

Abstract geometric lines in the top left corner, consisting of several thin, dark lines forming overlapping, irregular shapes and polygons.

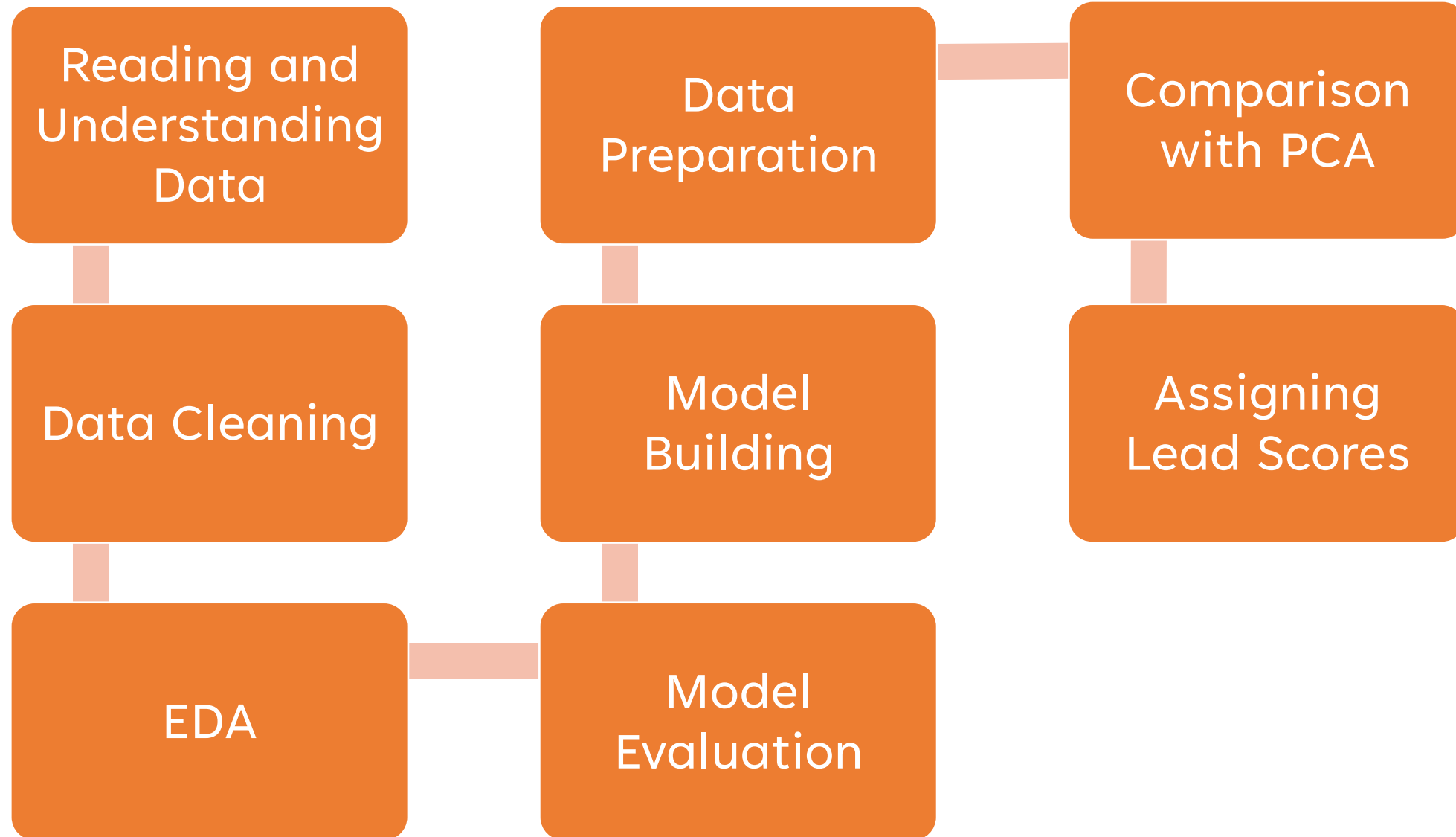
LEAD SCORING CASE STUDY

BUSINESS OBJECTIVE

To help X Education select most promising Hot leads that are most likely to convert into paying customers.

