Lead Score assignment

X Education enrolling Analysis

1	Lead Source	9204 non-null	object
2	Do Not Email	9240 non-null	object
3	Do Not Call	9240 non-null	object
4	Converted	9240 non-null	int64
5	TotalVisits	9103 non-null	float64
6	Total Time Spent on Website	9240 non-null	int64
7	Page Views Per Visit	9103 non-null	float64
8	Last Activity	9137 non-null	object
9	Country	6779 non-null	object
10	Specialization	7802 non-null	object
11	How did you hear about X Education	7033 non-null	object
12	What is your current occupation	6550 non-null	object
13	What matters most to you in choosing a course	6531 non-null	object
14	Search	9240 non-null	object
15	Magazine	9240 non-null	object
16	Newspaper Article	9240 non-null	object
17	X Education Forums	9240 non-null	object
18	Newspaper	9240 non-null	object
19	Digital Advertisement	9240 non-null	object
20	Through Recommendations	9240 non-null	object
21	Receive More Updates About Our Courses	9240 non-null	object
22	Tags	5887 non-null	object
23	Lead Quality	4473 non-null	object
24	Update me on Supply Chain Content	9240 non-null	object
25	Get updates on DM Content	9240 non-null	object
26	Lead Profile	6531 non-null	object
27	City	7820 non-null	object
28	Asymmetrique Activity Index	5022 non-null	object
29	Asymmetrique Profile Index	5022 non-null	object
30	Asymmetrique Activity Score	5022 non-null	float64
31	Asymmetrique Profile Score	5022 non-null	float64
32	I agree to pay the amount through cheque	9240 non-null	object
33	A free copy of Mastering The Interview	9240 non-null	object
34	Last Notable Activity	9240 non-null	object

