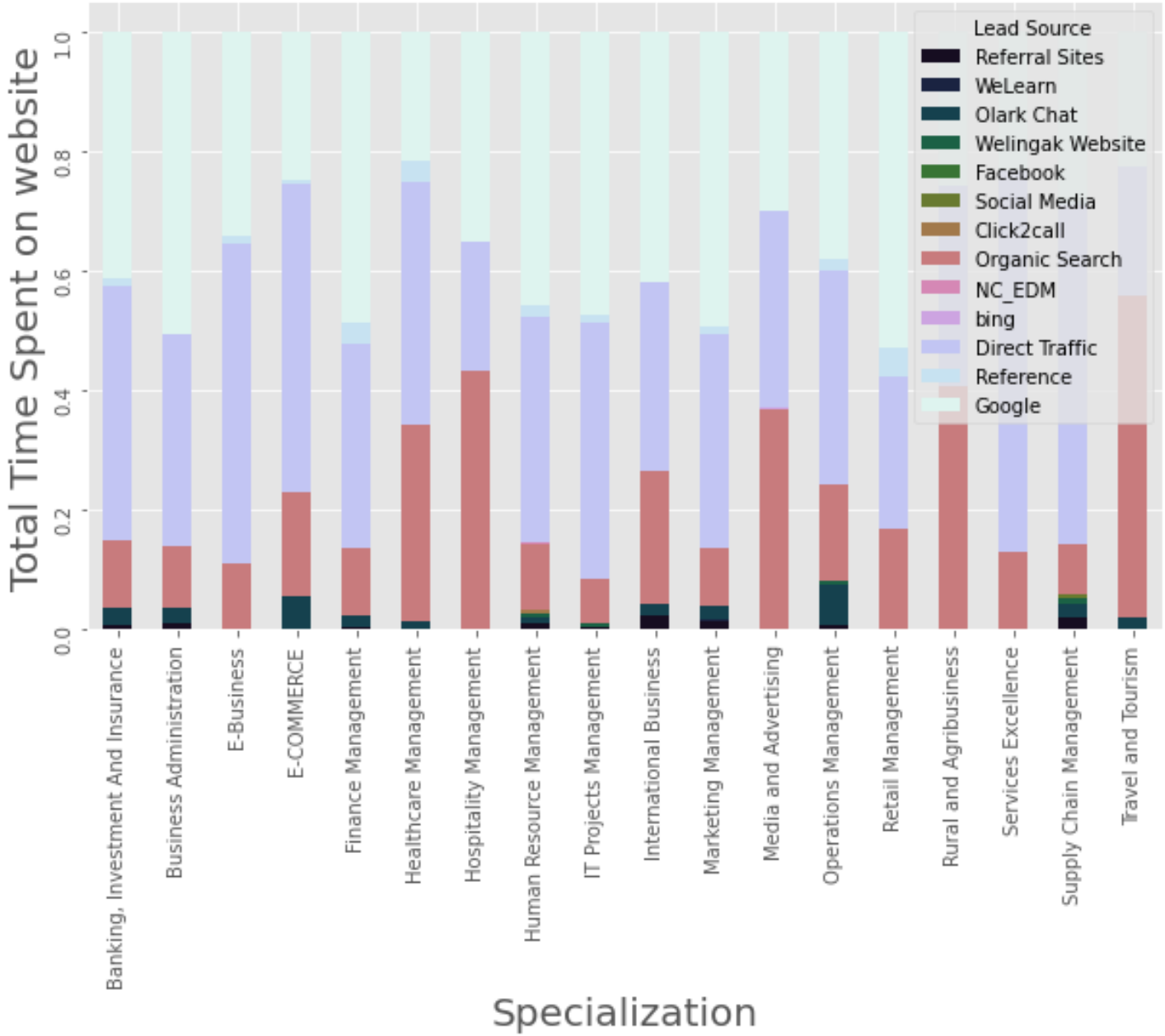
