



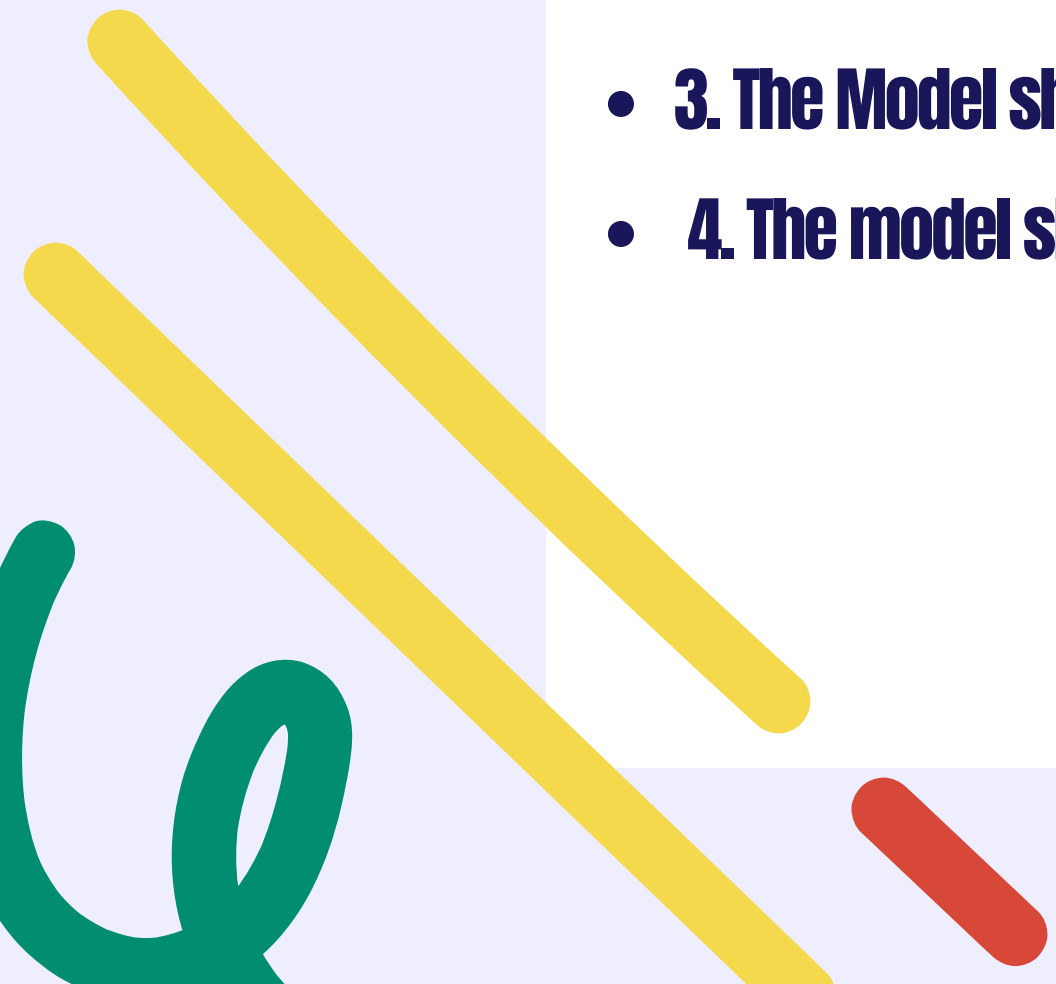
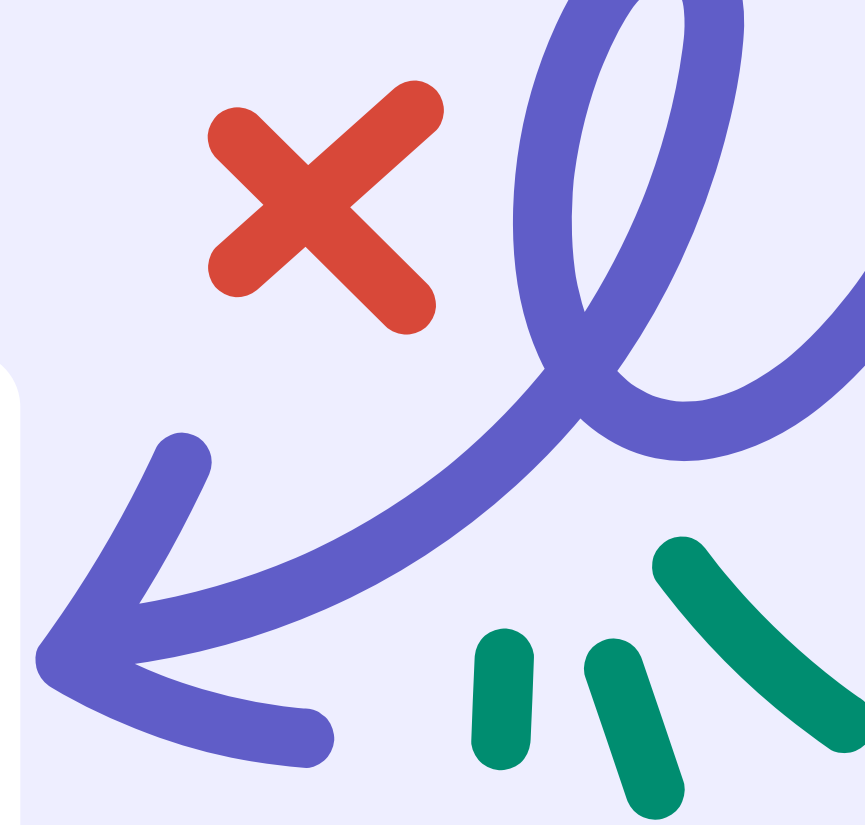
# LEAD SCORE CASE STUDY

