loan have **not been imputed or removed** as these “unknown” levels **may carry significance from a banking domain perspective**. The variable “pdays” has been converted from numerical to categorical since 999 depicts a bound category on an otherwise numeric variable.

One_week	0 to 7 days
two_weeks	8 to 14 days
three_weeks	15 to 21 days
four_week	22 to 28 days
Never	999 - Never contacted before

Preliminary Analysis

The predictor variables were categorized into 4 groups: client data, last contact of current campaign, other attributes and social & economic context attributes. A study was conducted to understand each predictor variable and its relationship with the response variable. A few predictor variables such as age, education and marital displayed a weak relationship with the response variable. Others such as “month”, “contact” and “poutcome” came across as strong predictors. The effect of some of these predictors is analyzed below.

1. Job

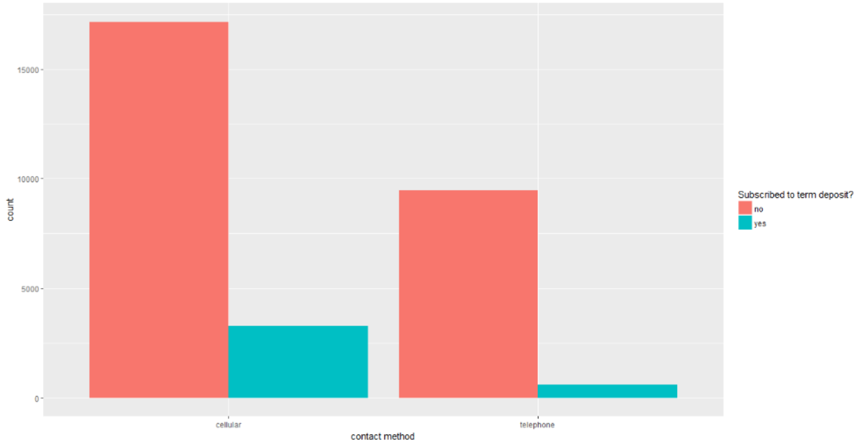


The highlighted fields show significant categories.

Row Labels	TD-No	TD-Yes	Grand Total	Yes %	No %
Admin	9070	1352	10422	12.973%	87.027%
Blue-collar	8616	638	9254	6.894%	93.106%
Entrepreneur	1332	124	1456	8.516%	91.484%
House-maid	954	106	1060	10.000%	90.000%
Mgmt.	2596	328	2924	11.218%	88.782%
Retired	1286	434	1720	25.233%	74.767%
Self-emp	1272	149	1421	10.486%	89.514%
services	3646	323	3969	8.138%	91.862%
Student	600	275	875	31.429%	68.571%
Technician	6013	730	6743	10.826%	89.174%
unemployed	870	144	1014	14.201%	85.799%
Unknown	293	37	330	11.212%	88.788%
Grand Total	**36548**	**4640**	**4118**		

2. Contact

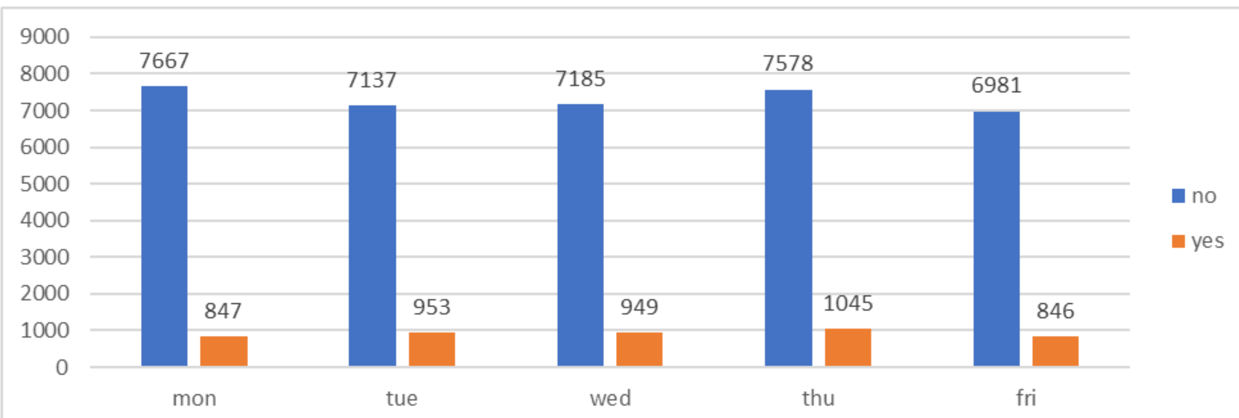
The **probability of reaching the customer is higher on cellular phone compared to landline telephone**. People who were contacted through cellular phone have been found more likely to subscribe to the term deposit than the people contacted through telephone. This is substantiated by the graph and table below.



