Lead Score
Case Study

~ Vasundhara L Kadekodi

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’.
- if they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

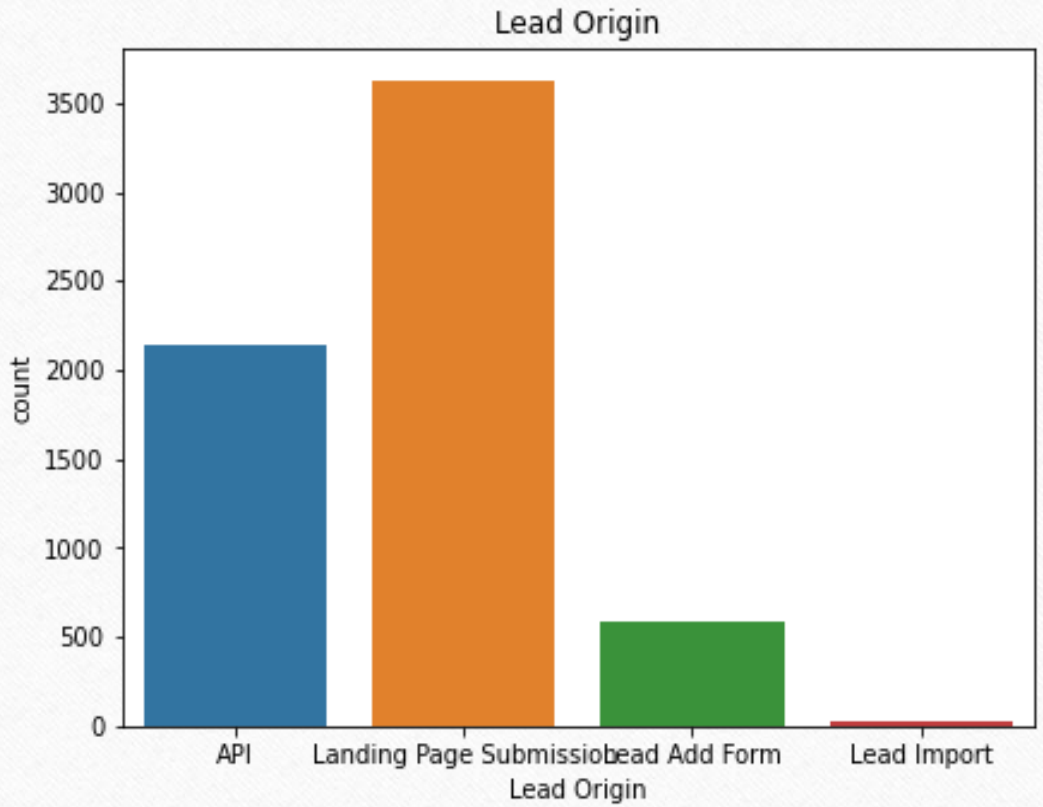
Business Objective

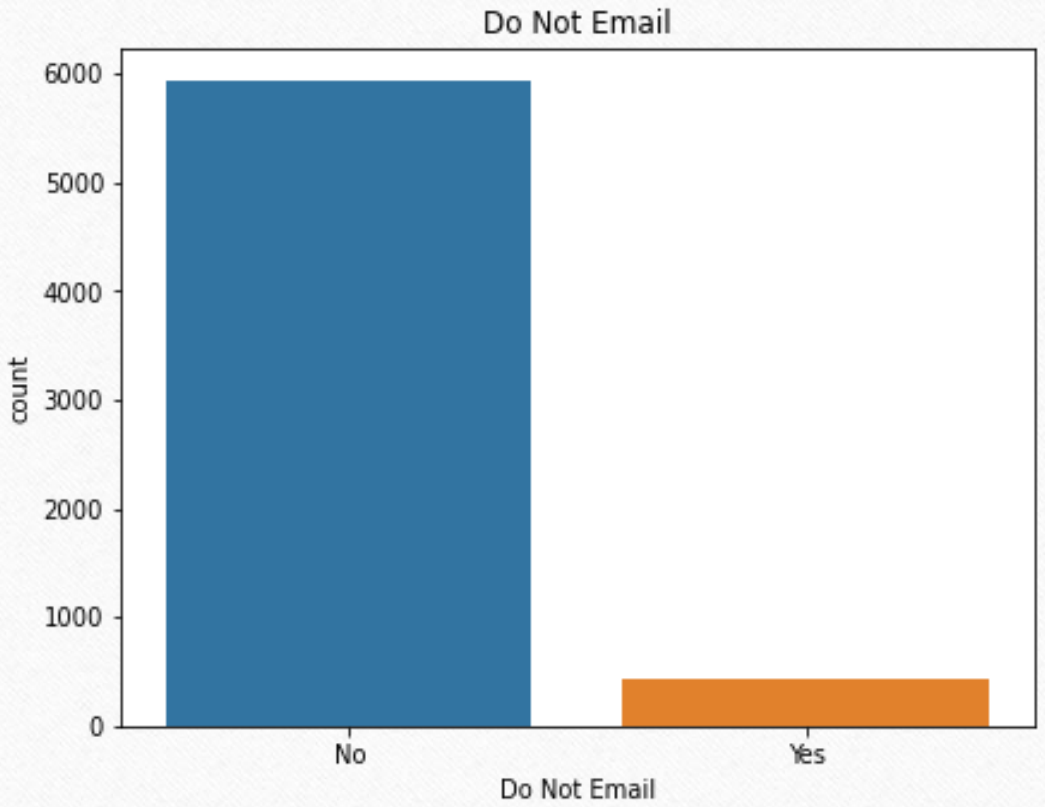
- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Solution Methodology

- Importing Data
- Data Understanding and Cleaning
- EDA
- Dummy Variables Creation
- Splitting the dataset into Train - Test Datasets
- Rescaling the features using Min-Max method
- Model building
- Model Evaluation
- Conclusion and recommendation to improve the business

EDA (Univariate Analysis)