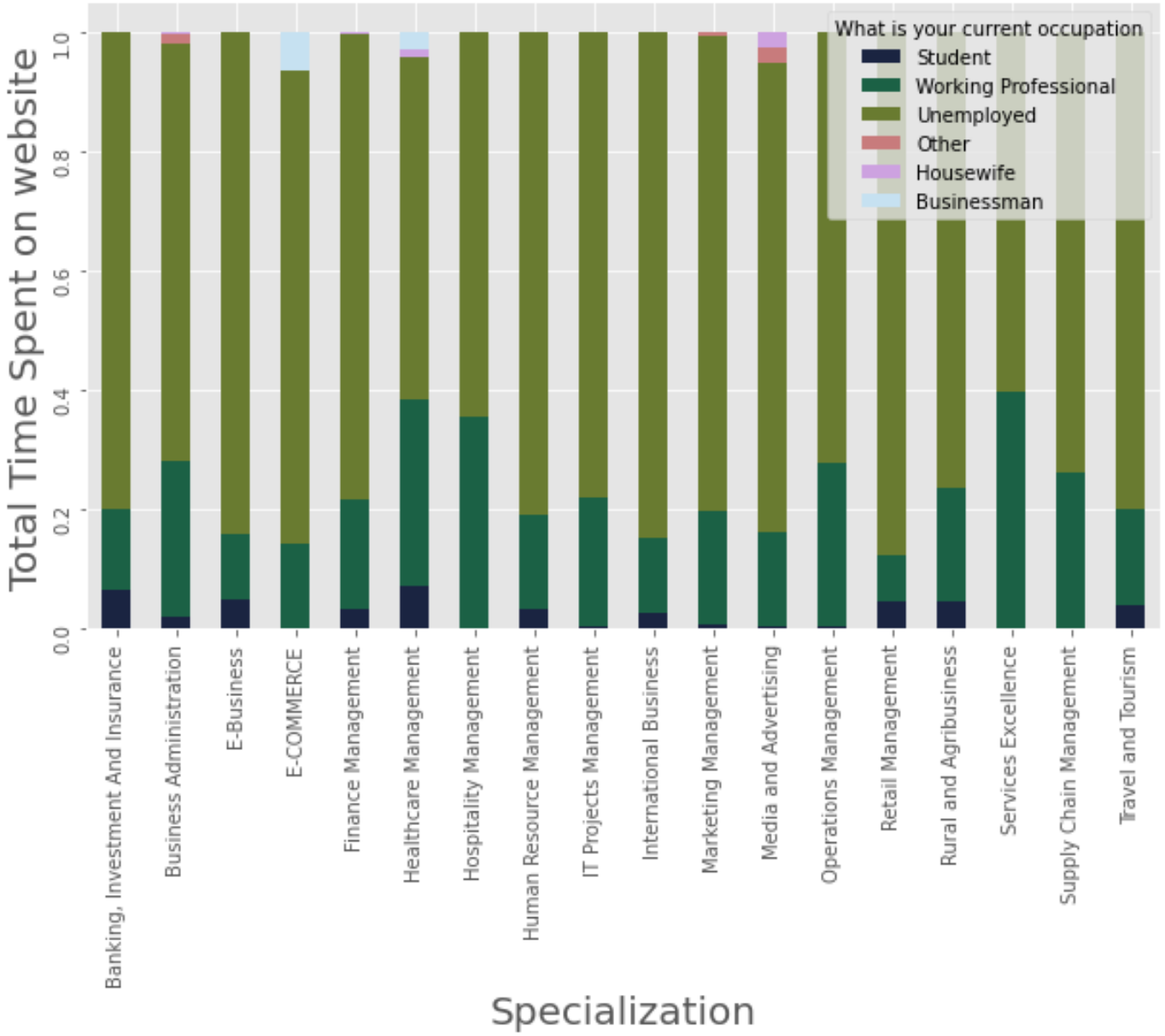
Lead – Case Study Visualization

