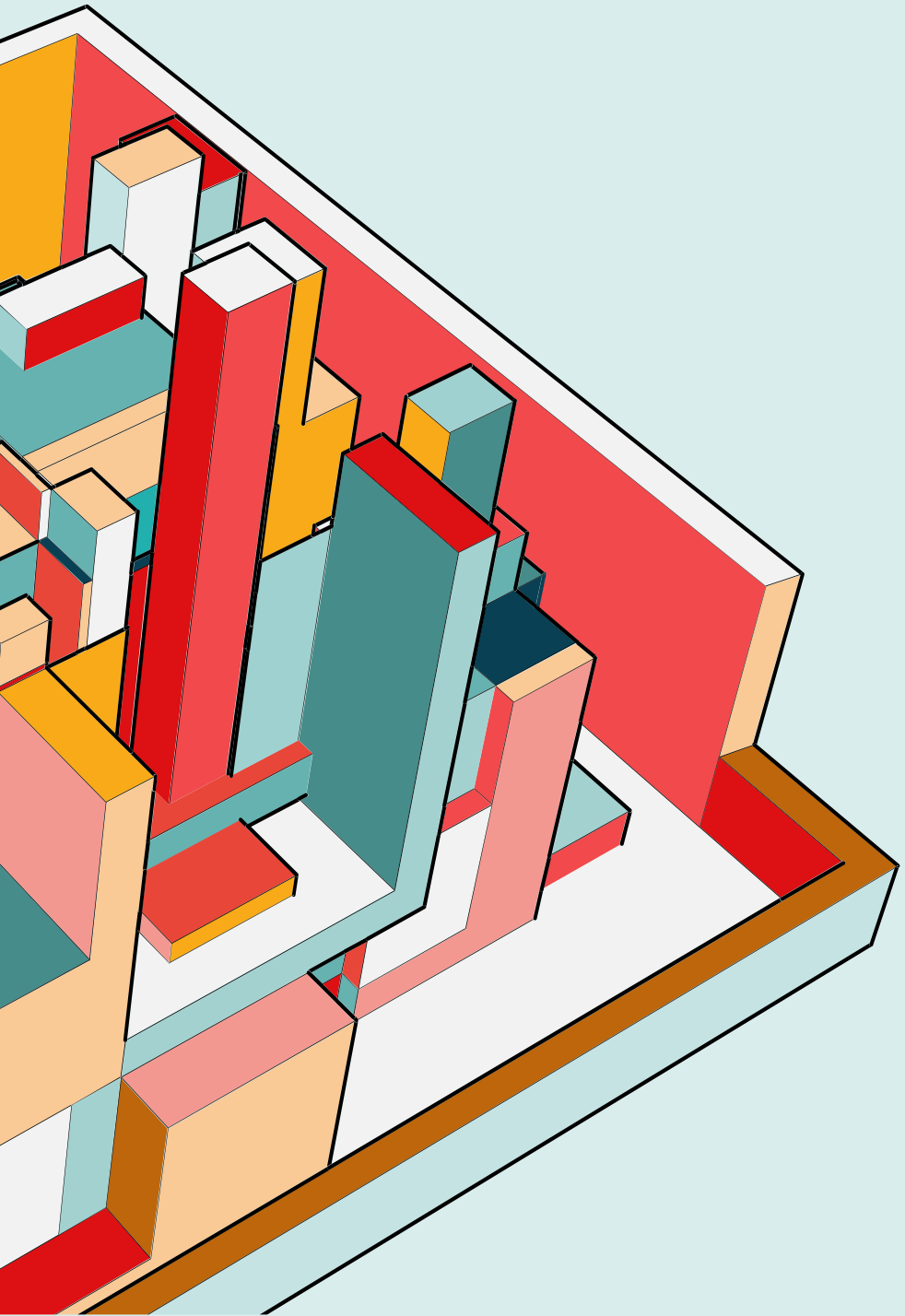
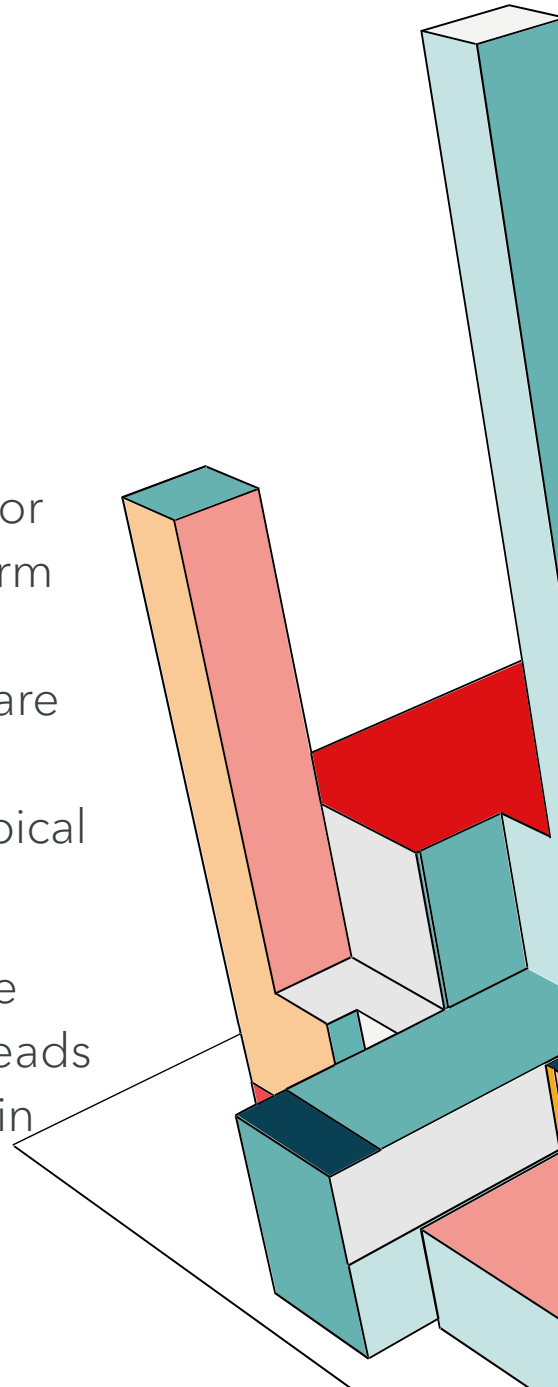


