

Lead Score Assignment

- Jharna Rajani

Problem Statement:

X Education Company wants to find out the leads that they can contact who will take up the courses on their platform. They want to make calls in such a way that they have maximum people who will take the courses. These potential leads are called as Hot Leads.

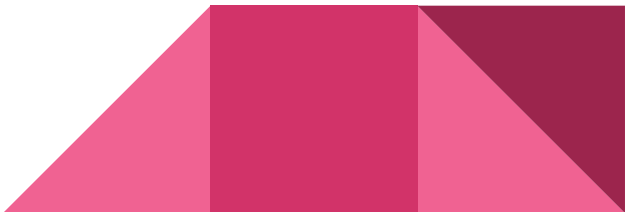


Data Understanding:

Data provided is a .csv files of all the leads that were contacted and if they were converted.

The `converted` column is our target variable.

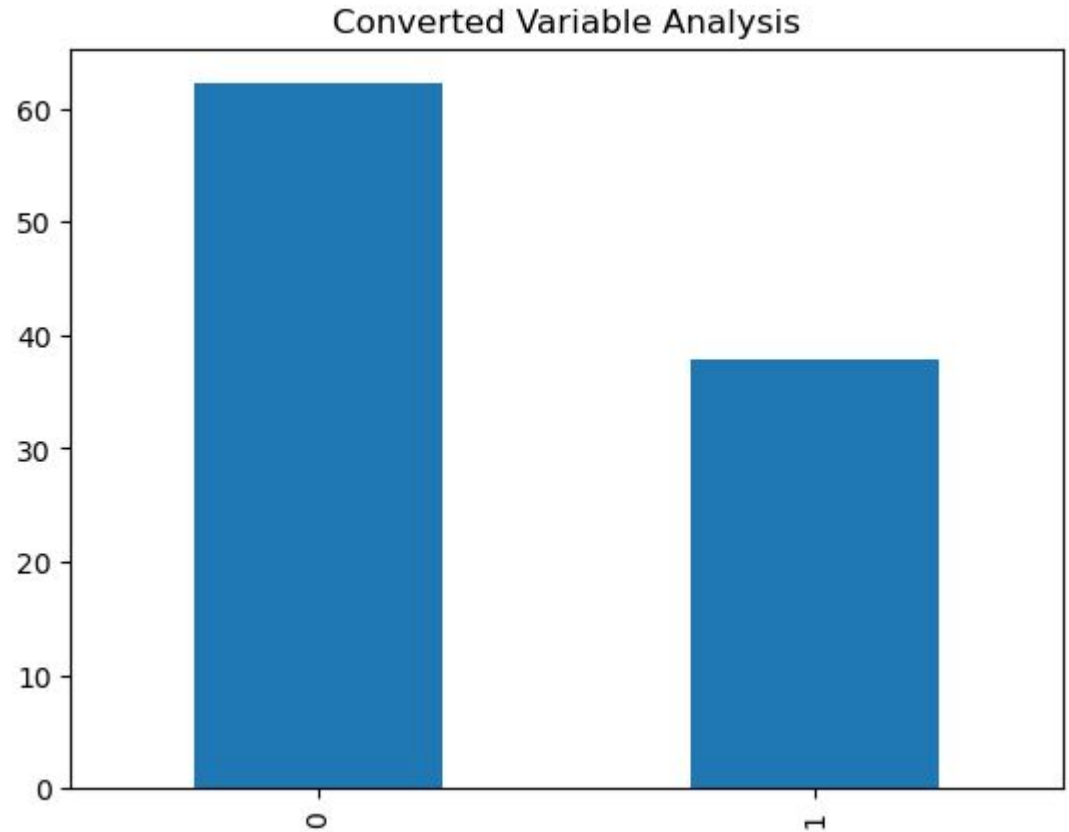
Process Followed:

- Data understanding and finding the target variable.
 - Handling missing information.
 - EDA
 - Based on the EDA, choosing the columns for model building
 - Diving the data into test and train data
 - Model Building
 - Model Evaluation
 - Making predictions
 - Findling the hot leads.
- 

Exploratory Data Analysis

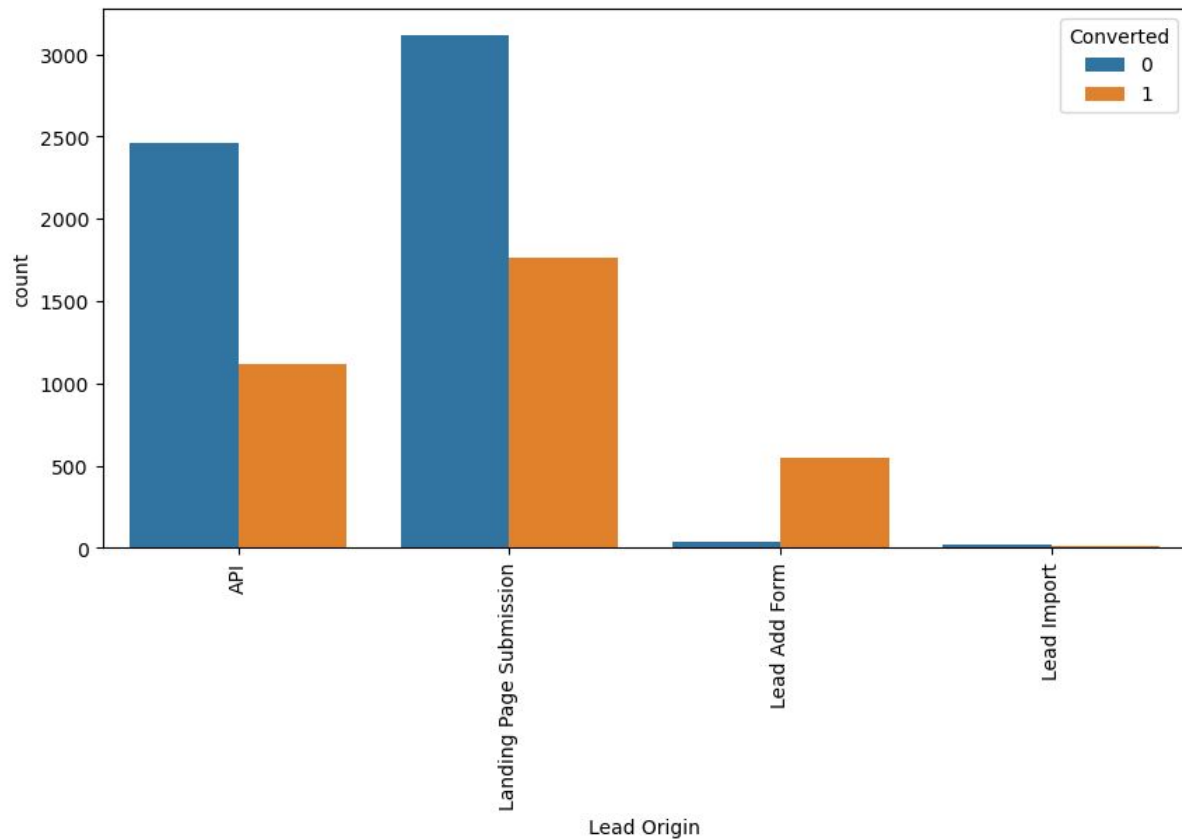
Converted Column (Target Column)