- The analysis of conversion data for users which are influenced by categorical data such as Lead source, Specialization, and Occupation is made and the data is visualised for the better understanding.
- The data is separated into converted/non-converted users and analysed with the categorical variables and then the data is prepared for modelling.

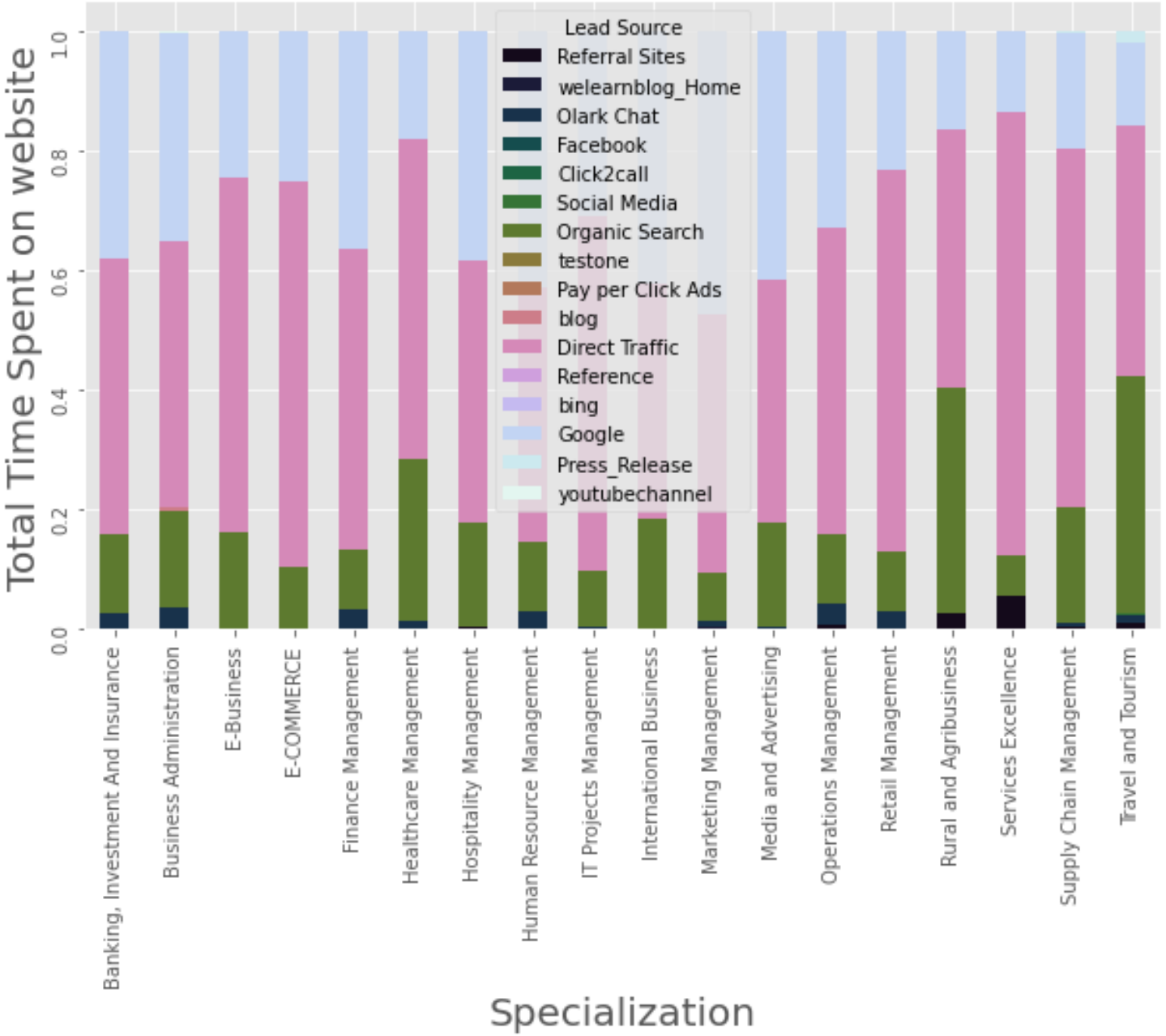
Lead Source vs Specialization Analysis on User Activity - Converted