Row Labels	Count	Percentage
Cellular	**20435**	
No	17163	83.99%
Yes	3272	16.01%
Telephone	**10043**	
No	9457	94.17%
Yes	586	5.84%
Grand Total	**30478**	

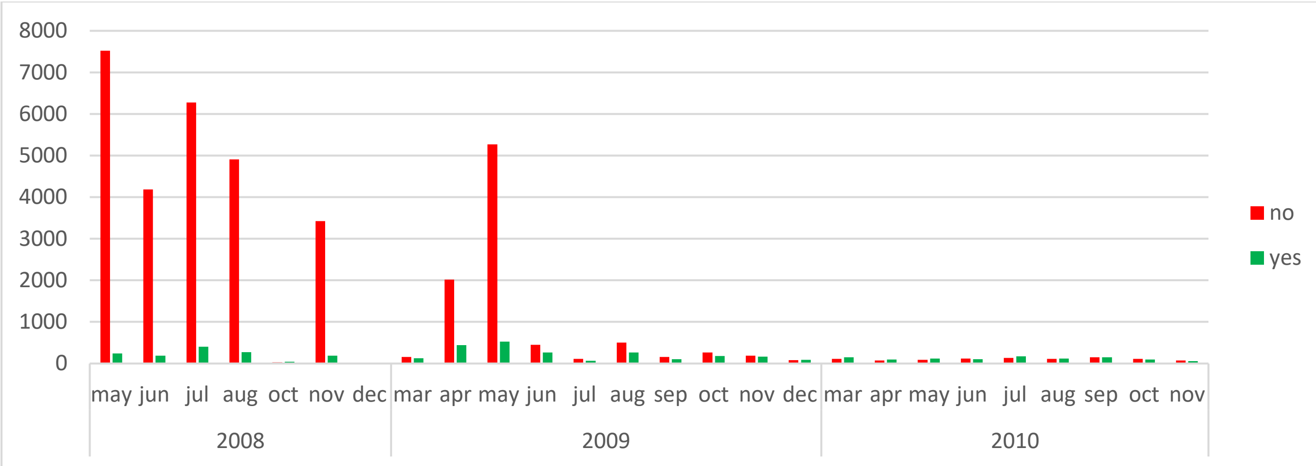
3. day_of_week

Since the number of people subscribing or declining the term deposit seems to be fairly consistent across all days of the week, this parameter needs to be checked for statistical significance in the model.



4. Month

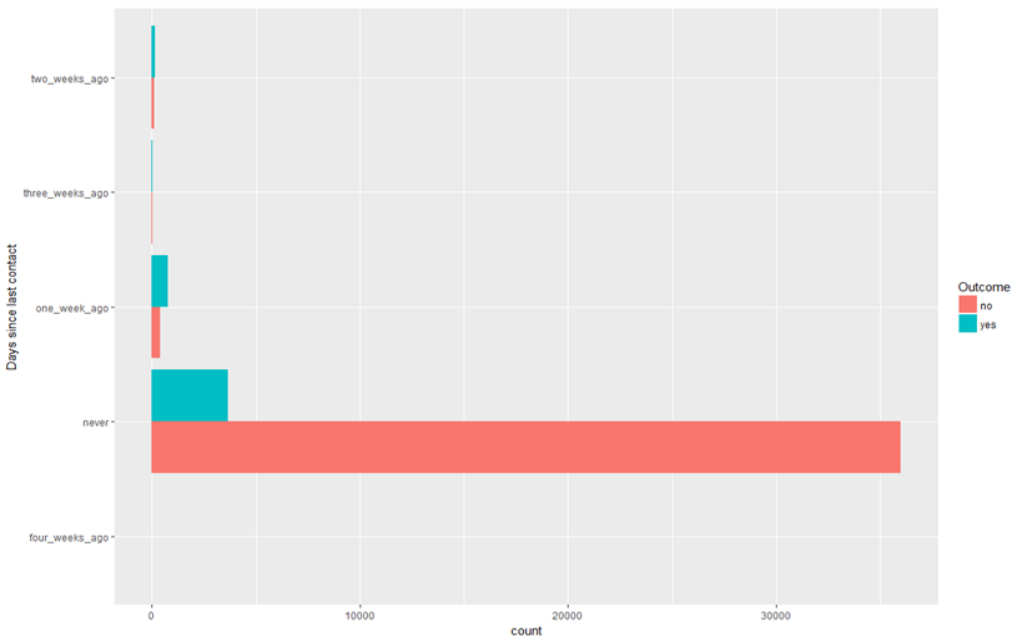
For a better understanding of the month variable we created an additional variable 'Year', since the data provided was already sorted by year. From the graph below, it is evident that the number of people who declined the term deposit was maximum during year 2008 and during mid-2009. **This can be explained by the economic recession that occurred during the same period and subsided in June 2009**. Due to this fluctuation, 'month' was diagnosed as the most significant factor during the initial run of the model. However, **after the occurrence of this uncommon event, the number of people subscribing or declining the term deposit remained fairly consistent**. Hence, the 'month' parameter cannot be considered as a parameter in the model for a realistic outcome.



5. Duration

The **duration of the call cannot be known before placing the call**, hence would not be considered for modelling.

6. pdays




```
read.csv("C:/Users/david/Downloads/latest_bank_v7.csv")
%>% group_by(pdays,y) %>%
summarise(count = n()) %>%
ggplot(aes(x=as.factor(pdays),y
=count,fill=as.factor(y))) +
geom_bar(stat =
"identity",position =
position_dodge()) +
labs(x='Days since last contact')
+ scale_fill_discrete(name =
"Outcome") + coord_flip()
```