We have 38% lead conversion rate.



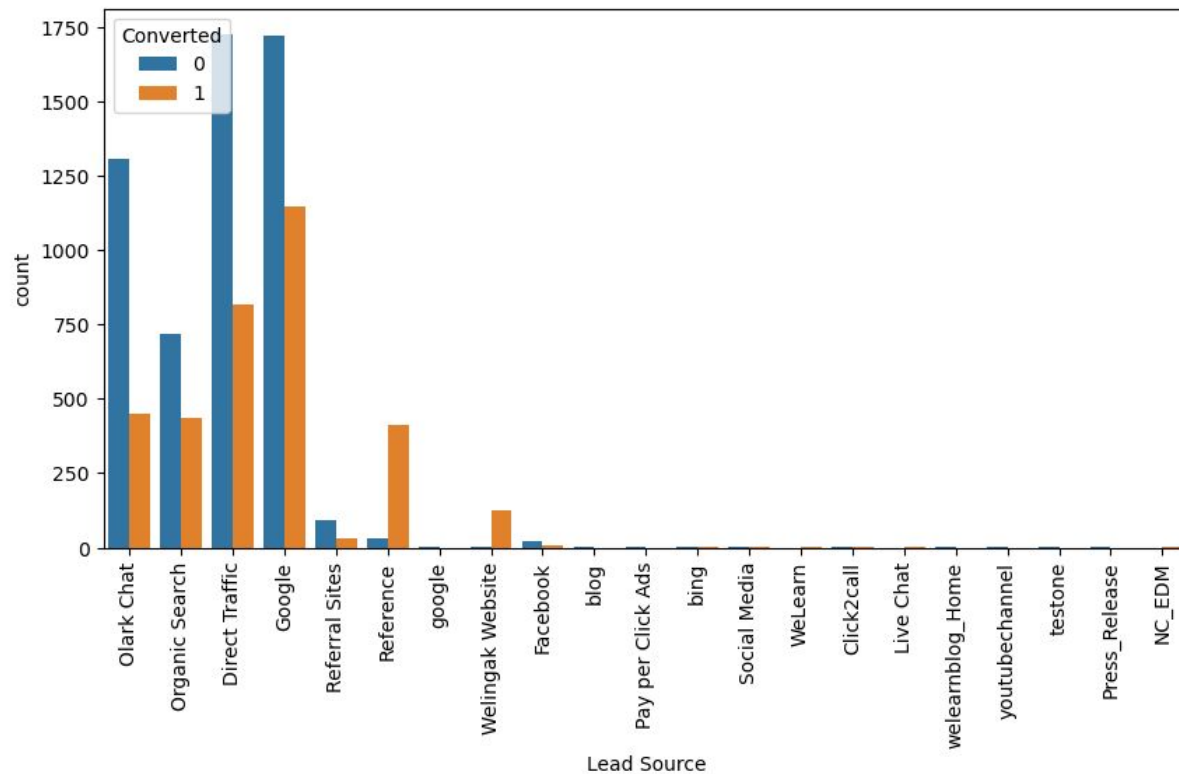
Lead Origin Column

Landing page Submission has the highest lead conversion rate followed by API.



