

The background of the slide is a light gray gradient. It is decorated with numerous realistic water droplets of various sizes. Some droplets are large and prominent, while others are small and subtle. They are scattered across the slide, with a higher concentration in the top left and bottom right corners. The droplets have highlights and shadows, giving them a three-dimensional appearance.

X EDUCATION - LEAD SCORING CASE STUDY

IDENTIFICATION OF HOT LEADS AND OPTIMIZE
RESOURCES USED IN LEAD CONVERSIONS


Vamsidhar Nookala

BACKGROUND

- X EDUCATION , AN EDUCATION COMPANY NAMED SELLS ONLINE COURSES TO INDUSTRY PROFESSIONALS
- MANY INTERESTED PROFESSIONALS LAND ON THEIR WEBSITE
- THE COMPANY MARKETS ITS COURSES ON SEVERAL WEBSITES LIKE GOOGLE. ONCE THESE PEOPLE LAND ON THE WEBSITE, THEY MIGHT BROWSE THE COURSES OR FILL UP A FORM FOR THE COURSE OR WATCH SOME VIDEOS




BACKGROUND

- WHEN THESE PEOPLE FILL UP A FORM PROVIDING THEIR EMAIL ADDRESS OR PHONE NUMBER, THEY ARE CLASSIFIED TO BE A LEAD
 - ONCE THESE LEADS ARE ACQUIRED, EMPLOYEES FROM THE SALES TEAM START MAKING CALLS, WRITING EMAILS, ETC. THROUGH THIS PROCESS, SOME OF THE LEADS GET CONVERTED WHILE MOST DO NOT
 - THE TYPICAL LEAD CONVERSION RATE AT X EDUCATION IS AROUND 30%
- 




PROBLEM STATEMENT

- X EDUCATION GETS A LOT OF LEADS BUT ITS LEAD CONVERSION RATE IS VERY POOR
 - TO MAKE THIS PROCESS MORE 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 CONVERSION RATE SHOULD GO UP AS THE SALES TEAM WILL NOW BE FOCUSING MORE ON COMMUNICATING WITH THE POTENTIAL LEADS RATHER THAN MAKING CALLS TO EVERYONE
- 



PROBLEM STATEMENT

- WE WILL HELP THEM TO SELECT THE MOST PROMISING LEADS, I.E. THE LEADS THAT ARE MOST LIKELY TO CONVERT INTO PAYING CUSTOMERS.
 - WE ARE REQUIRED TO BUILD A MODEL WHEREIN WE NEED TO ASSIGN A LEAD SCORE TO EACH OF THE LEADS SUCH THAT THE CUSTOMERS WITH HIGHER LEAD SCORE HAVE A HIGHER CONVERSION CHANCE
 - THE CEO, IN PARTICULAR, HAS GIVEN A BALLPARK OF THE TARGET LEAD CONVERSION RATE TO BE 80%.
- 

Lead – Conversion Process



Lead Conversion Process - Demonstrated as a funnel

STAGE1: INITIAL POOL OF LEADS RECEIVED THROUGH VARIOUS PLATFORMS LIKE CALLING, SOCIAL MEDIA MARKETING


STAGE2: LEAD NURTURING OR CONVERSION OF INITIAL POOL TO HOT LEADS BASIS SOME CUSTOMER RESPONSES

STAGE3: LEADS CONVERSION WHEREIN ACTUAL REVENUE GETS GENERATED FOR THE COMPANY

AS OF NOW IN STAGE2, CONVERSION OF INITIAL POOL TO HOT LEADS HAS BEEN MANUAL PROCESS BASIS CUSTOMER RESPONSES WHICH CAN BE DATA DRIVEN THROUGH LOGISTIC REGRESSION MODEL



PROPOSED SOLUTION

- IN STAGE2 OF LEAD NURTURING, ALL THE INITIAL POOL OF LEADS SHALL BE PASSED THROUGH LOGISTIC REGRESSION MODEL AND SHALL BE FILTERED BASIS THE POTENTIALITY OF LEAD CONVERSION
 - MODEL SHALL BE BASED ON PARAMETERS LIKE LEAD SOURCE, TIME SPENT ON THE WEBSITE, TOTAL VISITS, ETC.,
 - BASIS THIS FILTERS COMPANY CAN GET HOT LEADS WHICH HAVE A HIGHER RATE OF LEAD CONVERSION AND CAN EFFECTIVELY USE THE MANPOWER RATHER THAN TRYING THOSE LEADS WHICH HAVE A VERY LESS CHANCE OF CONVERSION
- 



SOLUTION

FOR OUR PROBLEM SOLUTION, THE CRUCIAL PART IS TO ACCURATELY IDENTIFY HOT LEADS.

THE MORE ACCURATE WE OBTAIN THE HOT LEAD, THE MORE CHANCE WE GET OF HIGHER
SELECTION OF HOT LEADS
CONVERSION RATIO.

SINCE WE HAVE A TARGET OF 80% CONVERSION RATE, WE WOULD WANT TO OBTAIN A HIGH
ACCURACY IN OBTAINING HOT LEADS.