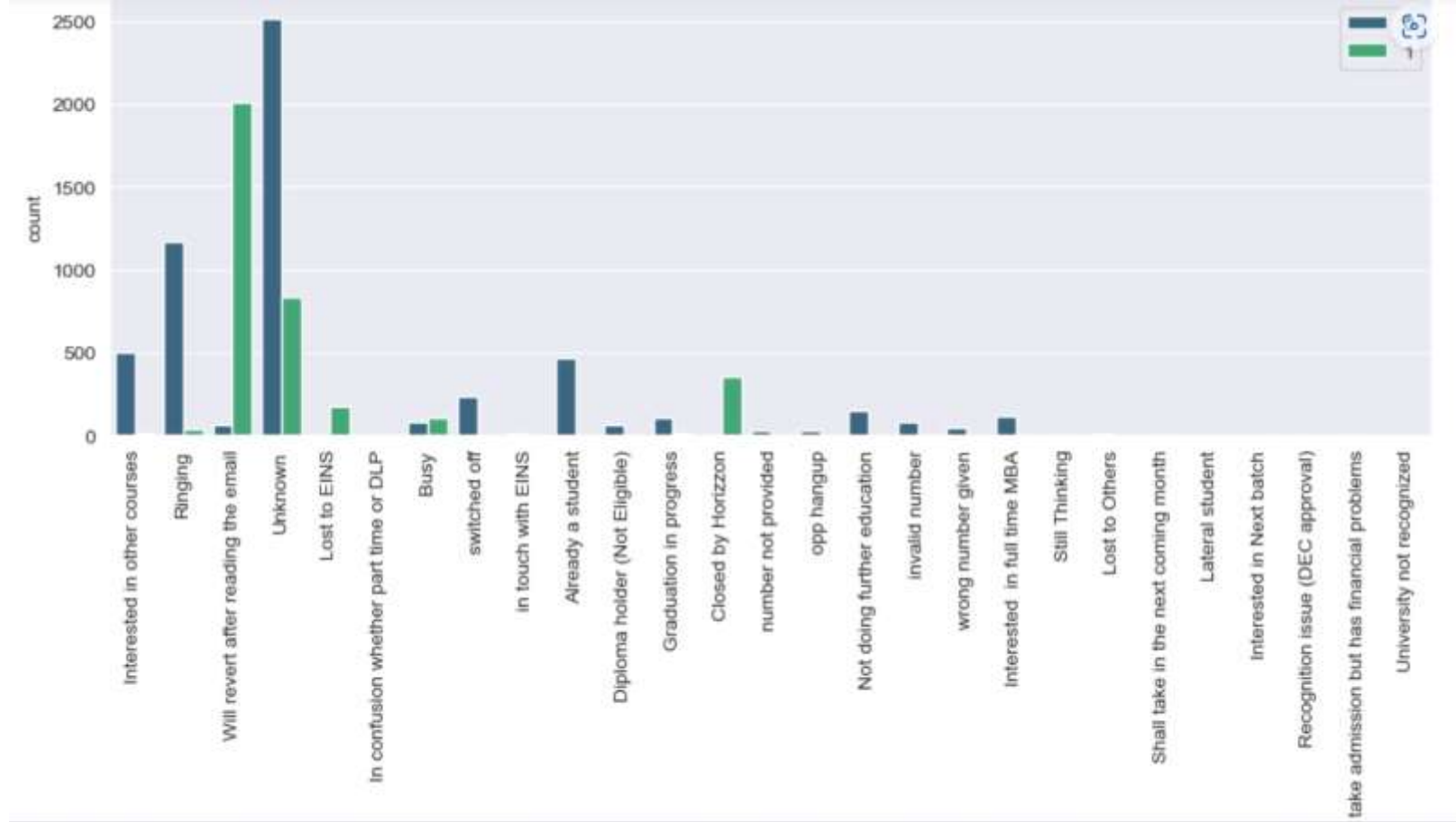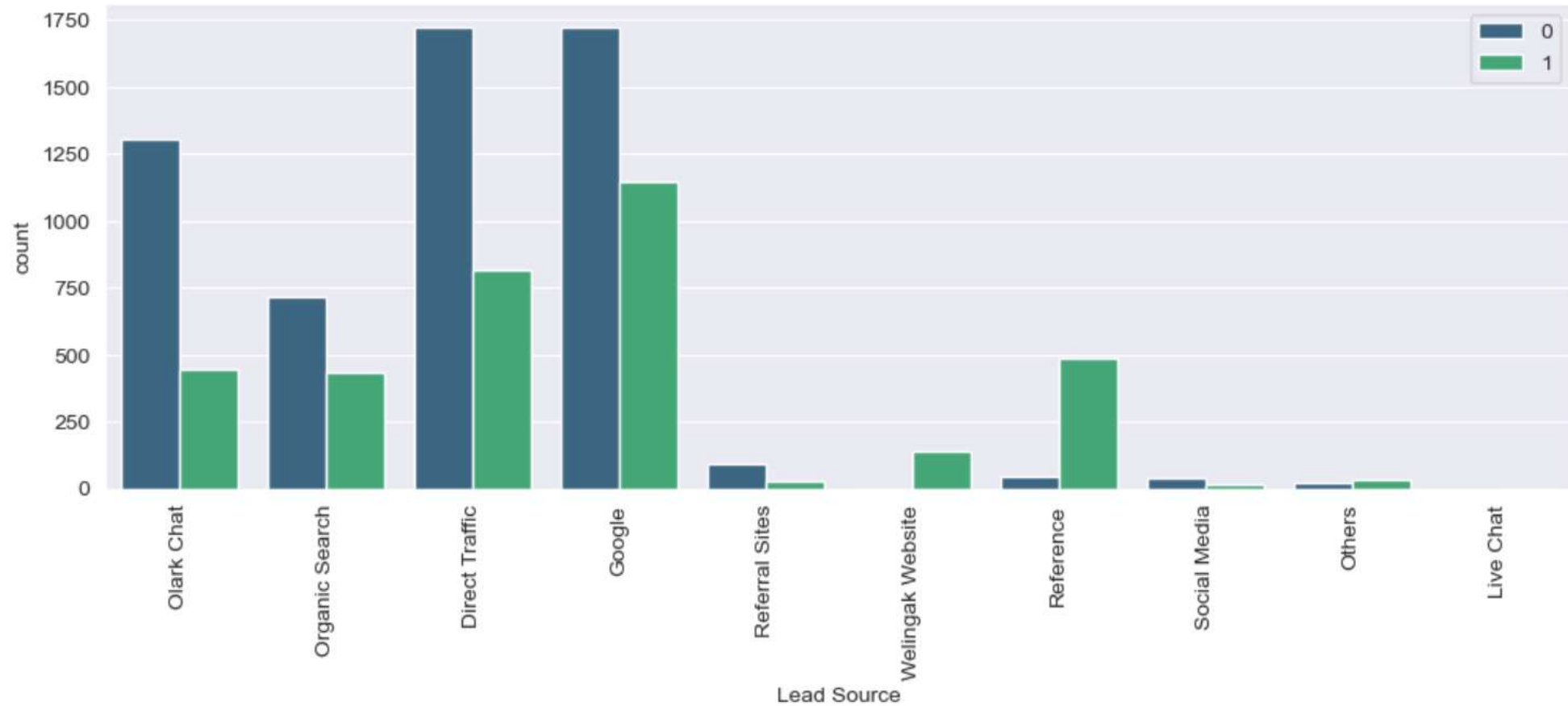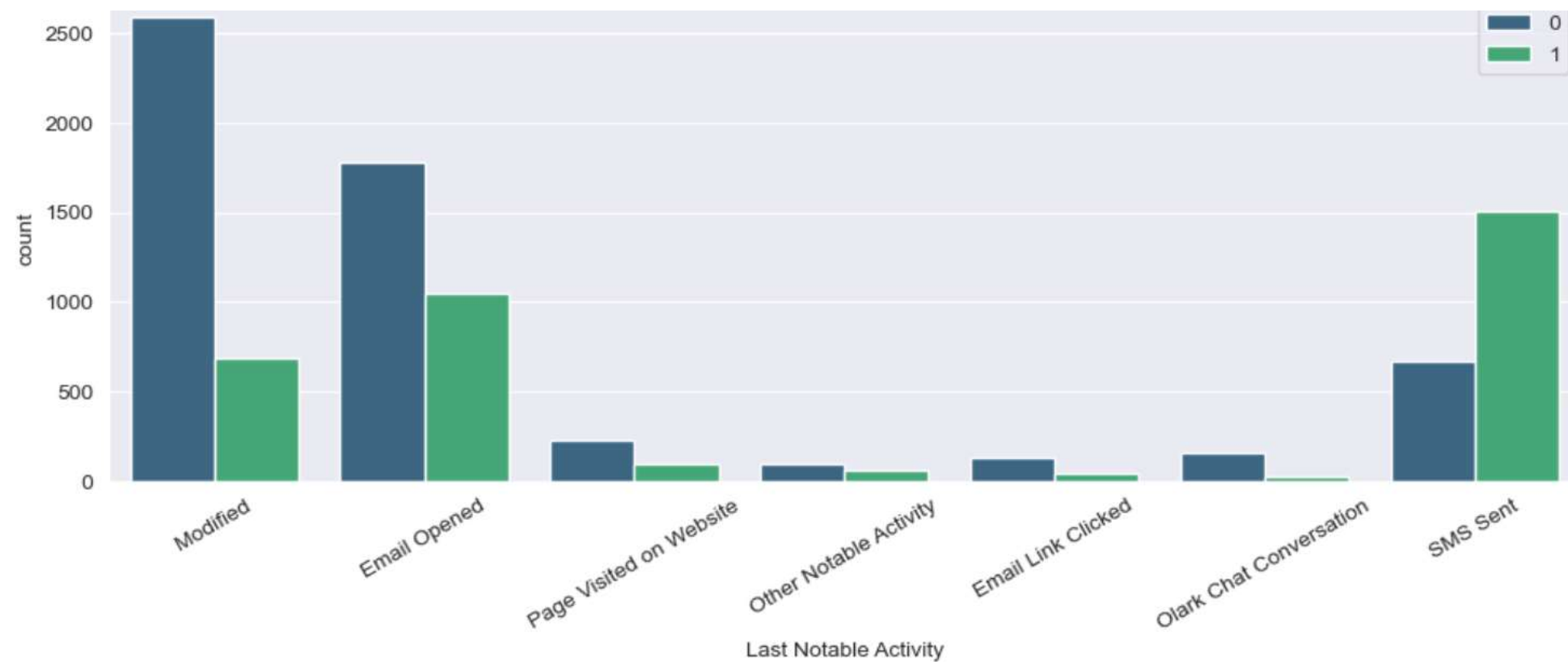