Done by: S.Sai puja tanmayi



# BUSINESS OBJECTIVES:

- **1. How to select the most promising leads, i.e leads that high chance to convert into customers.**
- **2. Identify customer with higher lead score, high conversion chance and lower score have lower chance of conversion.**
- **3. The Model should identify hot leads.**
- **4. The model should adapt to future usage as well.**



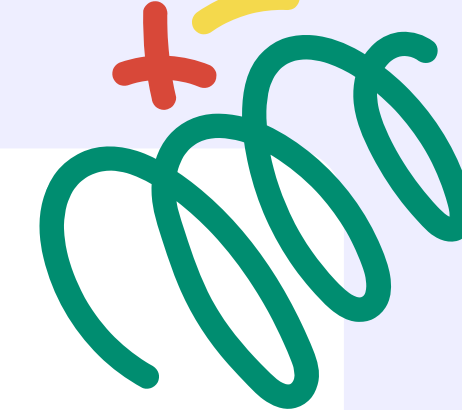
## **Approach:**

Step-1: Loading and Understanding the data: The data has 9240 rows and 37 columns in the beginning.

Step-2: Cleaning the data: After cleaning the data we are left with 9240 rows and 21 columns.

(we dropped columns that have so many null values which can effect analysis). cleaning involves converting, imputing null values, dropping unnecessary columns.

Step-3:Exploratory Data Analysis: It helps us to understand the trends in the data by visualizing it . We can clearly see the difference or relation in the variables in this step. Below are the graphs or plots of the analysis that I have performed. It will help us to identify the required and unrequired parts quickly .

