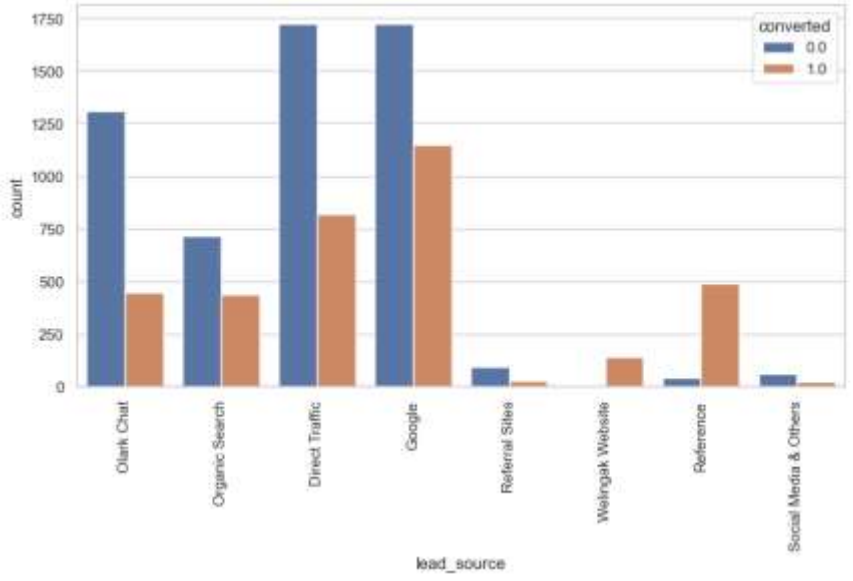
LEAD SCORING CASE STUDY

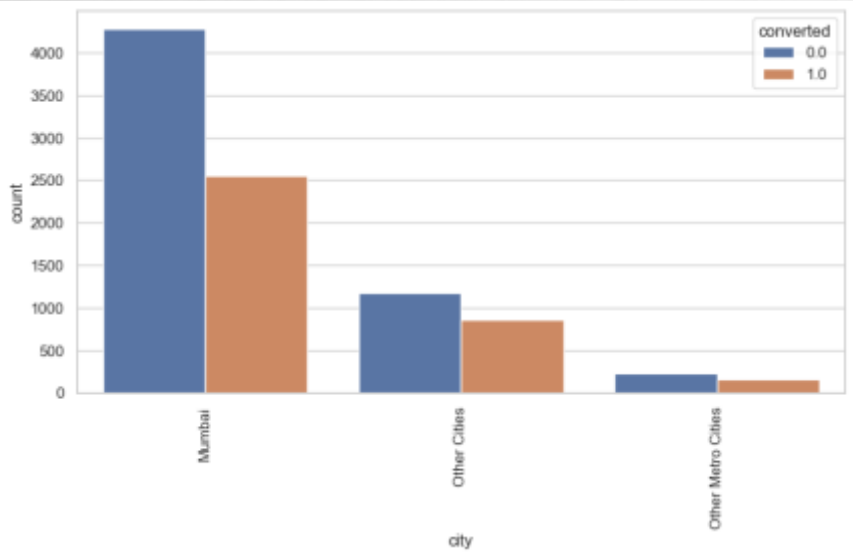
AIM: TO IDENTIFY THE MOST POTENTIAL LEADS, ALSO KNOWN AS ‘HOT LEADS’

