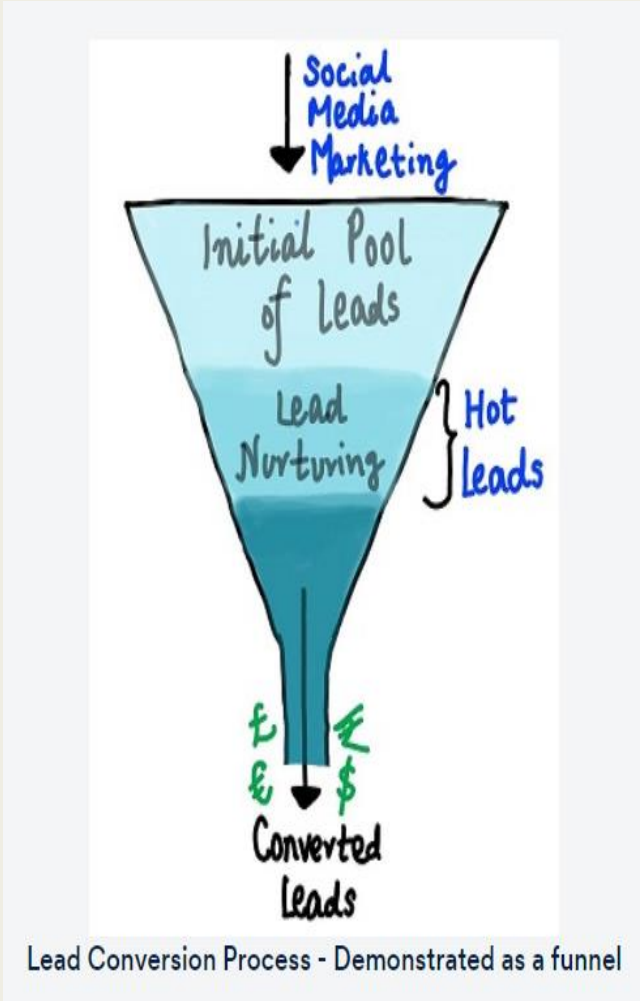
LEAD SCORING CASE STUDY

BY RAHUL NANDI
ANURAG TIWARI
RAHUL GARIGANTI

PROBLEM STATEMENT



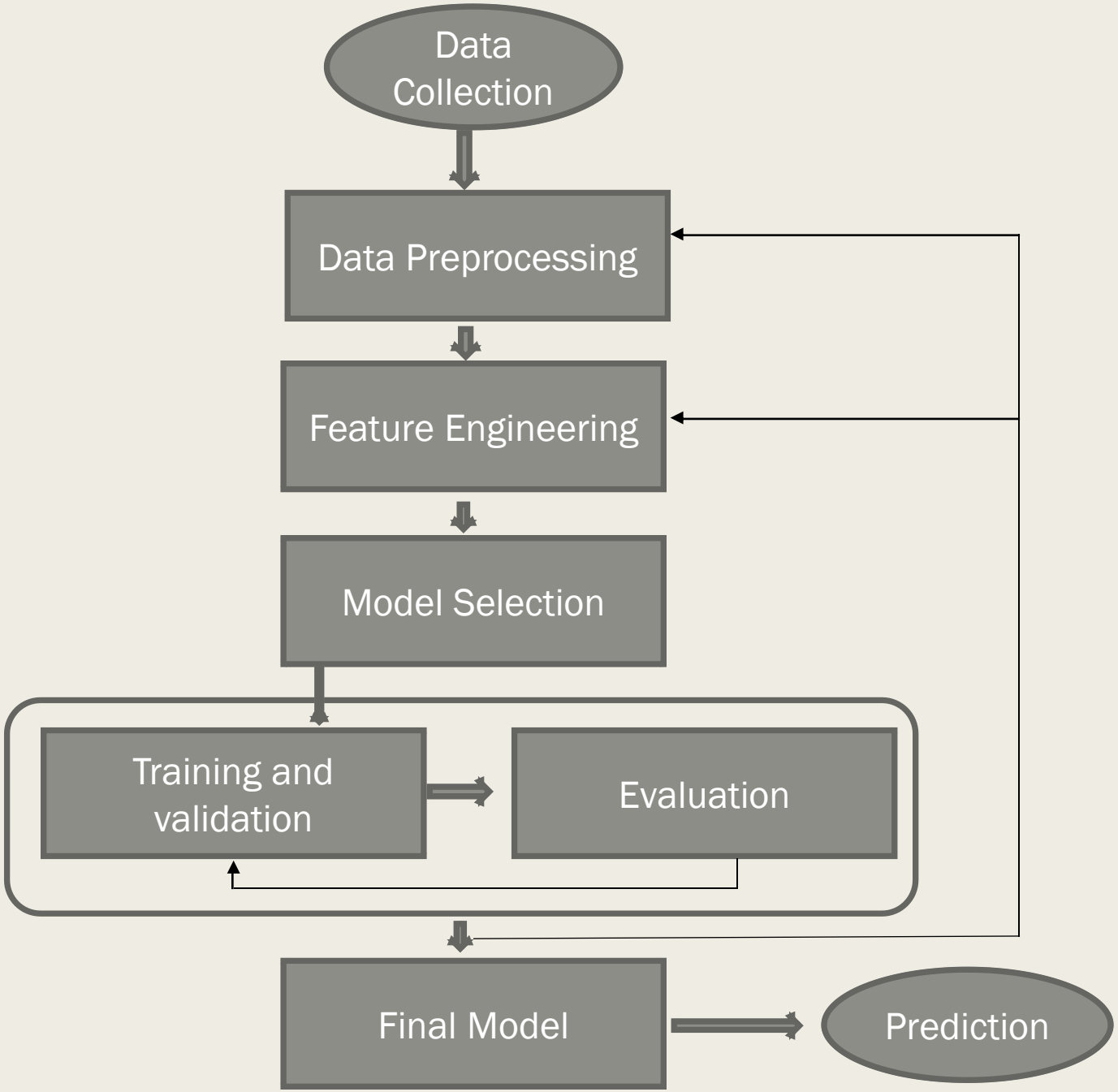
Lead Conversion Process - Demonstrated as a funnel