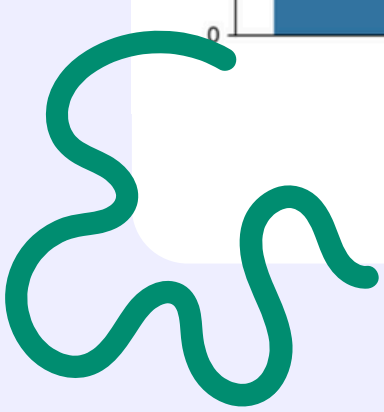
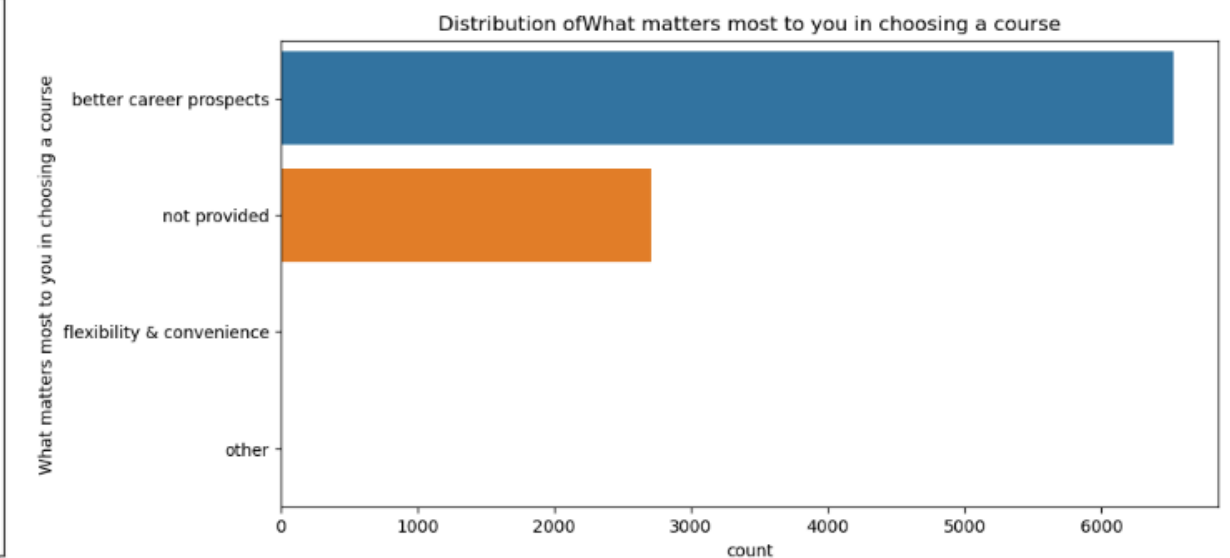
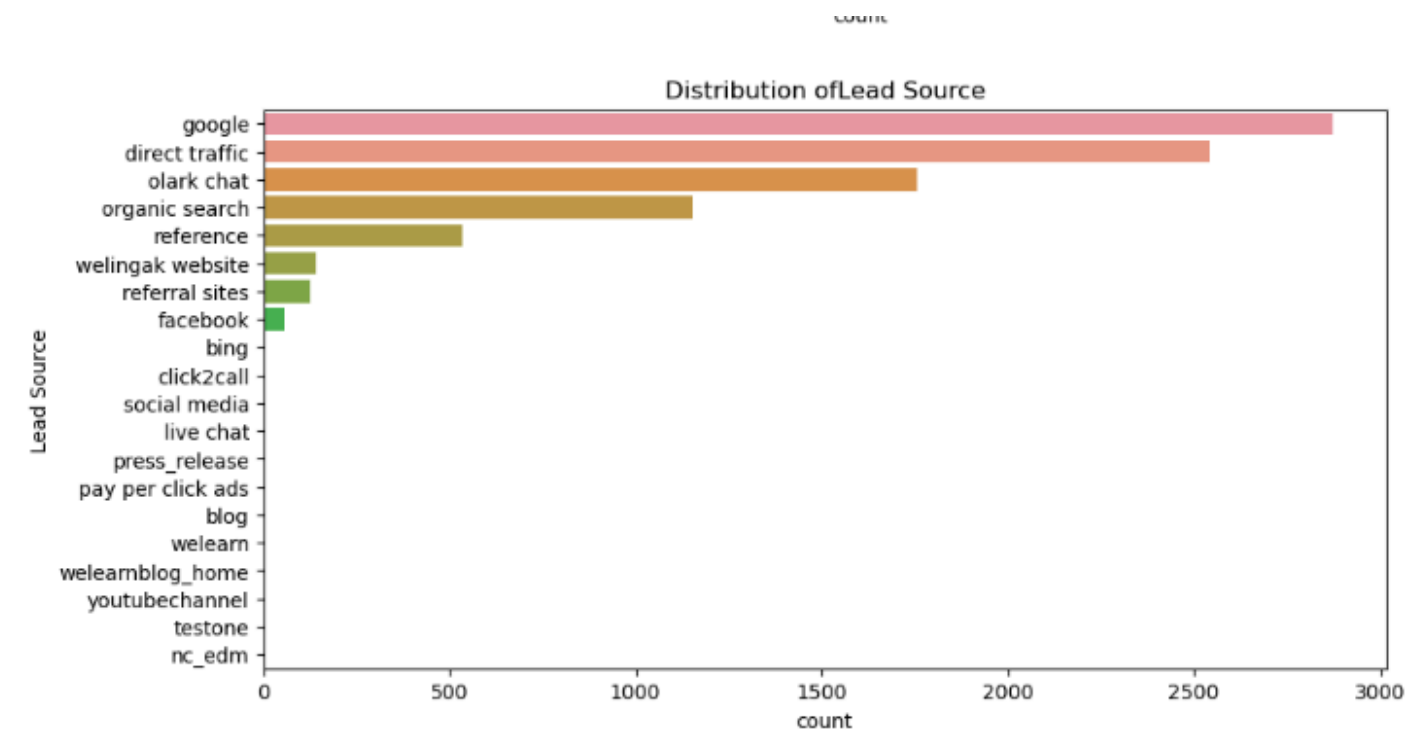
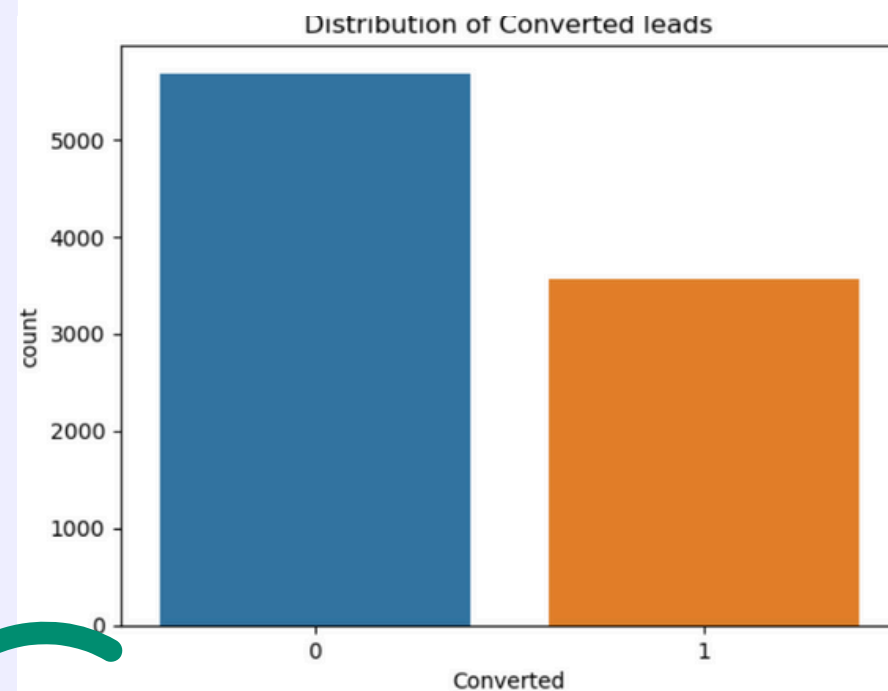
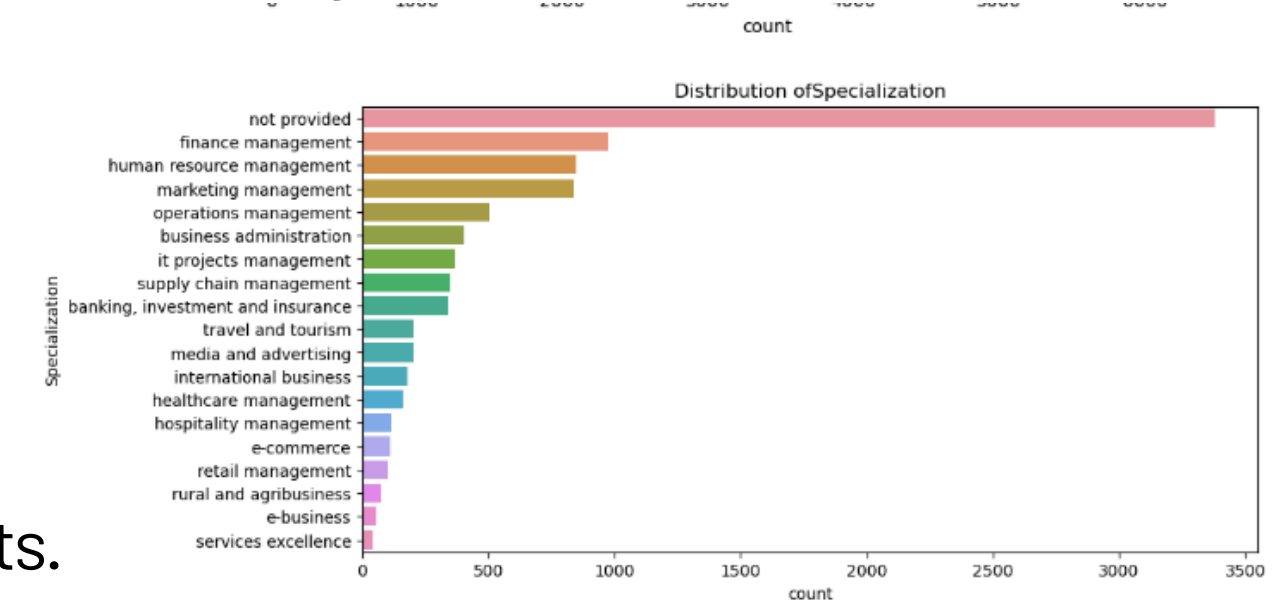