INITIAL FLOW OF LEADS WITH REQUISITE DATA POINTS

DATA GATHERING AND UNDERSTANDING

- INITIALLY FAMILIARIZE WITH ALL THE PARAMETERS
- TARGET VARIABLE IN THE DATA IS CONVERTED
- UNDERSTAND THE EFFECT OF THE VARIABLES ON TARGET VARIABLE

DATA CLEANING

- MISSING VALUE TREATMENT
- IMPUTING MISSING VALUES
- REMOVING NULL VALUES IF PARAMETER IS NOT EFFECTIVE
- OUTLIER TREATMENT
- FINDING OUTLIERS BASIS BOX PLOTS
- REMOVING TOP AND BOTTOM 1% OUTLIERS BASIS BUSINESS REQUIREMENTS

MODEL BUILDING

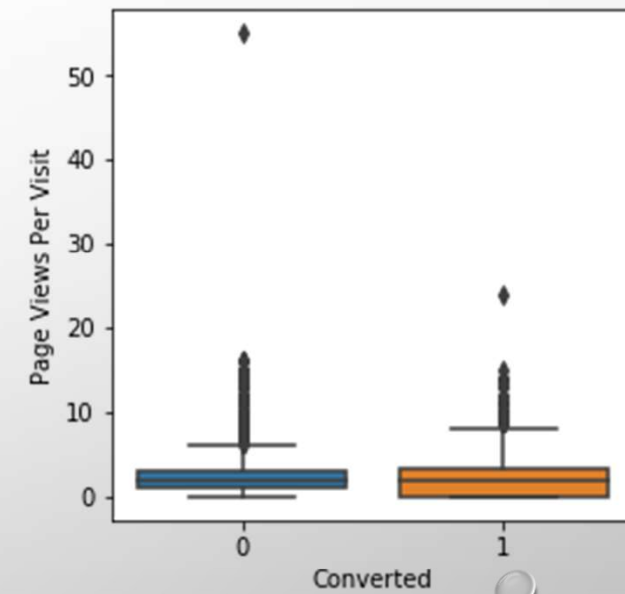
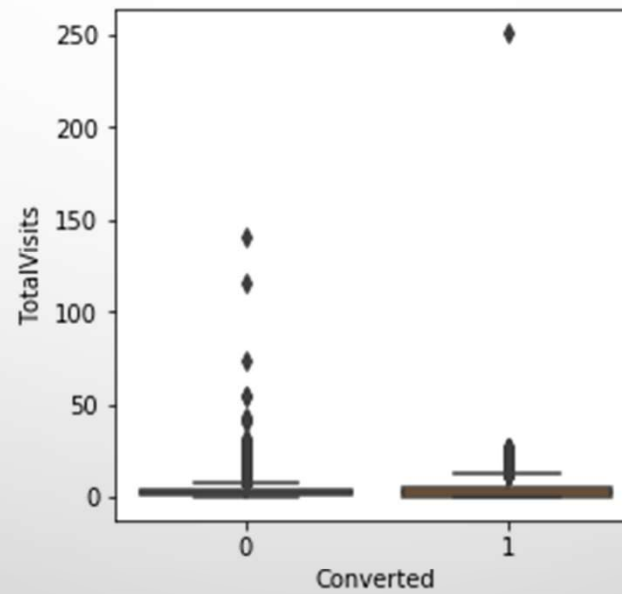
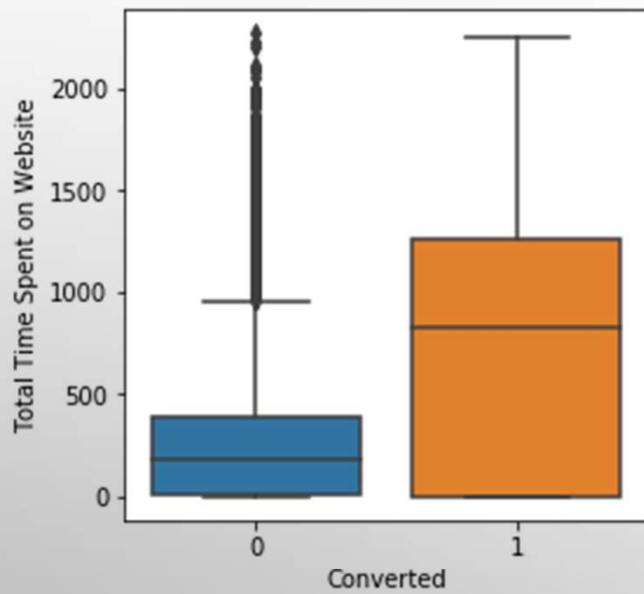
- ONCE DATA CLEANING IS DONE RFE MODEL TO BE BUILT WHEREIN SOME PARAMETERS WITH MULTICOLLINEARITY OR LESS DEPENDENCY ARE REMOVED
- FURTHER LOOKING AT P VALUE AND VIF FURTHER TUNING OF THE MODEL TO BE DONE