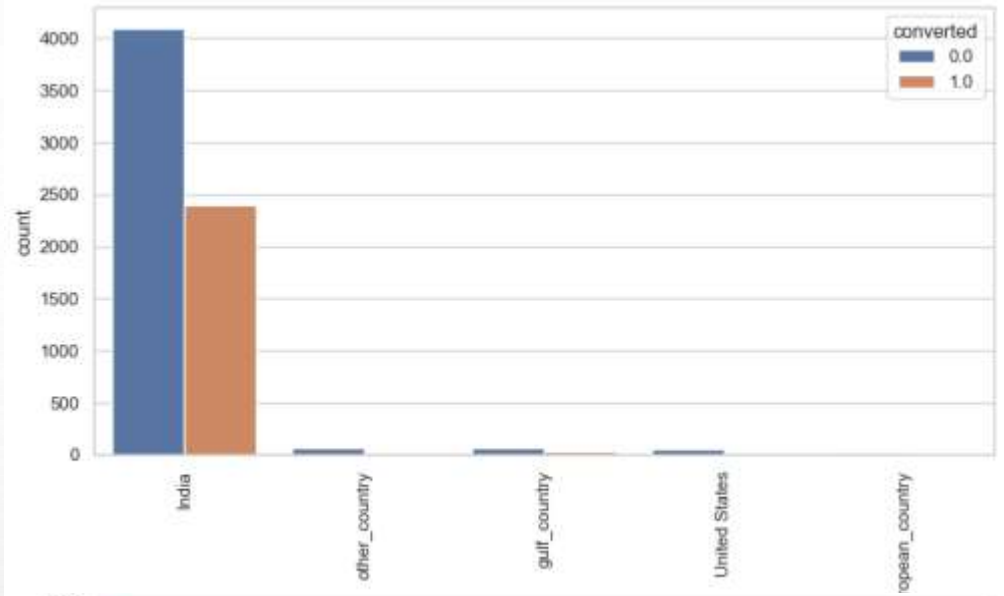
OBJECTIVE: WITH ABOUT 37 PREDICTOR VARIABLES PROVIDED WE HAD TO BUILD A LOGISTIC REGRESSION MODEL. THAT WILL HELP THE SALES TEAM FOCUS THEIR ENERGY AND TIME ON ONLY THE HIGHLY COVERABLE CANDIDATES. THEREBY INCREASING THE CONVERSION RATE WHICH AT PRESENT IS ONLY 30%.

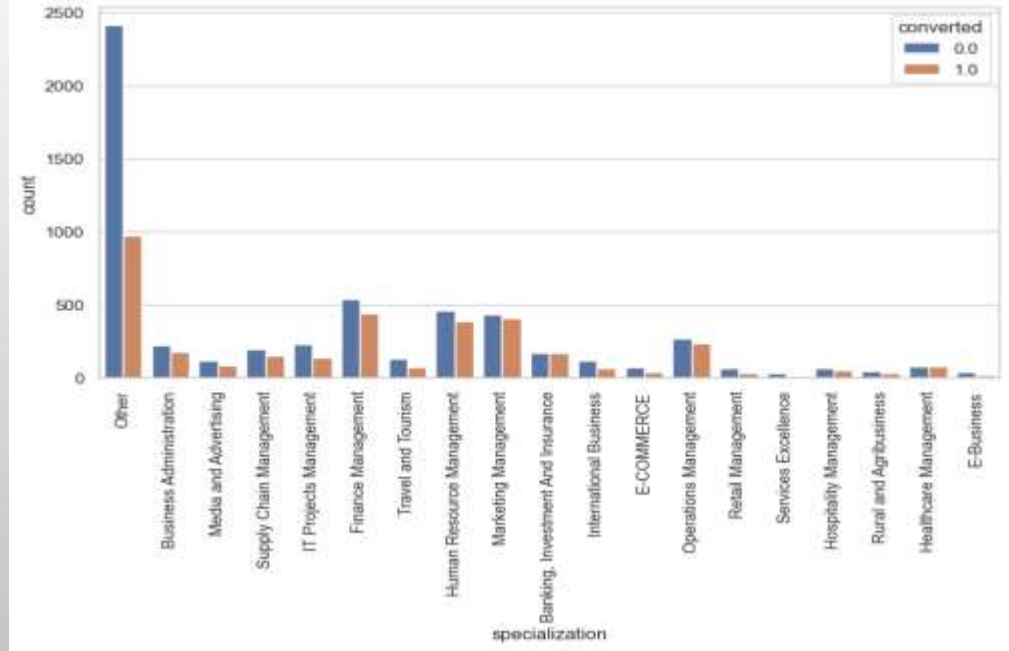
PROJECT BY: AKASH GAIKWAD

RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"