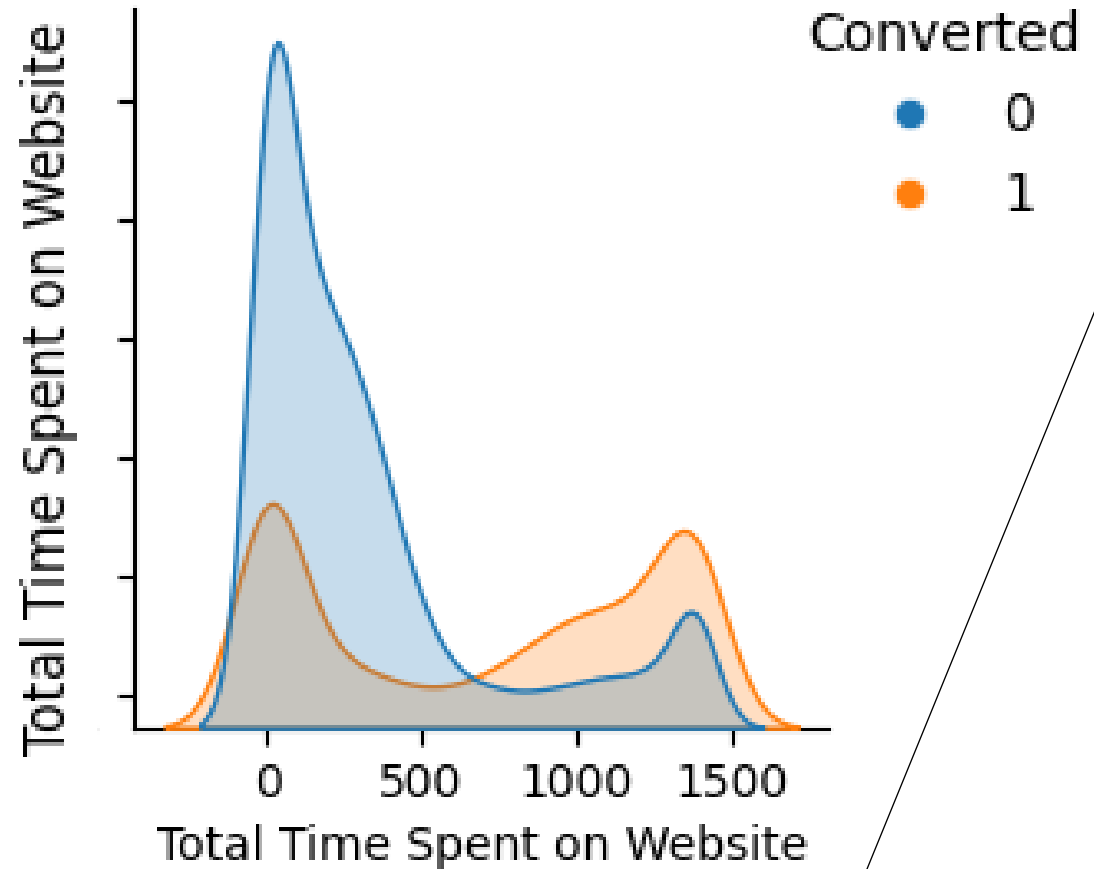
METHODOLOGY

To build a Logistic Regression Model that assigns lead scores to all leads such that the customers with higher lead scores have a higher conversion chance.