The graph helps us to see the difference much clearly. As we can see in the Converted graph

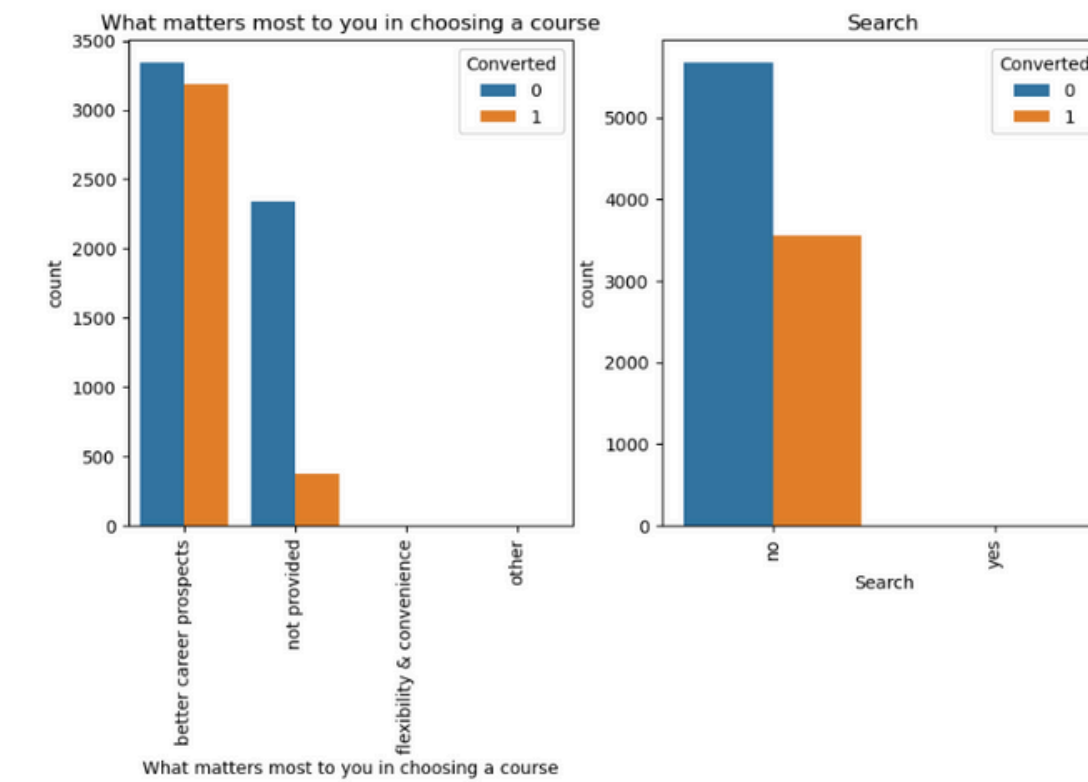
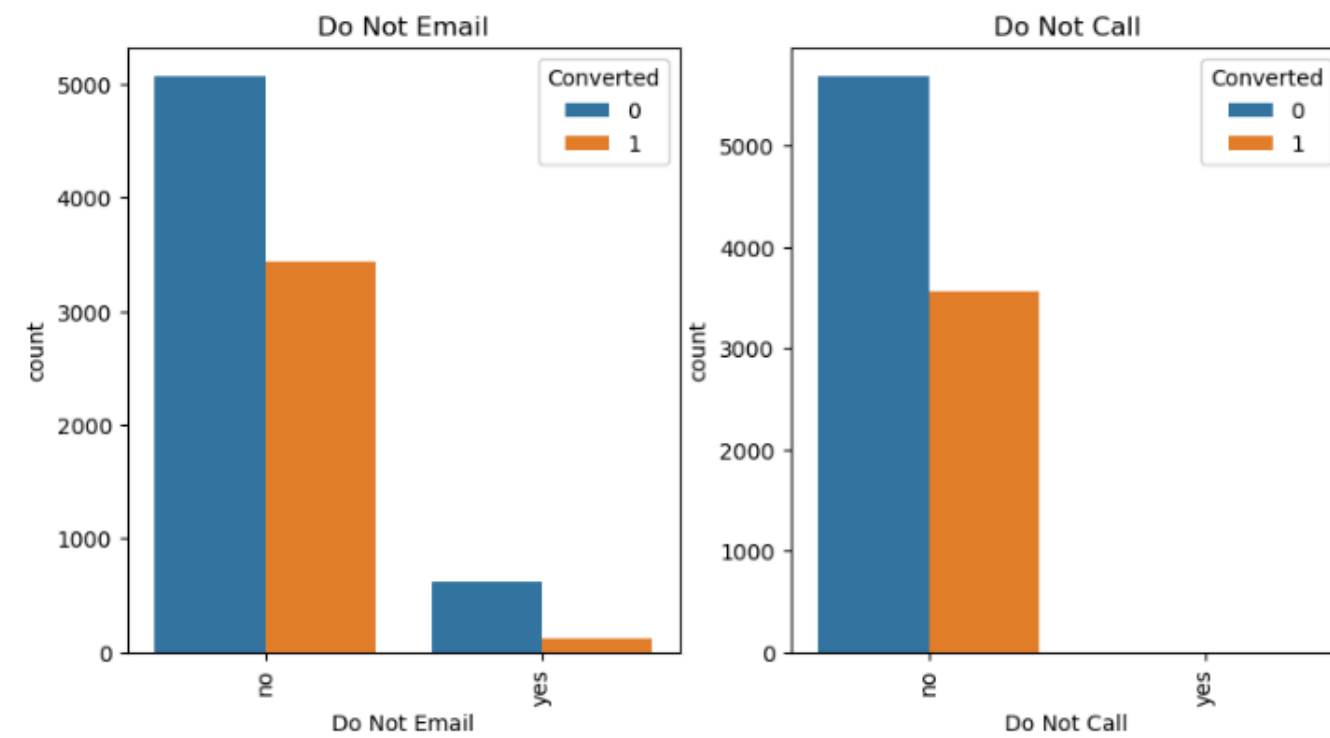