X Education is an online education company that sells courses to industry professionals. The company has a high lead generation rate, but only 30% of leads convert into paying customers. The CEO wants to develop a model and strategy to increase the lead conversion rate.

DATA INFORMATION

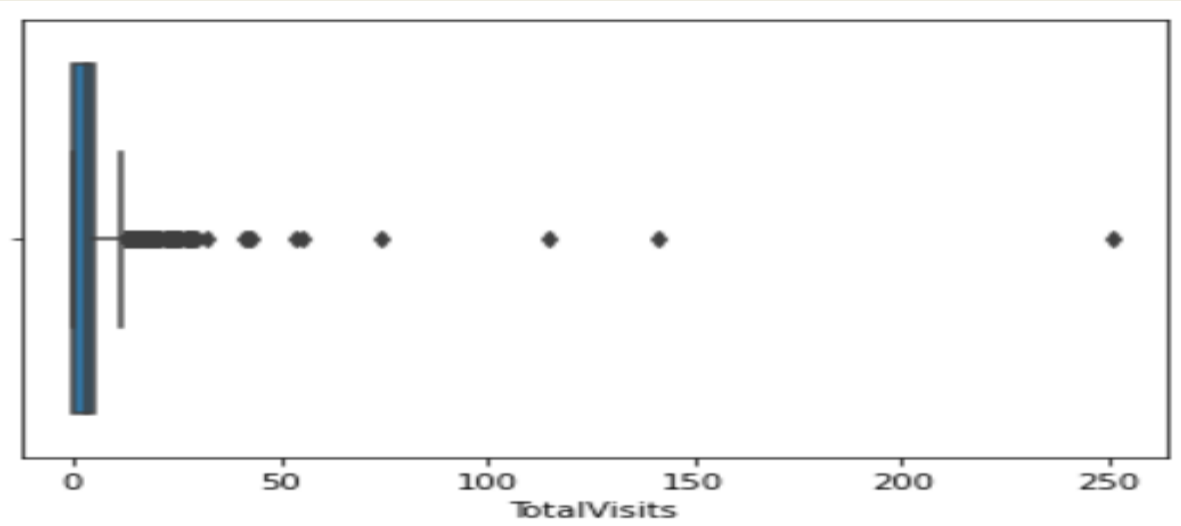
- Raw Data collected from multiple sources: Marketing, websites, Forms, Calls, Referrals etc.
- We got data for 9240 unique customers.
- The dataset consists of various attributes such as lead source, total spent on website, total visits, last activity etc.
- Target column is Converted.