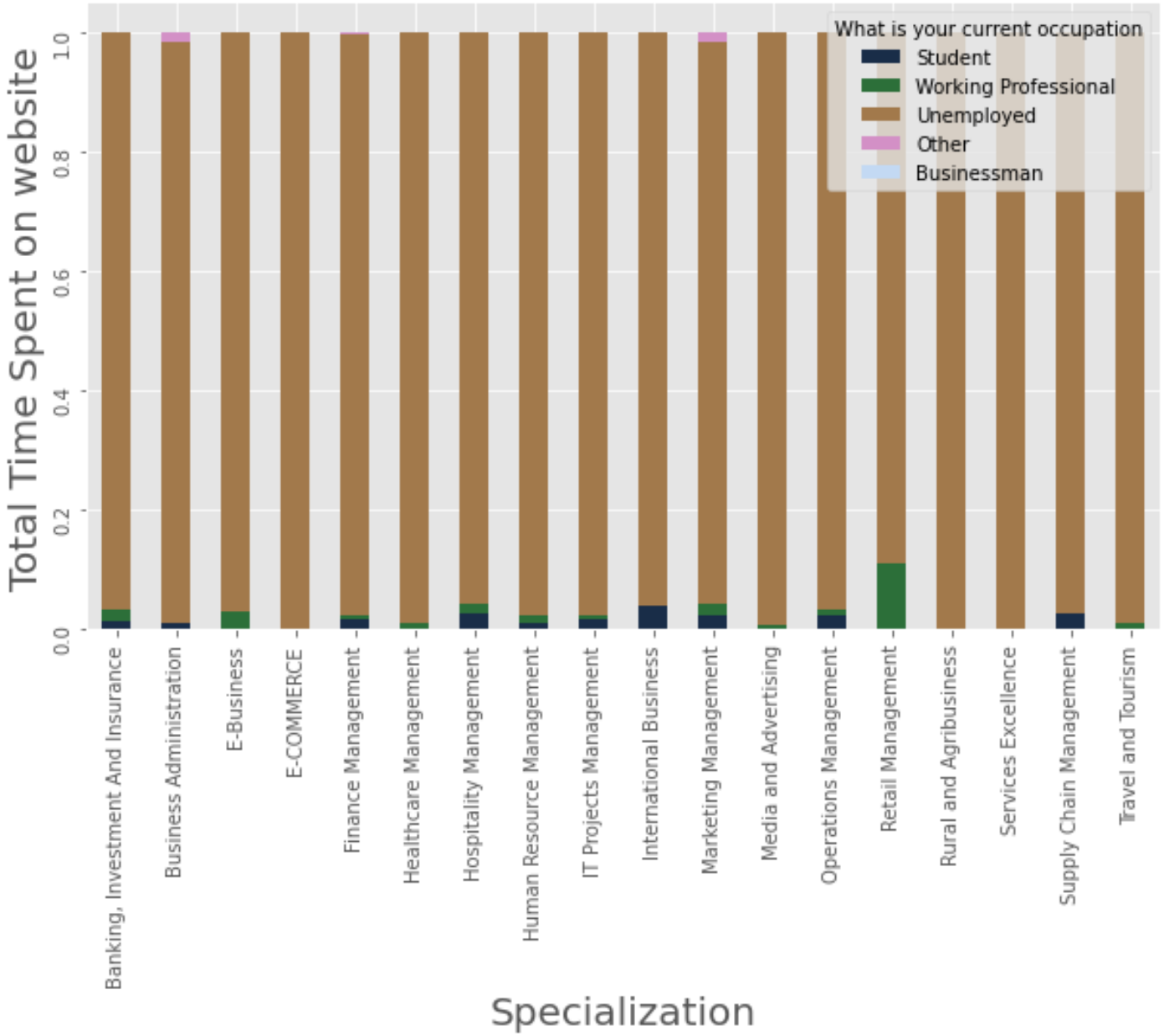
Current Occupation vs Specialization Analysis on User Activity - Converted