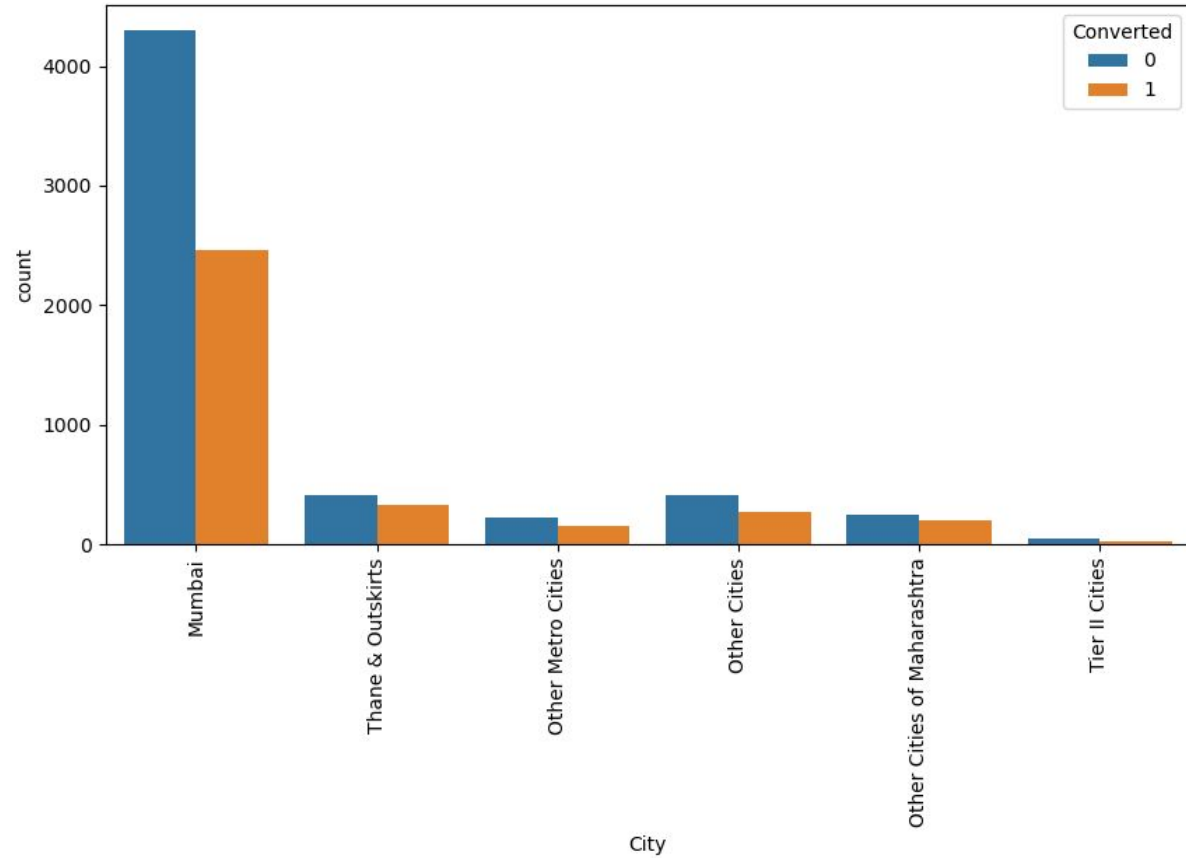
Lead Source Column

1. We notice that google is mentioned twice, so we need to change google to Google
2. A few sources can be aggregated under the name **advertisements** as they fall under that category



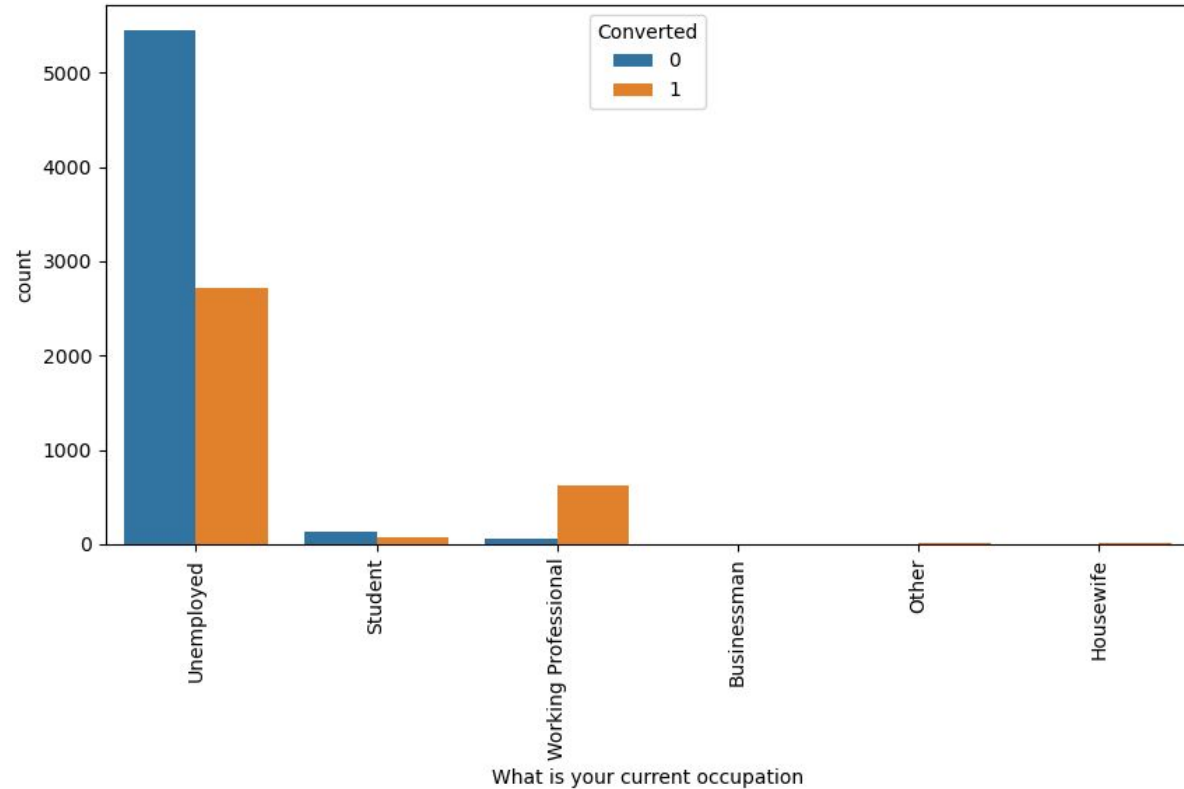
City Column

Mumbai has the maximum lead conversions.



