LEAD SCORING CASE STUDY

DONE BY:

ALEX K BABU

BUSINESS OBJECTIVES

- To build a model where we assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- Increasing the lead conversion rate from 30% to 80% by focusing on target hot leads rather than everyone.

Lead Conversion Process - Demonstrated as a funnel

SOLUTION METHODOLOGY

- Exploratory data analysis(handling the null values, Select(unfilled) columns and outliers, Dummy variable creation, Test-Train split and Scaling
- Model Building using Logistic Regression(P-value and VIF)
- Model Evaluation and selecting optimal cut-off point(ROC Curve)
- Making predictions on Test Data set
- Precision and Recall trade-off, finally generating a lead score between 0 and 100.

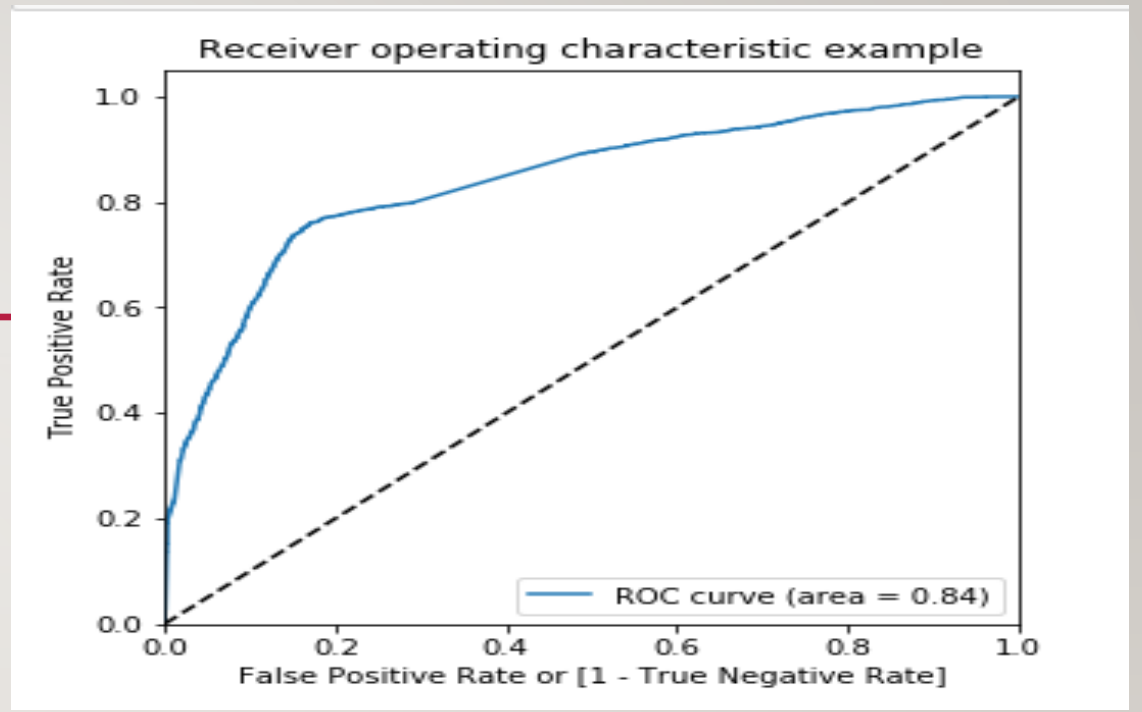
LOGISTIC REGRESSION MODEL

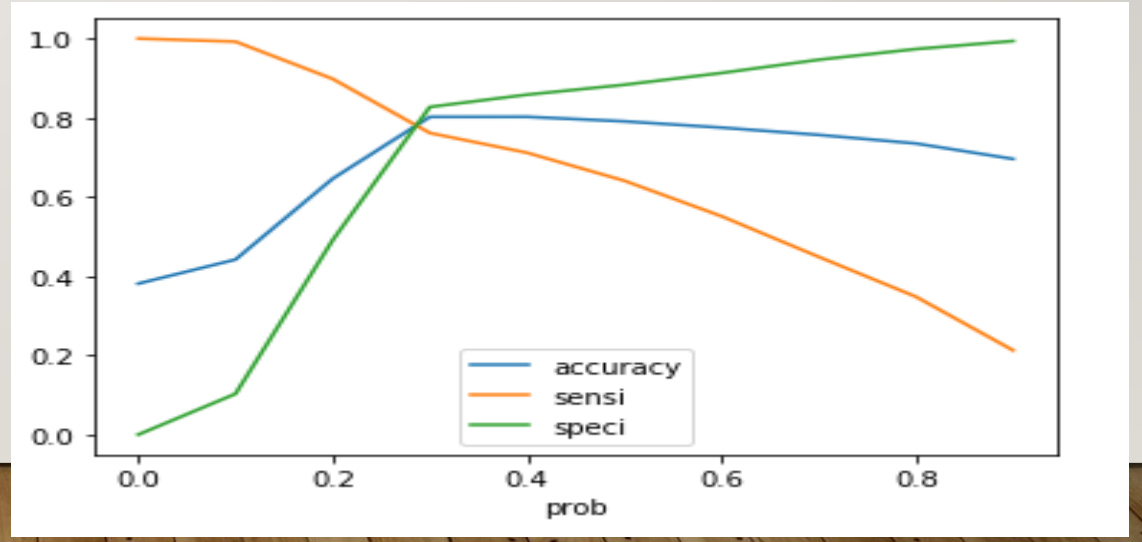
- At the end of building logistic regression model result obtained include following variables. Variables where removed using RFE and manual approach
- VIF < 3 and P-value closer to zero was the criteria used for manual approach