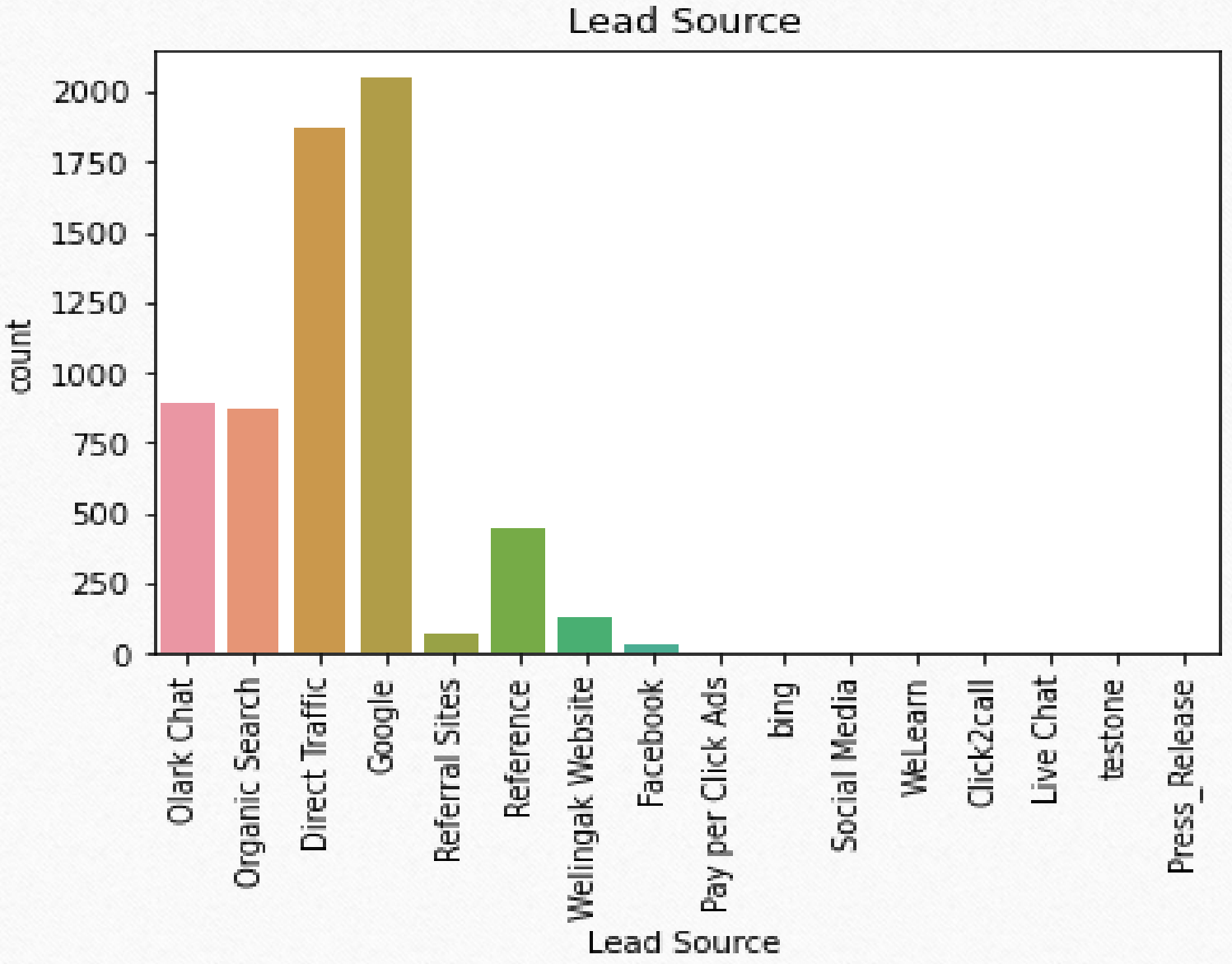




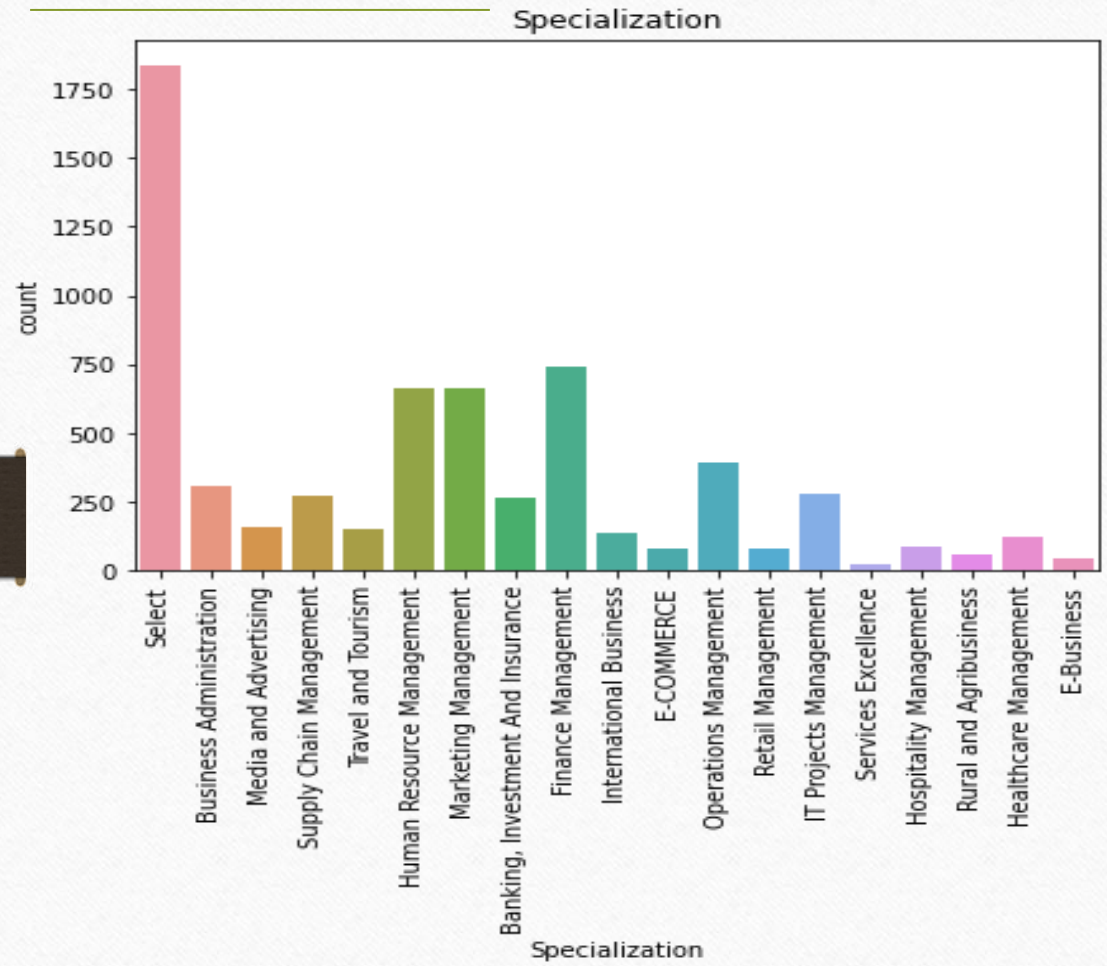
From the graphs we can say that, the count of ‘Landing page submission’ is very high in ‘Lead Origin’ and maximum people are opted to not to email.