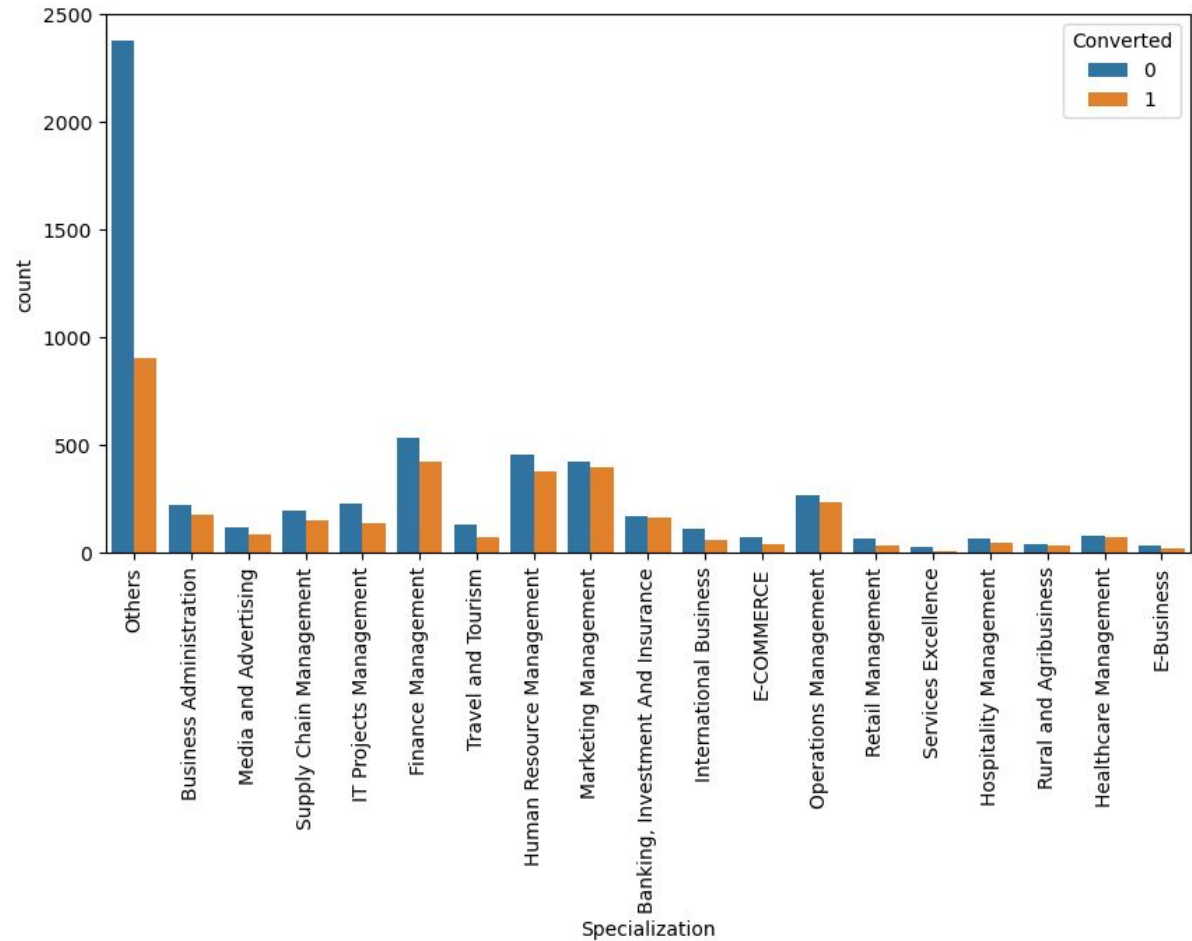
What is your current occupation column

1. Unemployed people are highly likely to become a lead
2. Working professionals are more likely to convert into lead



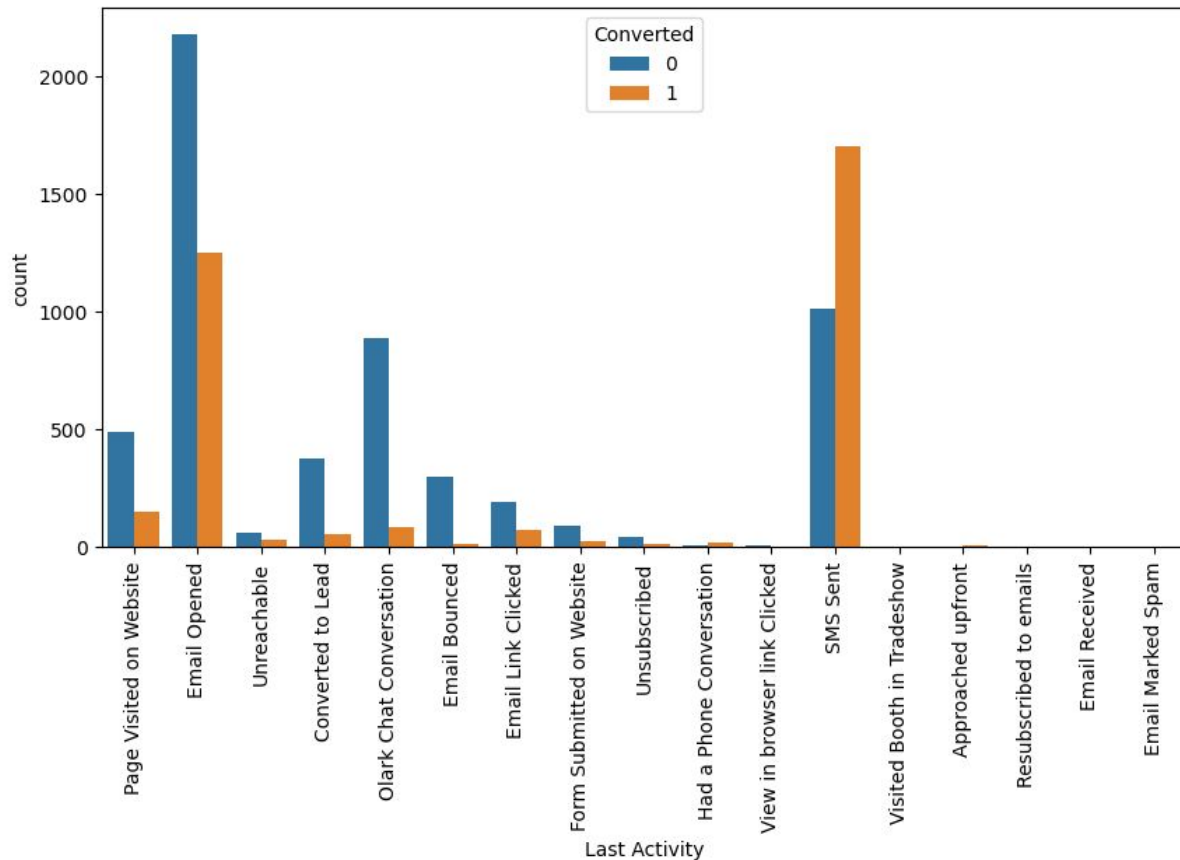
Specialization

Max Lead conversion is for leads with
`Other` specialization.