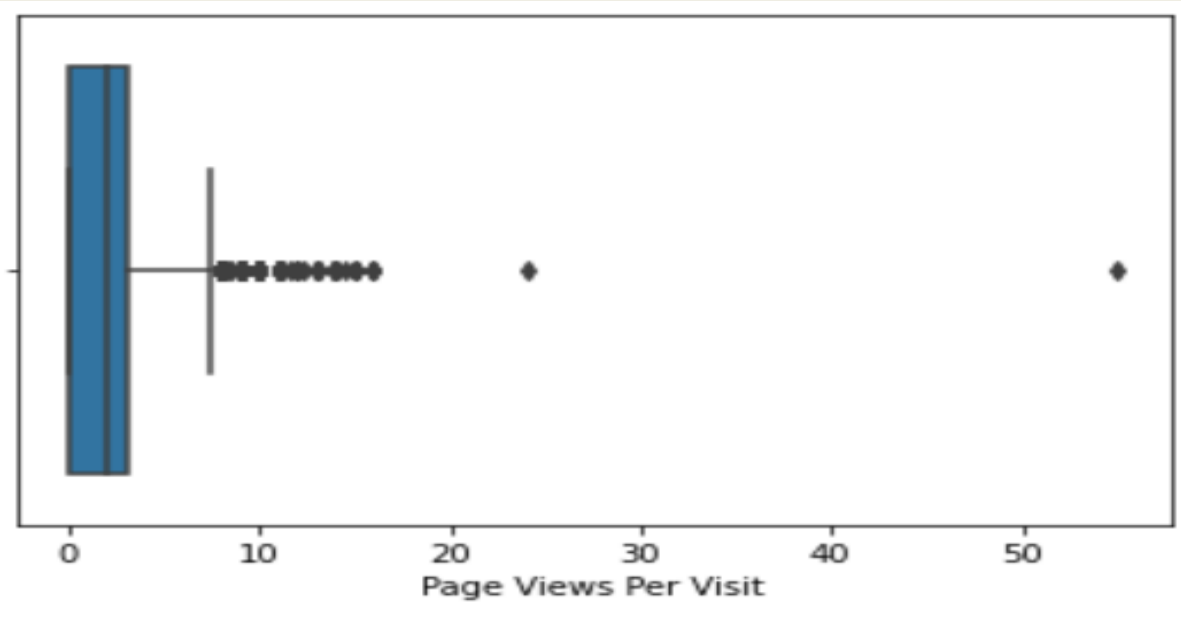
MODEL FLOW CHART



DATA CLEANING

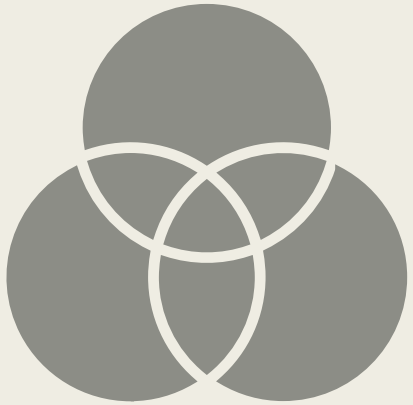
- Huge amount of missing value present on categorical columns, To avoid information loss we imputed missing value as “others”.
- Lots of features are not important so we dropped that features.
- Dummy variable created for categorical columns.
- Feature scaling did for numerical columns for normalize the range.



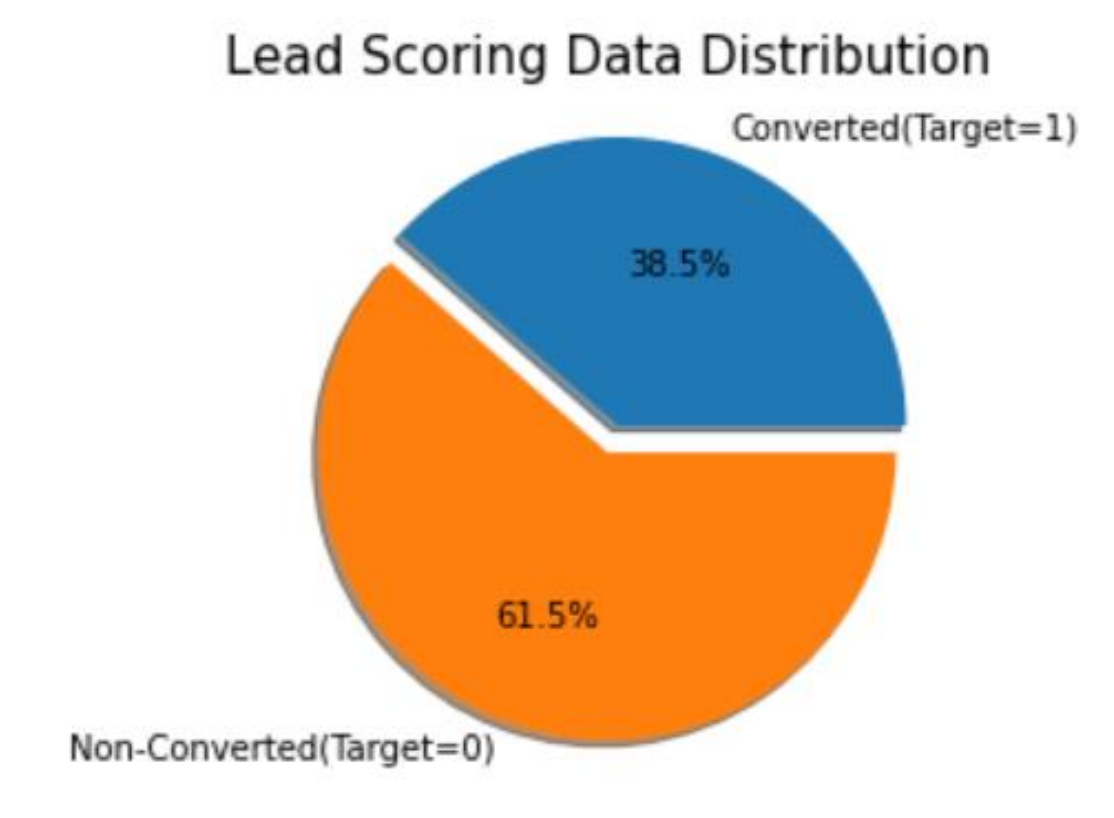


OUTLIERS

- In some cases, If the value is beyond the upper or lower range, then I treated it as an outlier.
- Formula – Upper range Q3+ (1.5*IQR) & Lower range Q1-(1.5*IQR)
- Total Visits and Page views per visit features has outliers.