MODEL EVALUATION

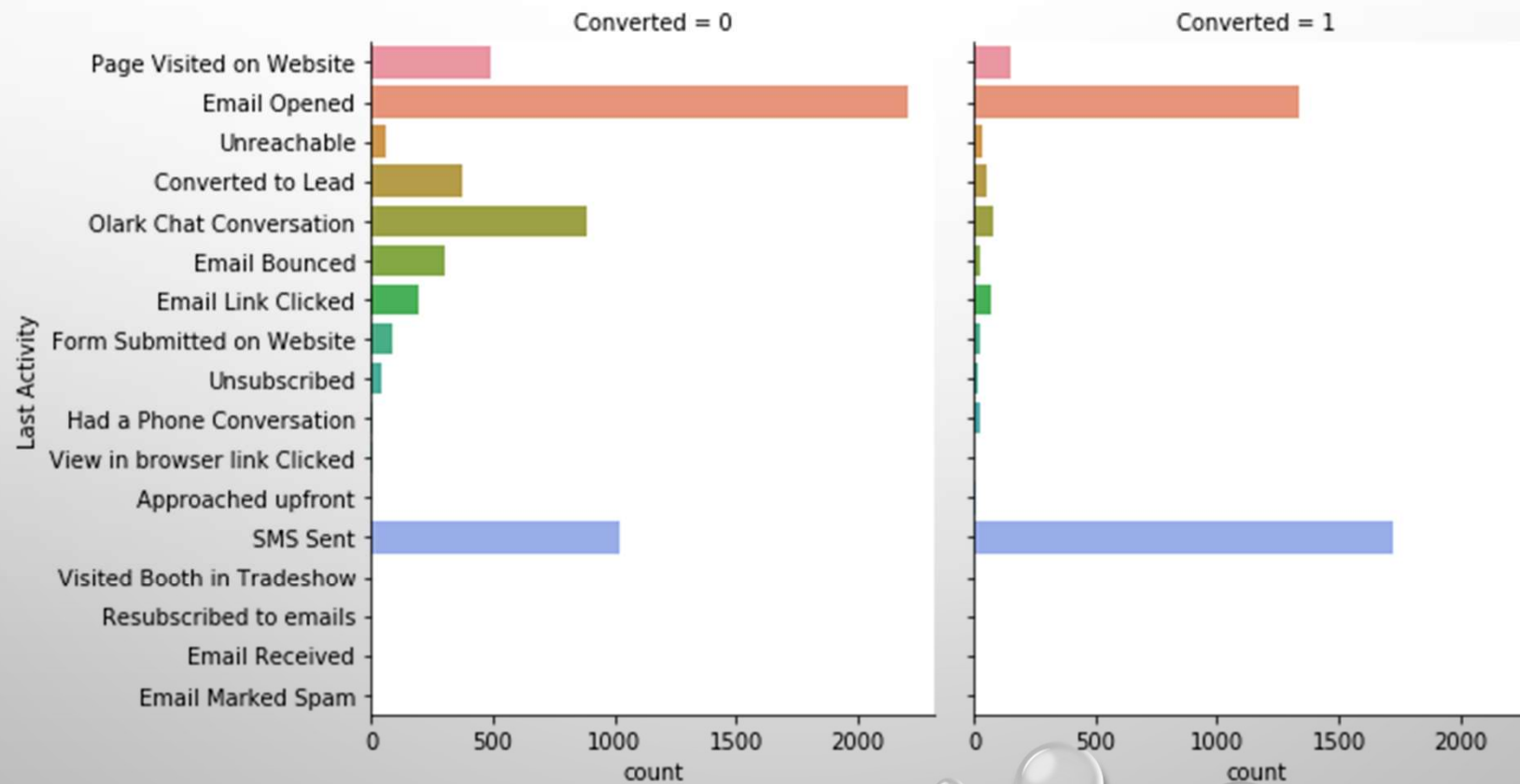
- ONCE MODEL IS BUILT THE SAME IS RUN THROUGH THE TEST SPLIT
- BASIS THE RESULTS ONE CAN ANALYSE ACCURACY, SPECIFICITY AND SENSITIVITY
- FURTHER ANALYSIS TO BE DONE BASIS PRECISION AND RECALL TO BE DERIVE AT OPTIMUM CUTOFF