# Overall view of country and specialization
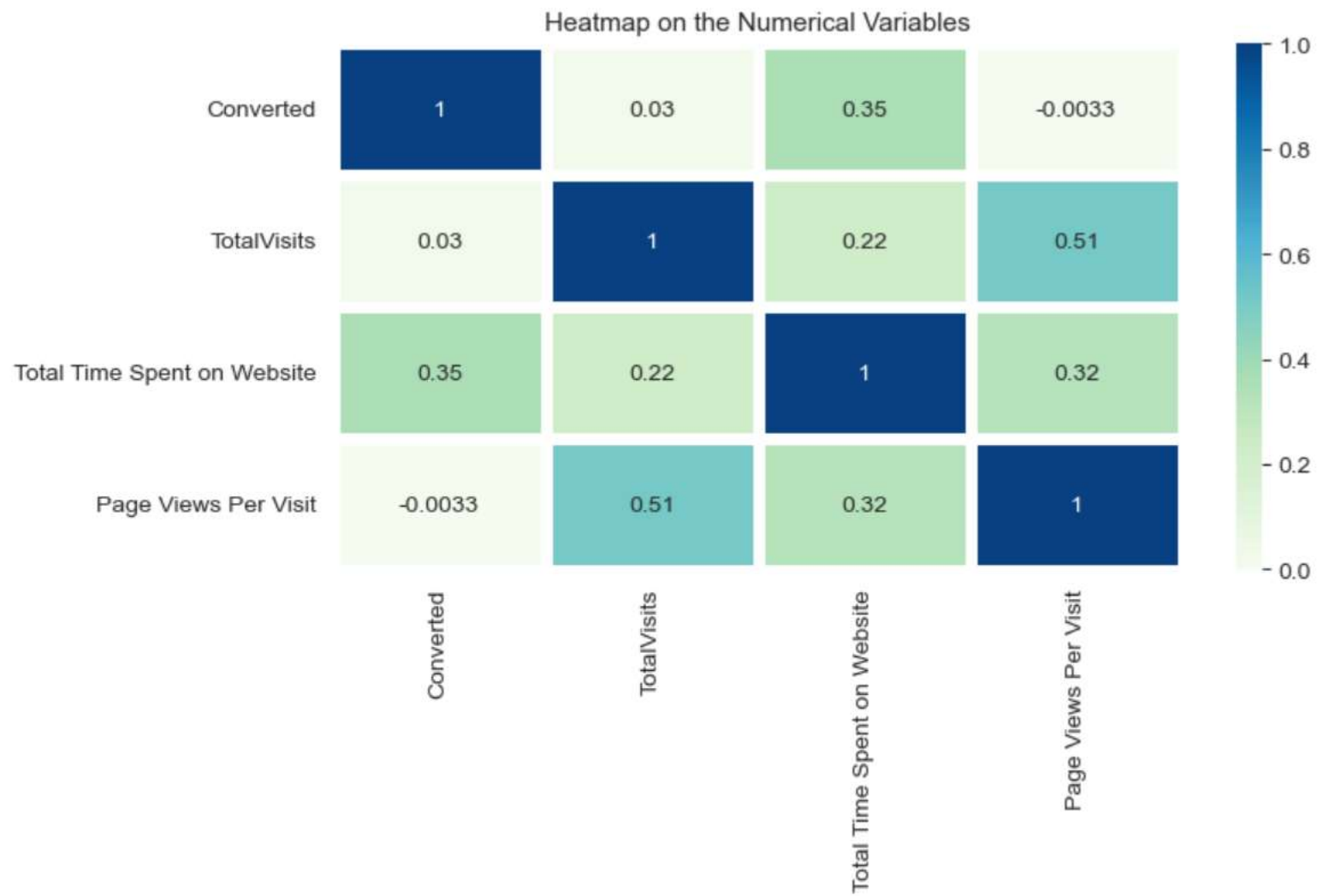
Country