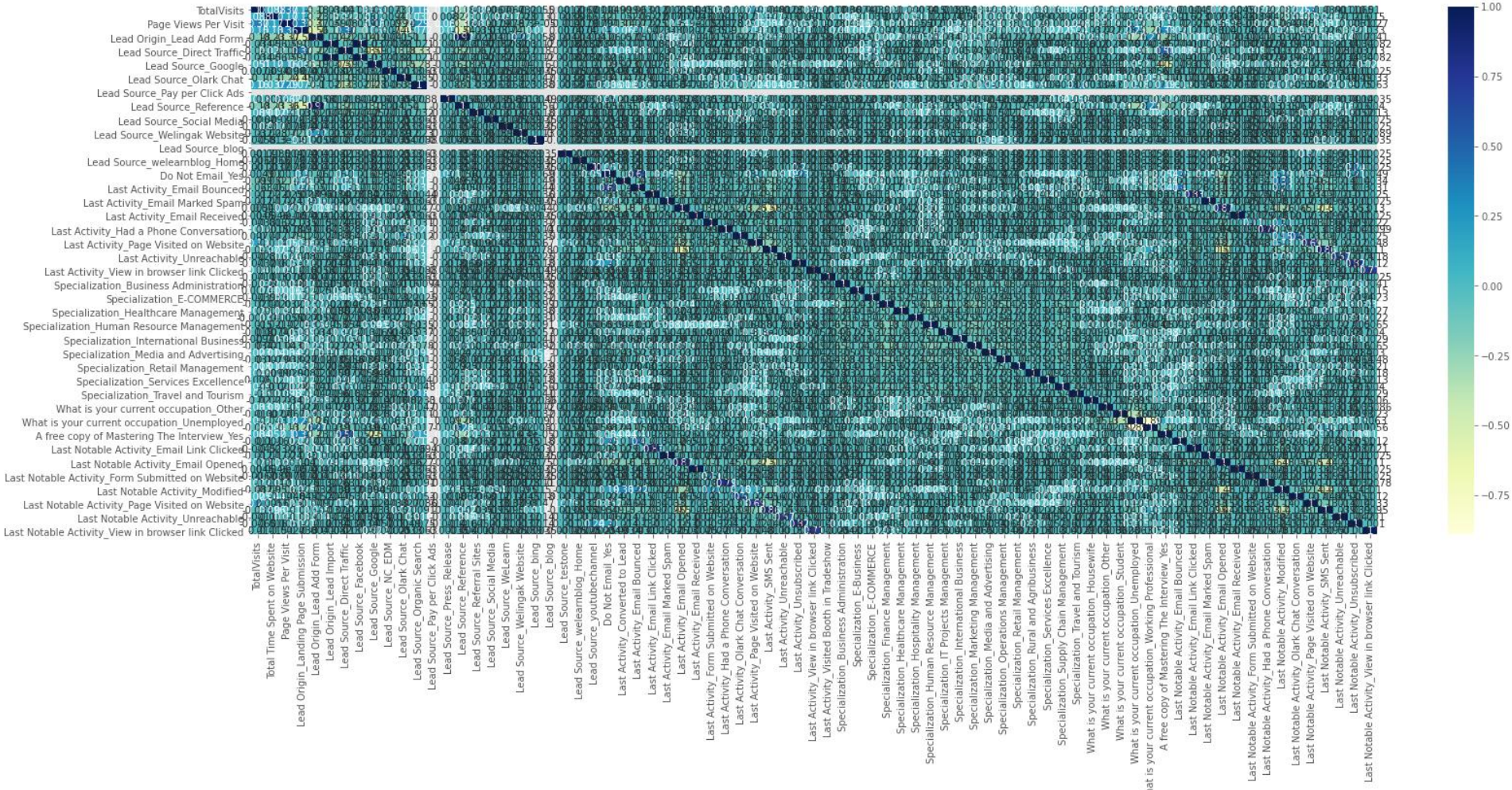
Lead Source vs Specialization Analysis on User Activity - Not Converted



Current Occupation vs Specialization Analysis on User Activity – Not Converted



Correlation-Matrix On Train Data (before RFE)