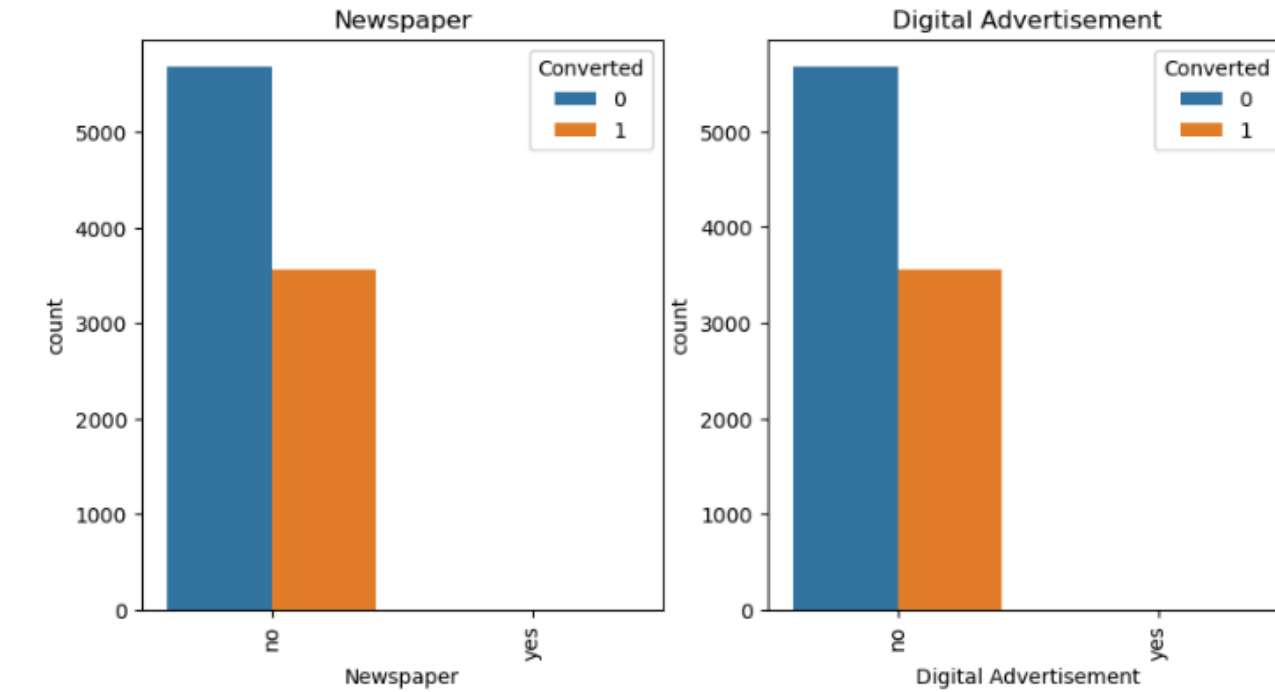
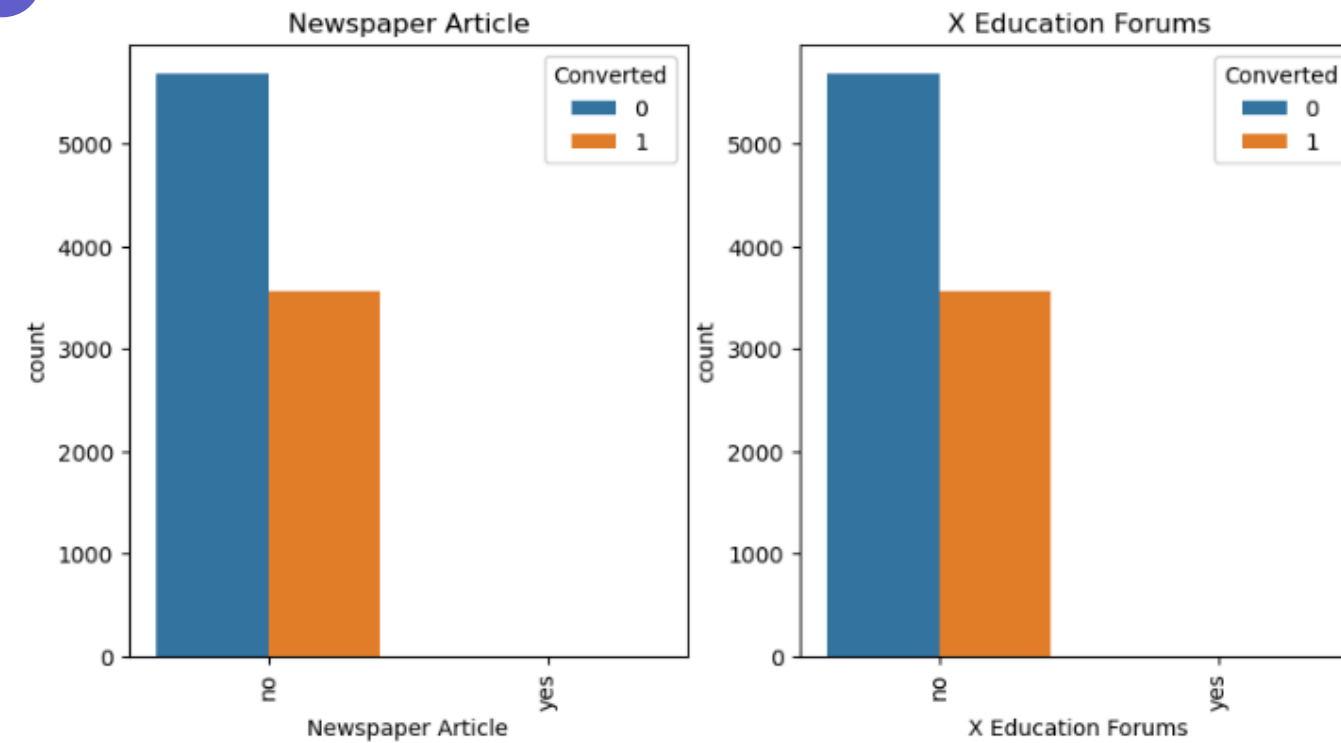
We can see the distribution of lead source to find where most leads are coming from.

