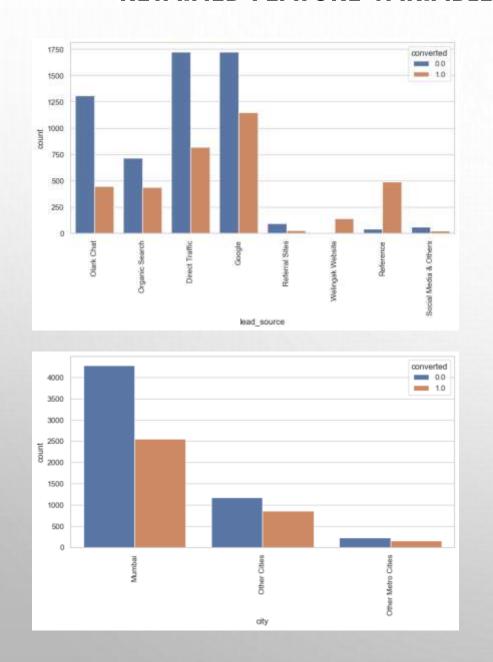
# LEAD SCORING CASE STUDY

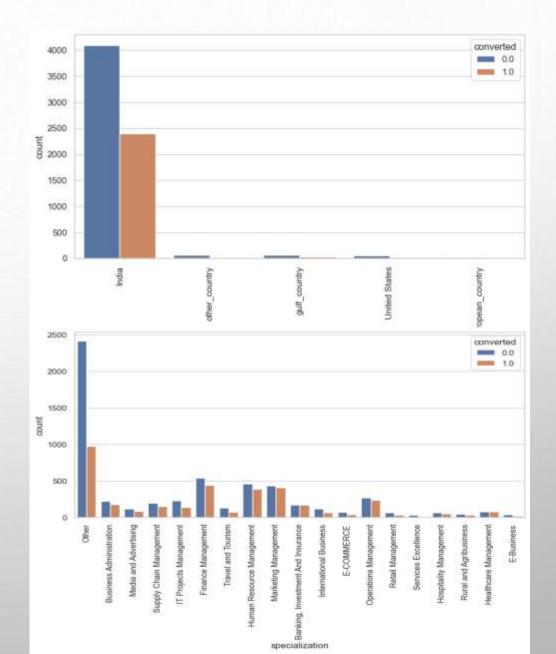
AIM: TO IDENTIFY THE MOST POTENTIAL LEADS, ALSO KNOWN AS 'HOT LEADS'

**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%.

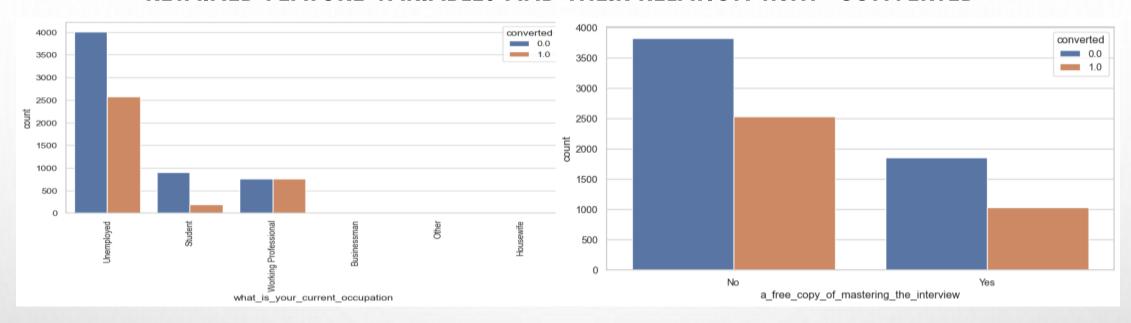
PROJECT BY: AKASH GAIKWAD

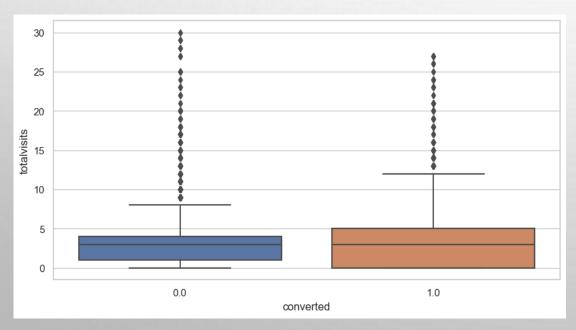
# RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"