Trained Model

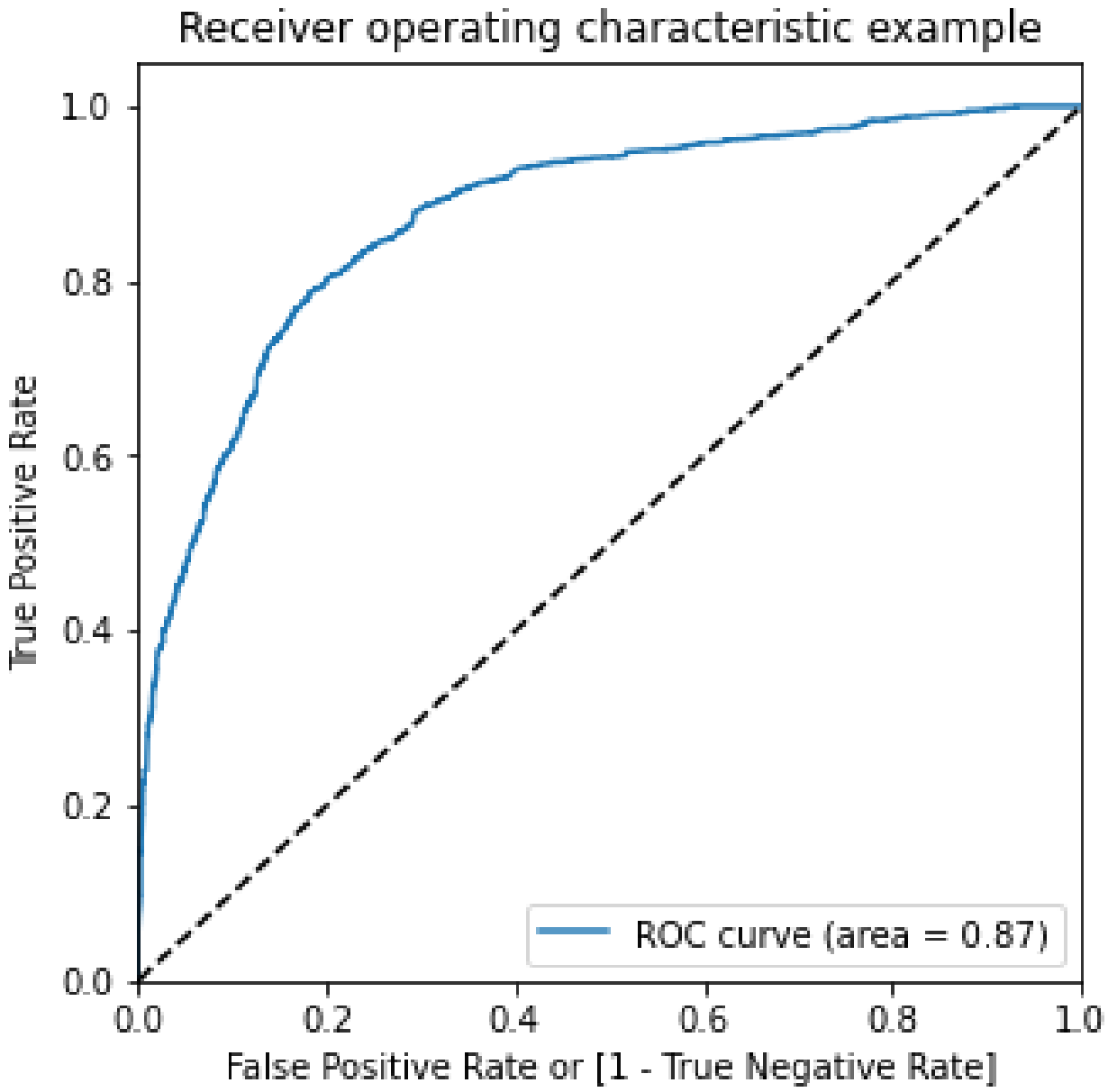
	coef	std err	z	P>|z|	[0.025	0.975]
const	1.3307	0.136	9.807	0.000	1.065	1.597
Total Time Spent on Website	1.0478	0.045	23.352	0.000	0.960	1.136
Lead Origin_Landing Page Submission	-1.3043	0.127	-10.280	0.000	-1.553	-1.056
Lead Source_Reference	3.2334	0.355	9.109	0.000	2.538	3.929
Lead Source_Referral Sites	1.8968	0.578	3.282	0.001	0.764	3.030
Do Not Email_Yes	-1.5132	0.235	-6.441	0.000	-1.974	-1.053
Last Activity_Email Bounced	-1.0966	0.487	-2.249	0.024	-2.052	-0.141
Last Activity_Form Submitted on Website	-1.0067	0.359	-2.802	0.005	-1.711	-0.303
What is your current occupation_Working Professional	2.9186	0.224	13.052	0.000	2.480	3.357
Last Notable Activity_Email Link Clicked	-1.1845	0.312	-3.801	0.000	-1.795	-0.574
Last Notable Activity_Email Opened	-0.9877	0.099	-9.930	0.000	-1.183	-0.793
Last Notable Activity_Had a Phone Conversation	2.3337	1.086	2.149	0.032	0.205	4.462
Last Notable Activity_Modified	-1.4268	0.111	-12.865	0.000	-1.644	-1.209

Correlation-Matrix On Train Data (after RFE)