LEAD SCORE CASE STUDY

BY KRITHIKA CHANDRASEKARAN

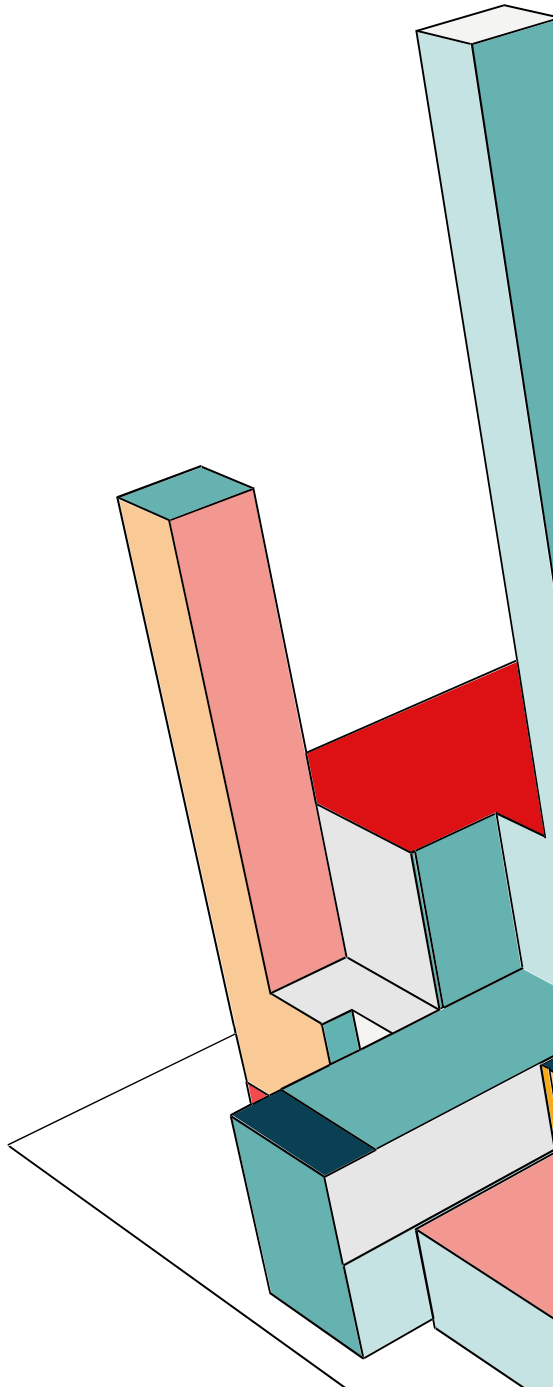
PROBLEM STATEMENT:

- An education company named X Education sells online courses to industry professionals
- The company markets its courses on several websites and search engines 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. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. 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%.
- There are a lot of leads generated in the initial stage, but only a few of them come out as paying customers. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.



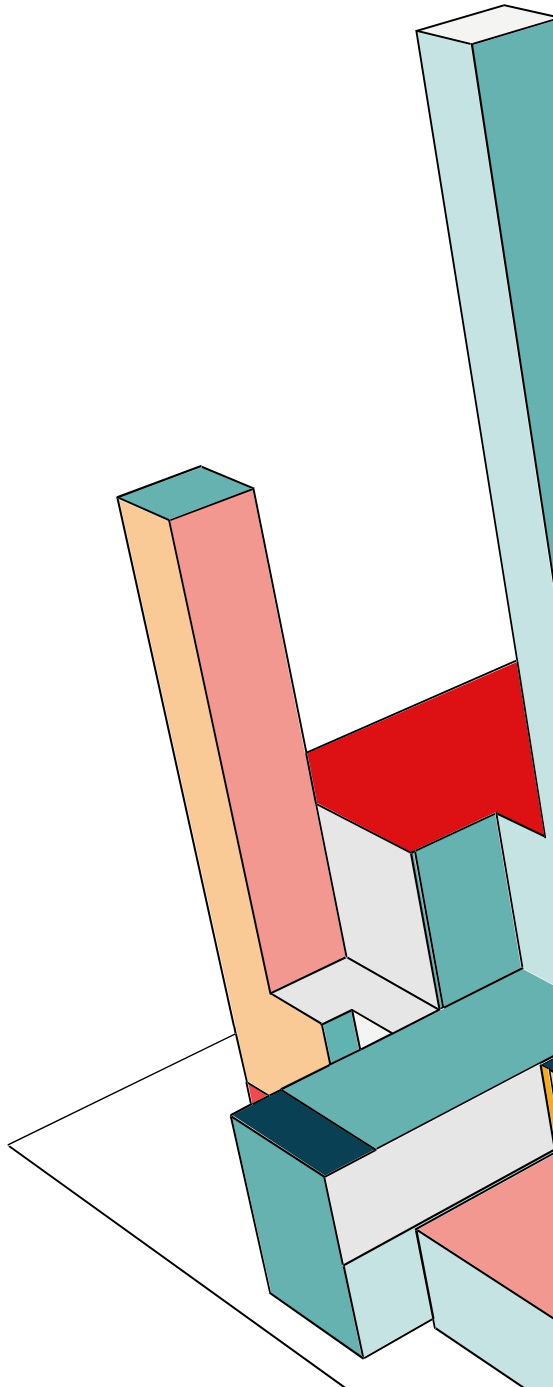
SOLUTION METHODOLOGY

- Understand the Problem Statement
- Data Collection
- Data Preprocessing
    - Data Cleaning
    - Data Transformation
    - Feature Engineering
- Exploratory Data Analysis (EDA)
- Model Selection
- Model Evaluation

