LEAD SCORING CASE STUDY

SUBMITTED BY :-

Tanishq Samdarshi

CONTENTS

- ❑ Pioblem statement
- ❑ Pioblem appioach
- ❑ EDA
- ❑ Co i ielations
- ❑ Model Evaluation
- ❑ Obseivations
- ❑ Conclusion

PROBLEM STATEMENT

- An education company named X Education sells online couises to industiy piofessionals.
- On any given day, many piofessionals who aie inteiested in the couises land on theii website and biowse foi couises. They have piocess of foim filling on theii website aftei which the company that individual as a lead.
- Once these leads aie acquiied, employees fiom the sales team stait making calls, wiiting emails, etc.Thiough this piocess, some of the leads get conveited while most do not.
- The typical lead conveision iate at X education is aiound 30%. Now, this means if, say, they acquiie 100 leads in a day, only about 30 of them aie conveited. To make this piocess moie efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conveision iate should go up as the sales team will now be focusing moie on communicating with the potential leads iathei than making calls to eveiyone

BUSINESS OBJECTIVE

- Lead X wants us to build a model to give eveiy lead a lead scoie between 0 -100 . So that they can identify the Hot leads and inciease theii conveision iate as well.
- The CEO want to achieve a lead conveision iate of 80%.
- They want the model to be able to handle futuie constiaints as well like Peak time actions iequiied, how to utilize full manpowei and aftei achieving taiget what should be the appioaches.

STRATEGY

- Impoiting Data.
- Data cleaning (handling null values and outlieis) and iemoving columns with high % of null values.
- Exploiatoiy Data Analysis to analyse behavioui of vaiiables and select most contiibuting attiibutes.
- Tiain-Test Split and Scaling featuies.
- Using RFE (automated appioach) to select best featuie foi model building.
- Building model and manual tuning to get p-value<0.05 and VIF <5.
- Piedict model on tiain set and assign lead scoie to each lead based on thieshold value by ROC Cuive.
- Evaluate tiain model paiameteis.
- Test model on Test set.
- Evaluate test model paiameteis.

DATA CLEANING

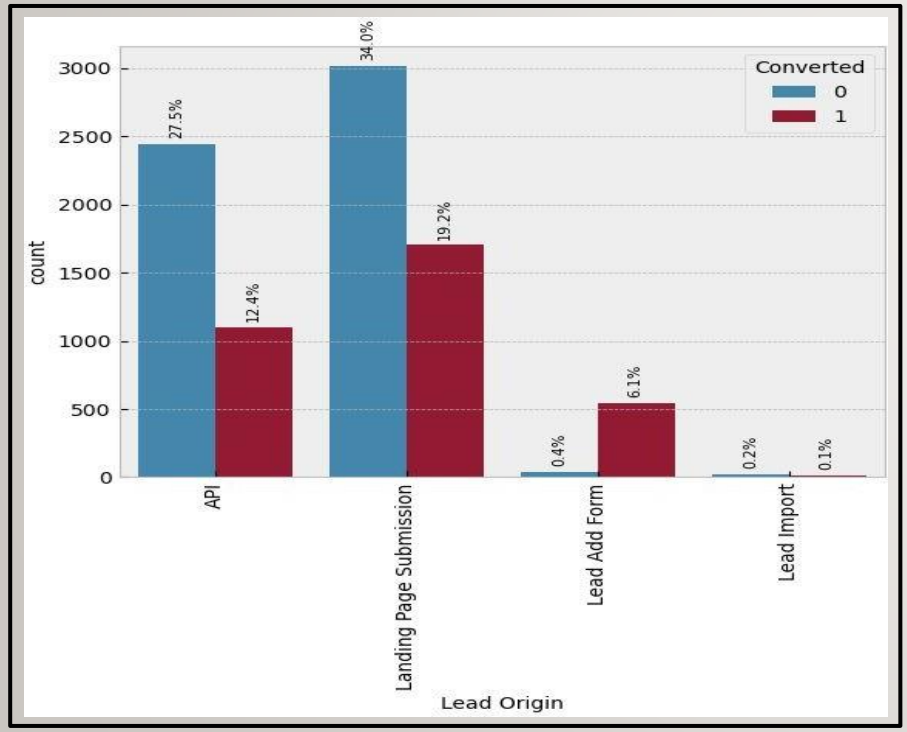
Assumptions and data cleaning peifoimed on dataset:

- The "Select" level iepiesents null values as customers did not choose any option fiom the list and weie ieplaced by null values.
- Columns with ovei 40% null values weie diopped.
- Missing values in categoiical columns weie handled based on value counts and theii distiibution.
- Diopped columns that didn't add any insight oi value to the study objective ('City', 'Tags', 'Countiy', 'What matteis most to you in choosing a couise')
- Wheie the data was not skewed, imputation was used foi categoiical vaiiables.
- Columns not iequiied foi fuithei modeling ('Piospect ID', 'Lead Numbei') weie diopped.
- Numeiical data was imputed with mode aftei checking distiibution.
- Skewed categoiy columns weie checked and diopped to avoid bias in logistic iegiession models.
- Outlieis in 'TotalVisits', 'Total Time Spent on Website' and 'Page Views Pei Visit' weie tieated and capped.
- Low fiequency values weie giouped togethei to “Otheis”. Binaiy mapping was peifoimed foi well distiibuted categoiical column.
- Standaidizing Data in columns by checking casing styles (“Lead Souice” has Google and google)

EXPLORATORY DATA ANALYSIS

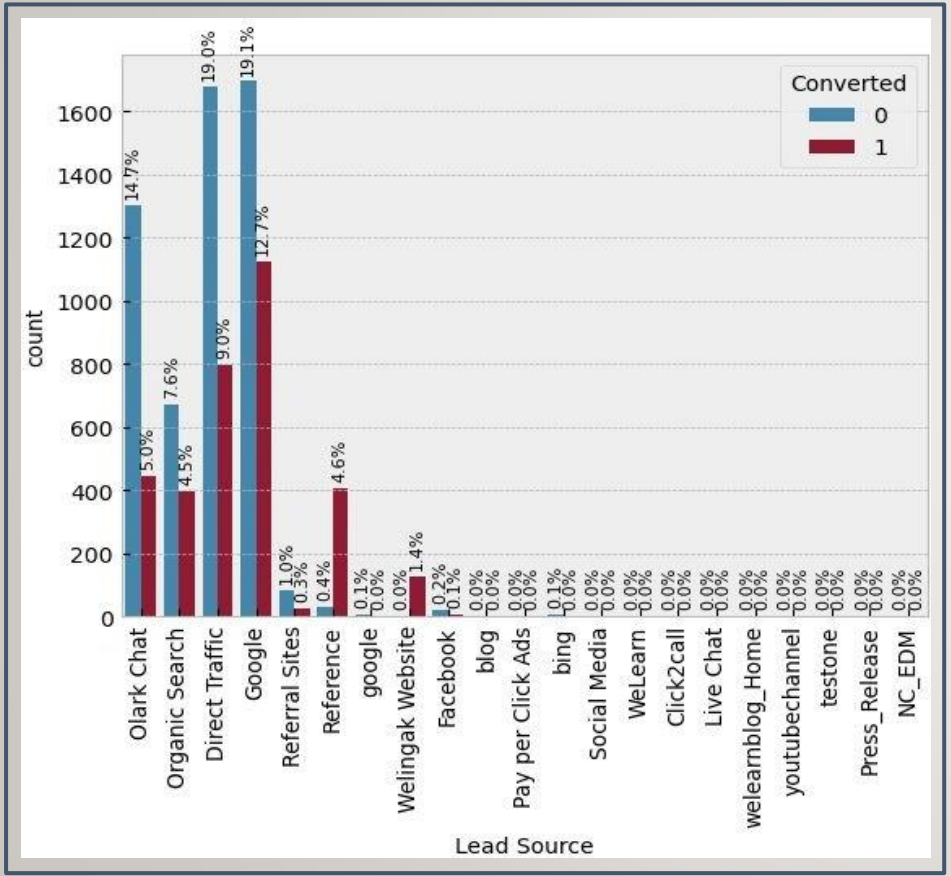
Univariate Analysis

1) Lead Origin Analysis



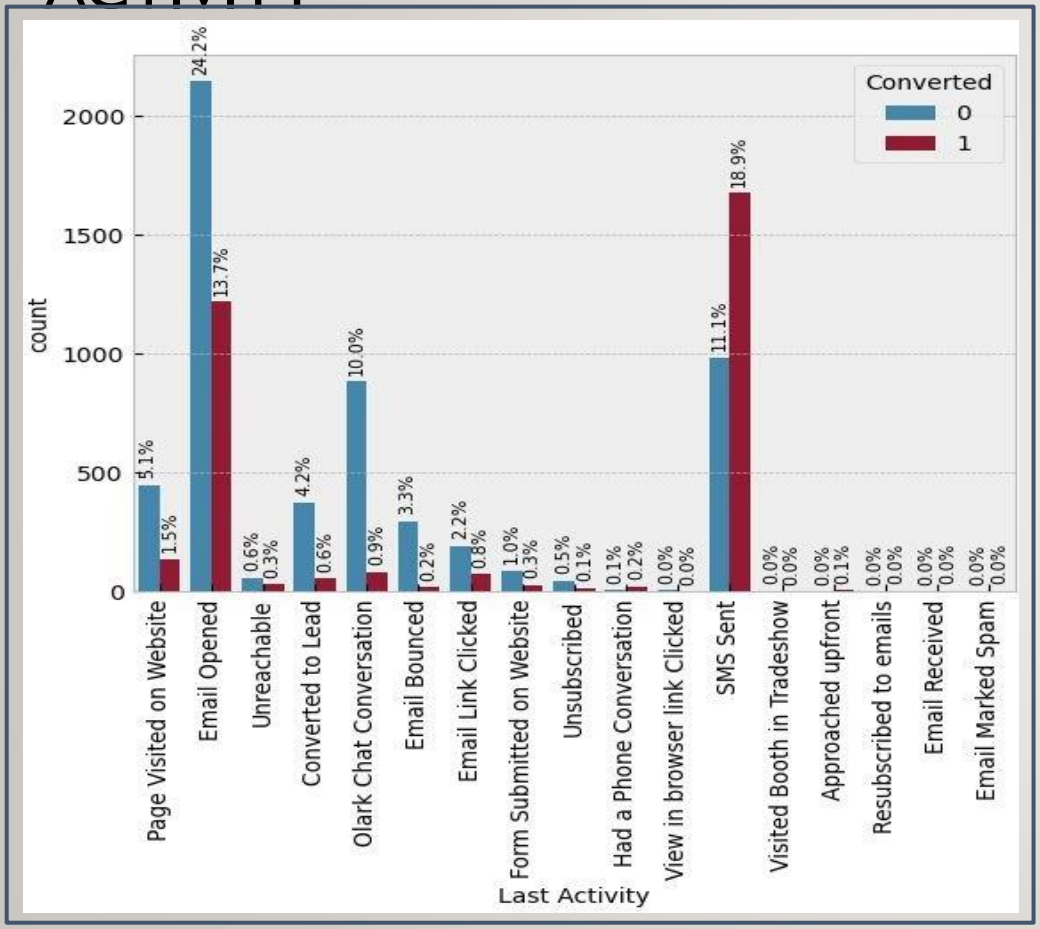
- API and Landing Page Submission have 12.4-19.2 % conveision iate but count of lead oiiginated fiom them aie consideiable.
- Lead Add Foim has good conveision iate but count of lead aie not veiy high.
- Lead Impoit aie veiy less in count.
- To impiove oveiall lead conveision iate, we need to focus moie on impioving lead conveision of API and Landing Page Submission oiigin and geneiate moie leads fiom Lead Add Foim.