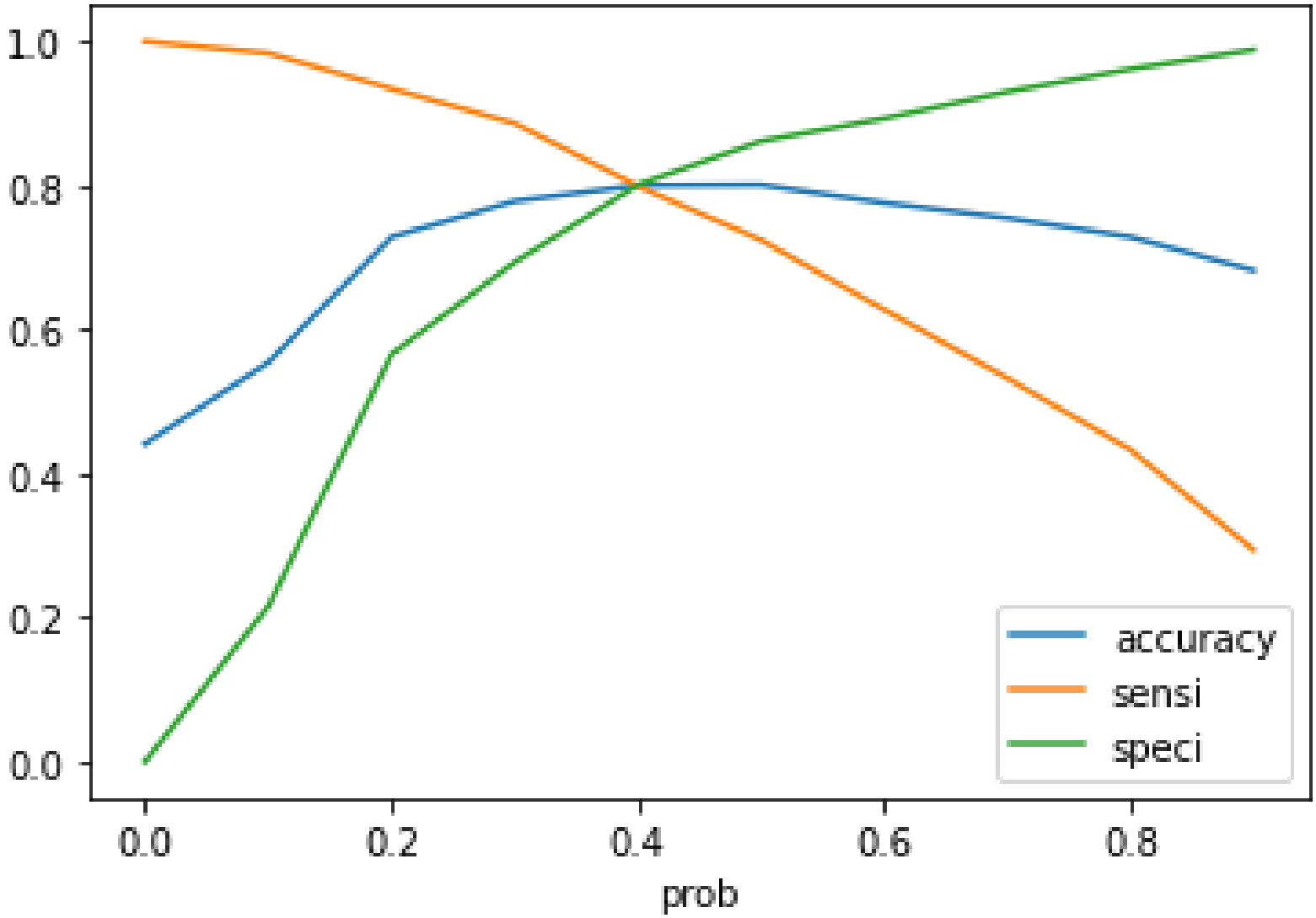
VIF for the train model

	Features	VIF
0	const	11.26
5	Do Not Email_Yes	1.67
6	Last Activity_Email Bounced	1.63
3	Lead Source_Reference	1.52
12	Last Notable Activity_Modified	1.46
2	Lead Origin_Landing Page Submission	1.45
10	Last Notable Activity_Email Opened	1.40
8	What is your current occupation_Working Profes...	1.14
1	Total Time Spent on Website	1.10
7	Last Activity_Form Submitted on Website	1.06
9	Last Notable Activity_Email Link Clicked	1.04
4	Lead Source_Referral Sites	1.01
11	Last Notable Activity_Had a Phone Conversation	1.01