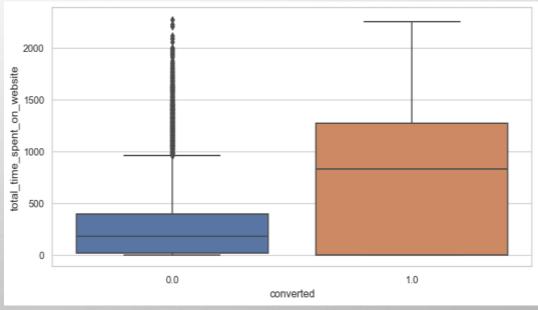




# RETAINED FEATURE VARIABLES AND THEIR RELATION WITH "CONVERTED"







#### MAJOR HURDLES DURING DATA CLEANING

MOST OF THEM BEING CATEGORICAL IN NATURE.

```
1) THE DATASET HAD MANY MISSING VAL 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 DISCREPANCE 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
other country
european country
United States
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** "LINE THE **SELECTION** -0.25

#### 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

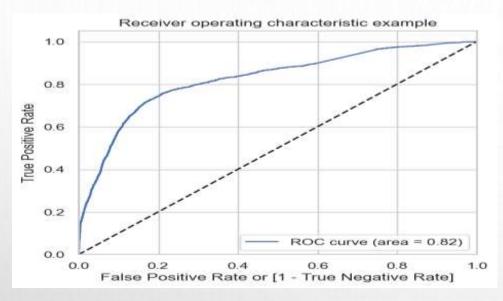
do_riot_email	+	-0.037	-0.000	0.011	0.023	0.00043	0.014	-0.02	0.07	0.027	-0.0036	-0.0063	0.027	0.075	0.0051	0.04	0.44	-0.036	0.0019	
Model_More_spent_con_website	-0.007	14.5	0.042	0.002	-0.01	0.078	0.02	0.07	0.025	0.06B	0.036	0.05	0.025	0 0036	0.052	-0.076	0.28	-0.14	-0.082	
cialization_Banking; investment And insurance	0.022	0.042	4	-0.049	-0.022	-0.066	-0.025	-0:062	0.039	0.002	-0.03	-0.047	0.018	0.012	0.038	-0.013		(6.01	-0 024	
specialization_Musiness Admirestration	0.011	0.062	-0.043	14	-0.024	-0.076	0.029	0.069	-0.044	0.07	0.033	0.062	-0.021	0.013	0.043	0.028		0.012	-0.014	
specialization_E-COMMERCE	0.023	-0.01	-0.022	0.024	ST .	-0.038	-0.015	o oas	-0.022	-0.036	-0.017	0.027	-0.01	-0.0068	0.022	-0.015	0.004	-0.021	-0.014	
specialization, Finance Management	0.00043	0.078	-0.066	-0.075	0.036		0.045	-0.11	0.068	0.11	0.052	0.062	0.032	-0.021	0.067	-0.015	0.19	0.016	-0.037	
specialization_HealPowers Management	0.014	0.02	-0.028	0.029	0.016	0.046	1,0	0.042	0.026	-0.042	0.02	-0.031	0.012	-0.008	-0.026	0.0033	0.087	0.022	0.0058	
agrecialization_Human Resource Management	-0.09	0.07	-0.062	0.069	o oas	-0.44	-0.1942	9	0.063	-0.1	-0.048	-0.076	-0.00	-0.019	-0.002	-0.000	0.10	0.038	0.016	
specialization_IT Projects Management	0.07	0.025	-0.039	-0.044	-0.022	-o.osa	-0.026	-0.063	296	0.064	-0.031	-0.048	-0.019	-0.012	0.039	0.00042	0.44	-0.036	-0.017	ı
specialization_Marketing Management	-0.027	0.088	-0.062	-0.07	-0.036	30031	-0.042	-0.1	0.064	11	-0.049	-0.077	-0.03	0.05	-0.065	-0.028	0.13	0.058	-D 039	
spenialization_Master and Advertising	0.0038	0.038	-0.08	0.033	0.017	-0.062	0.02	-0.048	0.031	0.049	*	-0.038	-0.014	-0.0003	-0.03	0.011	9310	-0.036	-0.018	
apecialization_Operations Management	0.0002	0.05	0.047	-0.052	-0.027	-0.002	0.001	0.076	0.048	0.077	0.030		-0.022	0.015	0.047	0.032	0.12	-0.014	-0.023	-