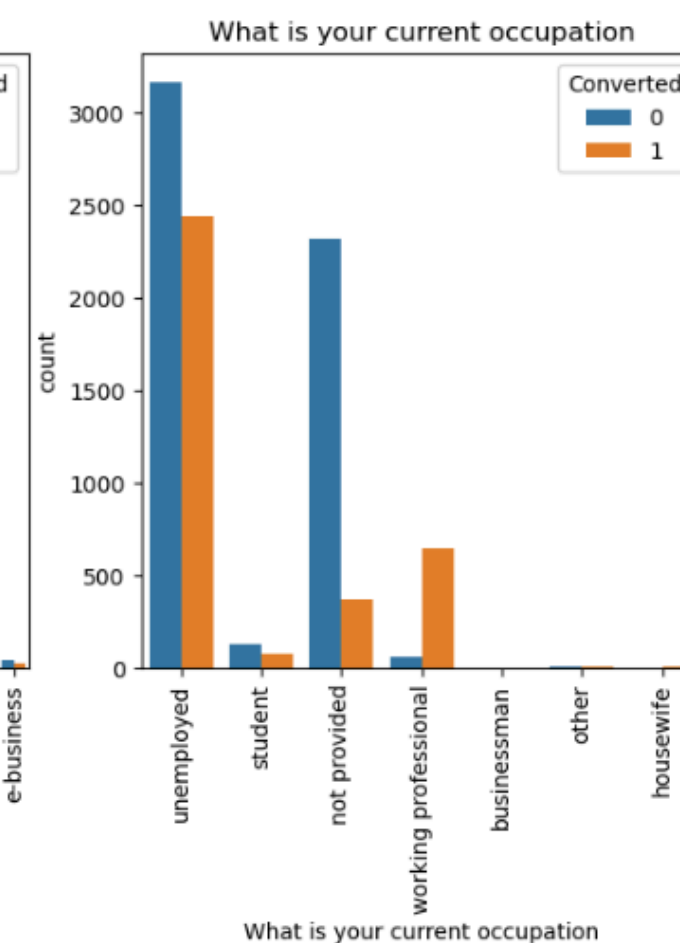
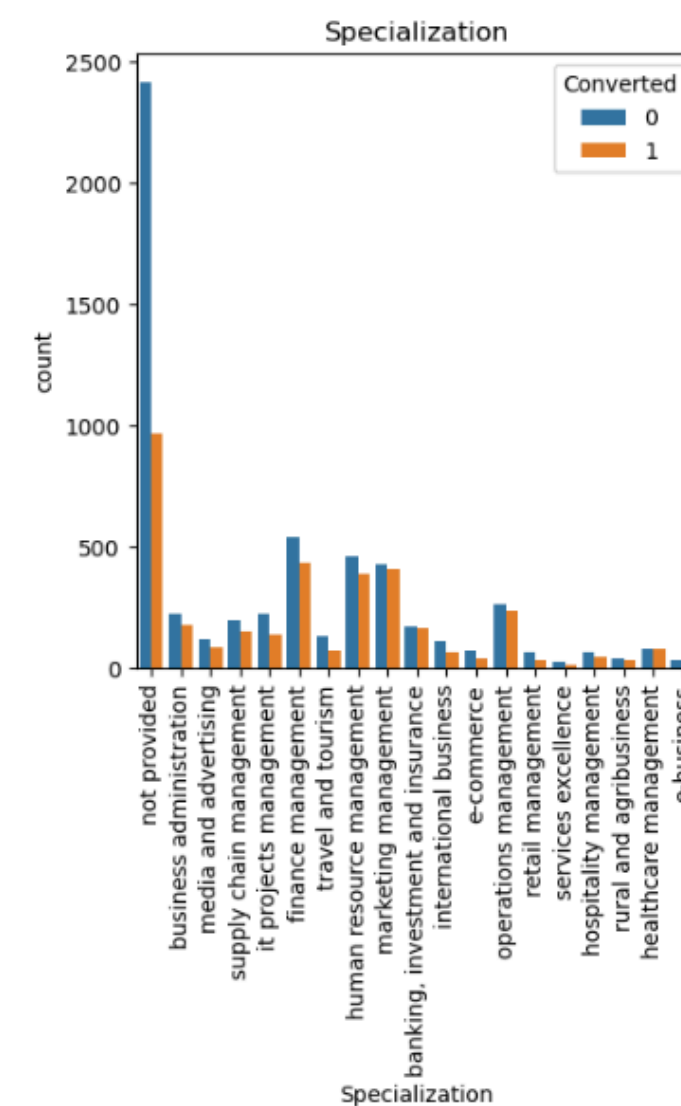
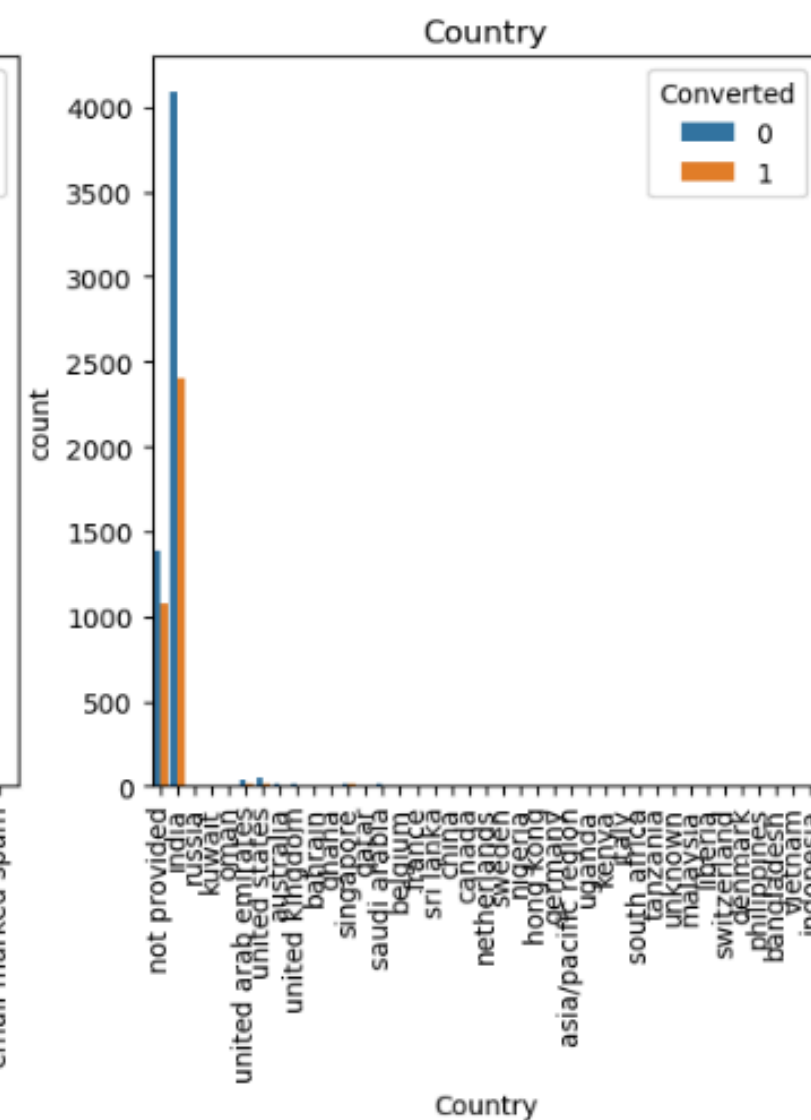