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6750	0.180	3.759	0.000	0.323	1.027
TotalVisits	5.6015	1.834	3.055	0.002	2.008	9.195
Total Time Spent on Website	4.5825	0.155	29.593	0.000	4.279	4.886
Lead Origin_Lead Add Form	3.8828	0.188	20.613	0.000	3.514	4.252
Lead Source_Olark Chat	0.8249	0.096	8.599	0.000	0.637	1.013
Lead Source_Welingak Website	2.1360	0.742	2.880	0.004	0.683	3.590
Do Not Email_Yes	-1.4051	0.158	-8.913	0.000	-1.714	-1.096
Specialization_Hospitality Management	-0.8441	0.307	-2.745	0.006	-1.447	-0.241
Specialization_Rural and Agribusiness	0.5953	0.365	1.632	0.103	-0.120	1.310
What is your current occupation_Other	-3.2917	0.785	-4.192	0.000	-4.831	-1.753
What is your current occupation_Student	-2.6224	0.274	-9.554	0.000	-3.160	-2.084
What is your current occupation_Unemployed	-2.8208	0.175	-16.127	0.000	-3.164	-2.478

MODEL EVALUATION(ROC CURVE)

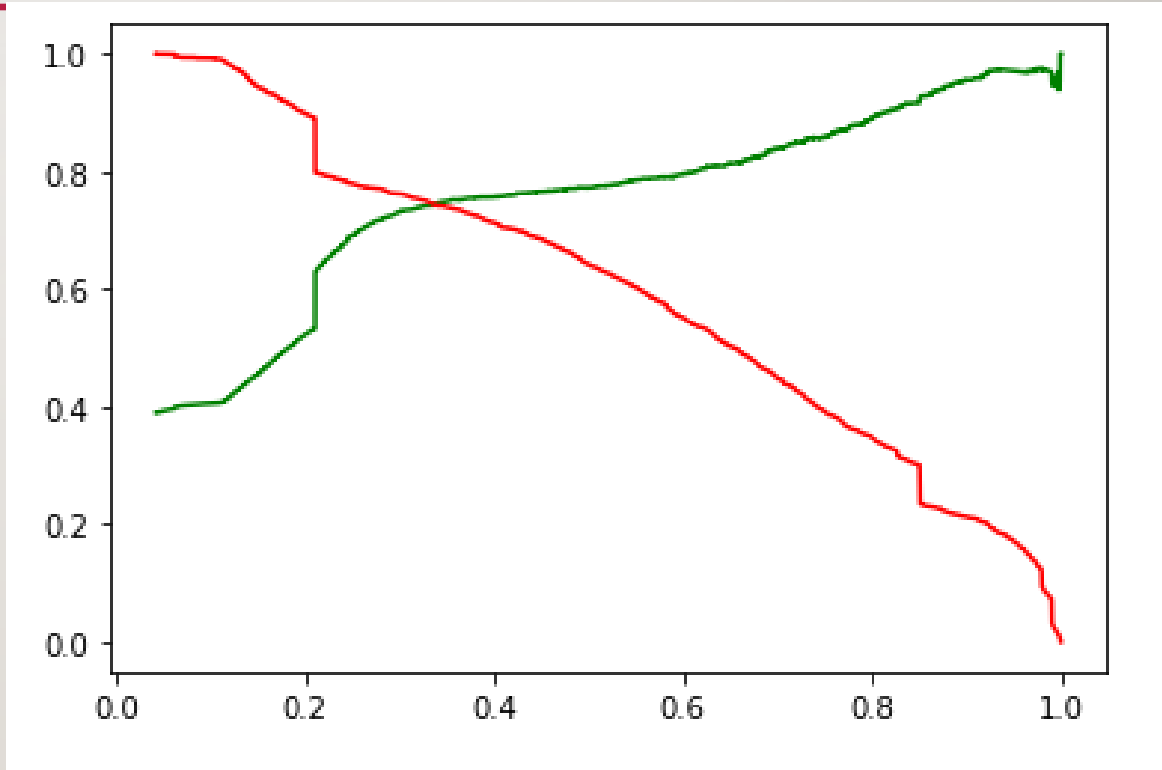
- For our model the area under the curve is 0.84 and it is almost close to the upper left corner, which implies model is good.
- The optimal cut off Cut-off point is 0.29, where all the 3 points coincide as seen in figure.
- Specificity : 76.27%
- Sensitivity: 82.03%
- Accuracy: 79.83%





MODEL EVALUATION: PRECISION AND RECALL

- As per business requirements we have chosen 0.29 as the optimal cut-off value
- Accuracy : 78.9%
- Precision : 77.19%
- Recall : 64.11%



PREDICTIONS ON FINAL DATASET

By choosing cut-off point as 0.29, We obtain

- sensitivity = 75%
- specificity = 81%
- Accuracy = 78.93%
- false positive rate = 0.18783542039355994
- positive predictive value = 0.7239263803680982
- Negative predictive value = 0.835070508890251 4

Actual/Predicted	Not converted	converted
Not converted	1362	315
converted	269	826

By choosing cut-off point as 0.29, We obtain Precision value almost near 78% , therefore we can use the model for achieving our objective of increasing the conversion rate to 80%.

TOP 3 FACTORS

- Top 3 most contributing variables towards the probability of a lead getting converted include
  - Total Visits
  - Total time spent on website
  - Lead Origin
    - Lead Add Form
- Top 3 categorical/dummy variables in the model which should be focused the most to increase the probability of lead conversion include
  - Lead Origin_Lead Add Form
  - Lead Source_Olark Chat
  - Lead Source_Welingak Website

CONCLUSION

- The model is prepared for prediction of the conversion of the leads.
- The probability value are generated by the model.
- The cut-off decided for the model is 0.29/29.
- All leads whose probability above this threshold value(lead score) 29 can be classified as a Hot Lead.

THANK YOU