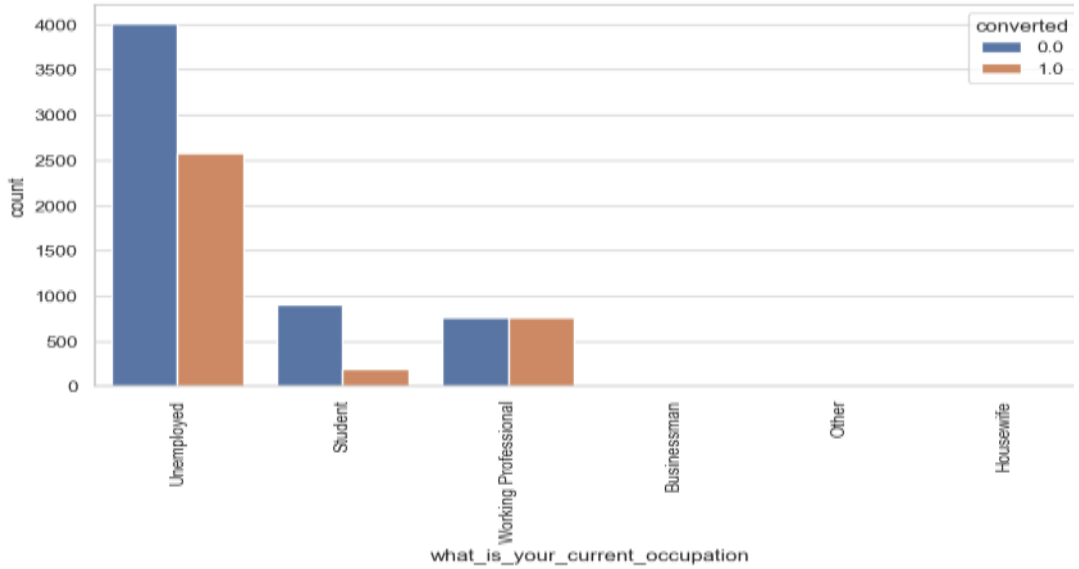


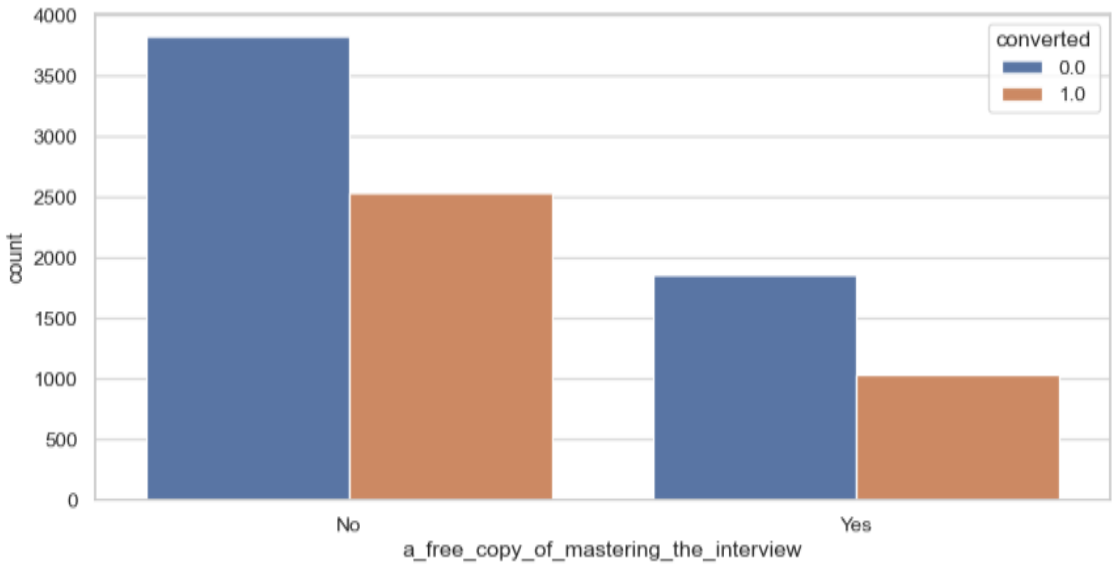


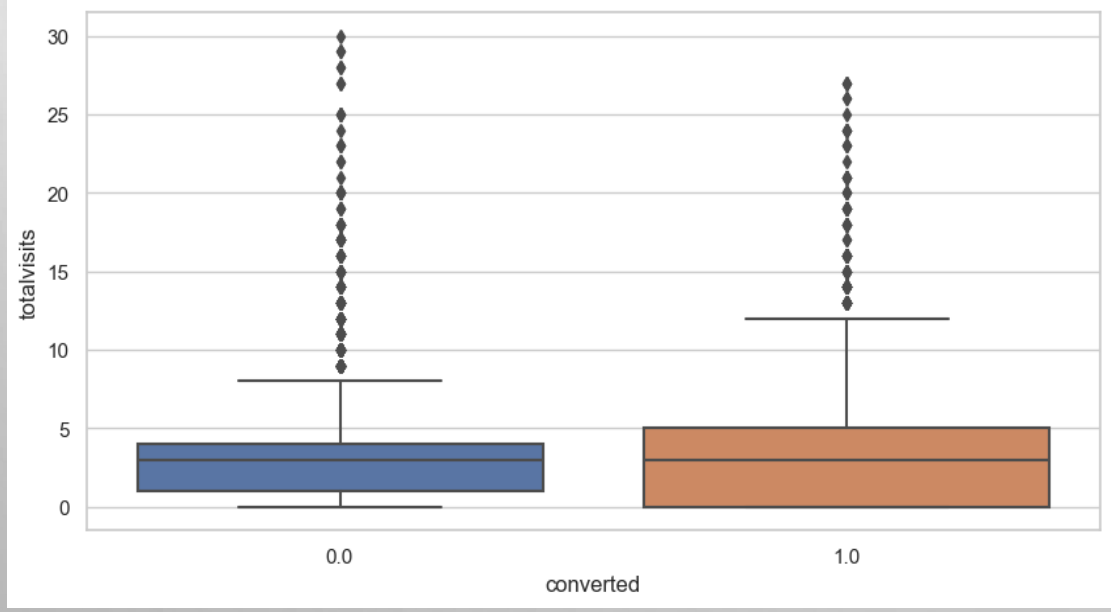


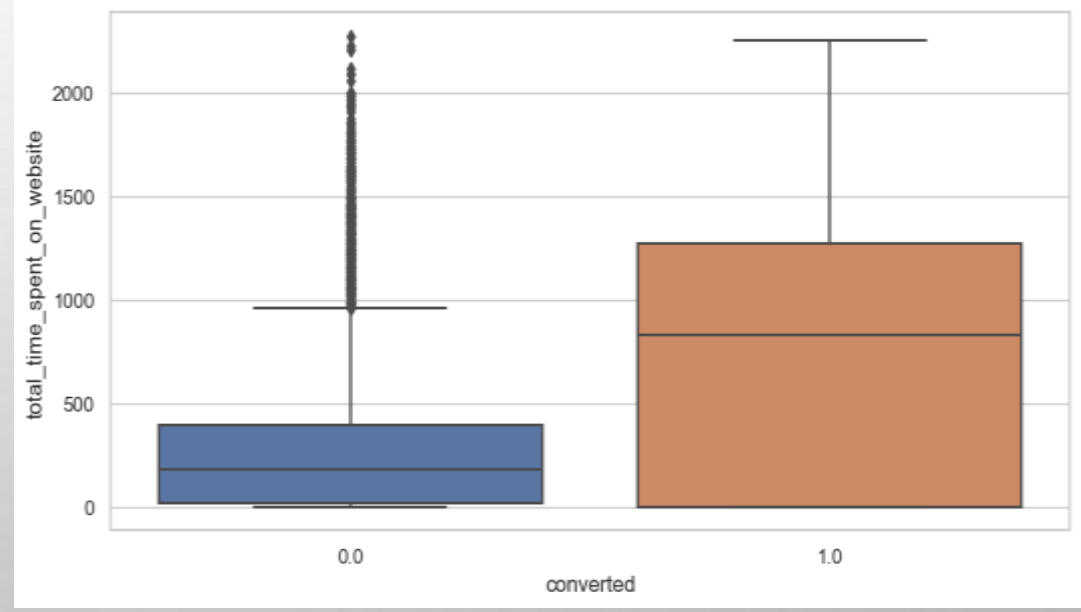


RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"









MAJOR HURDLES DURING DATA CLEANING

1) THE DATASET HAD MANY MISSING VAL

MOST OF THEM BEING CATEGORICAL IN

NATURE.

```
df1.country[df1.city=='Mumbai'].value_counts()
```

```
India                2918
gulf_country           32
United States          30
other_country          20
european_country       11
Name: country, dtype: int64
```

2) THERE WERE MAJOR DATA DISCREPANC

IN COUNTRY AND CITY

```
df1.country[df1.city=='Other Cities of Maharashtra'].value_counts()
```

3) MOST OF THE COLUMNS IN THE DATA

HAD VERY SKEWED DATA.

```
df1.country[df1.city=='Other Cities of Maharashtra'].value_counts()
```