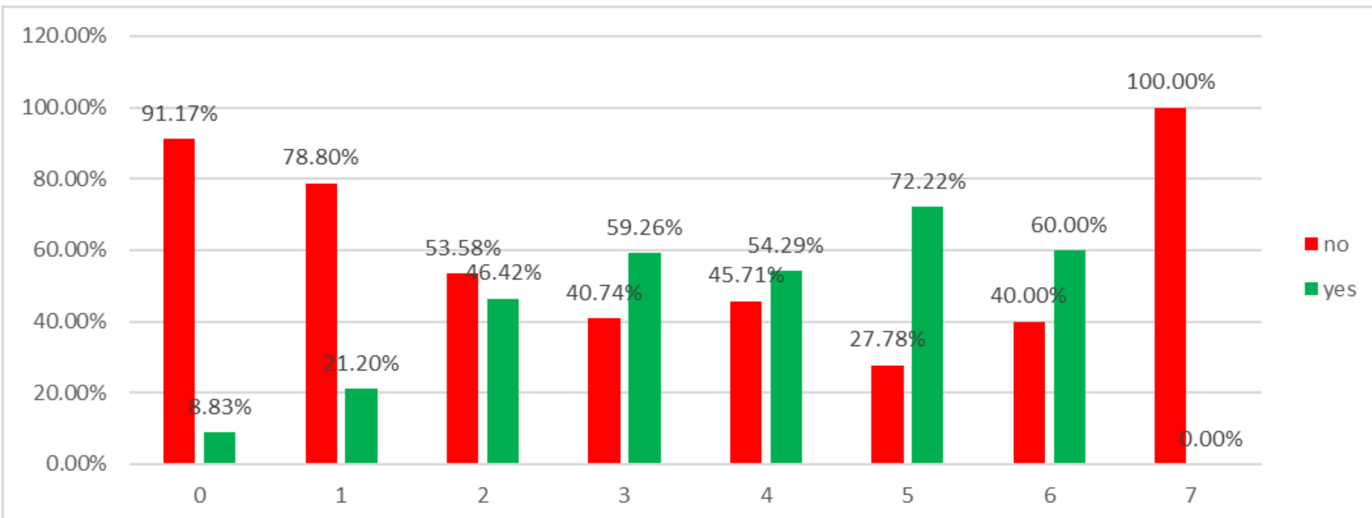
It is seen that clients who were never contacted before are unlikely to subscribe to the term deposit and clients who were contacted earlier are more likely to subscribe.

7. Campaign

The number of contacts made which resulted in ‘No’ is more spread out than ‘Yes’, as evident from the following box-plot. **This shows that repetitive persuasion doesn’t result in subscription to term deposit**.

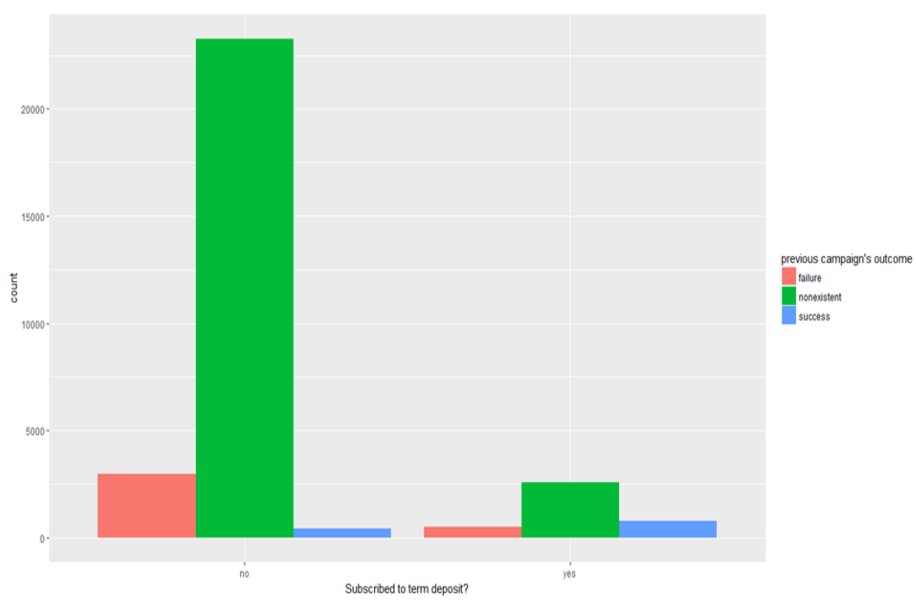
8. previous

The clients who have been previously contacted, are more likely to subscribe for the term deposit. As evident from the graph, most of the clients who did not subscribe to the term deposit are the ones who have been recently contacted.



9. poutcome

It is evident from the graph that majority of the clients who had subscribed to term deposit during the earlier campaign have opted for it again. While for the clients who did not take the term deposit earlier haven't subscribed to this time as well. **This shows that 'poutcome' is a significant parameter in predicting the outcome of current campaign.**




```
ggplot(data, aes(data$y,
..count..)) + geom_bar(aes(fill =
data$poutcome), position =
"dodge") + labs(x = "Subscribed
to term deposit?") +
scale_fill_discrete(name =
"previous campaign's
outcome")
```

10. emp.var.rate

The employment variation rate is the quarterly employment indicator over the course of a year. It shows the trend of employees being hired or fired due to shifts in the condition of the economy. When the variation rate is negative, it means the rate of hiring is low and people will be more conservative with their money. When the rate is positive, it denotes the rate of hiring is high and people are more likely to opt for term deposit.