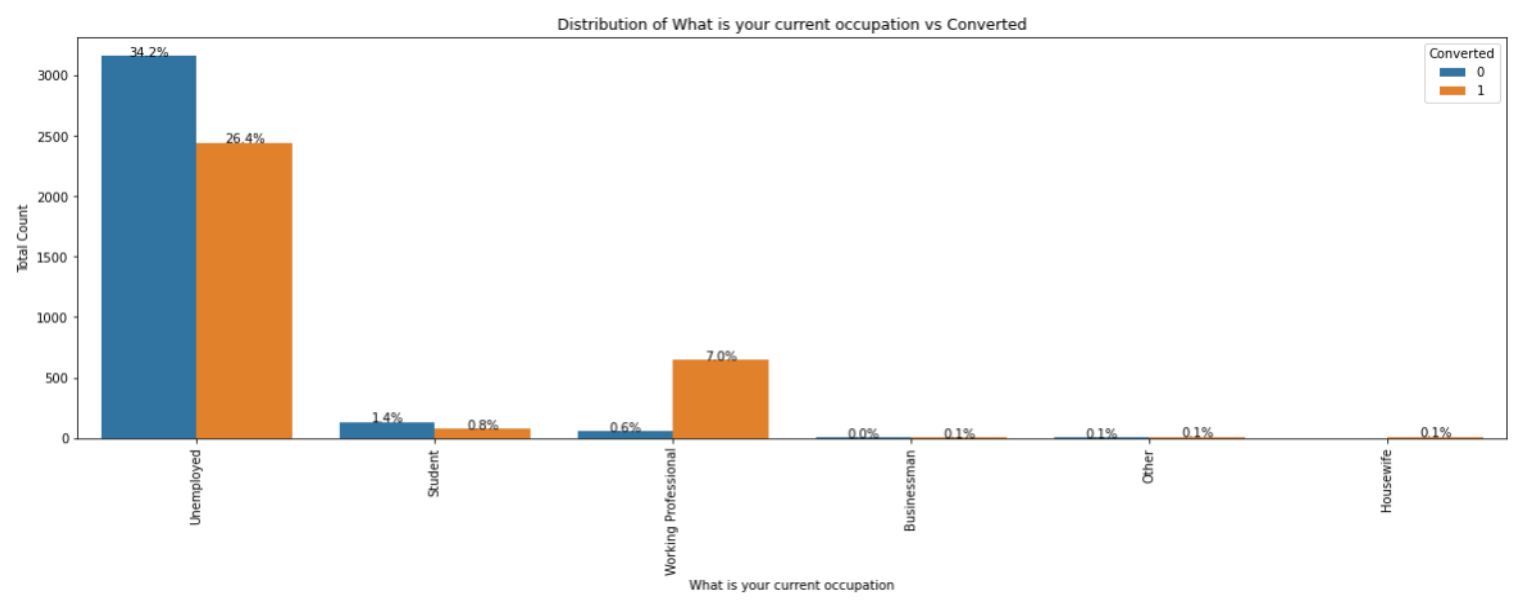


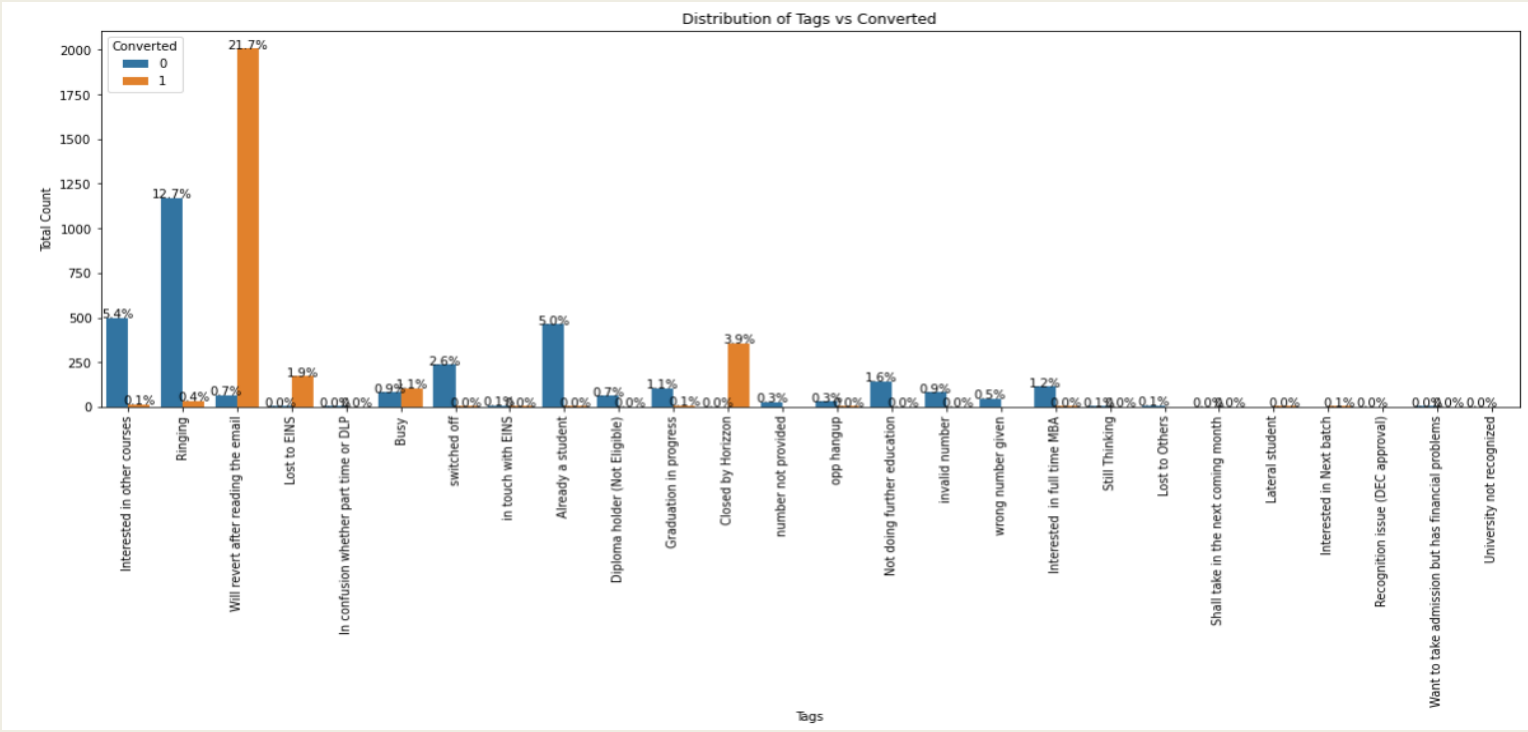
UNIVARIATE / BIVARIATE ANALYSIS



TARGET COLUMN

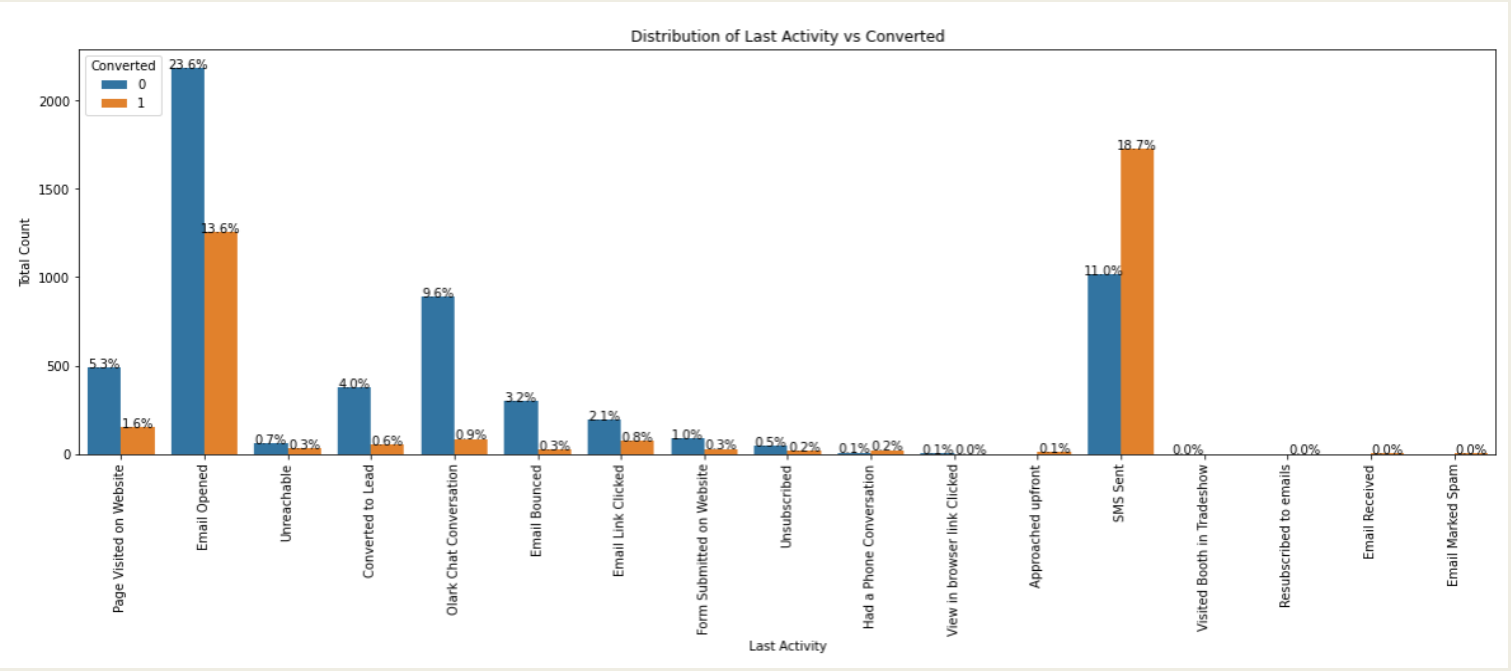
- Converted is target feature.
- This is Balanced Data, Around 38.5% of customers is converted, and the remaining 61.5% of customers is not converted.