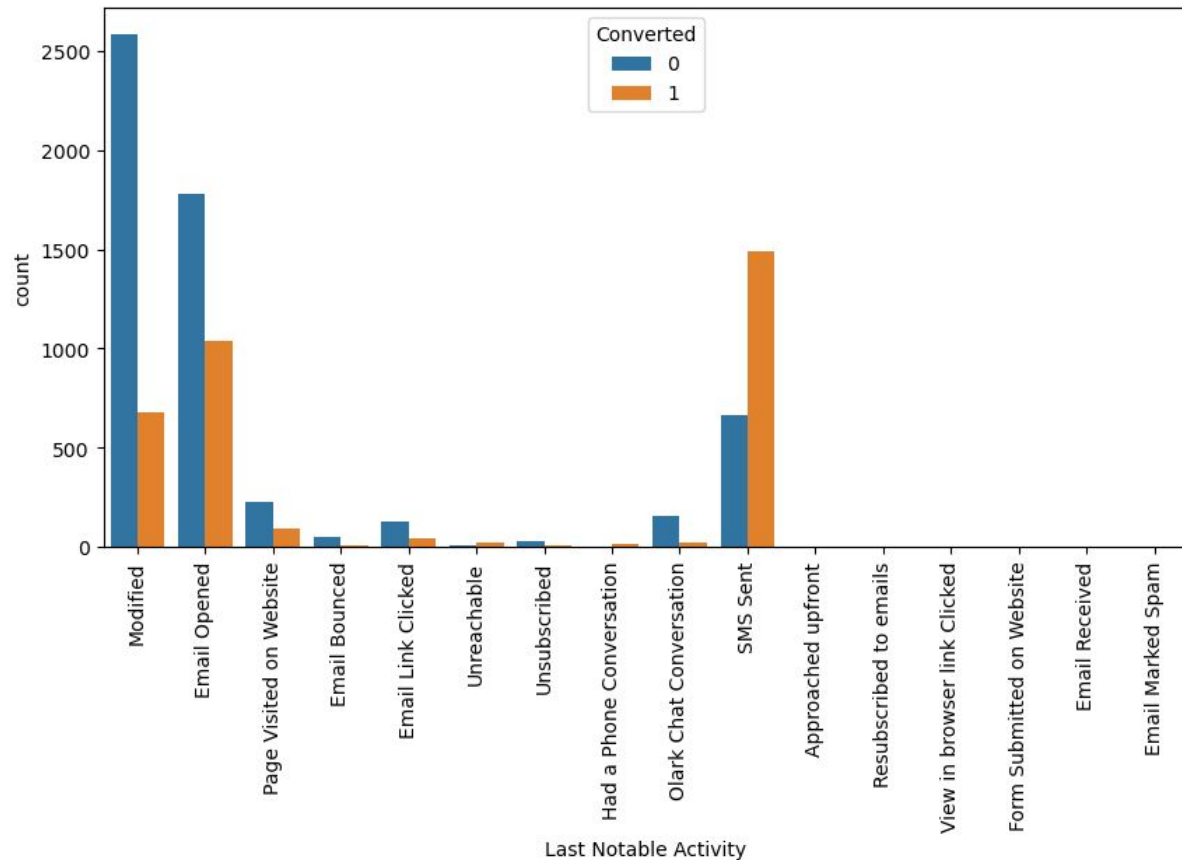
Last Activity Column

Max Lead conversion is for leads with SMS Sent as last activity



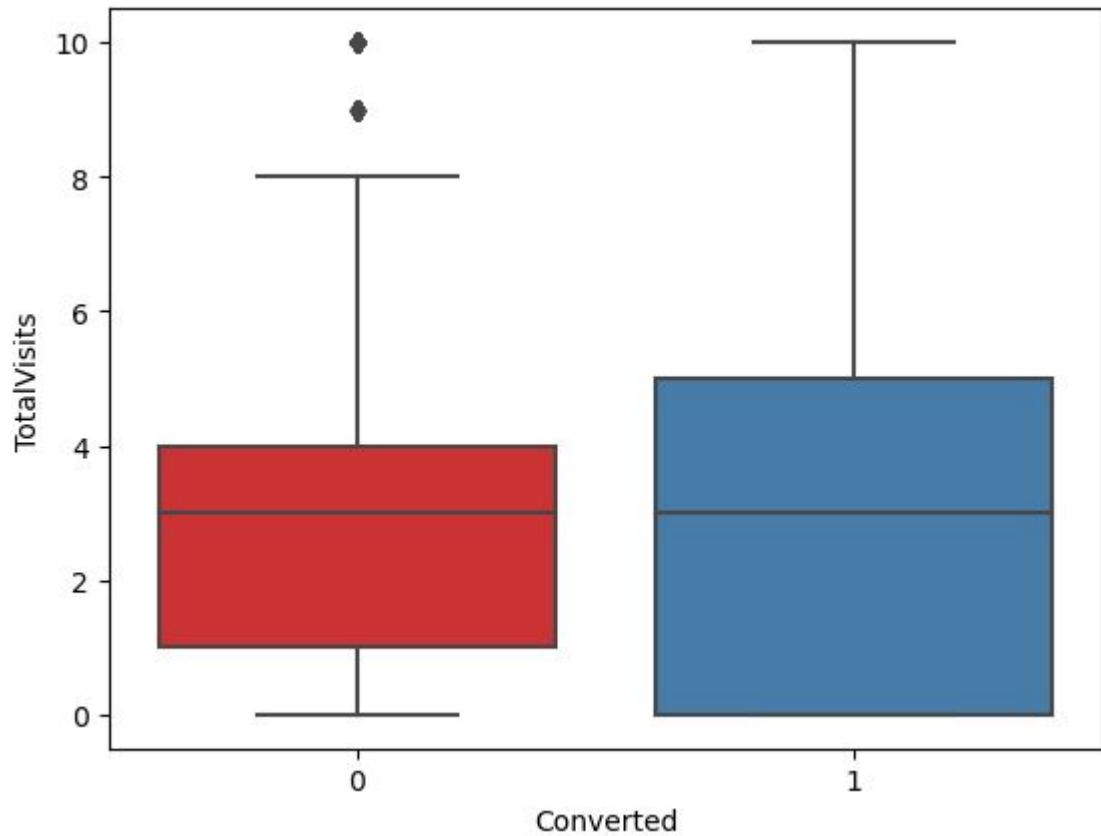
Last Notable Activity

Max Lead Conversion is for leads with last notable activity as SMS Sent.