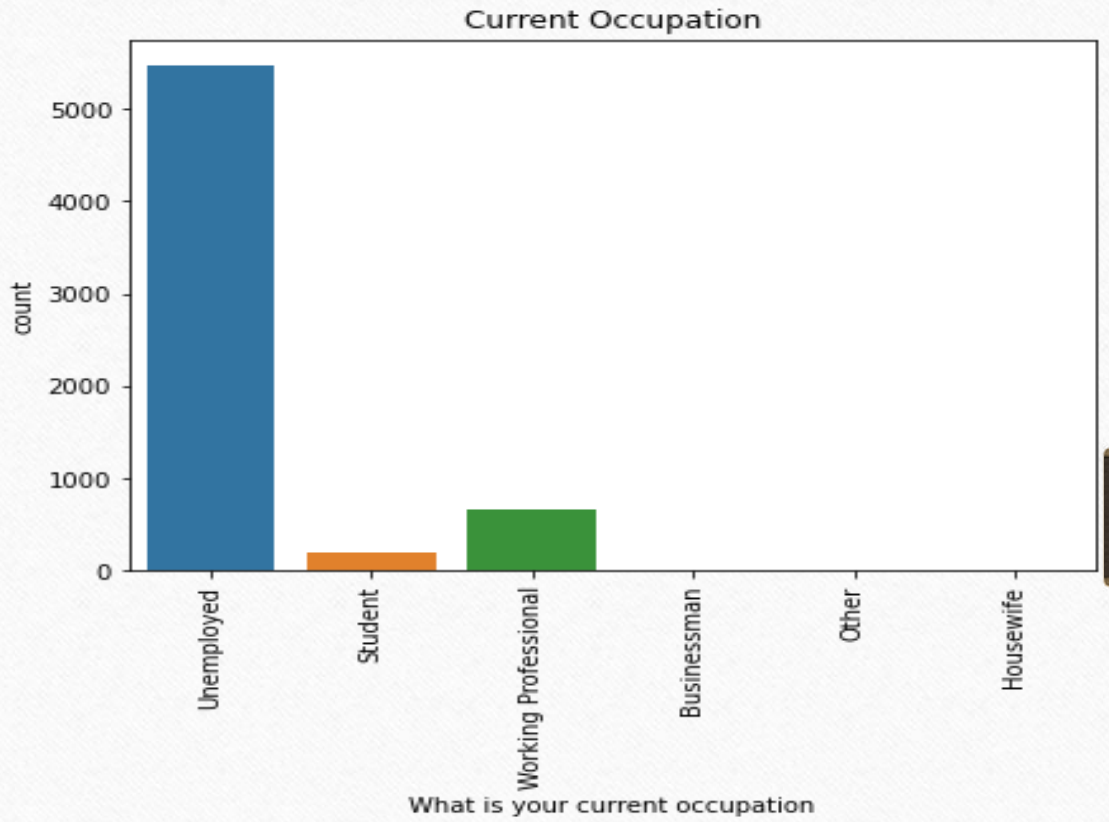
Univariate Analysis

From the graph we can observe that, the maximum source of the lead is from ‘Google’ and ‘Direct Traffic’