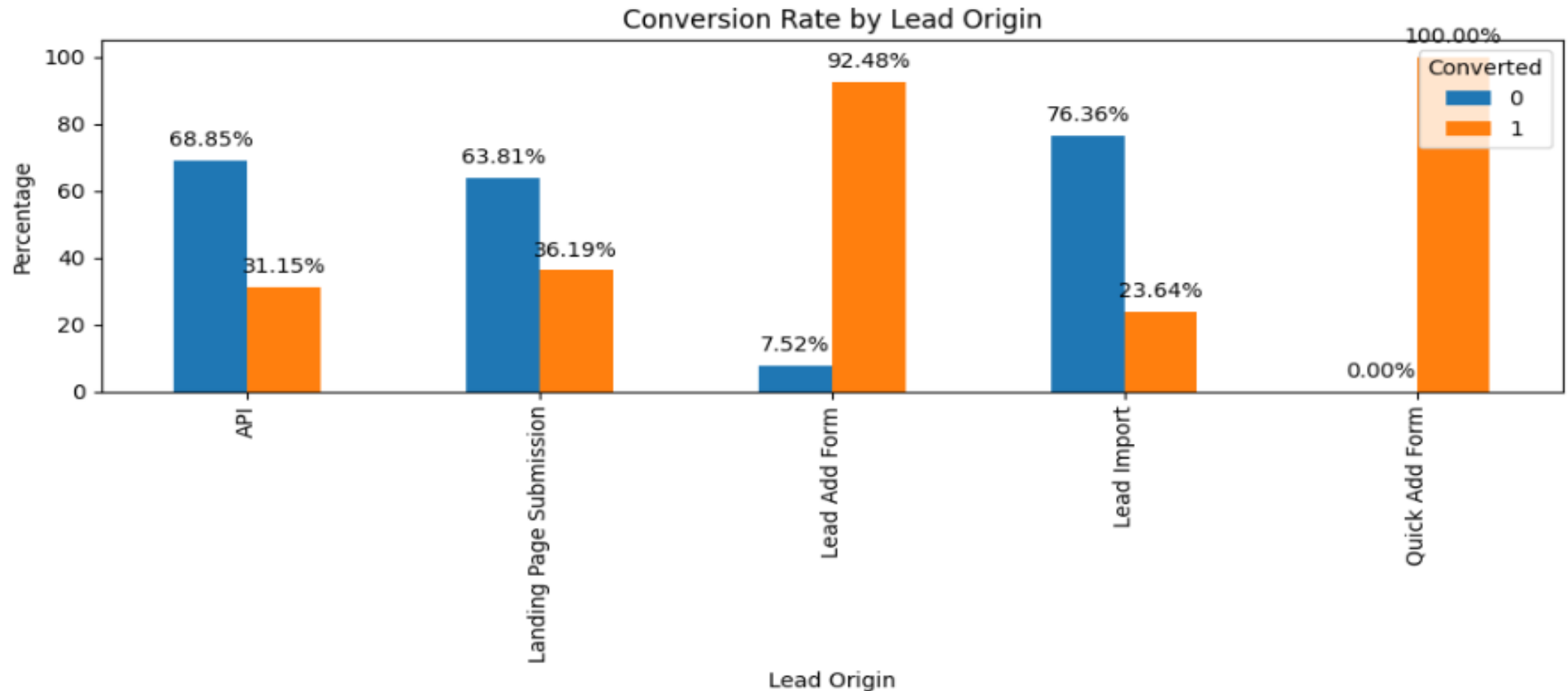


DATA VISUALIZATION

The more time spent on the Website significantly increases the likelihood of conversion.



Lead Origin_Lead Add Form has a strong positive impact on lead conversion.