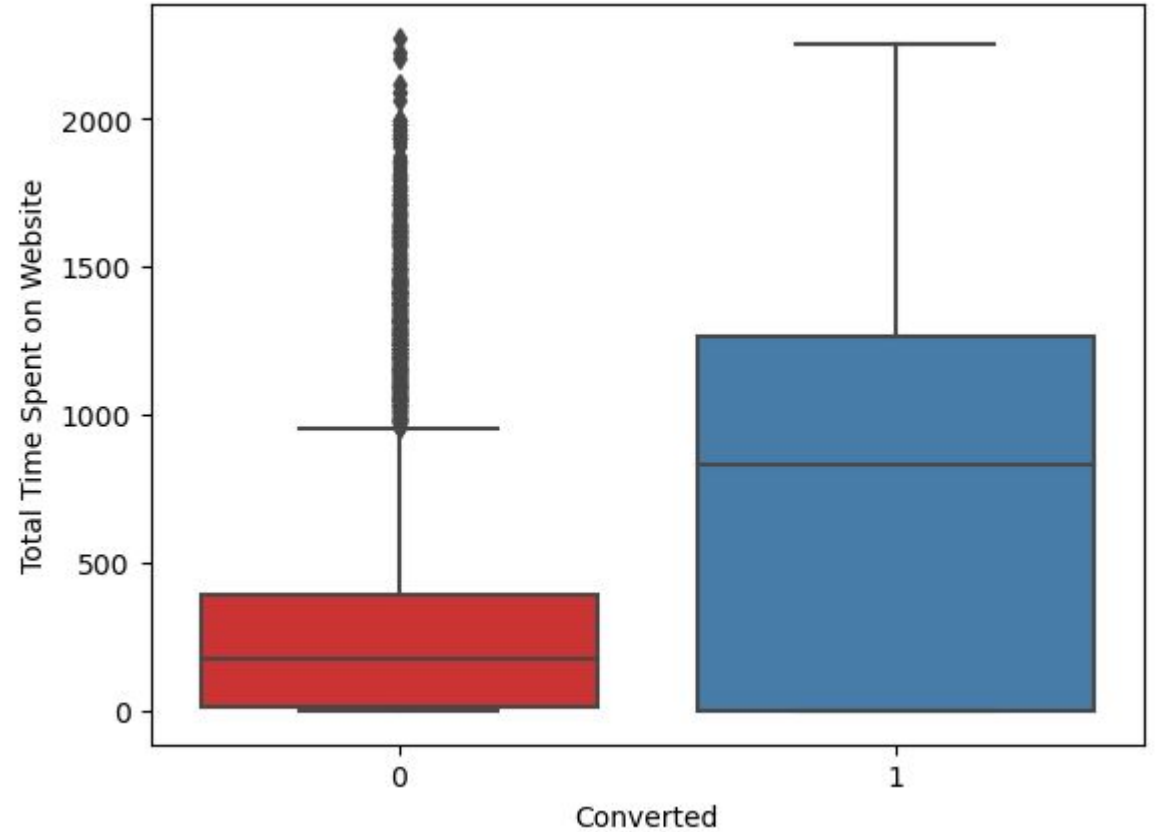
Total Visits Column

High chances of lead conversion if the total number of visits are 0-5



Total Time spent in website column

Leads spending more time on the website are highly likely to get converted





Model Building and Evaluation

VIF of Finalised Features

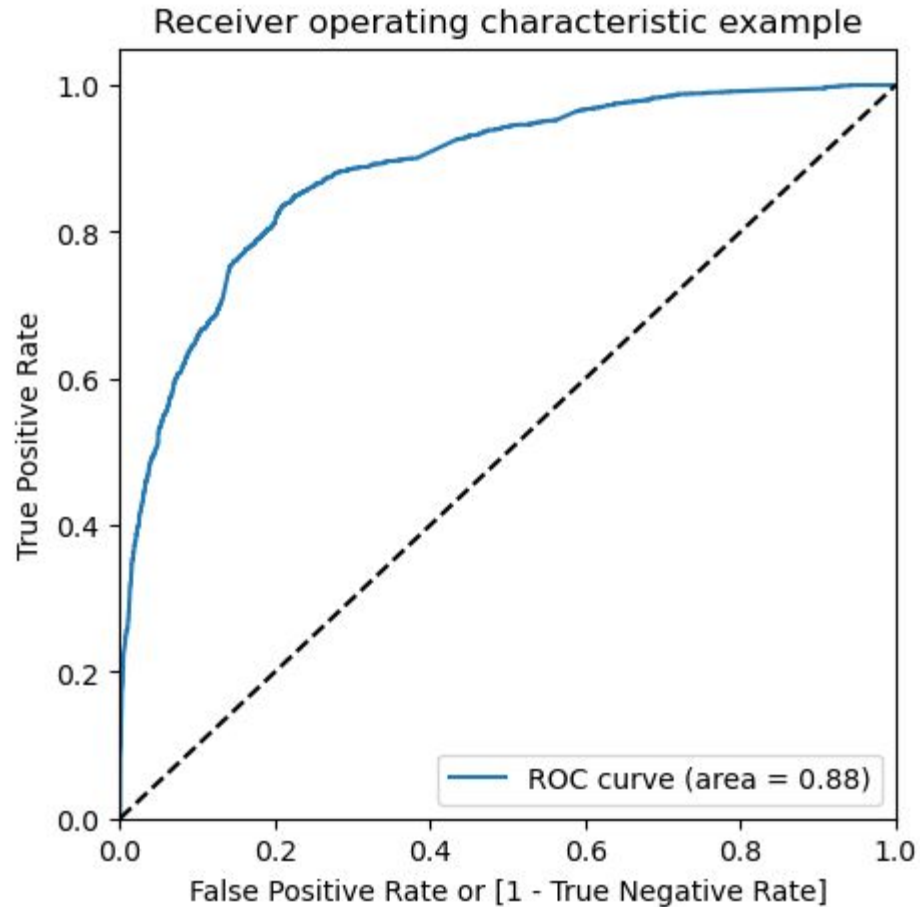
vif

]:

| | Features | VIF |
|----|---|------|
| 2 | Lead Origin_Landing Page Submission | 3.41 |
| 15 | Last Notable Activity_Modified | 2.71 |
| 4 | Lead Source_Direct Traffic | 2.14 |
| 11 | Specialization_Others | 2.14 |
| 10 | Last Activity_Olark Chat Conversation | 1.94 |
| 0 | Do Not Email | 1.86 |
| 14 | Last Notable Activity_Email Opened | 1.85 |