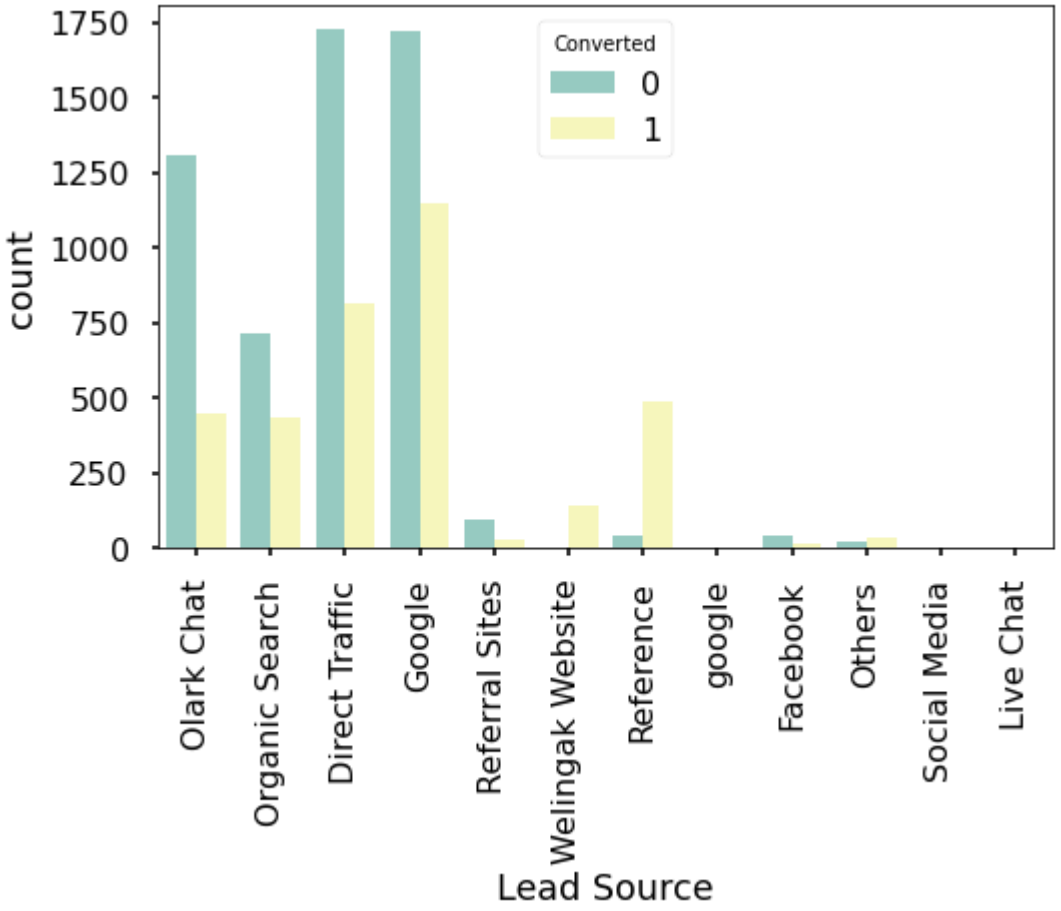


DATA PROCESSING

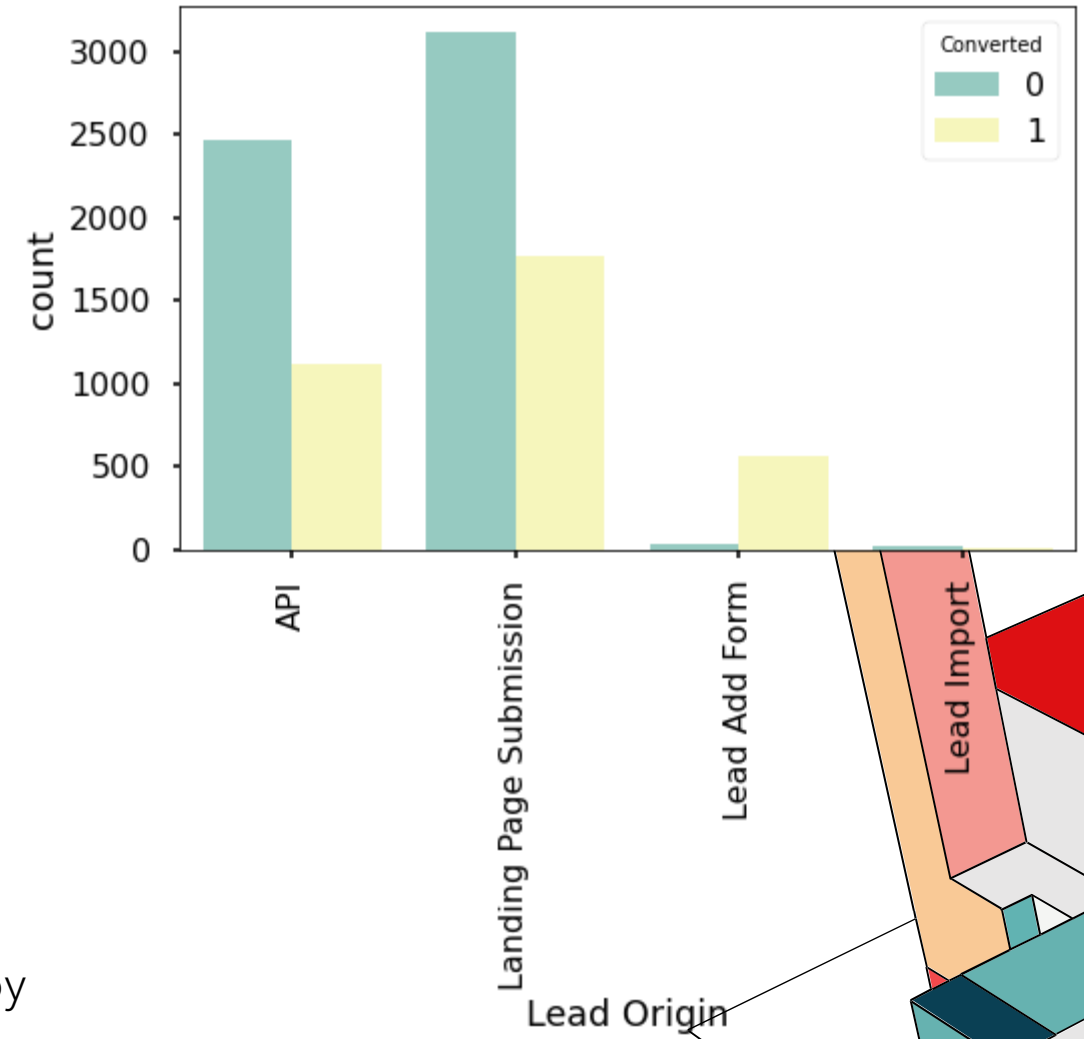
- Dataset contain 37 column and 9240 rows include all NaN
- Converted the categorical field NaN as 'select’
- Observed and dropped the column has more than 45 % of NaN
- Dropped the column has Imbalanced data such as ('Through Recommendations', 'Receive More Updates About Our Courses’, etc).
- Removed “Prospect ID” and “Lead Number” for analysis



EXPLORATORY DATA ANALYSIS (EDA)