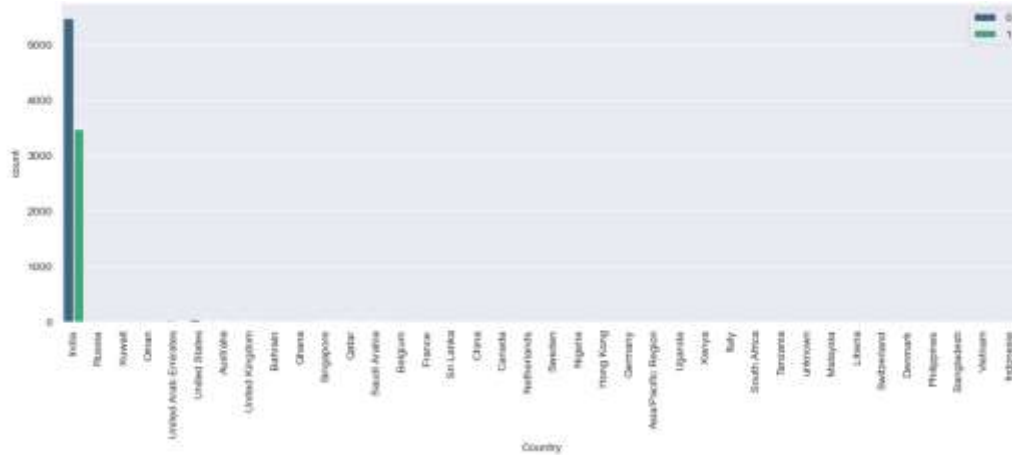
Specialization
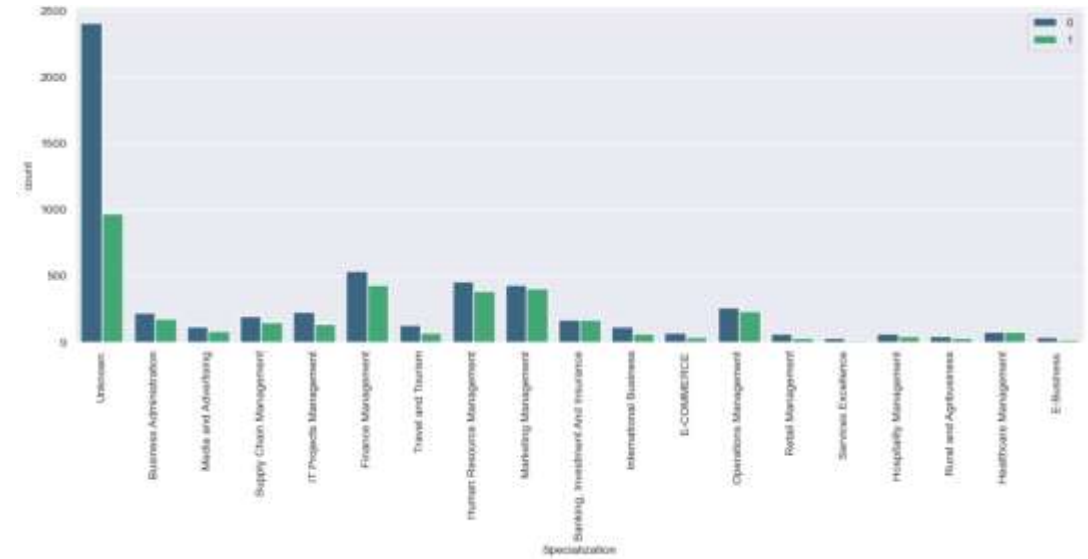
Why they are interesting