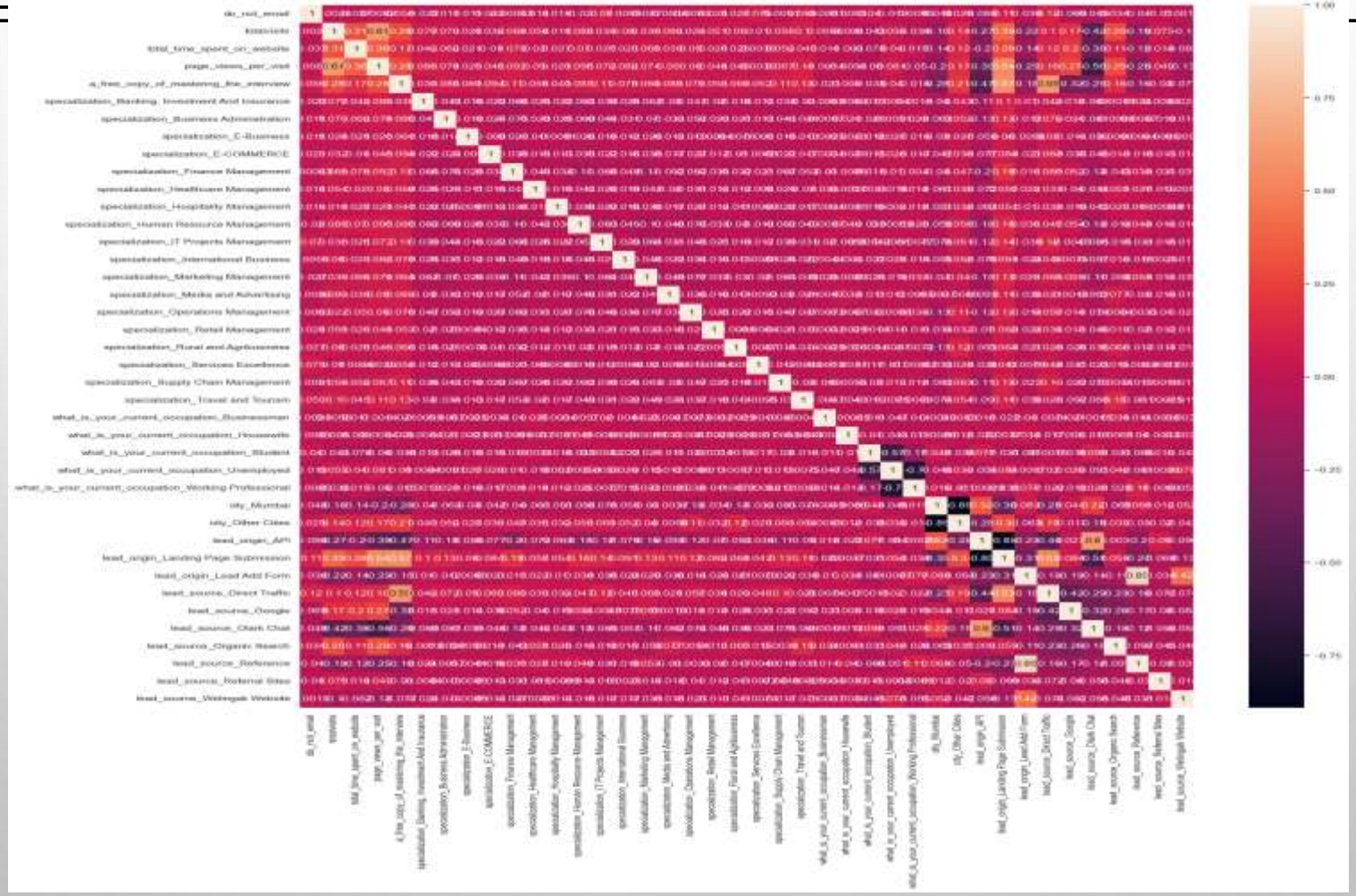
```
India                391
gulf_country           7
other_country          6
european_country       5
United States          5
Name: country, dtype: int64
```

4) SOME OF THE KEY VARIABLES USED IN THE

PREDICTION MODEL HAD OUTLIERS.