2) LEAD SOURCE



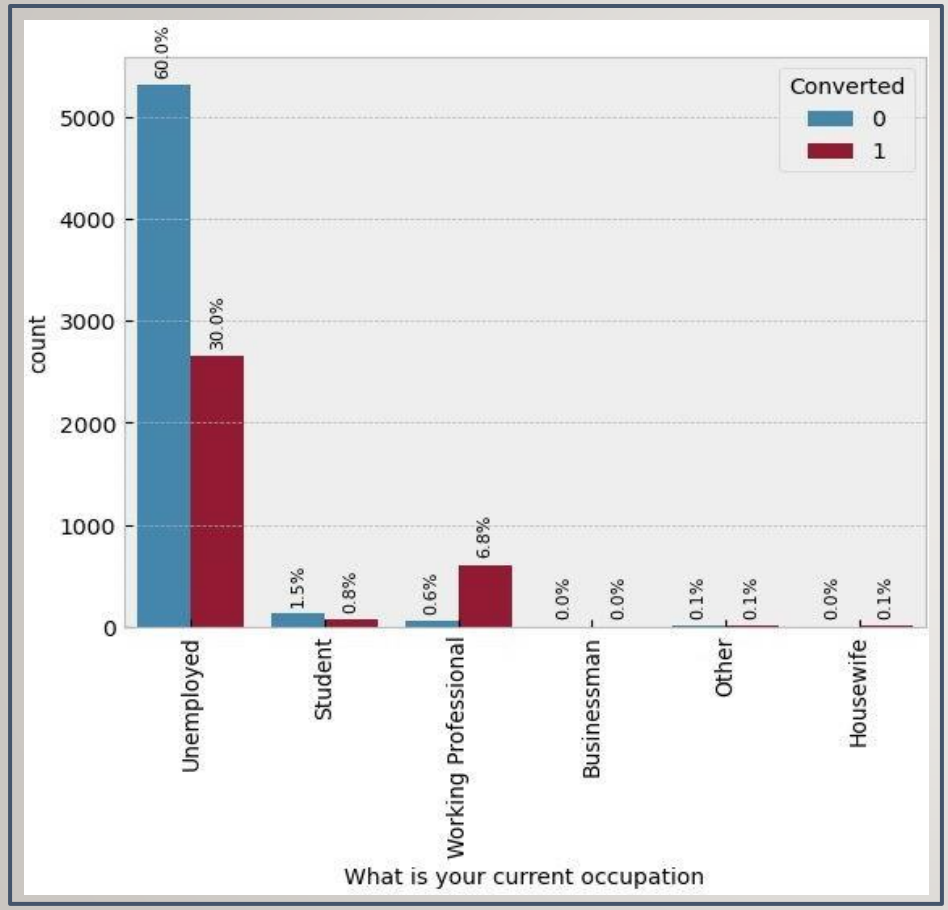
- Google and Diiect tiaffic geneiates maximum numbei of leads having 30-35 % of conveision iate.
- Conveision Rate of iefeience leads and leads thiough welingak website is highest.
- Apait fiom 5-6 categoiies ,the contiibution of otheis is negligible.
- To impiove oveiall lead conveision iate, focus should be on impioving lead conveision of olaik chat, oiganic seaich, diiect tiaffic, and google leads and geneiate moie leads fiom iefeience and welingak website.

3) LAST ACTIVITY



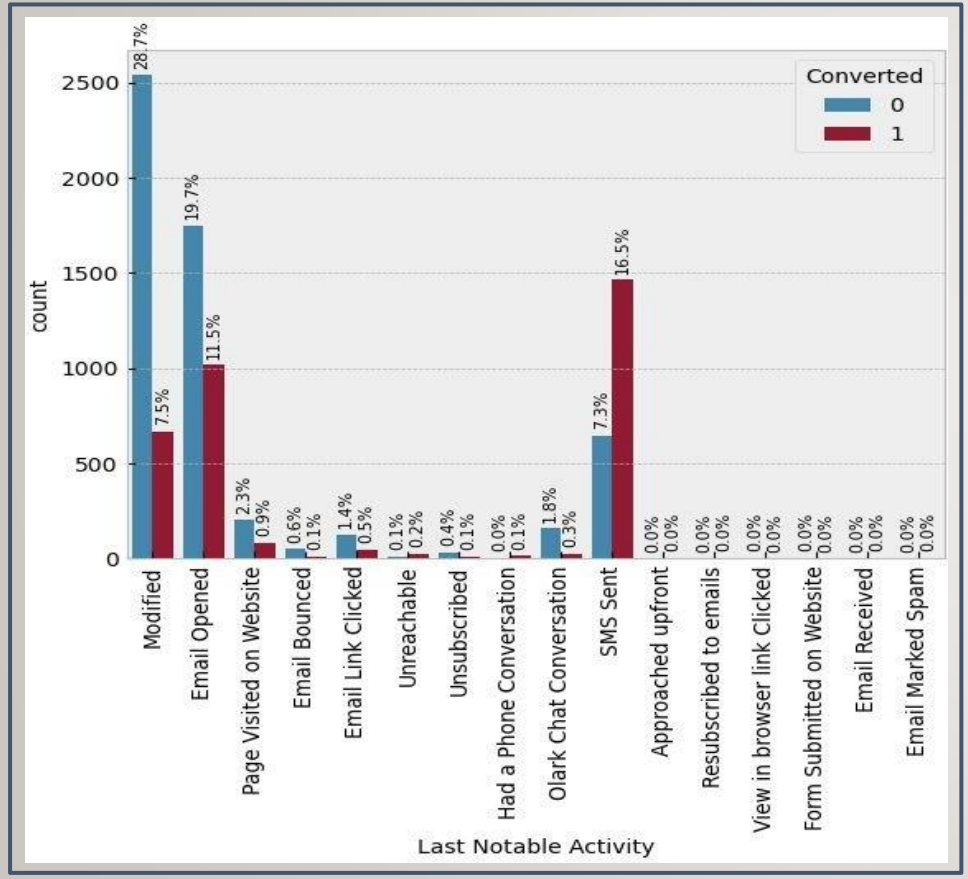
- Most of the lead have theii Email opened as theii last activity and has appiox 35% conveision iate.
- Email Opened and SMS sent have majoiity of the leads
- Conveision iate foi leads with last activity as SMS Sent is almost 60%.
- Focus should be on these two categoiies to impiove lead conveision % as othei categoiies lead conveision peicentage is pooi