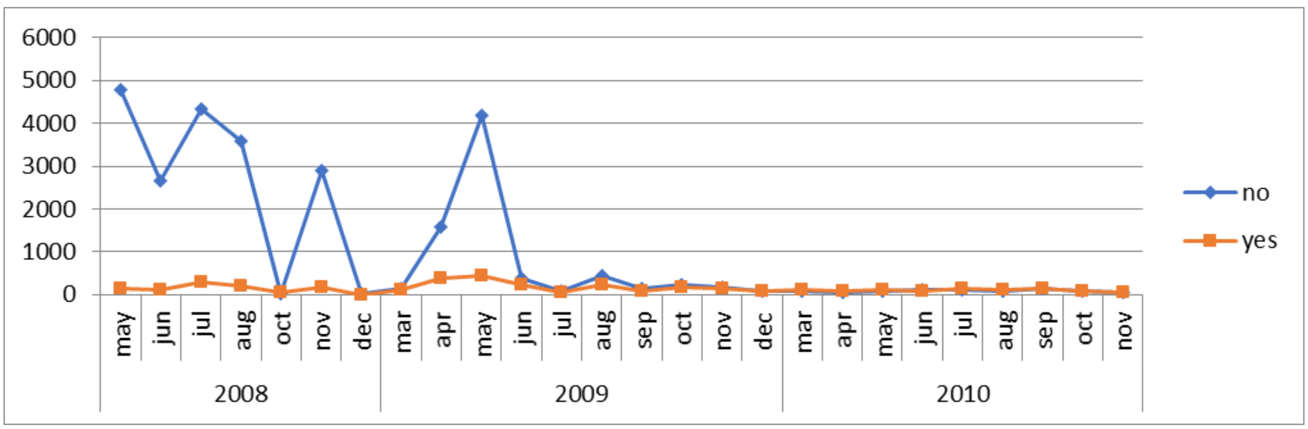
We expect this variable might be significant as the employment variation rate should have an impact on customer's choice in opting for a term deposit.

11. cons.price.idx

Fluctuations in Consumer Price Index(CPI) might have an influence on whether a customer is likely to subscribe to a term deposit. This is because CPI is an indicator of inflation. As inflation increases, prices increase and as a result consumer spending increases. This could lead to lesser savings and fewer subscriptions to term deposits. On the other hand, as CPI reduces, price of goods decreases and consumers would have more propensity to save and subscribe to products such as term deposits.

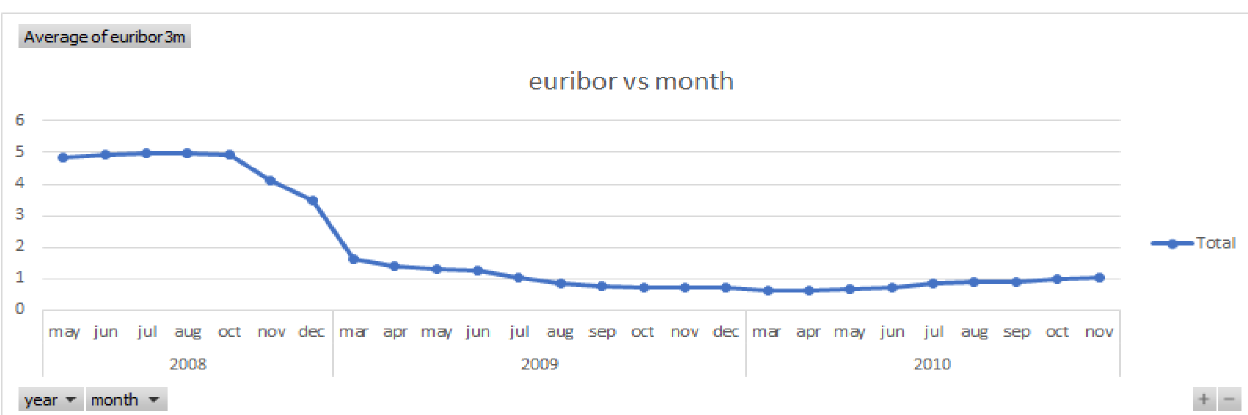
12. cons.conf.idx

Looking at the below graph it is evident that consumer confidence index was low during 2008 to mid-2009, hence the number of people subscribing to the term deposit is very low. After mid-2009 the number of people who subscribed or declined for the term deposit seems to be fairly constant.



13. euribor3m

In general, Euribor rate could directly affect the interest rate obtained on a term deposit. The influence of Euribor is even more pronounced if the term deposit offers a floating interest rate. However, in this case, Euribor might not be a significant variable in the prediction model. This is because after the drop in 2008, Euribor has remained constant. It was also noticed that Euribor exhibits strong correlation with other economic indicators. Plotted below is the average monthly Euribor rate.



14. nr.employed

Number of employees is a quarterly indicator which denotes the number of employees being employed by an organization. Subscription of term deposit logically should not be dependent on the number of employees as this is an operational level attribute.

Correlation Matrix

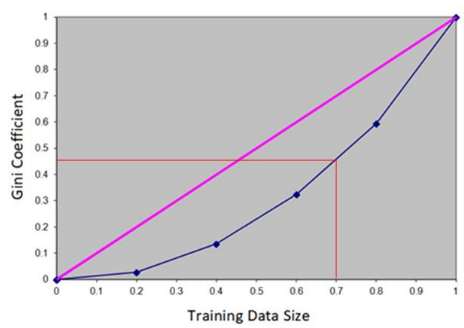
Correlations

	age	campaign	previous	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
age	1.0000	0.0046	0.0244	-0.0004	0.0008	0.1293	0.0107	-0.0177
campaign	0.0046	1.0000	-0.0791	0.1508	0.1278	-0.0137	0.1351	0.1441
previous	0.0244	-0.0791	1.0000	-0.4205	-0.2031	-0.0509	-0.4545	-0.5013
emp.var.rate	-0.0004	0.1508	-0.4205	1.0000	0.7753	0.1960	$\boxed{0.9722}$	$\boxed{0.9070}$
cons.price.idx	0.0008	0.1278	-0.2031	0.7753	1.0000	0.0590	0.6882	0.5220
cons.conf.idx	0.1293	-0.0137	-0.0509	0.1960	0.0590	1.0000	0.2777	0.1005
euribor3m	0.0107	0.1351	-0.4545	0.9722	0.6882	0.2777	1.0000	$\boxed{0.9452}$
nr.employed	-0.0177	0.1441	-0.5013	0.9070	0.5220	0.1005	0.9452	1.0000