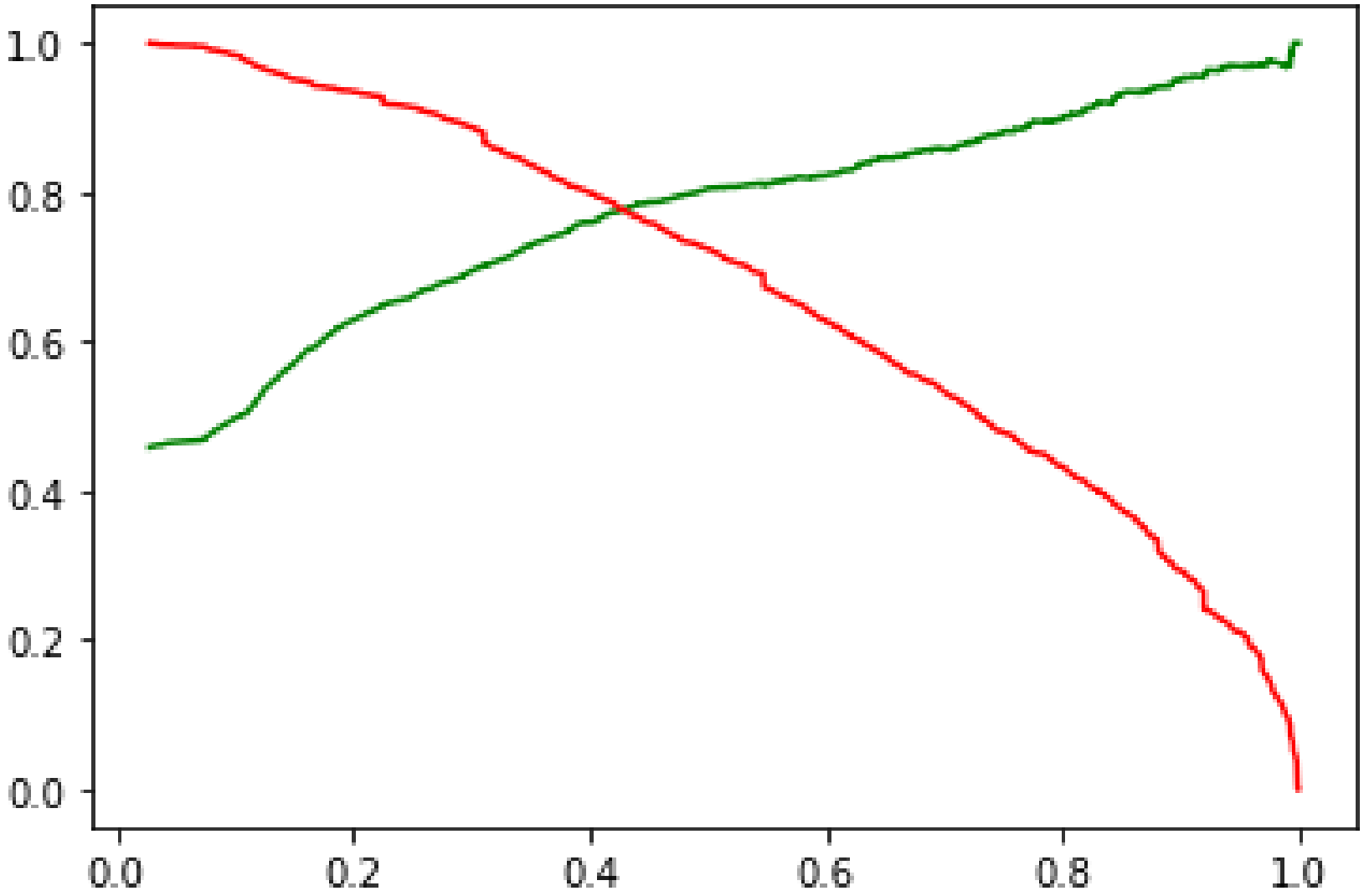
ROC Curve


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.87)

Probability Cut-off curve for varying probabilities



Precision – Recall Curve



Inferences

- To build a efficient and precise model, the RFE is used and top 15 variables are considered. By eliminating the variables with high p-values and VIF values, we generated a better model and proceed with our analysis.
- After building the final model, we had generated an ROC curve which had an area score of 0.87, which is a very good score which means we have good accuracy.
- We had calculated the accuracy, sensitivity and specificity and found out that the probability cut-off value is 0.4 (approx).
- The precision and recall values are 0.76 and 0.77.
- The accuracy of the predicted model is found to be around 0.8.
- The variables to be focused for conversion are,
  - Total Time Spent on Website
  - Lead Origin_Landing Page Submission(Direct Traffic)
  - Lead Source_Reference