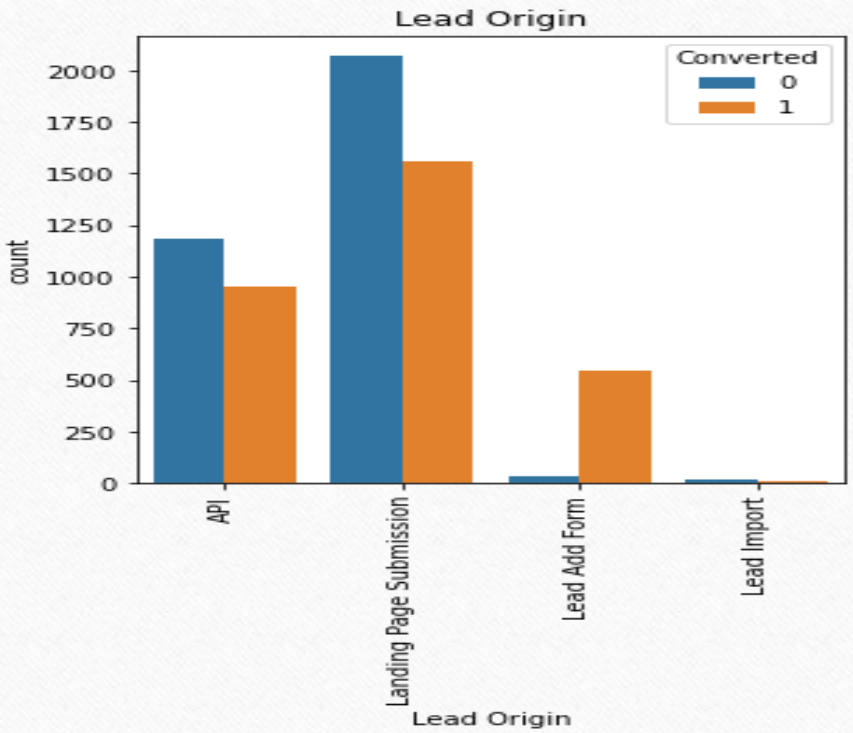
Univariate Analysis

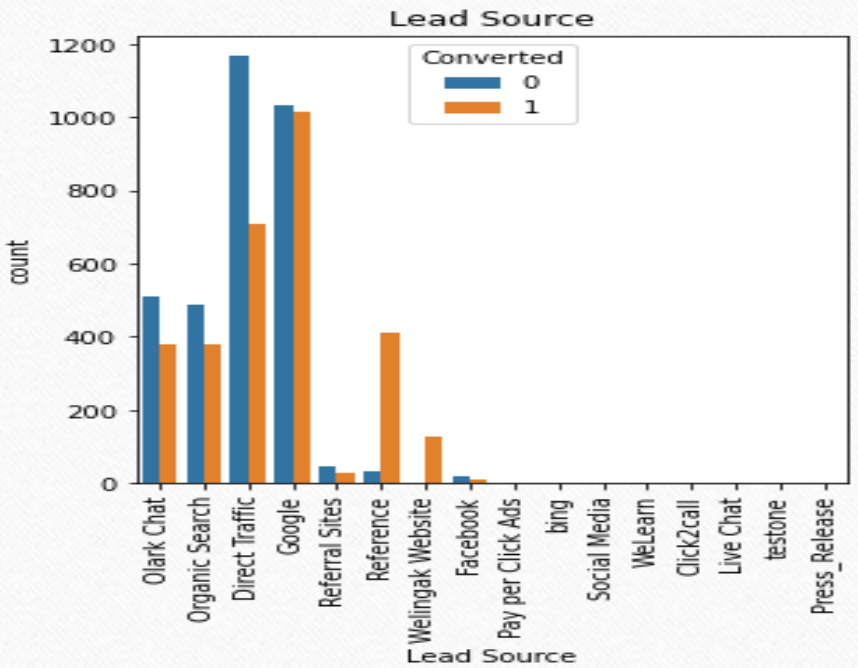




As we can see that maximum count of applicants for the course is from ‘Finance Management’ Sector when it comes to already working professionals but overall 90% of the applicants who are ‘Unemployed’.