MACHINE LEANRING MODEL MAKING

AFTER CLEANING THE DATA, MANIPULATING IT, AND CREATING DUMMY VARIABLES, WE STA RTED THE MODEL BUILDING PROCESS USING NEARLY 40 VARIABLES. IT WAS ALMOST IMPOSSIBLE TO READ FROM THERE WE GO DIRECTLY TO REF TO STREAMLINE THE SELECTION



MACHINE LEANRING MODEL MAKING

RFE FEATURE SELECTION WITH 20 CUT-OFFS GOT US THE 20 MOST RELEVANT VARIABLES WITH THE LOWEST P-VALUE AND LOWEST VIF. VIF WAS USED TO STUDY THE MULTI-COLLINEARITY OF THE FEATURE VARIABLES.

'do_not_email',

'total_time_spent_on_website',

'specialization_Banking, Investment And Insurance',

'specialization_Business Administration',

'specialization_E-COMMERCE',

'specialization_Finance Management',

'specialization_Healthcare Management',

'specialization_Human Resource Management',

'specialization_IT Projects Management',

'lead_origin_Lead Add Form',

'specialization_Marketing Management',

'specialization_Media and Advertising',

'specialization_Operations Management',

'specialization_Rural and Agribusiness',

'specialization_Services Excellence',

'specialization_Supply Chain Management',

'what_is_your_current_occupation_Housewife',

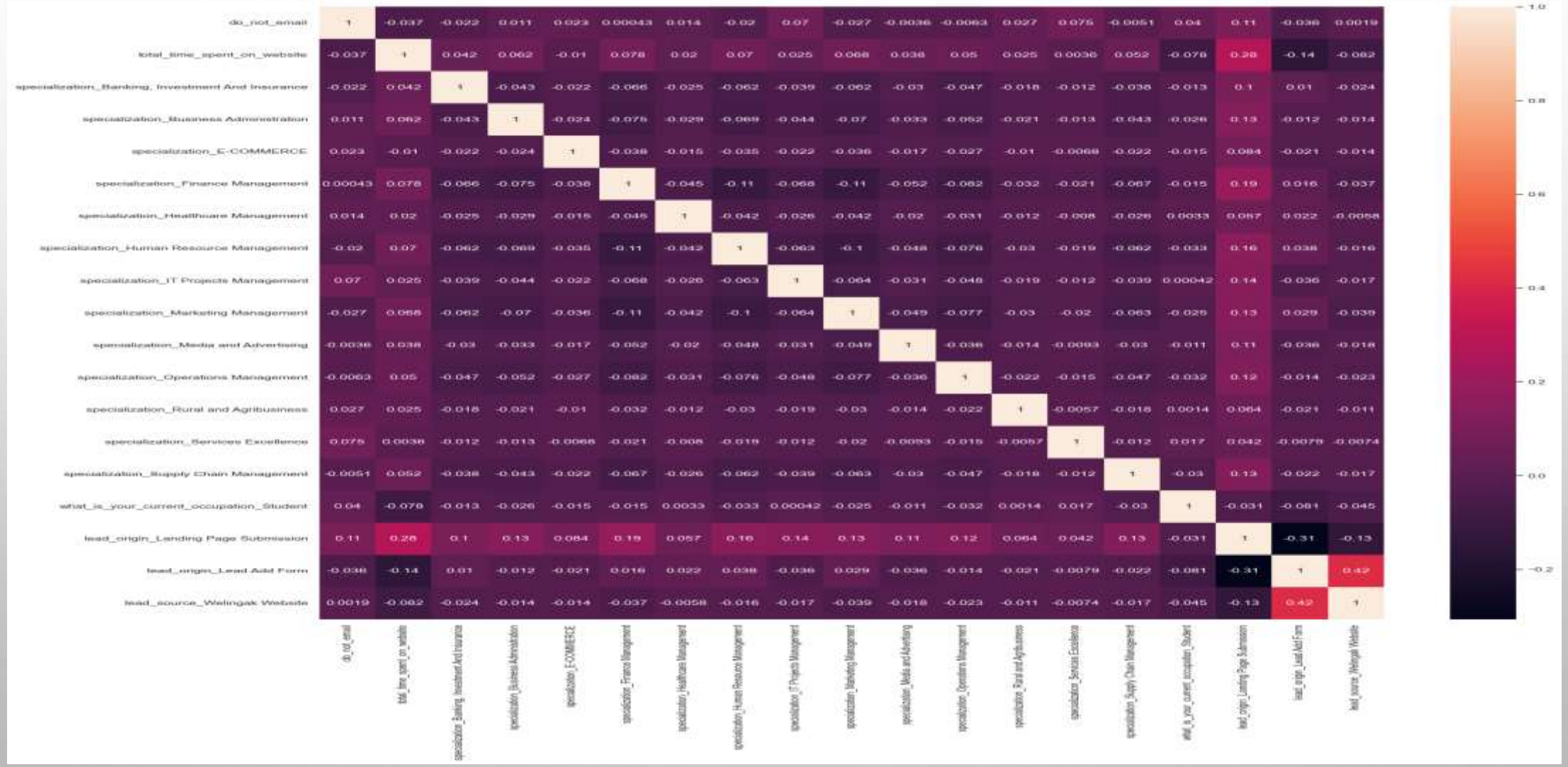
'what_is_your_current_occupation_Student',