FINAL MODEL SUMMARY

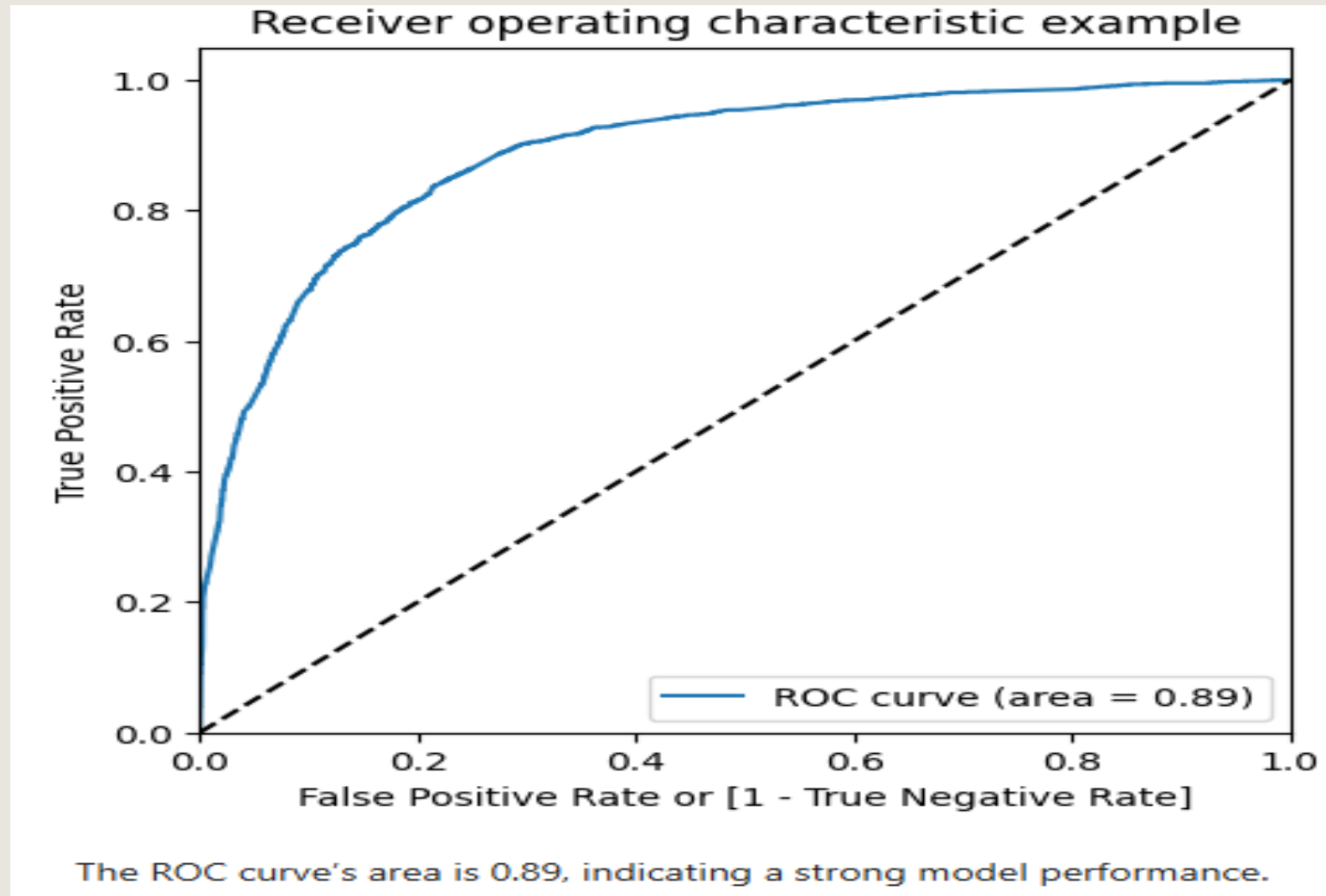
Generalized Linear Model Regression Results

=====						
Dep. Variable:	Converted	No. Observations:	6468			
Model:	GLM	Df Residuals:	6455			
Model Family:	Binomial	Df Model:	12			
Link Function:	Logit	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-2629.4			
Date:	Sat, 16 Nov 2024	Deviance:	5258.7			
Time:	12:56:41	Pearson chi2:	6.91e+03			
No. Iterations:	7	Pseudo R-squ. (CS):	0.4033			
Covariance Type:	nonrobust					
=====						
		coef	std err	z	P> z	[0.025 0.975]