| 9 | Last Activity_Email Bounced | 1.76 |
| 3 | Lead Origin_Lead Add Form | 1.54 |
| 7 | Lead Source_Welingak Website | 1.37 |
| 16 | Last Notable Activity_Olark Chat Conversation | 1.37 |
| 5 | Lead Source_Organic Search | 1.29 |
| 8 | Last Activity_Converted to Lead | 1.24 |
| 1 | Total Time Spent on Website | 1.19 |
| 12 | What is your current occupation_Working Profes... | 1.16 |
| 17 | Last Notable Activity_Page Visited on Website | 1.10 |
| 13 | Last Notable Activity_Email Link Clicked | 1.06 |
| 6 | Lead Source_Referral Sites | 1.03 |

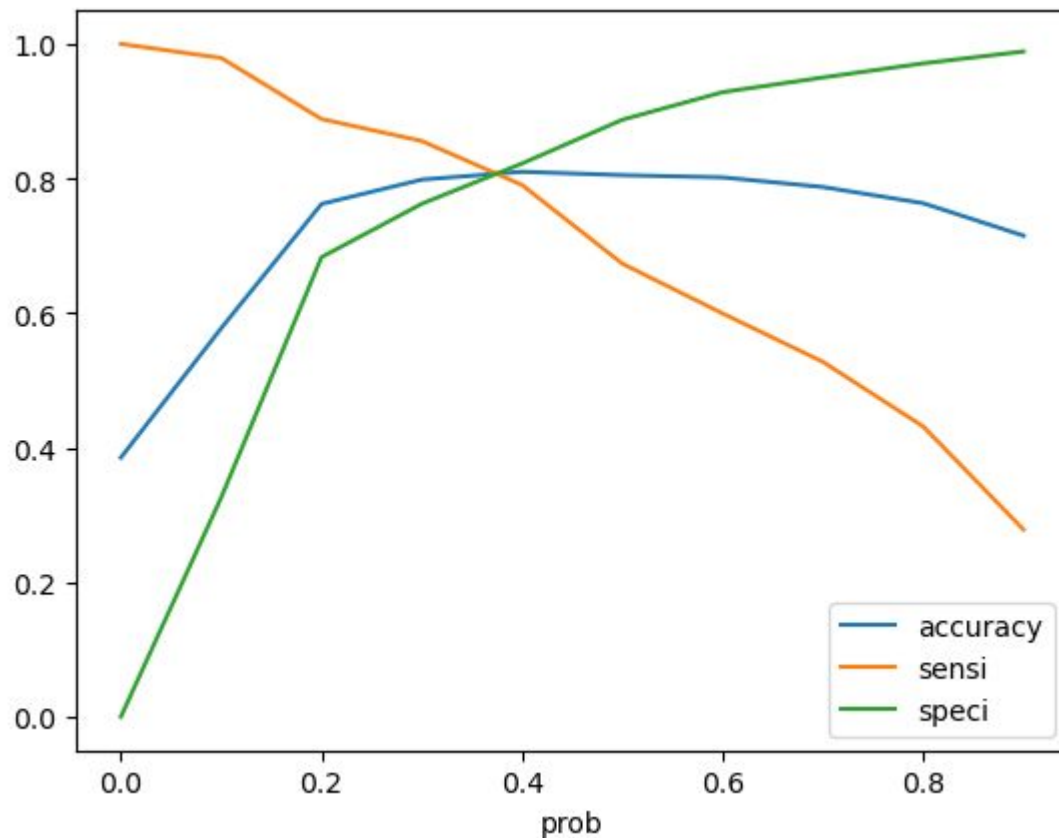
ROC Curve

We have 88% area under the curve.
Hence the model created is good.