EDA (Bivariate Analysis)



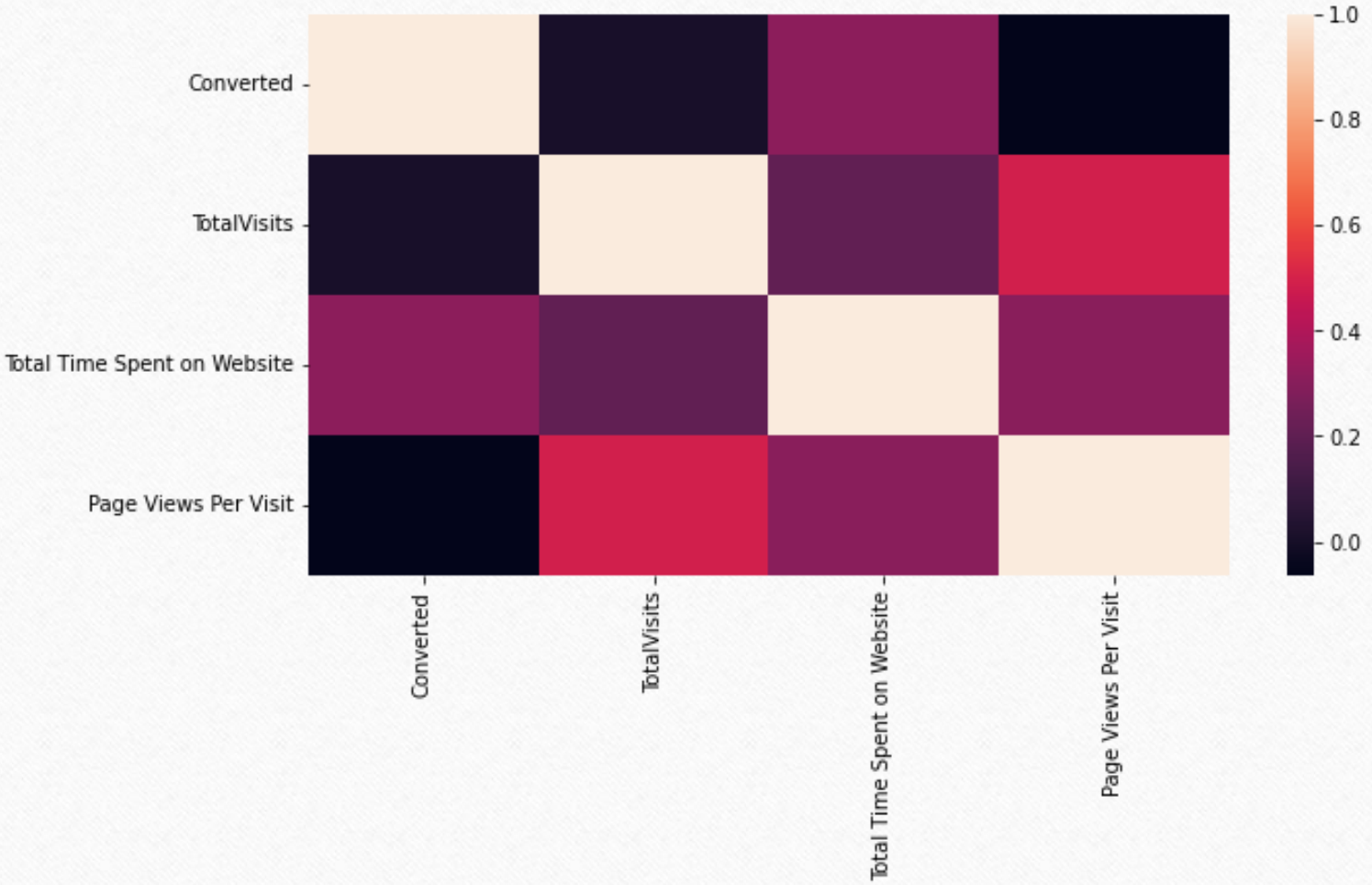


From the above graph we can say that, In Lead Origin API, Landing Page Submission, Lead import having approximately 40%of conversion But in Lead Add form even count is less but the conversion rate is very high about 90%.