As evident from the above table, emp.var.rate is correlated with euribor3m as well as with nr.employed. The information about number of employed personnel is already captured in emp.var.rate and hence nr.employed is being dropped. Also, as discussed in point 13 of preliminary analysis, euribor3m is seen to remain constant post 2008-09 recession. Hence euribor3m is also being dropped.

Test/ Train Data Split

For our train and test data split, we initially followed the pareto principle of 80/20 to split our data. However, doing so yielded an unbalanced data split with a Gini coefficient of 60%, thus we turned to a split ratio of 70/30 as it yielded a Gini coefficient of 45% which is more moderately balanced and below the 50% mark. By taking 30% of the data for our test set, we have 12,356 observations for our final model. This is a comfortable number of observations for a model with 9 variables with about a 1373:1 ratio for observations to variables.



Model Building

Iteration 0: Removing month, duration as per our pre-analysis and nr.employed and euribor3m based on correlation study

Effect Summary

Source	LogWorth	PValue
emp.var.rate	284.683	0.00000
cons.price.idx	103.815	0.00000
contact	43.672	0.00000
cons.conf.idx	25.218	0.00000
poutcome	6.202	0.00000
pdays	5.469	0.00000
job	4.282	0.00005
day_of_week	3.654	0.00022
campaign	2.927	0.00118
default	2.599	0.00252
education	1.766	0.01713
loan	0.668	0.21467
previous	0.434	0.36784
marital	0.333	0.46485
housing	0.200	0.63070
age	0.189	0.64646

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2045.251	43	4090.502	<.0001*
Full	8019.970			
Reduced	10065.221			

RSquare (U)	0.2032
AICc	16130.1
BIC	16501.7
Observations (or Sum Wgts)	28575

Iteration 1: Removed age

Effect Summary

Source	LogWorth	PValue
emp.var.rate	285.880	0.00000
cons.price.idx	104.137	0.00000
contact	43.877	0.00000
cons.conf.idx	25.702	0.00000
poutcome	6.195	0.00000
pdays	5.452	0.00000
job	4.879	0.00001
day_of_week	3.640	0.00023
campaign	2.926	0.00119
default	2.555	0.00279
education	1.745	0.01801
loan	0.666	0.21555
previous	0.434	0.36833
marital	0.301	0.50054
housing	0.202	0.62769

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2045.146	42	4090.291	<.0001*
Full	8020.075			
Reduced	10065.221			

RSquare (U)	0.2032
AICc	16128.3
BIC	16491.6
Observations (or Sum Wgts)	28575

Iteration 2: Removed Housing

Effect Summary

Source	LogWorth	PValue
emp.var.rate	285.870	0.00000
cons.price.idx	104.147	0.00000
contact	43.865	0.00000
cons.conf.idx	25.767	0.00000
poutcome	6.188	0.00000
pdays	5.466	0.00000
job	4.874	0.00001
day_of_week	3.638	0.00023
campaign	2.925	0.00119
default	2.558	0.00276
education	1.742	0.01810
loan	0.606	0.24751
previous	0.436	0.36639
marital	0.301	0.49963

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2045.094	41	4090.188	<.0001*
Full	8020.127			
Reduced	10065.221			

RSquare (U)	0.2032
AICc	16124.4
BIC	16471.2
Observations (or Sum Wgts)	28575

Iteration 3: Removed Marital

Effect Summary

Source	LogWorth	PValue
emp.var.rate	288.107	0.00000
cons.price.idx	104.533	0.00000
contact	43.981	0.00000
cons.conf.idx	25.603	0.00000
poutcome	6.221	0.00000
pdays	5.445	0.00000
job	5.292	0.00001
day_of_week	3.642	0.00023
campaign	2.913	0.00122
default	2.678	0.00210
education	1.902	0.01254
loan	0.608	0.24666
previous	0.440	0.36343

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2043.910	38	4087.82	<.0001*
Full	8021.311			
Reduced	10065.221			

RSquare (U)	0.2031
AICc	16120.7
BIC	16442.8
Observations (or Sum Wgts)	28575

Iteration 4: Removed previous

Effect Summary

Source	LogWorth	PValue
emp.var.rate	293.915	0.00000
cons.price.idx	109.366	0.00000
contact	43.886	0.00000
cons.conf.idx	25.424	0.00000
poutcome	19.982	0.00000
pdays	5.461	0.00000
job	5.272	0.00001
day_of_week	3.616	0.00024
campaign	2.914	0.00122
default	2.671	0.00213
education	1.891	0.01284
loan	0.608	0.24648

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2043.497	37	4086.994	<.0001*
Full	8021.724			
Reduced	10065.221			