BASIS
OPTIMUM
CUTOFF
LEADS ARE
FILTERED TO
HOT LEADS

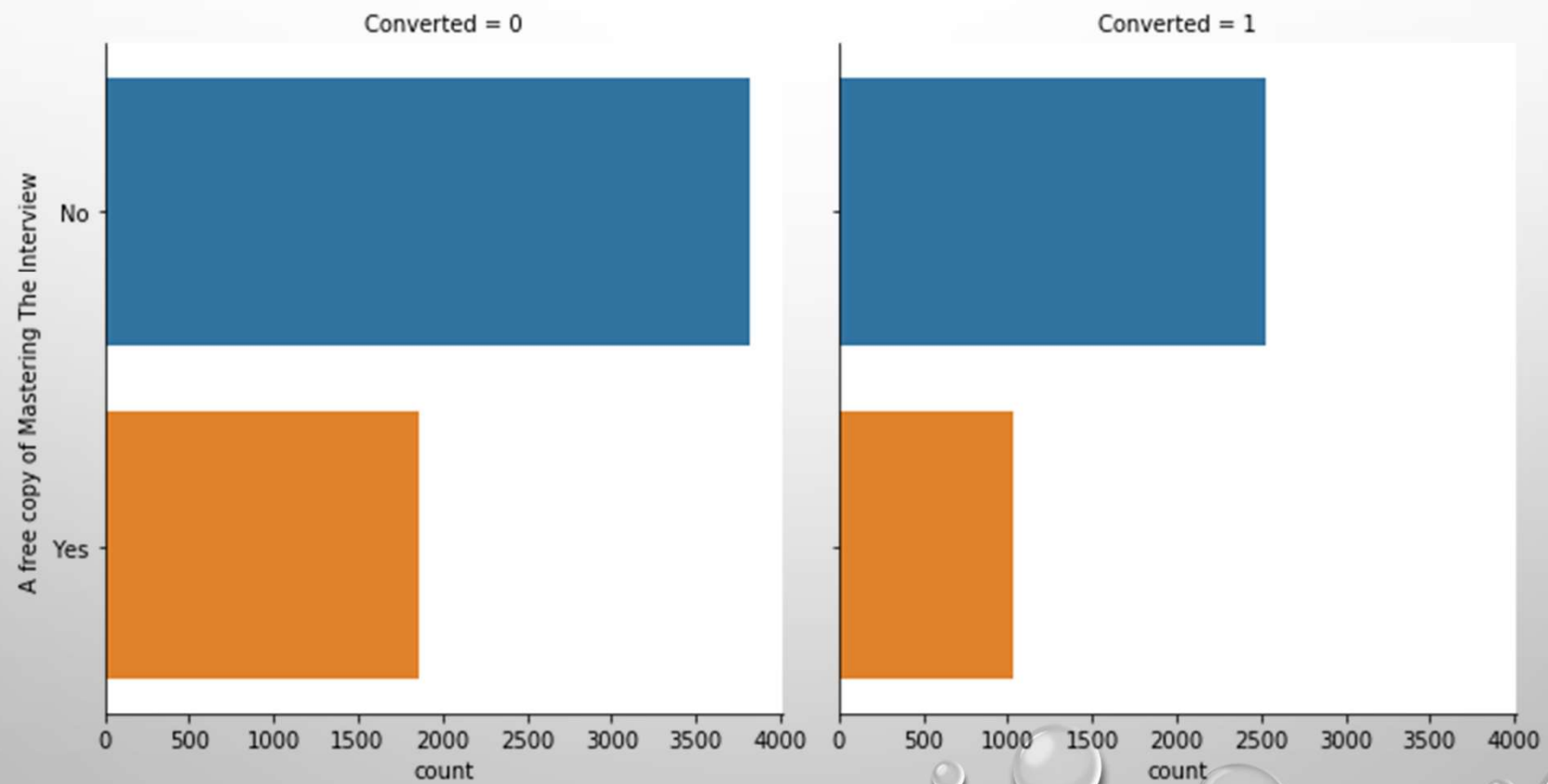
BIVARIATE ANALYSIS ON NUMERICAL DATAPOINTS



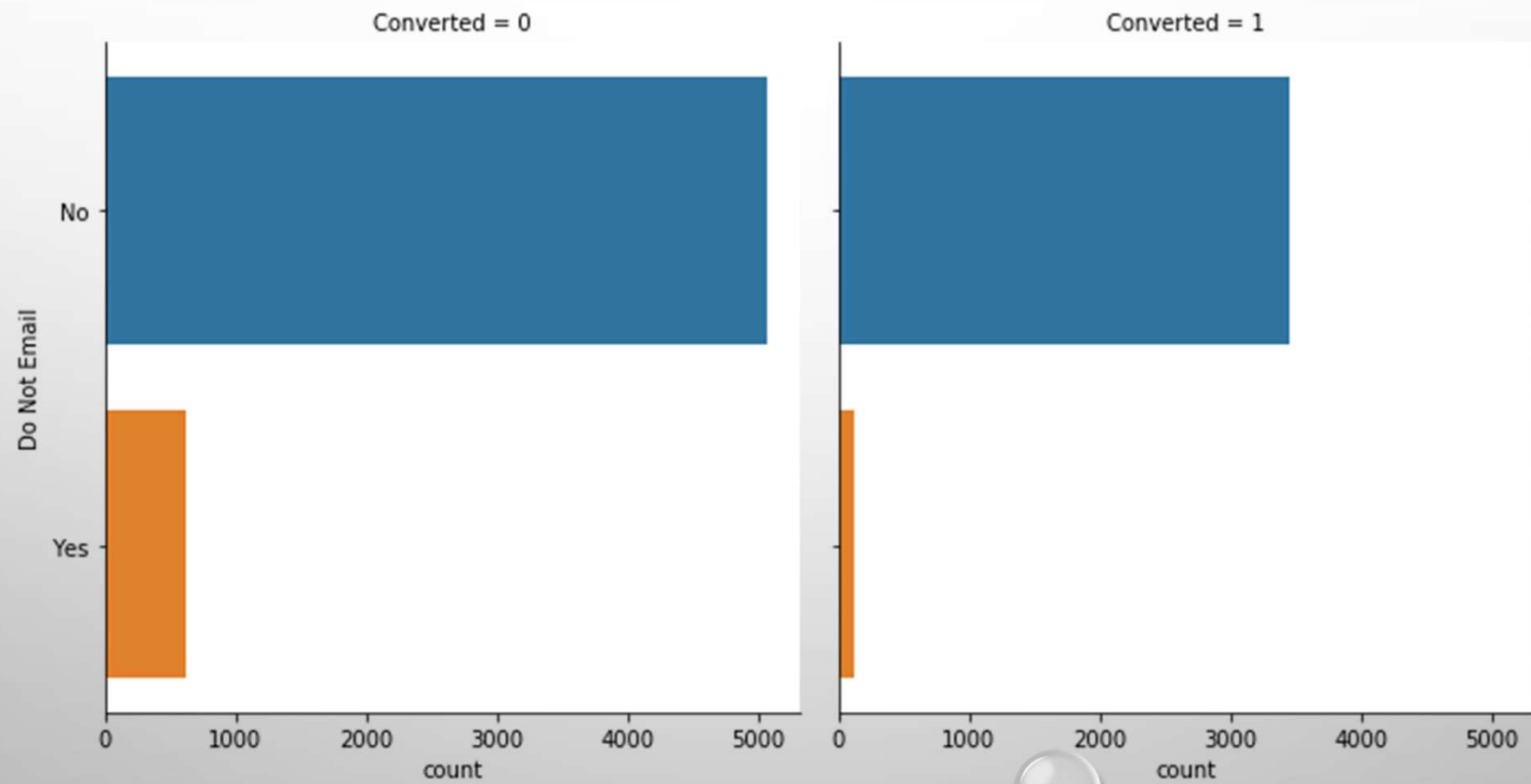
BIRVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS (Last Activity)