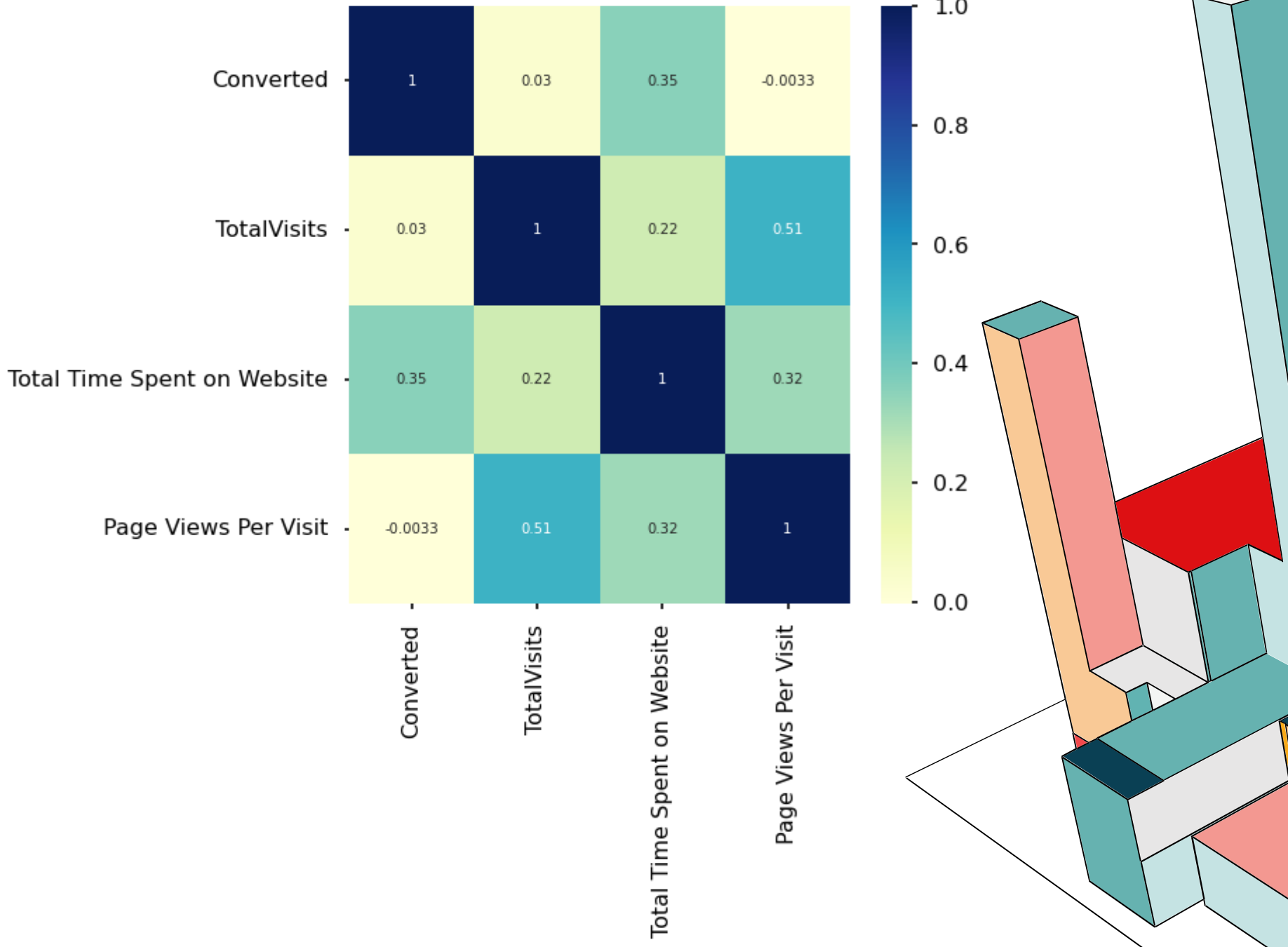


- Max number of lead from Google link and followed by Direct traffic
- Most conversion of leads from the welingak website