specialization_Rural and Agrituainess	0.027	0.025	-0.018	0.021	-0.01	-0.032	0.012	-0.03	-0.019	0.03	-0.014	-0.022	*	-0.0057	-0.016	0.0014	0.064	-0.021	-0.011	
specialization_Services Excellence	0.078	0.0038	0.012	-0.013	0.0068	-0.021	0.008	0.010	-0.012	0.02	0.0000	-0.018	0.0067	2.	0.012	0.017	0.042	0.0078	-0.0074	
specializátion_Bupply Chain Management	0.0051	0.052	-0.028	0.013	0.022	0.067	-0.026	-0.062	0.039	0.063	0.00	-0.047	-0.018	-0.019	*	0.03	0.10	-0.022	-0.017	ı
what_is_your_current_occupation_Student	0.04	-0.078	-0.013	-0.026	-0.015	-0.015	0.0033	0.000	0.00042	0.025	0.011	-0.032	0.0014	0.017	0.03	*	-0.031	-0.081	-0.045	
feed_cingln_Lending Page Submission	0.11	(0.20)		0.13	0.084		0.067	0.10	9.14				0.064	0.042	9.15	-0.031	10	-0.31	-0(10)	
head_origin_Lead Add Parm	0.038	-0.34	0.01	0.012	-0.021	0.016	0.055	0.038	0.030	0.029	0.000	0.014	0.021	0.0070	0.055	-0.081	-9.31	. 1	a as	ı
lead_source_Welingak Website	0.0019	-0.002	-0.024	0.014	-0.014	-0.007	-0.0058	-0.016	-0.017	-0.039	-0.018	-0.023	-0.011	-0.0074	-0.017	-0.045	-0.13	0.42	9)	
	B 10 00 00	dals firm spert on website	color Saint, Institut Ad human	peciation Bares Amendadon	SCHIMOL FOMBECE	specification France Management	specializator, Healthcan Management	ciatulos Huna Resuco Masgerent	(pesistrion (FPsjert) Management	spootcater Maketog Management	gestable Mels and Abertang	posizio Qualità l'Impret	specialistics. Fars and Agricumoss	specials Sease Ecolons	speciation Supp. Clan Massperer	etal, s. you cutest computer (Sobert	led_ovpr_Laming Page Submacon	me opp (see Adrian	had yours. Heliopa Wessie	

# FINAL MACHINE LEARNING MODEL FOR X SALES TEAM

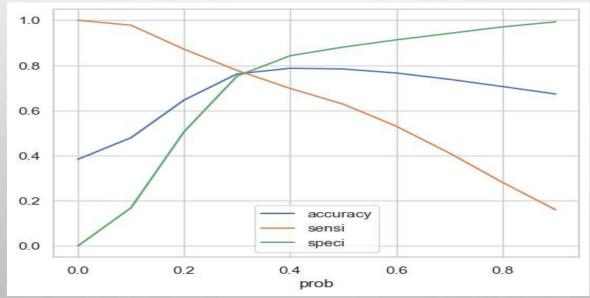
WITH A PROBABILITY THRESHOLD OF 0.30, WE CAN GET THE FOLLOWING SCORE WHICH IS SIMILAR TO THE TRAINING DATA WHEN TESTED ON THE TEST DATA. INDICATING THAT IT IS A RELIABLE MODEL.

TRAI	N DATA		TEST DATA								
Confusion matrics:	2982	983	Confusion matrics:	1261 437	′						
	554	1916		233 828							
Accuracy score:	76%		Accuracy score:	75.71%							
Sensitivity score:	77.57%		Sensitivity score:	78.03%							
Specificity score:	75.20%		Specificity score:	74.26%							
False Positive Rate:	24.79%										

# FINAL MACHINE LEARNING MODEL FOR X EDUCATION SALES TEAM







# **SUGGESTION**

# **General Suggestion**

- 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!