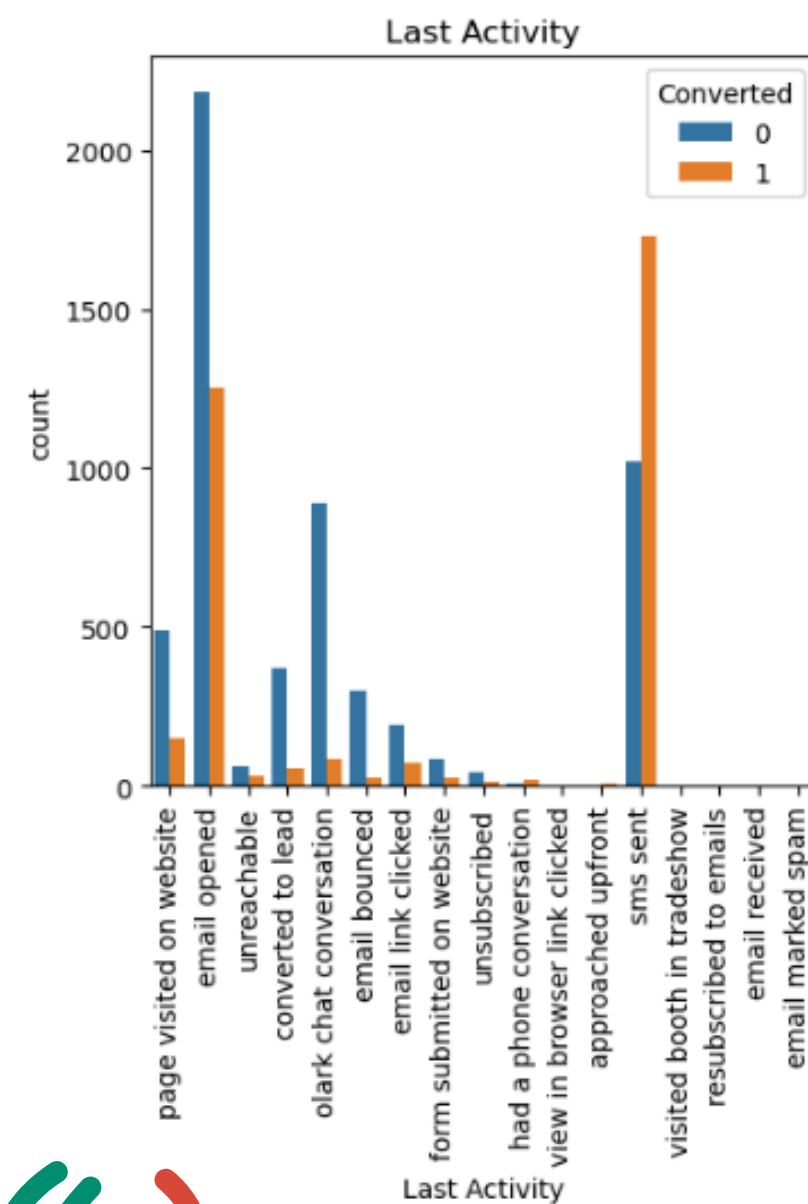
And Distribution of Specialization to find which background leads come from here, financial management.