'lead_origin_Landing Page Submission',

'lead_source_Welingak Website'

MACHINE LEANRING MODEL MAKING

HEATMAP REFLECTING THE CO-RELATION BETWEEN EACH OF THE VARIABLES



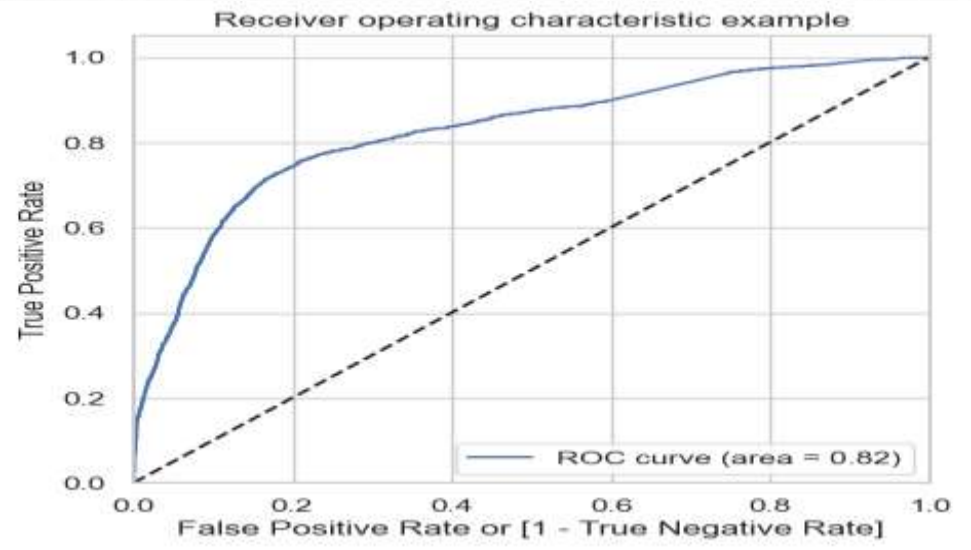
FINAL MACHINE LEARNING MODEL FOR X SALES TEAM

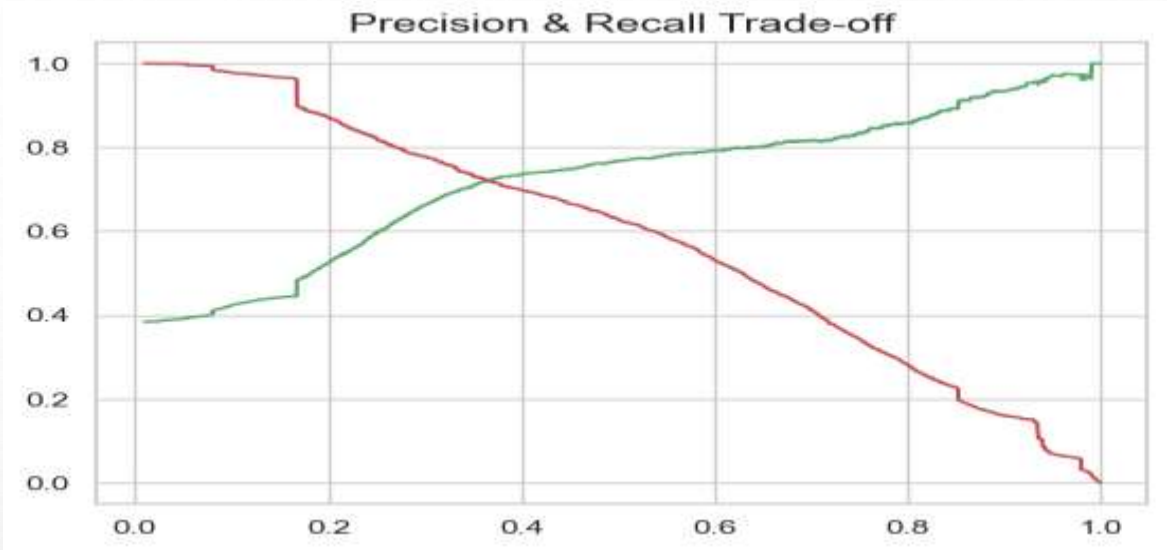
WITH A PROBABILITY THRESHOLD OF 0.30, WE CAN GET THE FOLLOWING SCORE WHICH IS SIMILAR TO THE TRAINING DATA WHEN TESTED ON THE TEST DATA . INDICATING THAT IT IS A RELIABLE MODEL.