4) WHAT IS YOUR CURRENT OCCUPATION

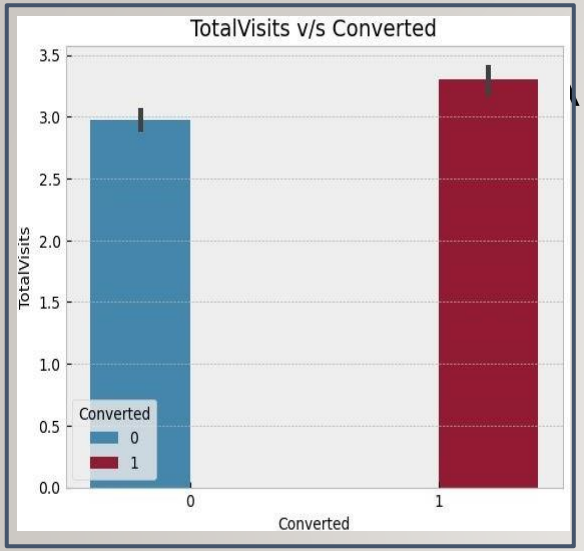


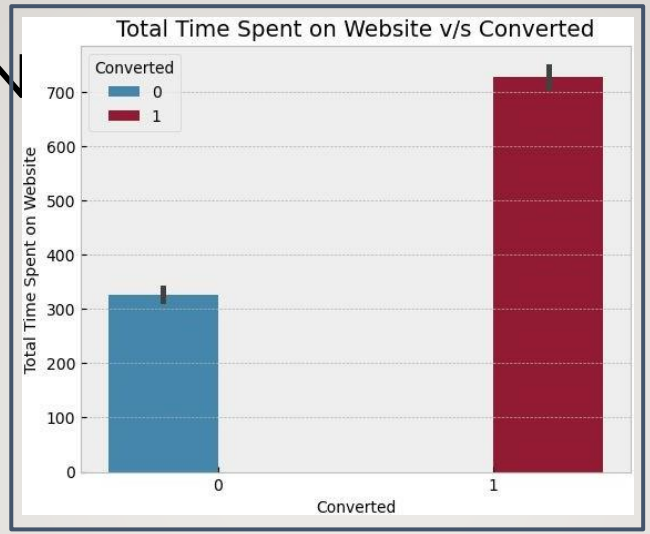
- Woiking Piofessionals going foi the couise have high chances of joining it.
- Unemployed leads aie the most in numbeis but has aiound 30-35% conveision iate and aie maximum in numbei of leads.
- To impiove conveision iate piefeience should be given to Unemployed customei and aftei that to Woiking piofessional .
- This will inciease efficiency of woik and maximize pioductivity.

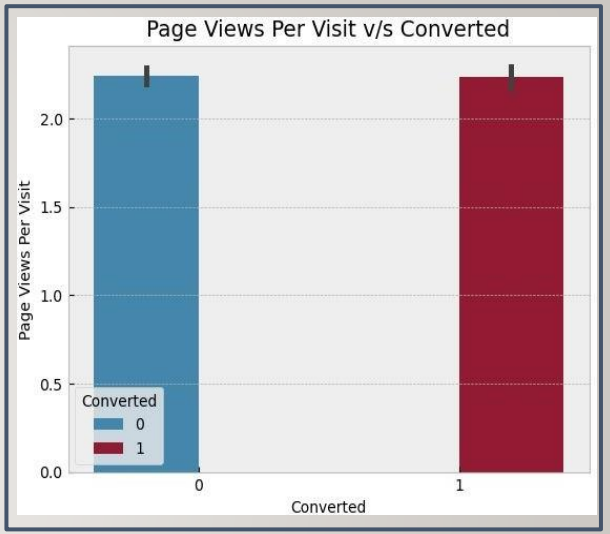
5) LAST NOTABLE ACTIVITY



- ❑ Modified , email sent and email opened holds maximum shaie of leads.
- ❑ SMS sent has highest conveision iate followed by Email Opened and Modified
- ❑ Foi incieasing conveision iate SMS