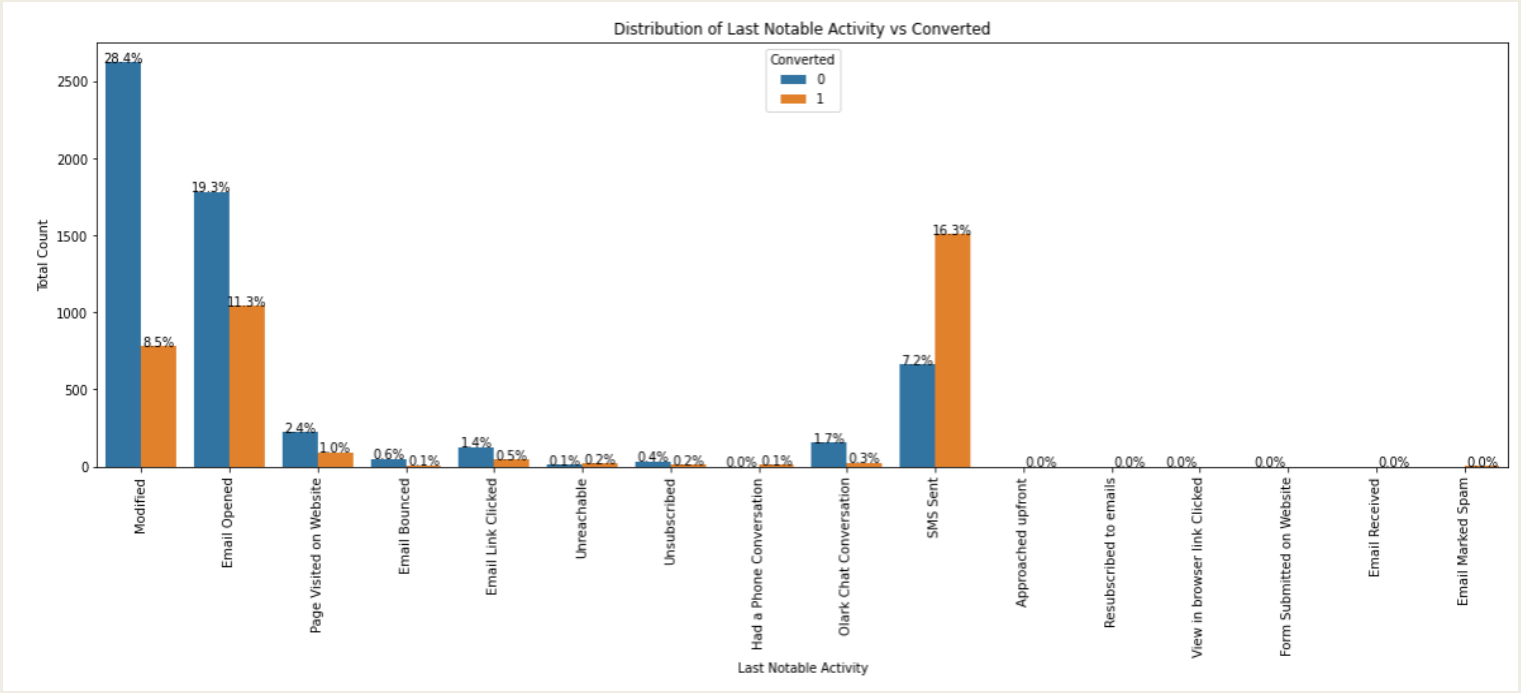




ANALYSIS W.R.T TARGET COLUMN

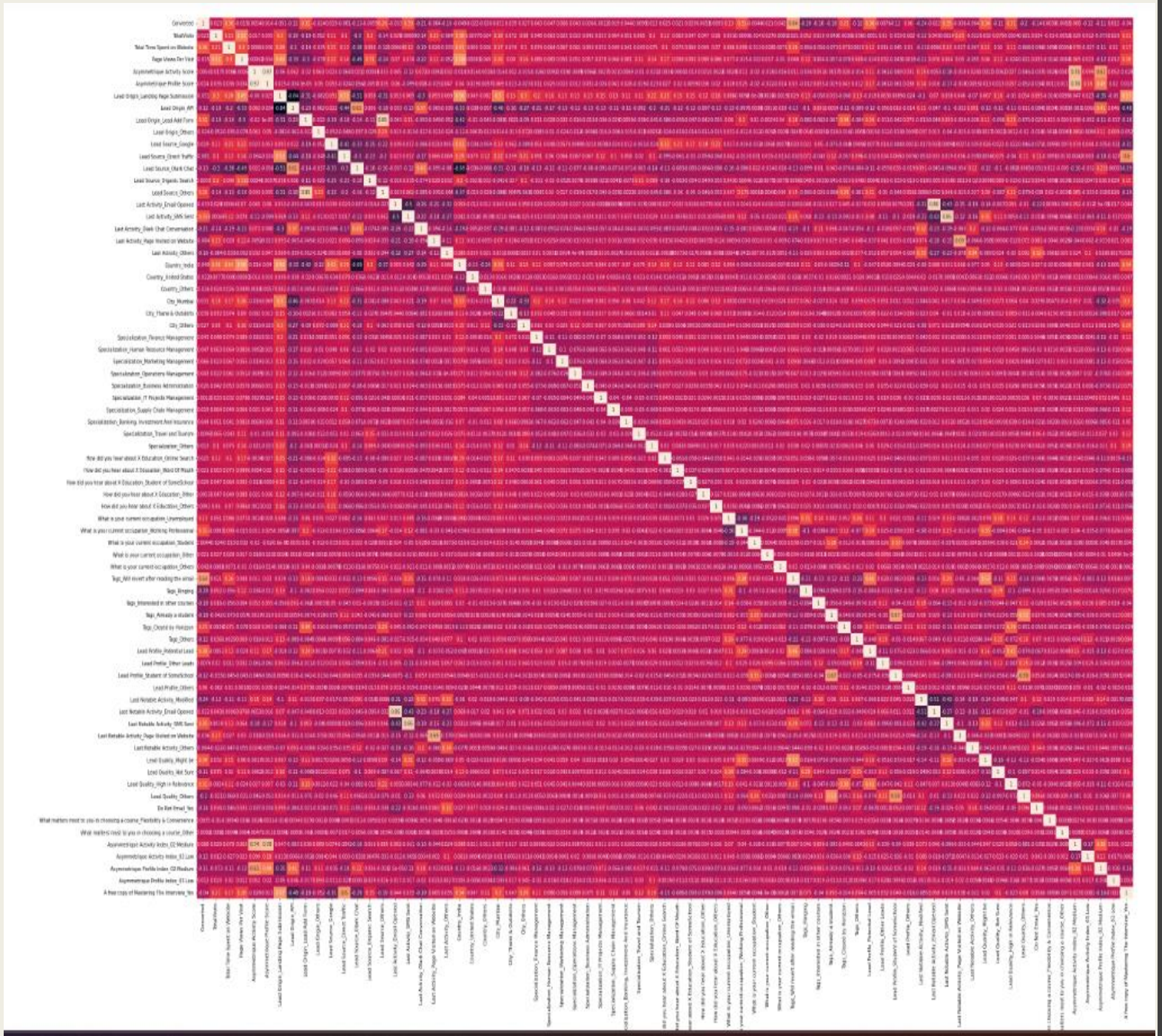
- Working professional has high chance to take a course.
- Unemployed also taking the course but need to look closely, as they won't have money to buy courses.
- “Will revert after reading the mail” has very high chance to take the course.