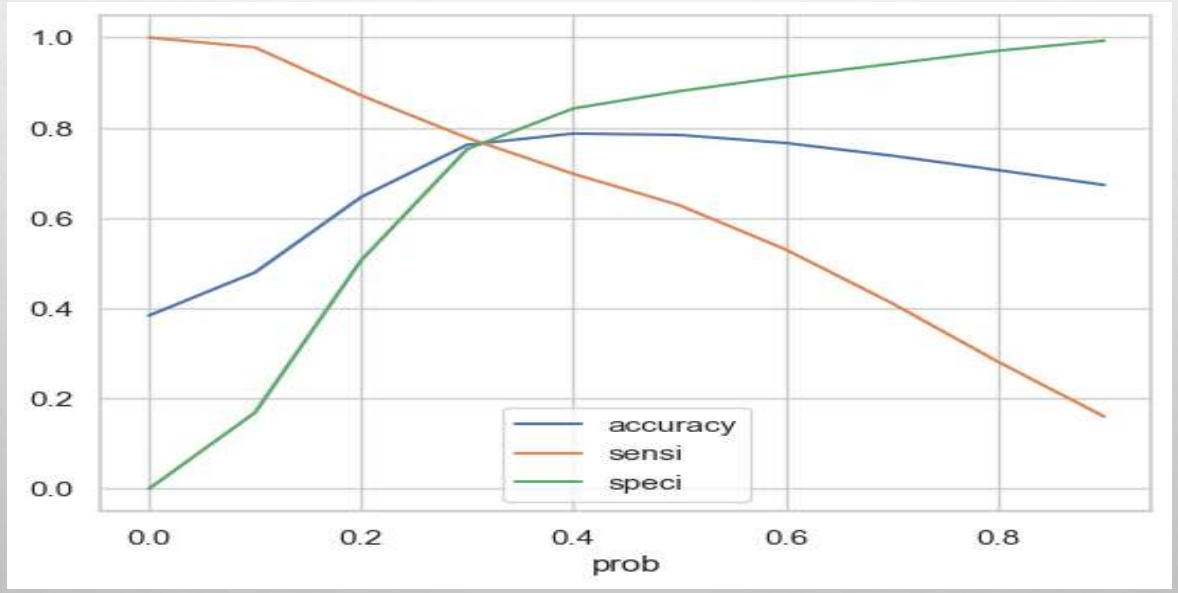
TRAIN DATA		
Confusion matrics:	2982	983
	554	1916
Accuracy score:	76%	
Sensitivity score:	77.57%	
Specificity score:	75.20%	
False Positive Rate:	24.79%	

TEST DATA		
Confusion matrics:	1261	437
	233	828
Accuracy score:	75.71%	
Sensitivity score:	78.03%	
Specificity score:	74.26%	

FINAL MACHINE LEARNING MODEL FOR X EDUCATION SALES TEAM







SUGGESTION

General Suggestion

1. The initial data collection form needs some major rework
2. There should be many options to provide free readings or bring your information to attract customers and interests.
3. Even though customers select 'NO' for call & e-mail they are highly convertible.

SPECIALIZATION BASED SUGGESTION

1. Banking, Investment And Insurance 1.1235
2. Business Administration 0.8662
3. E-Commerce 0.8385
4. Finance Management 0.8982
5. Healthcare Management 1.1608
6. Human Resource Management 0.9038
7. IT Projects Management 0.8173
8. Marketing Management 0.9700
9. Media and Advertising 0.8955
10. Operations Management 1.0477
11. Rural and Agribusiness 0.8370
12. Services Excellence 1.2927
13. Supply Chain Management 0.9176

Thank You!