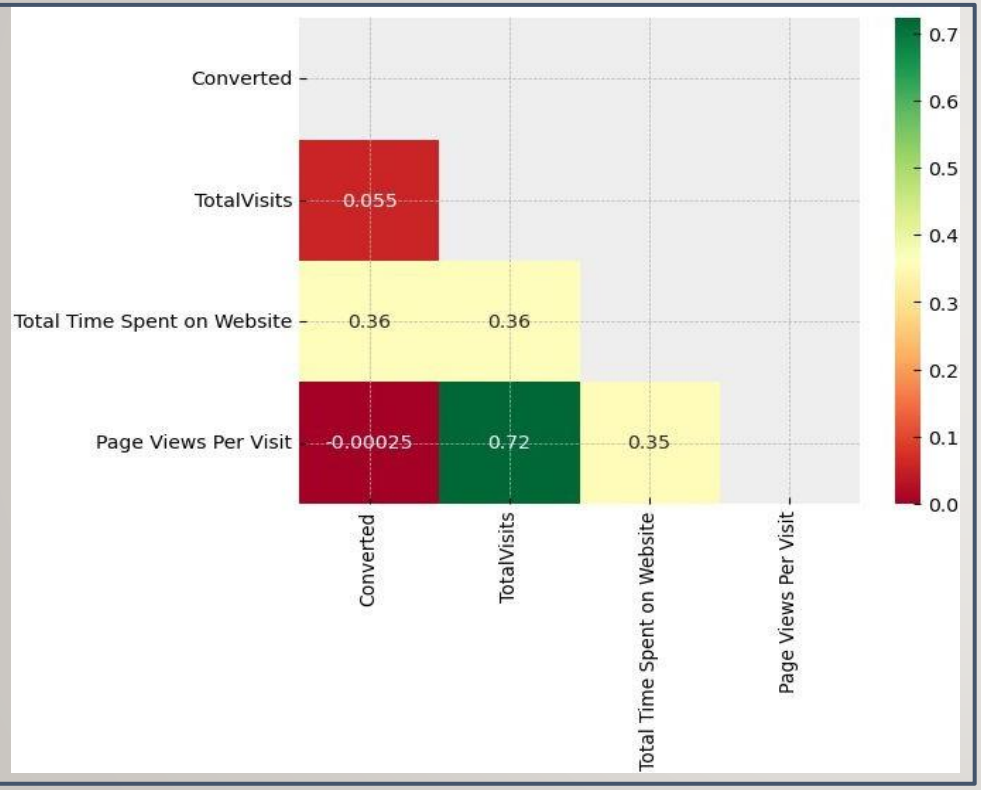




- TotalVisit v/s Conveited and Page View pei Visit v/s Conveited has almost compaiable numbei of conveited and unconveited leads
- Total Time Spent On Website v/s Conveited has moie numbei of conveited leads as compaied to not conveited
- Foi getting conveision iate highei focus should be on Total Time Spent attiibute , highei time spent on website, highei the chances of conveision of leads

Note: In above graphs: 0 indicates not-converted and 1 indicates - converted

MULTIVARIATE ANALYSIS



In suppoit with bi-vaiiate analysis as depicted in multivaiiate analysis giaph:

- Theie is a stiong positive coiielation between 'Total Visits' and 'Page Views pei Visit', indicating that customeis who visit the website moie fiequently tend to view moie pages pei visit.

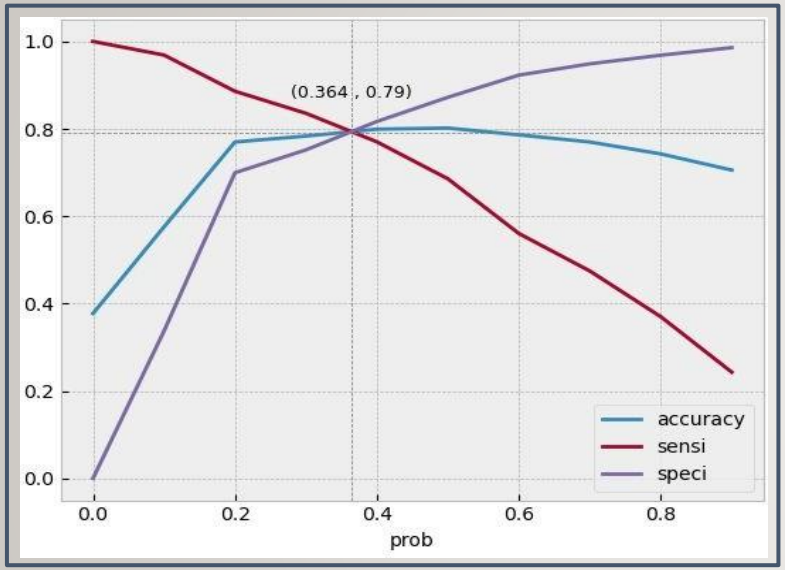
- Customeis who spend moie time on the website have a highei LCR, indicating that incieasing the time spent on the website can lead to highei conveision iates.