In Lead Source, Reference and Weilingak Website have more conversion rate as compare to other variables.

Bivariate Analysis

From the heatmap, we can observe that variable ‘Total Visit’ is having strong correlation with ‘Page views per visit’ and poor correlation with the target variable ‘Converted’.



Model Building

- Spitted the main data into train-test data (70:30).
- Selected 15 features which are contributing more to the analysis using RFE method.
- In final model we are left with 11 features.
- The result of the model is as follows:
  - Optimal Cut-off point = 0.42
  - Overall Accuracy = 78%
  - Recall = 78%
  - Precision = 77%

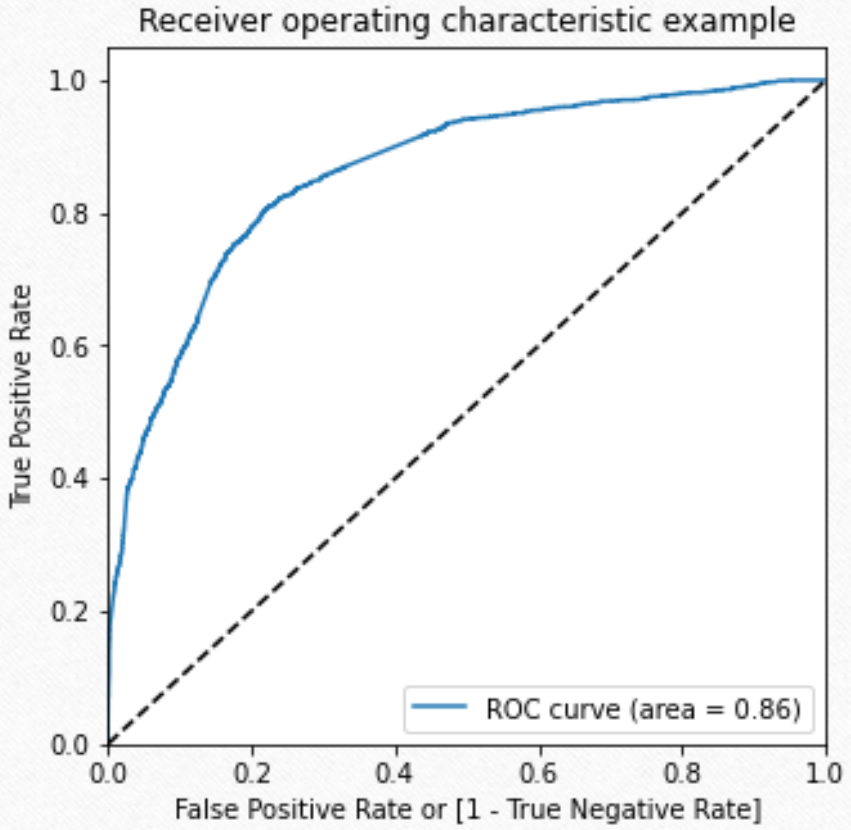
Final Model Results

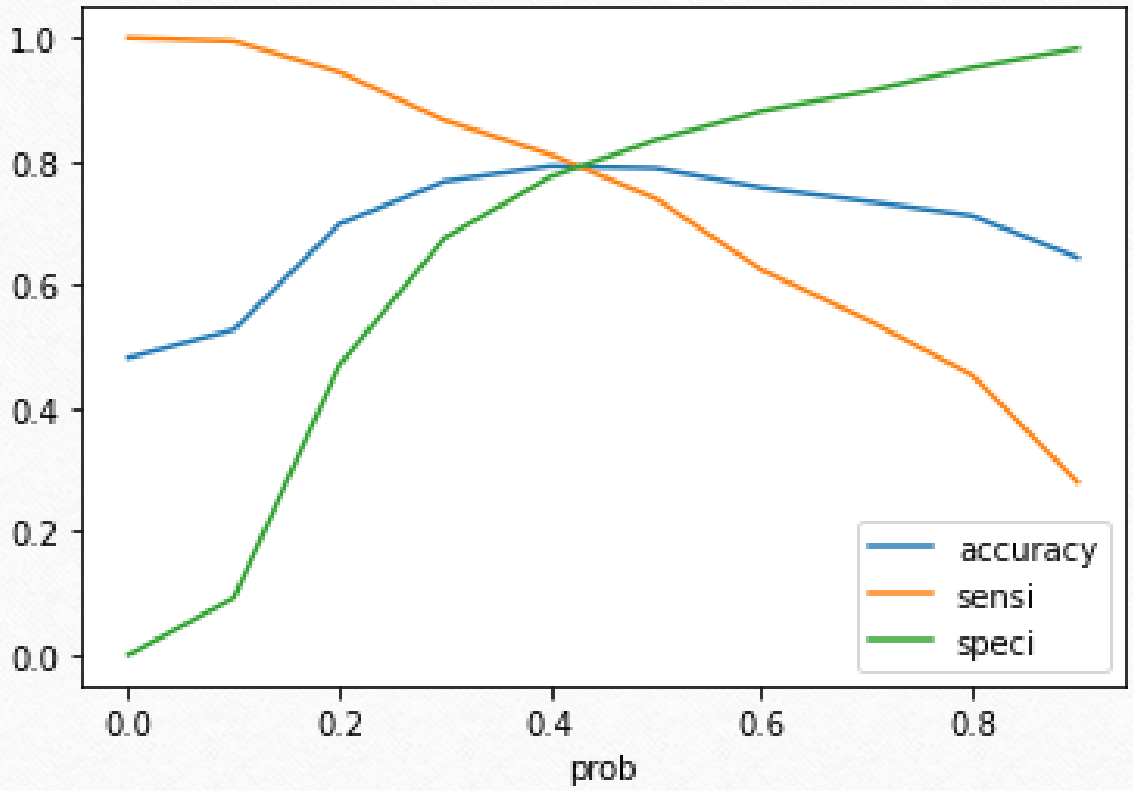
Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	4461
Model:	GLM	Df Residuals:	4449
Model Family:	Binomial	Df Model:	11
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2079.1
Date:	Tue, 13 Sep 2022	Deviance:	4158.1
Time:	18:10:40	Pearson chi2:	4.80e+03
No. Iterations:	7		
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2040	0.196	1.043	0.297	-0.179	0.587
TotalVisits	11.1489	2.665	4.184	0.000	5.926	16.371
Total Time Spent on Website	4.4223	0.185	23.899	0.000	4.060	4.785
Lead Origin_Lead Add Form	4.2051	0.258	16.275	0.000	3.699	4.712
Lead Source_Olark Chat	1.4526	0.122	11.934	0.000	1.214	1.691
Lead Source_Welingak Website	2.1526	1.037	2.076	0.038	0.121	4.185
Do Not Email_Yes	-1.5037	0.193	-7.774	0.000	-1.883	-1.125