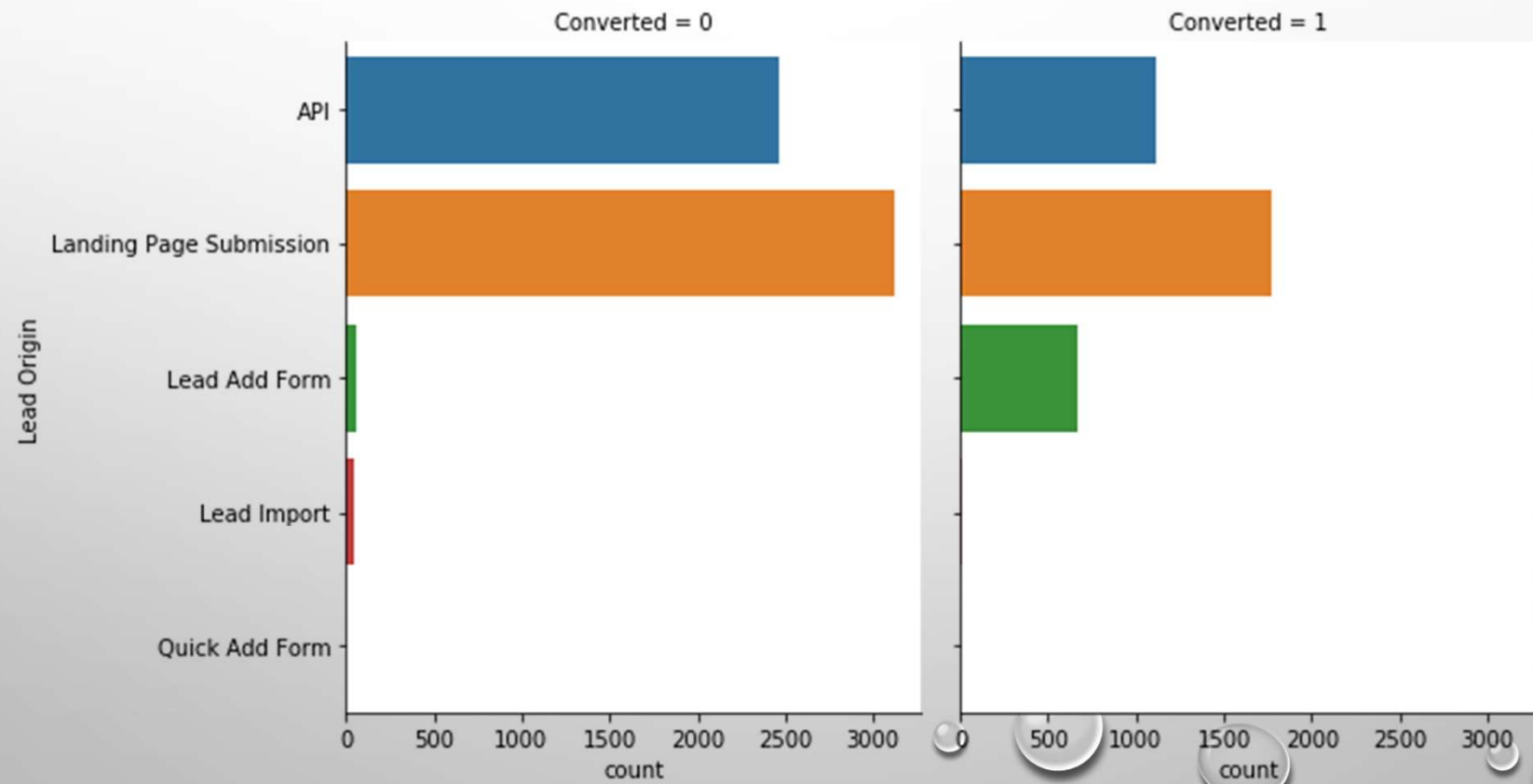
BIVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS (A free copy)