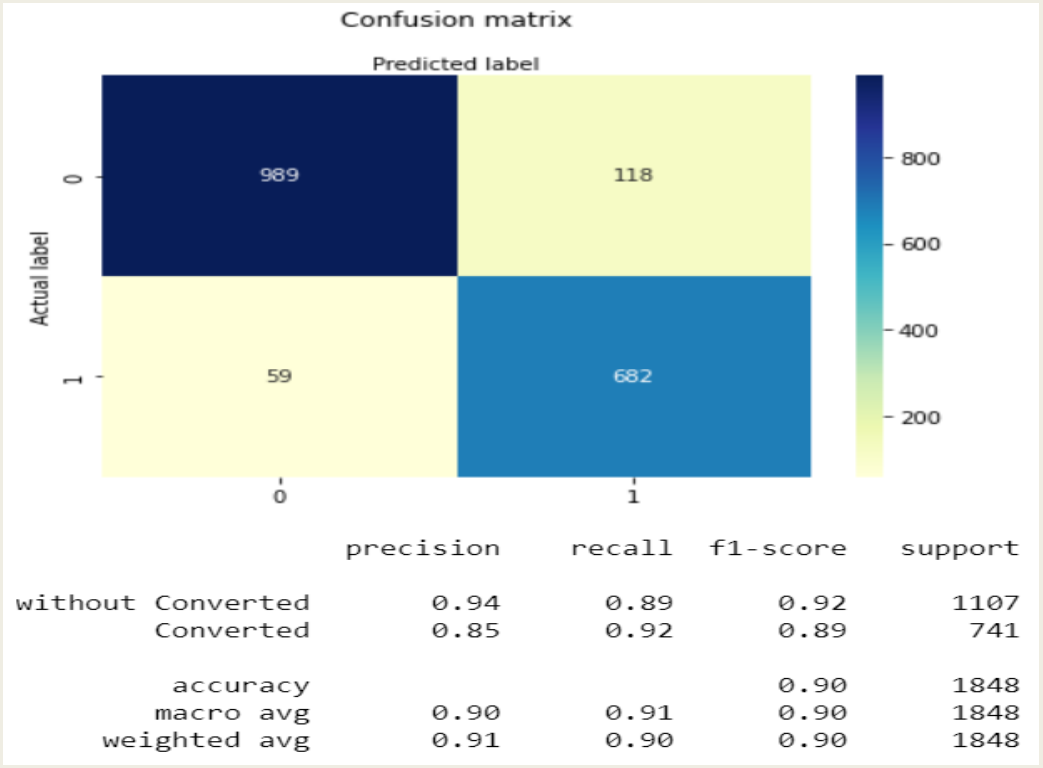
ANALYSIS W.R.T TARGET COLUMN

- Email opened is good signal that customer is interested on this, or customer is active.
- During the marking company must focus on more on more mail and SMS sent to customers.
- If customer modified the last notable activity, then its good lead to company that customer interested.

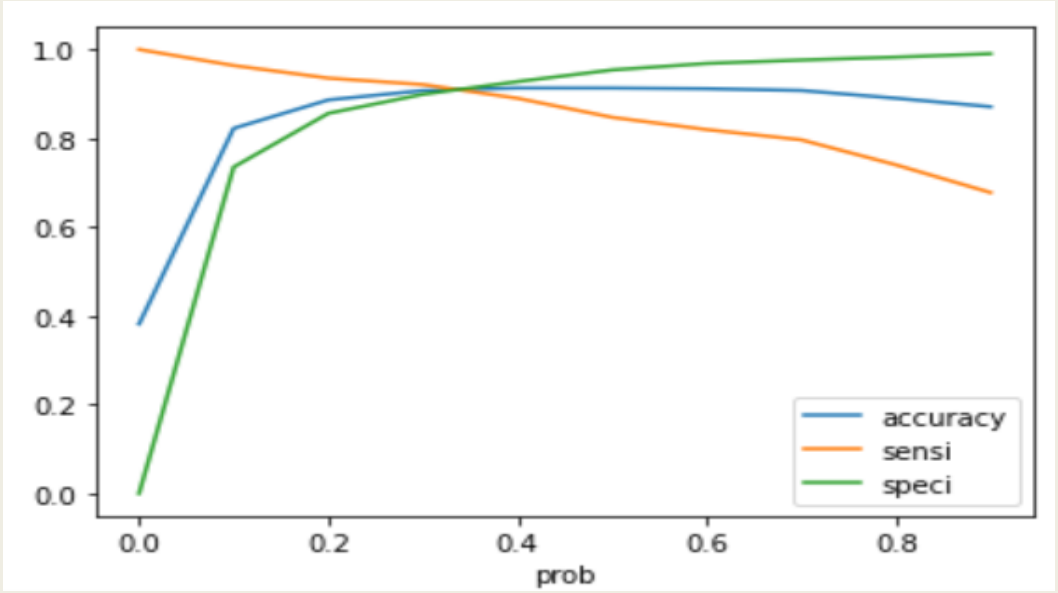


CORRELATION PLOT ON TRAINING DATA

- Few Columns are highly correlated.
- Highly corelated features dropped to avoid multicollinearity.
- Few columns are well corelated with target columns.



	precision	recall	f1-score	support
without Converted	0.94	0.89	0.92	1107
Converted	0.85	0.92	0.89	741
accuracy			0.90	1848
macro avg	0.90	0.91	0.90	1848
weighted avg	0.91	0.90	0.90	1848



MODEL EVALUATION ON TEST DATA

- Evaluated our model using a confusion matrix and focused on recall.
- We obtained a cutoff value of 0.3, which resulted in a recall of 92%, a precision of 85%, an F1 score of 89%, and an accuracy of 90% on the test data.
- The high recall value indicates that our model is good at identifying positive leads, even if it is less precise.

CONCLUSION

Our model has achieved a recall of 92% on the test dataset, which means that it correctly predicts 92% of the positive leads. This meets the CEO's expectation of a lead conversion rate of around 80%.

We should focus on the following customer segments for lead conversion:

- Working professionals
- Customers who spend more time on the X Education website
- Customers who visit the website regularly and have the highest page views per visit
- Customers who have responded to previous emails

We should not focus on the following customer segments:

- Customers who are interested in other courses
- Students

Thank You

BY RAHUL NANDI
ANURAG TIWARI
Rahul GARIGANTI