RSquare (U)	0.2030
AICc	16119.6
BIC	16433.3
Observations (or Sum Wgts)	28575

Iteration 5: Removed loan

Parameter Estimates

Term		Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept		119.060842	777.65427	0.02	0.8783
job[admin.]		-0.0384623	0.0528631	0.53	0.4669
job[blue-collar]		0.11534381	0.0656815	3.08	0.0791
job[entrepreneur]		0.11340815	0.1152995	0.97	0.3253
job[housemaid]		-0.0115296	0.1344523	0.01	0.9317
job[management]		0.10078705	0.0814169	1.53	0.2157
job[retired]		-0.3379498	0.0838775	16.23	<.0001*
job[self-employed]		0.09785768	0.1128215	0.75	0.3857
job[services]		0.17069847	0.080733	4.47	0.0345*
job[student]		-0.4250053	0.0998313	18.12	<.0001*
job[technician]		-0.0126731	0.0653138	0.04	0.8461
job[unemployed]		0.0819091	0.1236912	0.44	0.5078
education[basic.4y]		0.1554506	0.158619	0.96	0.3271
education[basic.6y]		-0.0953389	0.166444	0.33	0.5668
education[basic.9y]		0.15281813	0.1548468	0.97	0.3237
education[high.school]		0.10115349	0.1507412	0.45	0.5022
education[illiterate]		-0.212937	0.9952742	0.05	0.8306
education[professional.course]		0.02657352	0.1561097	0.03	0.8648
education[university.degree]		-0.0839401	0.1501032	0.31	0.5760
default[no]	Unstable	-4.4883311	777.63758	0.00	0.9954
default[unknown]	Unstable	-4.2586683	777.63758	0.00	0.9956
contact[cellular]		-0.4272219	0.031337	185.86	<.0001*
day_of_week[mon]		0.16558992	0.042549	15.15	<.0001*
day_of_week[tue]		-0.0077362	0.0426351	0.03	0.8560
day_of_week[wed]		-0.0887482	0.0420306	4.46	0.0347*
day_of_week[thu]		-0.1015861	0.0402899	6.36	0.0117*
campaign		0.03343804	0.0107688	9.64	0.0019*
pdays[four_weeks_ago]		-0.8437911	0.8881848	0.90	0.3421
pdays[never]		1.00021391	0.28286	12.50	0.0004*
pdays[one_week_ago]		-0.2231984	0.2541417	0.77	0.3798
pdays[three_weeks_ago]		0.34949127	0.3389458	1.06	0.3025
poutcome[failure]		0.42477061	0.0842778	25.40	<.0001*
poutcome[nonexistent]		-0.1682657	0.0921879	3.33	0.0680
emp.var.rate		0.75413405	0.0209277	1298.5	<.0001*
cons.price.idx		-1.2270602	0.0545736	505.55	<.0001*
cons.conf.idx		-0.0431122	0.0040919	111.01	<.0001*

Effect Summary

Source	LogWorth	PValue
emp.var.rate	293.759	0.00000
cons.price.idx	109.169	0.00000
contact	43.835	0.00000
cons.conf.idx	25.451	0.00000
poutcome	20.056	0.00000
pdays	5.388	0.00000
job	5.265	0.00001
day_of_week	3.614	0.00024
campaign	2.942	0.00114
default	2.663	0.00217
education	1.883	0.01311

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2042.096	35	4084.193	<.0001*
Full	8023.125			
Reduced	10065.221			

RSquare (U)	0.2029
AICc	16118.3
BIC	16415.6
Observations (or Sum Wgts)	28575

Based on the above chi-square value for education, we see that it has only 1 significant level and hence acts as a constant category. Therefore, excluding it from the model.

Iteration 6: Removed education

Parameter Estimates

Term		Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept		119.542747	777.6551	0.02	0.8778
job[admin.]		-0.076431	0.0489766	2.44	0.1186
job[blue-collar]		0.17121769	0.0582482	8.64	0.0033*
job[entrepreneur]		0.08989598	0.1148909	0.61	0.4340
job[housemaid]		0.03721343	0.131856	0.08	0.7778
job[management]		0.02744589	0.0781278	0.12	0.7254
job[retired]		-0.2990457	0.0812996	13.53	0.0002*
job[self-employed]		0.04941307	0.1108936	0.20	0.6559
job[services]		0.21352548	0.0768681	7.72	0.0055*
job[student]		-0.4088116	0.0980825	17.37	<.0001*
job[technician]		-0.0313832	0.0589219	0.28	0.5943
job[unemployed]		0.10286881	0.1234344	0.69	0.4046
default[no]	Unstable	-4.4968812	777.63845	0.00	0.9954
default[unknown]	Unstable	-4.2610194	777.63845	0.00	0.9956
contact[cellular]		-0.4325102	0.0313075	190.85	<.0001*
day_of_week[mon]		0.16629555	0.0425215	15.29	<.0001*
day_of_week[tue]		-0.0055437	0.0426007	0.02	0.8965
day_of_week[wed]		-0.0873715	0.0420104	4.33	0.0375*
day_of_week[thu]		-0.1055708	0.0402402	6.88	0.0087*
campaign		0.03304922	0.0107377	9.47	0.0021*
pdays[four_weeks_ago]		-0.7902245	0.8860581	0.80	0.3725
pdays[never]		0.97967977	0.2825261	12.02	0.0005*
pdays[one_week_ago]		-0.2386055	0.2533216	0.89	0.3462
pdays[three_weeks_ago]		0.33102014	0.3384322	0.96	0.3280
poutcome[failure]		0.42978588	0.0841149	26.11	<.0001*
poutcome[nonexistent]		-0.166593	0.0920309	3.28	0.0703
emp.var.rate		0.75579337	0.0209317	1303.8	<.0001*
cons.price.idx		-1.2320115	0.0544854	511.29	<.0001*
cons.conf.idx		-0.0440865	0.0040768	116.94	<.0001*