Last Activity_Had a Phone Conversation	2.7552	0.802	3.438	0.001	1.184	4.326
Last Activity_SMS Sent	1.1856	0.082	14.421	0.000	1.024	1.347
What is your current occupation_Student	-2.3578	0.281	-8.392	0.000	-2.908	-1.807
What is your current occupation_Unemployed	-2.5445	0.186	-13.699	0.000	-2.908	-2.180
Last Notable Activity_Unreachable	2.7846	0.807	3.449	0.001	1.202	4.367

ROC Curve





From ROC curve we can observe that optimal cut-off point is approx. 0.42

Conclusion

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

1. Total number of visits.
2. The total time spend on the Website.
3. Lead Origin_Lead Add Form.
4. Last Notable Activity_Unreachable.
5. Last Activity_Had a Phone Conversation.
6. Lead Source_Welingak Website.
7. Lead Source_Olark Chat.
8. Last Activity_SMS Sent.
9. Do Not Email_Yes.
10. What is your current occupation_Student.
11. What is your current occupation_Unemployed.

Recommendations to improve the business

1. It is recommended to give utmost importance to the above variables while planning to achieve maximum registration/enrollment to the X Education.
2. As we can see Total number of visits and The total time spend on the Website are having very positive impact on the enrollment, we recommend to design UI/UX of the website more easy, eye catching and informative about the course.
3. And it is recommended to provide some attractive offers/discounts in the home page so that visitors will be curious to know more.

Thank You