Finding the best cutoff point

The best cutoff is 0.35 hence, shifting the cut off from 0.5 to 0.35



Comparing Train and Test Data

Compared the train and test data with the metrics and they are consistent.

Hence our model is good.

Results :

1) Comparing the values obtained for Train & Test:

Train Data:

- Accuracy : 80.7 %
- Sensitivity : 83 %
- Specificity : 79.3 %

Test Data:

- Accuracy : 80.2 %
- Sensitivity : 80.9 %
- Specificity : 79.8 %

ID's of prospective leads

The id's are of the prospective customers that have high chances of getting converted to leads.

The Prospect ID of the customers which should be contacted are :

```
: array([1490, 4216, 4223, 1946, 2461, 5822, 2684, 2010, 4062, 7696, 9049,
        1518, 4543, 4830, 4365, 3542, 7347, 2504, 7674, 8596, 4003, 4963,
        6947, 4807, 446, 789, 5805, 4478, 3758, 1561, 737, 9034, 6423,
        8286, 7174, 4461, 1436, 7552, 3932, 4080, 1475, 5785, 1867, 2860,
        7253, 4297, 5490, 1995, 6532, 4498, 5797, 8687, 7653, 4149, 76,
        6743, 3703, 3307, 3976, 5769, 1051, 1663, 3288, 8959, 7521, 8282,
        8213, 9063, 5292, 6913, 6015, 1481, 785, 3265, 3285, 7433, 2460,
        3858, 3810, 2009, 8106, 373, 3055, 7417, 4179, 8076, 8568, 7268,
        6784, 6754, 2960, 7753, 3983, 802, 8745, 4455, 4717, 505, 6094,
        4992, 7036, 2680, 7065, 112, 6149, 7175, 1675, 6999, 5932, 5826,
        8492, 6499, 2481, 3439, 4612, 7129, 4793, 1557, 4837, 2495, 822,
        2378, 5075, 7699, 5638, 2342, 8077, 2727, 720, 7489, 2961, 1542,
        5656, 2630, 8205, 6332, 8461, 2427, 5087, 174, 2674, 8065, 2095,
        1568, 8597, 4865, 4708, 6066, 6538, 5700, 1388, 5815, 7970, 7902,
        5804, 7805, 5042, 4081, 6684, 5440, 1927, 5824, 64, 2650, 5808,
        8610, 4578, 4803, 1470, 5810, 2473, 350, 2584, 7259, 3727, 3150,
        2118, 4403, 3194, 8475, 1200, 2575, 1299, 1525, 2166, 2913, 4613,
        4909, 8204, 4772, 1374, 8888, 8082, 4862, 1595, 1899, 8474, 3463,
        2022, 7893, 3248, 6486, 2281, 1729, 8620, 1190, 2486, 2158, 3355,
        5353, 2994, 4559, 8521, 7168, 4677, 7537, 493, 1563, 4860, 9076,
        8348, 2153, 1550, 5389, 5579, 1783, 2105, 1578, 6729, 1263, 2011,
        4330, 6252, 1820, 6760, 3015, 2285, 7091, 2598, 7962, 7018, 6290,
        5061, 356, 8271, 4285, 8540, 2854, 8375, 4505, 1979, 3532, 1444,
        4934, 8804, 1416, 7334, 2652, 7057, 5525, 2560, 3085, 7445, 3396,
        9062, 2943, 7690, 8198, 4233, 7750, 353, 4764, 8088, 7193, 7978,
        8928, 6567, 6685, 4378, 5455, 5363, 2354, 2714, 718, 2559, 5000,
        2664, 6040, 4068, 3570, 9043, 8090, 2483, 3762, 3717, 4112, 1407,
        6740, 6892, 5175, 662, 8452, 3207, 8505, 2988, 6175, 5561, 5633,
        8415, 3660, 2783, 3770, 220, 6994, 764, 4253, 3723, 6725, 746,
        8592, 3496, 5502, 4241, 6933, 4388, 7021, 3097, 3836, 4116, 6314,
        8322, 3165, 6723, 3817, 1360, 7053, 6944, 4671, 5877, 2673, 3146,
        2343, 5286, 745, 1950, 4382, 1682, 7240, 6375, 7941, 7285, 5293,
        3736, 7450, 2617, 6127, 6733, 4371, 1026, 8113, 6242, 1089, 2243,
        2841, 7136, 3477, 2763, 6890, 4734, 7823, 2870, 5337, 4879, 1467,
        3942, 8343, 8052, 1566, 6461, 5741, 6299])
```

Recommendations:

- The company **should make calls** to Working professionals
- The company **should make calls** where the lead source is Welingak Website
- The company **should make calls** where the lead origin is Lead Add Form
- The company **should make calls** to leads that have spent time on the website

The company should NOT make calls to:

- Leads that have lead source as direct traffic, organic search or referral sites
- Leads whose Last activity is converted to lead, Olark Chat Conversation, Email Bounced
- Leads whose lead origin is Landing page submission
- Leads whose last notable activity is Email opened, Olark Chat Conversation, Page Visited on Website, Email link clicked or modified
- Leads with Do Not Email