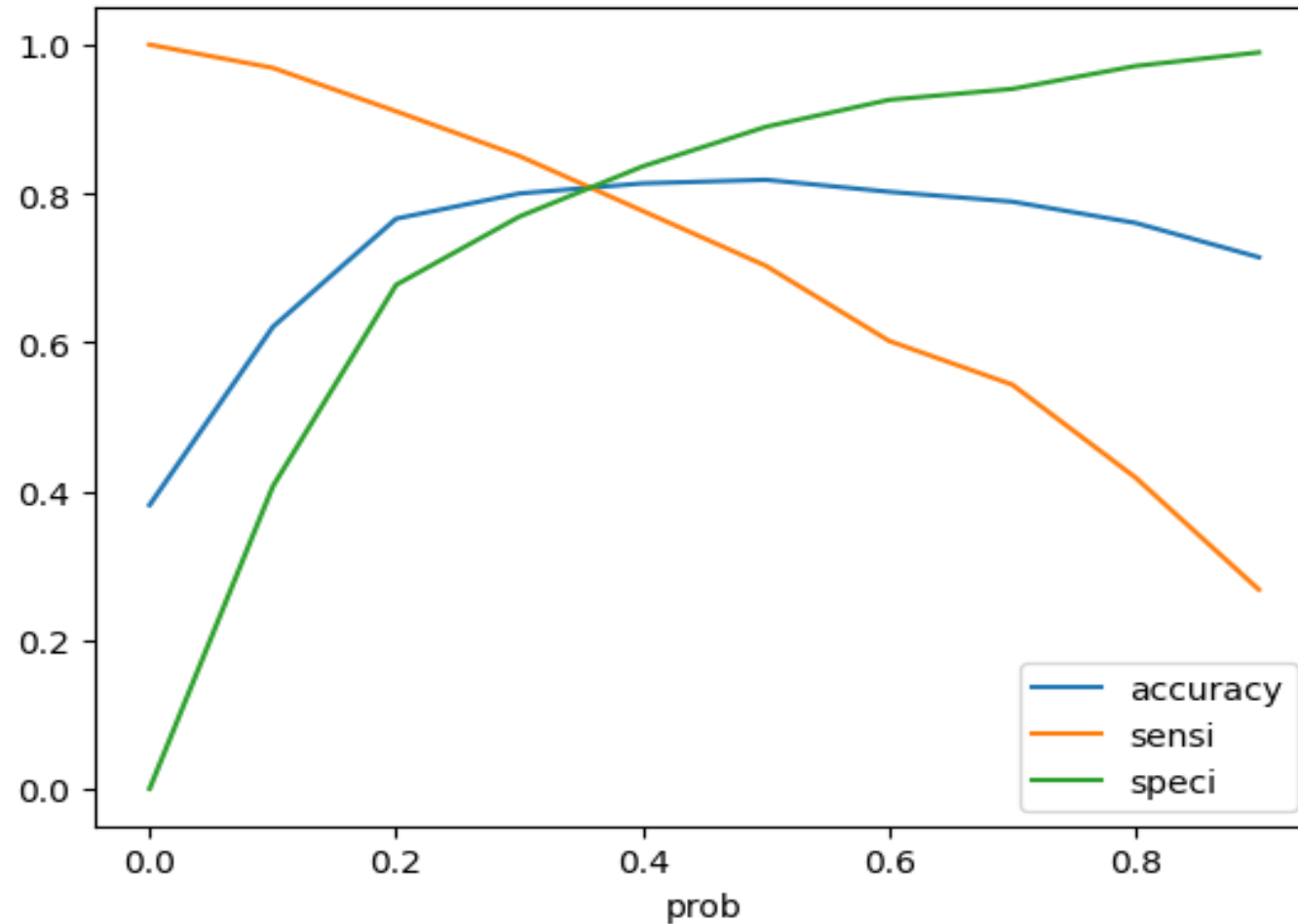
const		-0.0689	0.128	-0.537	0.591	-0.320 0.183
Total Time Spent on Website		1.1232	0.040	28.160	0.000	1.045 1.201
Lead Origin_Landing Page Submission		-1.0633	0.128	-8.290	0.000	-1.315 -0.812
Lead Origin_Lead Add Form		3.1042	0.204	15.200	0.000	2.704 3.504
Last Activity_Olark Chat		1.0825	0.122	8.839	0.000	0.842 1.323
Last Activity_Welingak Website		2.4249	0.745	3.253	0.001	0.964 3.886
Specialization_Hospitality Management		-1.0062	0.331	-3.044	0.002	-1.654 -0.358
Specialization_UnKnown		-1.0079	0.123	-8.186	0.000	-1.249 -0.767
What is your current occupation_UnKnown		-0.9813	0.088	-11.172	0.000	-1.153 -0.809
What is your current occupation_Working Professional		2.4057	0.191	12.572	0.000	2.031 2.781
Last Notable Activity_Modified		-0.7058	0.084	-8.363	0.000	-0.871 -0.540
Last Notable Activity_Olark Chat Conversation		-1.0782	0.346	-3.117	0.002	-1.756 -0.400
Last Notable Activity_SMS Sent		1.2604	0.085	14.756	0.000	1.093 1.428
=====						

p-values for all variables is less than 0.05

ROC CURVE



FINDING OPTIMAL THRESHOLD



From the curve above, 0.35 is the optimum point to take it as a cutoff probability.

INFERENCES

- The Optimal cutoff was selected based on precision and recall trade off score.
- The model also worked fine on test dataset.
- Overall, the model looks good and is able to identify the correct leads which has high chances of conversion using lead score prediction.

RECOMMENDATIONS

- By referring to the data visualizations focus on increasing the conversion rates for the categories generating more leads.
- Give relative importance to the features in the model and their positive or negative impact on the probability of conversion.
- Based on varying business needs modify the probability threshold value for identifying potential leads.



THANK YOU