For log odds of no/yes

Effect Summary

Source	LogWorth	PValue
emp.var.rate	294.825	0.00000
cons.price.idx	110.250	0.00000
contact	44.993	0.00000
cons.conf.idx	26.769	0.00000
poutcome	20.368	0.00000
job	6.314	0.00000
pdays	5.242	0.00001
day_of_week	3.698	0.00020
campaign	2.898	0.00127
default	2.857	0.00139

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2033.217	28	4066.434	<.0001*
Full	8032.004			
Reduced	10065.221			

RSquare (U)	0.2020
AICc	16122.1
BIC	16361.6
Observations (or Sum Wgts)	28575

Based on the above chi-square value for default, we see that it has only 1 significant level and hence acts as a constant category. Therefore, excluding it from the model. **Now we have our final model**.

Iteration 7: Removed default

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	116.294563	5.0893176	522.16	<.0001*
job[admin.]	-0.0934659	0.0487505	3.68	0.0552
job[blue-collar]	0.19229513	0.0578967	11.03	0.0009*
job[entrepreneur]	0.08599246	0.1147917	0.56	0.4538
job[housemaid]	0.04477319	0.1316386	0.12	0.7338
job[management]	0.014795	0.0780303	0.04	0.8496
job[retired]	-0.2941828	0.0811872	13.13	0.0003*
job[self-employed]	0.0441002	0.1108222	0.16	0.6907
job[services]	0.21197596	0.076799	7.62	0.0058*
job[student]	-0.4173607	0.097947	18.16	<.0001*
job[technician]	-0.0484566	0.0587166	0.68	0.4092
job[unemployed]	0.10129703	0.1232926	0.68	0.4113
contact[cellular]	-0.4382364	0.0312978	196.06	<.0001*
day_of_week[mon]	0.16793163	0.0425112	15.60	<.0001*
day_of_week[tue]	-0.0057818	0.0425983	0.02	0.8920
day_of_week[wed]	-0.0904533	0.0419984	4.64	0.0313*
day_of_week[thu]	-0.1053551	0.0402335	6.86	0.0088*
campaign	0.03303402	0.0107373	9.47	0.0021*
pdays[four_weeks_ago]	-0.7895052	0.8861758	0.79	0.3730
pdays[never]	0.97778377	0.2824594	11.98	0.0005*
pdays[one_week_ago]	-0.2378211	0.2533301	0.88	0.3478
pdays[three_weeks_ago]	0.32958427	0.3383888	0.95	0.3301
poutcome[failure]	0.43211453	0.0840544	26.43	<.0001*
poutcome[nonexistent]	-0.1656196	0.0919755	3.24	0.0718
emp.var.rate	0.76957637	0.0206379	1390.5	<.0001*
cons.price.idx	-1.2449616	0.0544397	522.97	<.0001*
cons.conf.idx	-0.0444171	0.0040767	118.71	<.0001*

Final value of estimates for our logit model

Effect Summary

Source	LogWorth	PValue
emp.var.rate	312.398	0.00000
cons.price.idx	112.513	0.00000
contact	46.201	0.00000
cons.conf.idx	27.159	0.00000
poutcome	20.497	0.00000
job	7.321	0.00000
pdays	5.214	0.00001
day_of_week	3.803	0.00016
campaign	2.896	0.00127

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2026.638	26	4053.277	<.0001*
Full	8038.583			
Reduced	10065.221			

RSquare (U)	0.2014
AICc	16131.2
BIC	16354.2
Observations (or Sum Wgts)	28575

Model Validation

Confusion Matrix

Training

Actual y	Predicted Count no	Predicted Count yes
no	25036	317
yes	2504	718

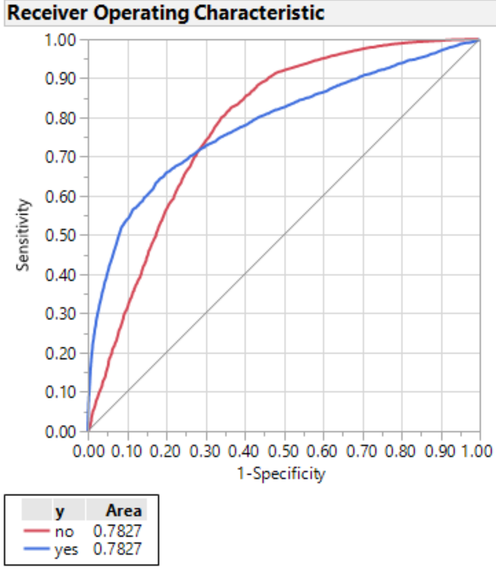
Validation

Actual y	Predicted Count no	Predicted Count yes
no	11033	161
yes	1126	292

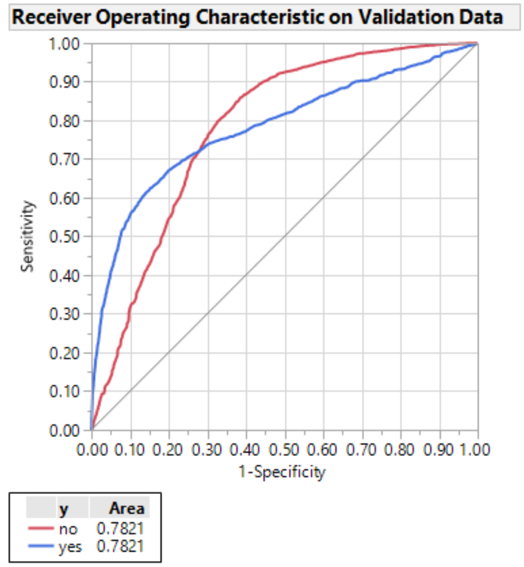
Training Dataset	
Accuracy	90.13%
True Positive Rate/Sensitivity or Recall	22.28%
False Positive Rate/ Type I error	1.25%
True Negative rate / Specificity	98.75%
False Negative rate/ Type II error	77.72%