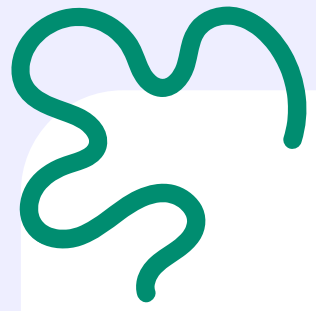
And distribution of why most leads choose the course , for better prospects.



# Exploratory data analysis (EDA):

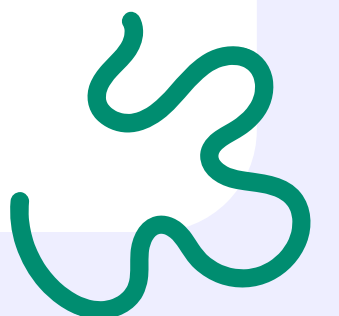






## Observations from EDA:

- Lead source is mostly from Google followed by direct traffic.
- Most people prefer to not receive mails.
- Everyone prefer not to receive calls either.
- people's last activity is email opened .
- most of the leads are from India.
- Most people prefer to not give specialization but if you desperately want to know followed by financial management
- Most leads are unemployed.
- Most leads choose this course for better career prospects.
- Not many leads saw the AD in newspaper, newspaper article, X Education forumns, Digital advertisement.
- Some leads got to know about them by distribution of free copy of mastering the interview.
- Leads last notable activity was modified.



Step-4: Dummy Variables: If there are many types in single columns like lead source: Google, direct, etc. We separate them using dummies. It helps for easier analysis.

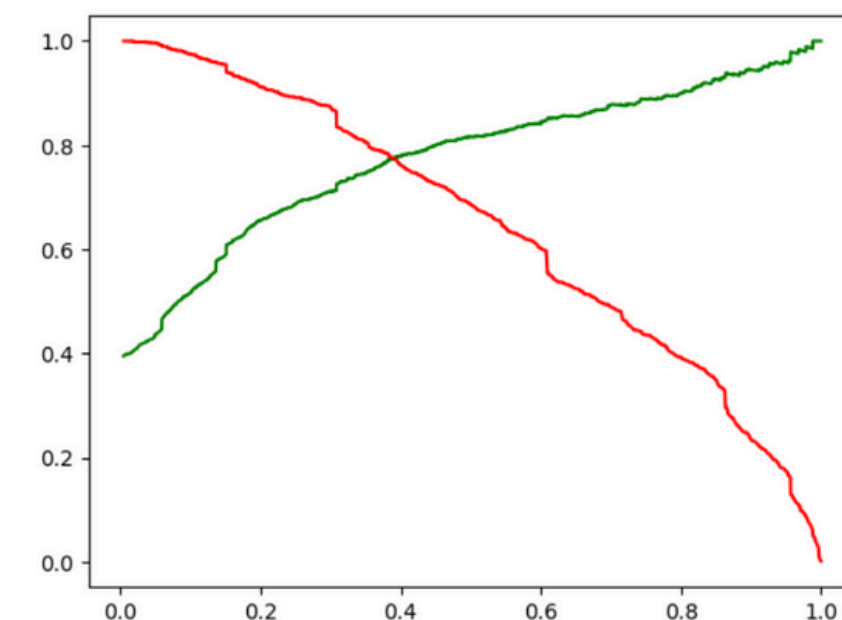
Step-5: Test-Train split and Building model: Here, We use a part of data to train the model. Then we use the rest of the part for evaluating the model.

Step-6: Model Evaluation: We use several methods to check the efficiency of the model. Here, we can also identify important variables and keep only them to increase the accuracy of the model.

Step-7: Making prediction: Here, we ask the model to make prediction to see how well it will align with the actual data to test the model further.

The graph will tell you how the final prediction is aligning with the actual data.

If you ask me it works pretty well. I'm not saying this because I made it though.







# OBSERVATIONS:



## **The Observations are made from the model:**

1. accuracy of train set is 79.6%, test set is 75.9% difference is around 0.4
2. Sensitivity train set is 78.6%, test set is 78.9% increased by 0.3%
3. Specificity(TN) of train: 82.2%, test: 83.6%
4. False Positive(FP) of train: 17.7%, test: 16.3%
5. True negative(TN) of train: 86.2%, test: 85.9% decreased by 0.3%
6. precision of train: 79.6%, test: 75.9%, decreased by 3.7%
7. recall: train: 69.8%, test: 78.9%, increased by 9.1%

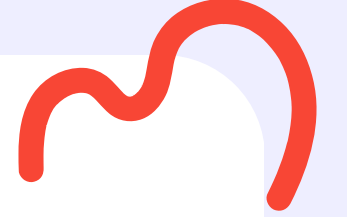
Overall the high recall and precision values indicates that it identifies positive cases while maintaining good number false positives. High specificity and low false positive rates further validate the model's reliability. Overall its a good model



## Conclusion:

- Important variables that contribute to the increase in probability:
  - 1.Total time spent on Website
  - 2.Total Visits
  - 3.lead source with google
- 3 top categories/ dummy variables :
  - 1)Lead source with Google
  - 2)Lead source with direct trafic
  - 3)Lead source with organic search
- Call Leads if:
  - 1)If they spend a lot of time on website. We can see people who spend a lot of time on website are mostly converted successfully





2) If Repeatedly comes back to website. There is a high possibility of them converting into successful leads.

3) If their last activity is through E-mail or SMS or Olark chat conversation.

4) If they are Working professionals or unemployed.

- Since most people don't like to receive calls as well there are other ways to reach. We can send them E-mails, as most of their notable last activity is opening e-mails. And sending SMS is also a good option. There is a high chance these strategies successfully working.