Approaching modes

Interacting

Heatmap on the Numerical Variables

```
                    Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:              Converted   No. Observations:                 6267
Model:                            GLM   Df Residuals:                     6253
Model Family:                Binomial   Df Model:                           13
Link Function:                  Logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -1263.3
Date:                Sun, 16 Apr 2023   Deviance:                       2526.6
Time:                        02:44:57   Pearson chi2:                 8.51e+03
No. Iterations:                     8   Pseudo R-squ. (CS):             0.6037
Covariance Type:            nonrobust
==============================================================================
                                          coef  std err       z   P>|z|   [0.025   0.975]
------------------------------------------------------------------------------
const                                  -1.1179    0.084 -13.382   0.000   -1.282   -0.954
Total Time Spent on Website             0.8896    0.053  16.907   0.000    0.786    0.993
Lead Origin_Lead Add Form               1.6630    0.455   3.657   0.000    0.772    2.554
Lead Source_Direct Traffic             -0.8212    0.127  -6.471   0.000   -1.070   -0.572
Lead Source_Welingak Website            3.8845    1.114   3.488   0.000    1.701    6.068
Last Activity_SMS Sent                  1.9981    0.113  17.718   0.000    1.777    2.219
Last Notable Activity_Modified         -1.6525    0.124 -13.279   0.000   -1.896   -1.409
Last Notable Activity_Olark Chat Conversation -1.8023 0.491 -3.669 0.000  -2.765   -0.839
Tags_Closed by Horizzon                 7.1955    1.020   7.053   0.000    5.196    9.195
Tags_Interested in other courses       -2.1318    0.406  -5.253   0.000   -2.927   -1.336
Tags_Lost to EINS                       5.9177    0.611   9.689   0.000    4.721    7.115
Tags_Other Tags                        -2.3737    0.206 -11.507   0.000   -2.778   -1.969
Tags_Ringing                           -3.4531    0.238 -14.532   0.000   -3.919   -2.987
Tags_Will revert after reading the email 4.5070   0.188  24.002   0.000    4.139    4.875
==============================================================================
```

Final Observation:

Let us compare the values obtained for Train & Test:

Train Data:

- Accuracy : 92.29%
- Sensitivity : 91.70%
- Specificity : 92.66%

Test Data:

- Accuracy : 92.78%
- Sensitivity : 91.98%
- Specificity : 93.26%

The Model seems to predict the Conversion Rate very well & we should be able to give the CEO confidence in making good calls based on this model.