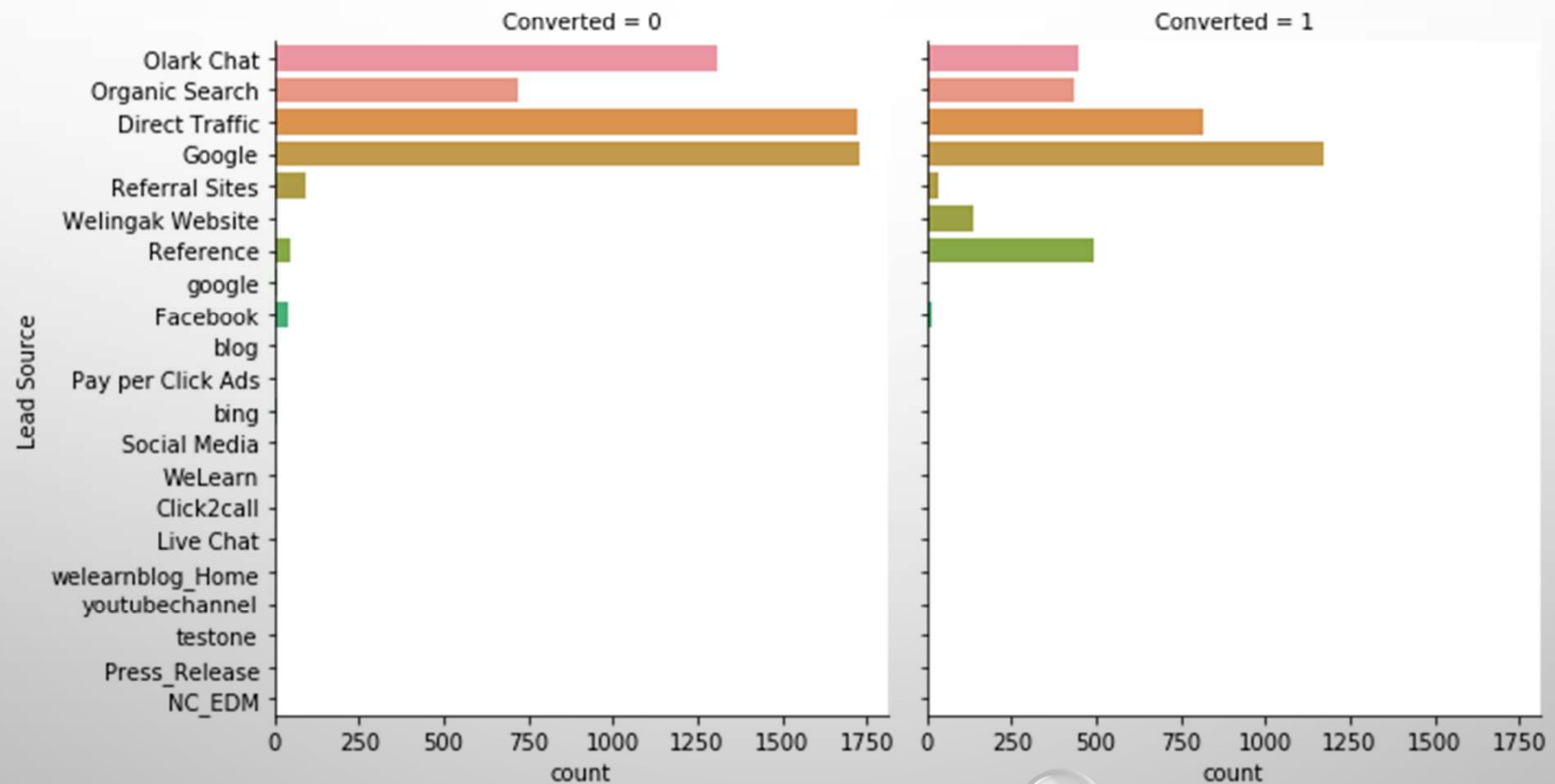
BIVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS



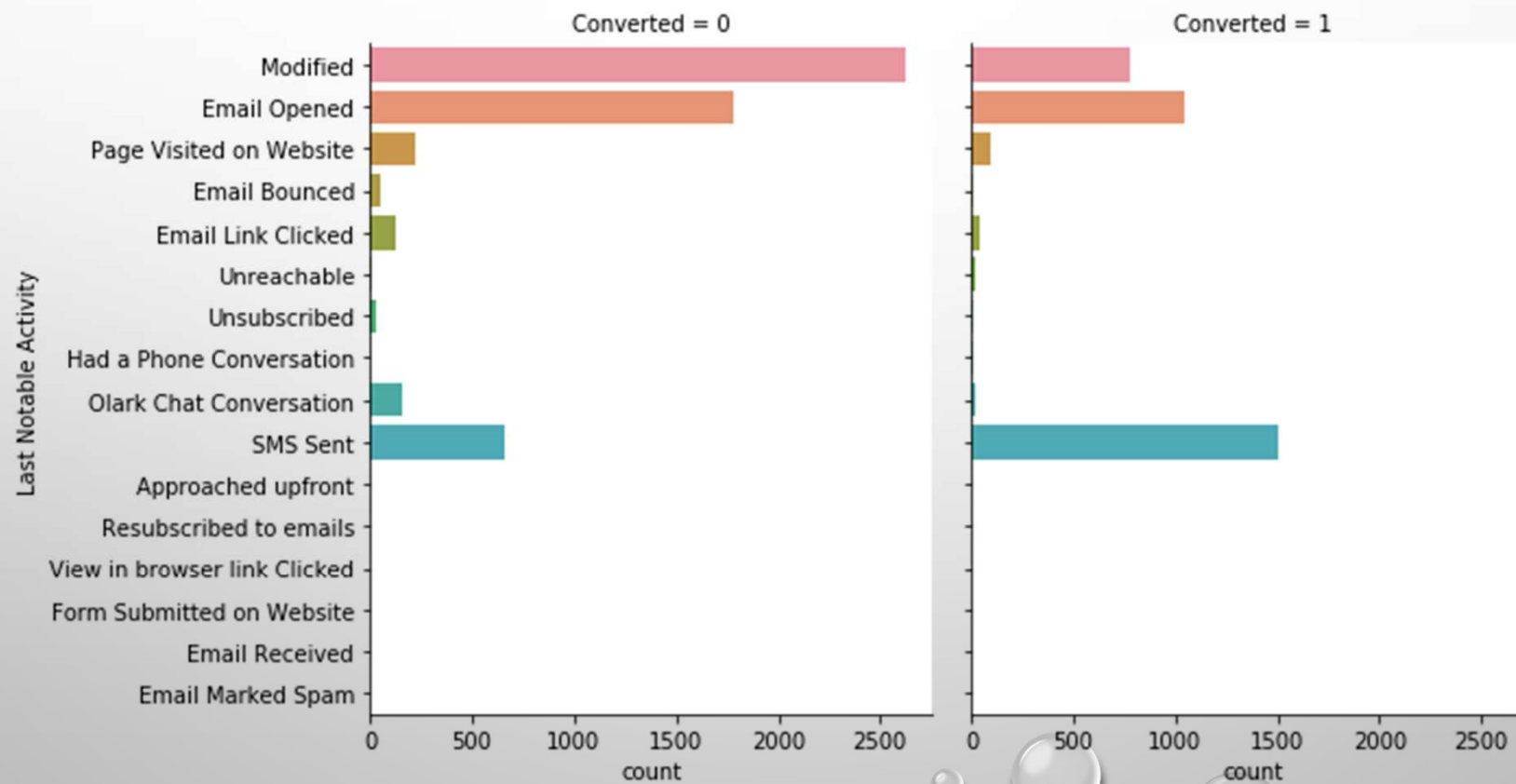
BIVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS (Lead Origin)