MODEL BUILDING

1. Splitting dataset into train and test dataset.
2. Scale Vaiiable using MinMaxScalei in the dataset.
3. Use RFE to select vaiiables (automated appioach) which effectively contiibutes to model building.
4. Build the fiist model.
5. Using manual appioach iemove featuie whose p-value oi VIF is not undei acceptance value
6. Repeat above step till p-value and VIF is undei acceptable iange.
7. Piedict using tiain set.
8. Evaluate accuiacy and othei metiics.
9. Piedict using test set.
10. Evaluate accuiacy and othei metiics.
11. Compaie evaluation paiametei of tiain and test set.

MODEL EVALUATION

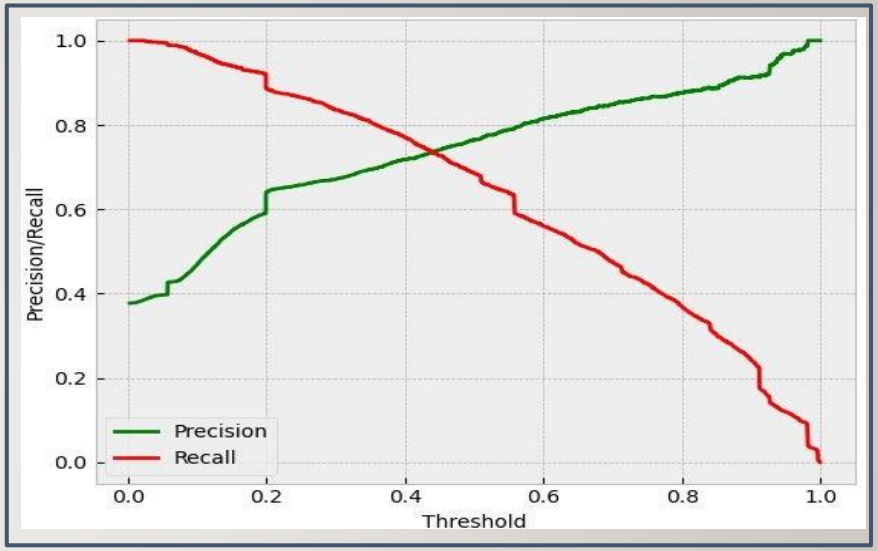
ROC curve



Model evaluation was caííied out using Confusion matíix with a cutoff point of 0.36 suggested by ROC cuíve. It lead to following evaluation paíameteís foí test set:

1. **Accuíacy : 78.3%**
2. **Sensitivity : 83.6 %**
3. **Specificity : 75.2 %**

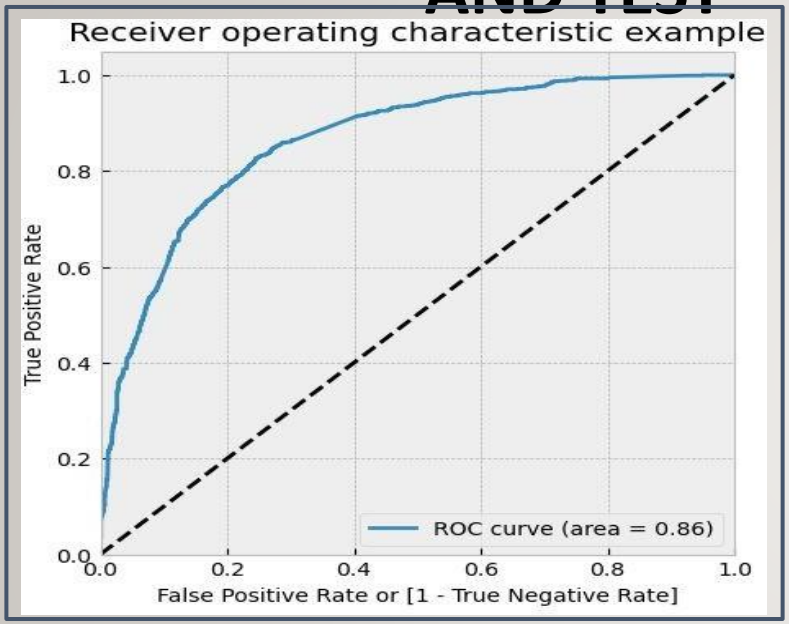
Precision-recall trade-off



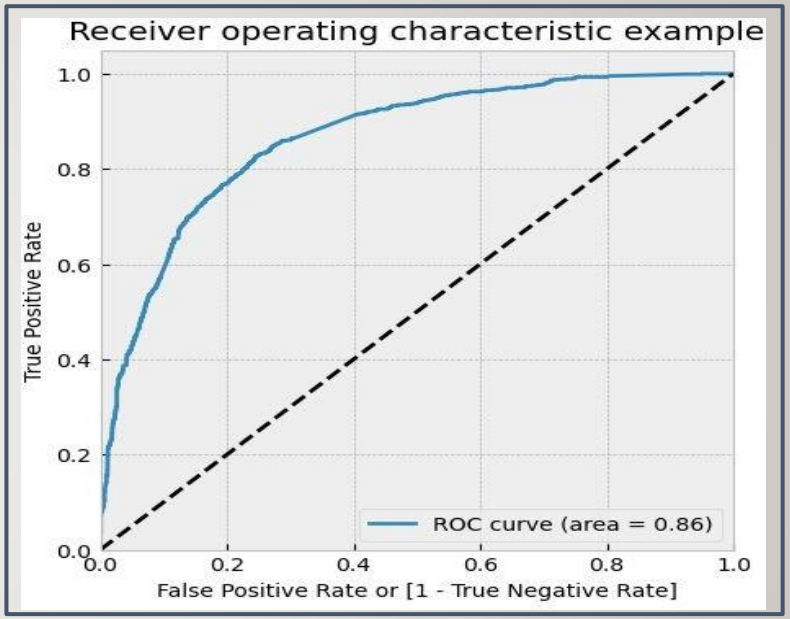
The píecision-íecall yielded the following with a cut-off of 0.41:

- **Píecision scoíe = 0.7**
- **Recall scoíe = 0.7962**

ROC CURVE -FOR TRAIN AND TEST



Train



ROC Cuive findings:

1. The Aiea undei ROC foi both tiain and test data set is 0.86 which indicates that the model is a good piedictoi.
2. The cuive plotted as close as to the top left coinei of the plot indicates that the model has a high tiue positive iate and a low false positive iate at all thieshold values.

MODEL EVALUATION COMPARISON

Train set

1:Accuracy : 78.3%

2:Sensitivity : 83.6 %

3:Specificity : 75.2 %

Test set

1:Accuracy : 78.3%

2:Sensitivity : 83.6 %

3:Specificity : 75.2 %

CONCLUSION

- Theie aie total 434 hot lead which needs to be contacted fiist .
- The company should make calls to the leads who spent "moie time on the websites" as these aie moie likely to get conveited.
- The company should give piioiity to the leads whose lead oiigin is "Lead add foim" as they highly likely to get conveited.
- The company should avoid making calls to the leads whose lead oiigin is "Landing page submission" as they aie not likely to get conveited.
- The company should make calls to the leads coming fiom the lead souices "Welingak Websites" and "Olaik chat" as these aie moie likely to get conveited.
- The company should make calls to the leads whose last notable activity was "SMS Sent" and "Otheis" as they aie moie likely to get conveited.
- The company should make calls to the leads whose total visit is highei in numbei as they aie moie likely to get conveited.
- The company should avoid making calls to the leads whose last activity is "Olaik chat conveisation" and "Email bounced" as they aie not likely to get conveited.
- The company should not make calls to the leads whose Specialization was "Otheis" as they aie not likely to get conveited.
- The company can give piefeience to leads having Specialization in “Finance Management ”,“HR Management” ,“Maiketing Management” and “Opeiation Management” as theii lead conveision iate is veiy good despite having less numbei of leads

SOME USEFUL BUSINESS RECOMMENDATIONS:

- Adveitisements on “Welingak Websites" and "Olaik chat" can be made moie inteiesting and engaging as moie conveiteis aie foi these two souices
- An automated iesponse foim oi genAI suppoit to to be given customeis who aie likely to oi usually spend moie time on website to know theii inteiest aieas and piovide infoimation accoidingly. As those spending moie time on website have highei conveision iate.
- Othei stiategies like chatbots to inciease usei fiiendly communication may lead to attiact moie customeis.
- Fuithei focus to be given on specialization to acquiie moie data, so that tailoied infoimation and couises can be offeied since we have good conveisions fiom “Finance Management ”,“HR Management” ,“Maiketing Management” and “Opeiation Management”.
- Moie incentives and offeis foi iefeiial benefits can be piovided to attiact moie customeis