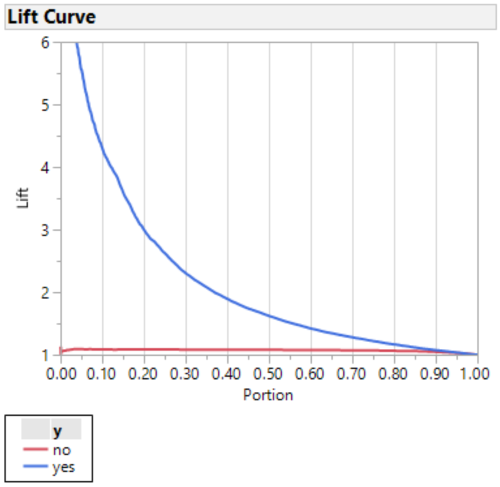
Validation Dataset	
Accuracy	89.80%
True Positive Rate/Sensitivity or Recall	20.59%
False Positive Rate/ Type I error	1.44%
True Negative rate / Specificity	98.56%
False Negative rate/ Type II error	79.41%

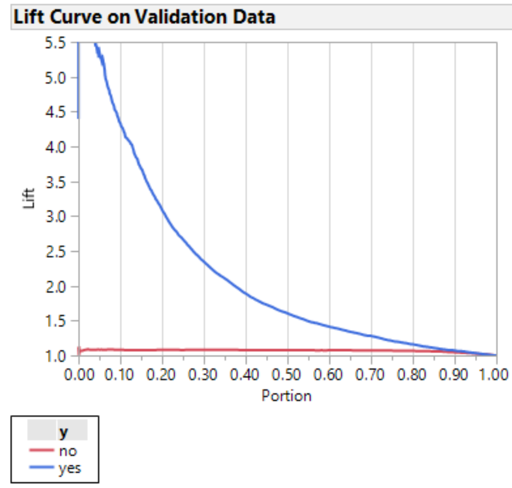
Receiver Operating Characteristic

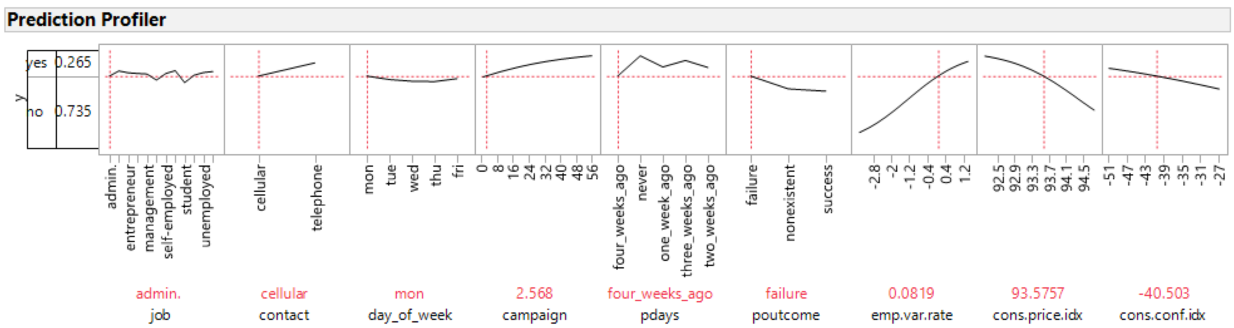


Receiver Operating Characteristic on Validation Data









Conclusion

A final model with a test data **accuracy of 89.80%** was built. It was noticed that the model had high specificity and low sensitivity. This is because the data is skewed with the proportion of No's being much larger than Yes's which was also reflected in the train dataset for building the model and test dataset for publishing model's performance (**The model had a larger number of NO instances to train from**). This was also illustrated in the preliminary analysis using a frequency plot of the target's values in the sample.

This model was built with an objective of facilitating the decision-making process of a marketing manager for running future campaigns.

Selected economic indicators such as employment variation rate, consumer price index and consumer confidence index were included in the model to understand the consumer sentiments and their interest towards subscribing to a term deposit. **In instances where the EVR, CPI and CCI fluctuate, the model would highlight the state of an average consumer in the economy**, which would help the manager set realistic expectations from the campaign.

Among the demographic attributes, job was the only parameter considered, as both preliminary analysis and final model suggested high significance of a few levels within this parameter. An inference can be made that students and retired people are more likely to subscribe to this term deposit and people in the blue-collar segment are unlikely to subscribe.

Parameters such as contact, poutcome and campaign were included to draw an accurate conclusion of the campaign attributes that would influence the subscription. **The model would be able to extract results from the earlier and current campaigns and this would help the manager tailor his/her strategy of approaching a client.**

Using this model, the manager would be able to set **'realistic'** targets and expectations for the staff and management.