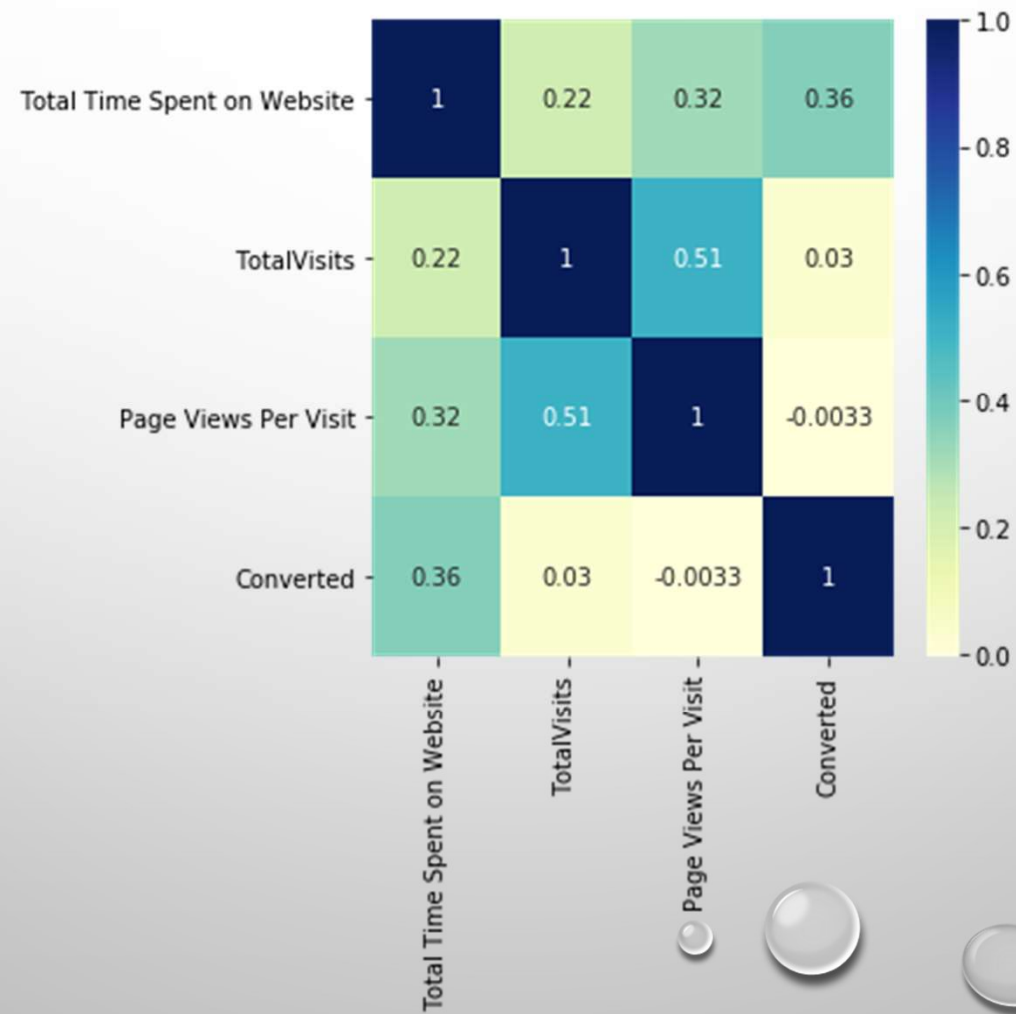
BIVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS (Lead Source)



BIVARIATE ANALYSIS ON CATEGORICAL DATAPOINTS (Last Notable Activity)

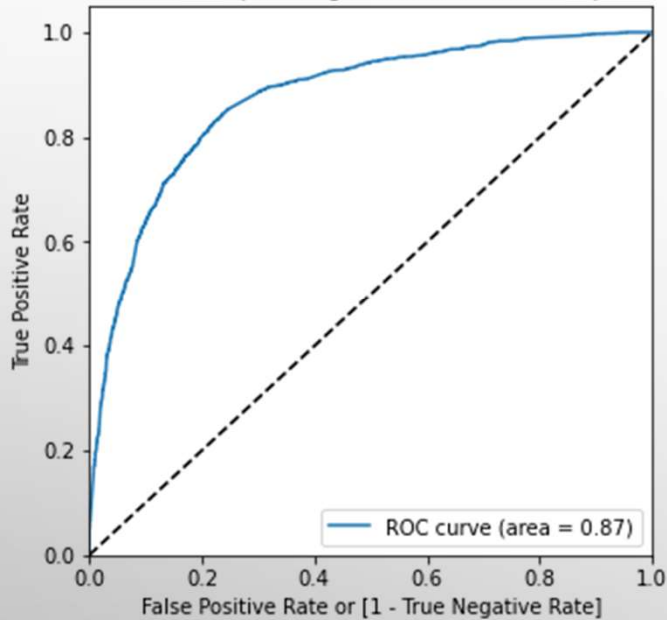


HEATMAP DEPICTING CORRELATION BETWEEN THE NUMERICAL VARIABLES

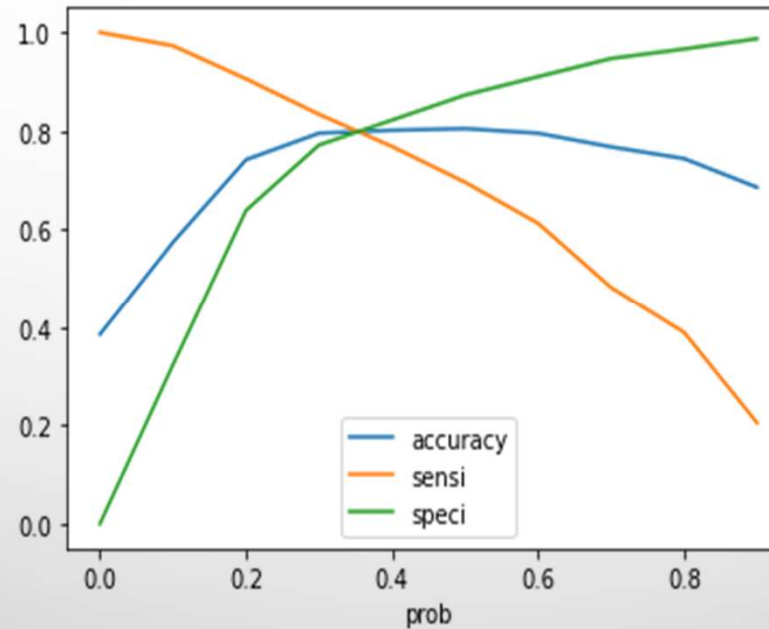


REGRESSION FINAL MODEL PERFORMANCE

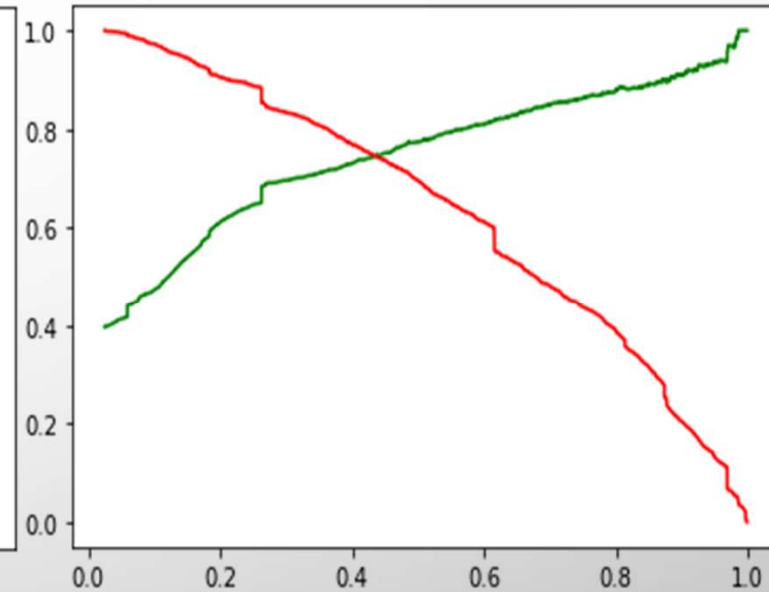
Receiver operating characteristic example



Area under ROC = 0.87



Cutoff as per accuracy, specificity
and sensitivity= 0.35



Final Cutoff basis Precision and
Recall= 0.42

INFERENCES FROM MODEL

TOP 3 VARIABLES IN MODEL, THAT CONTRIBUTE TOWARDS LEAD CONVERSION ARE:

- LEAD SOURCE_WELINGAK WEBSITE
- TOTAL TIME SPENT ON WEBSITE
- LEAD SOURCE_ REFERENCE

TOP NEGATIVELY IMPACTING VARIABLES

- DO NOT EMAIL_YES
- LAST NOTABLE ACTIVITY_EMAIL LINK CLICKED
- LAST NOTABLE ACTIVITY_MODIFIED

MODEL PERFORMANCE

ACCURACY – 78.89%

PRECISION – 71.62%

SENSITIVITY OR RECALL – 74.48%

SPECIFICITY – 81.64%

CONCLUSION

- WITH THE MODEL PERFORMANCE PARAMETERS IT CAN BE CONCLUDED THAT THE MODEL IS ACCURATELY DETERMINING THE POTENTIALITY OF LEADS TO GET CONVERTED WITH AN OVERALL ACCURACY OF AROUND 80%
- THIS IS SUBSTANTIAL INCREASE FROM THE PROBLEM STATEMENT WHEREIN COMPANY IS CURRENTLY ACHIEVING ONLY 30% CONVERSION OF LEADS
- HOT LEADS WILL GUIDE THE SALE TEAM TO FOCUS RATHER THAN WASTING TIME IN LEADS WITH LOW CONVERSION RATE
- SECONDLY IN CASE THE HOT LEADS ARE ALL CATERED AND IF TEAM IS FREE AND IN NEED OF MORE POTENTIAL LEADS, ONE CAN REDUCE THE CUTOFF FOR GETTING SOME MORE MEANINGFUL LEADS WITH LOWER CONVERSION RATE OF SAY 70-80%