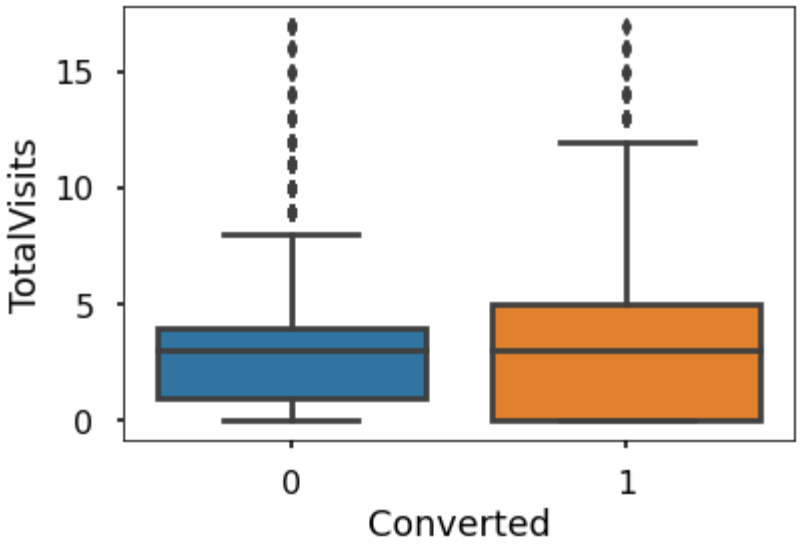


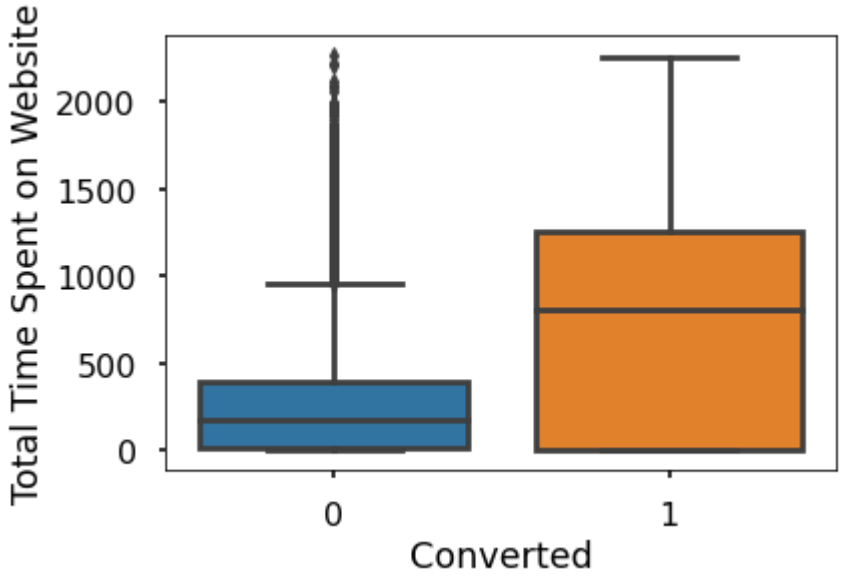
- API and Landing Page Submission bring higher number of leads as well as conversion.

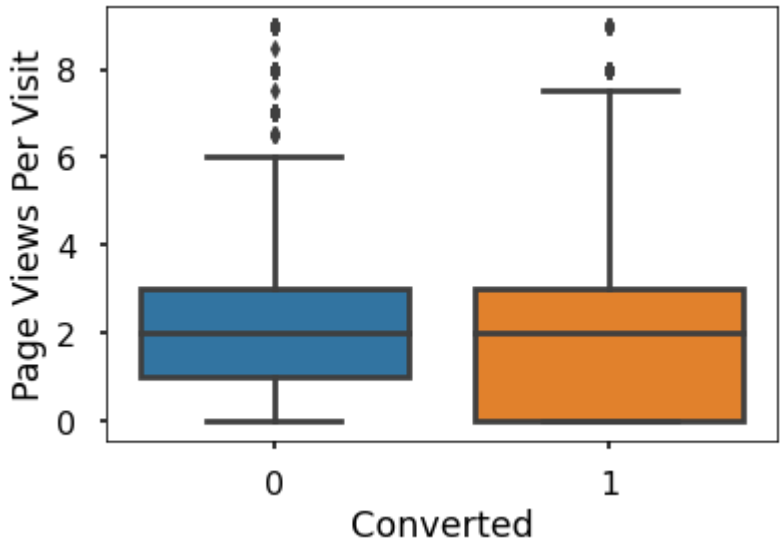
EDA CONTINUES ....

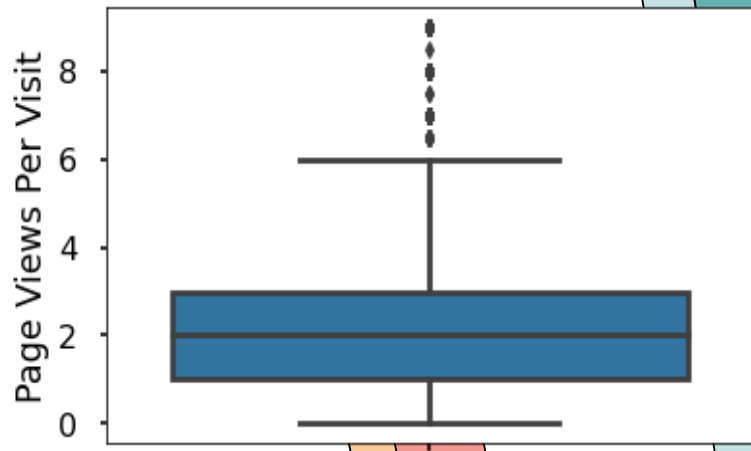


NUMERIC EDA ANALYSIS



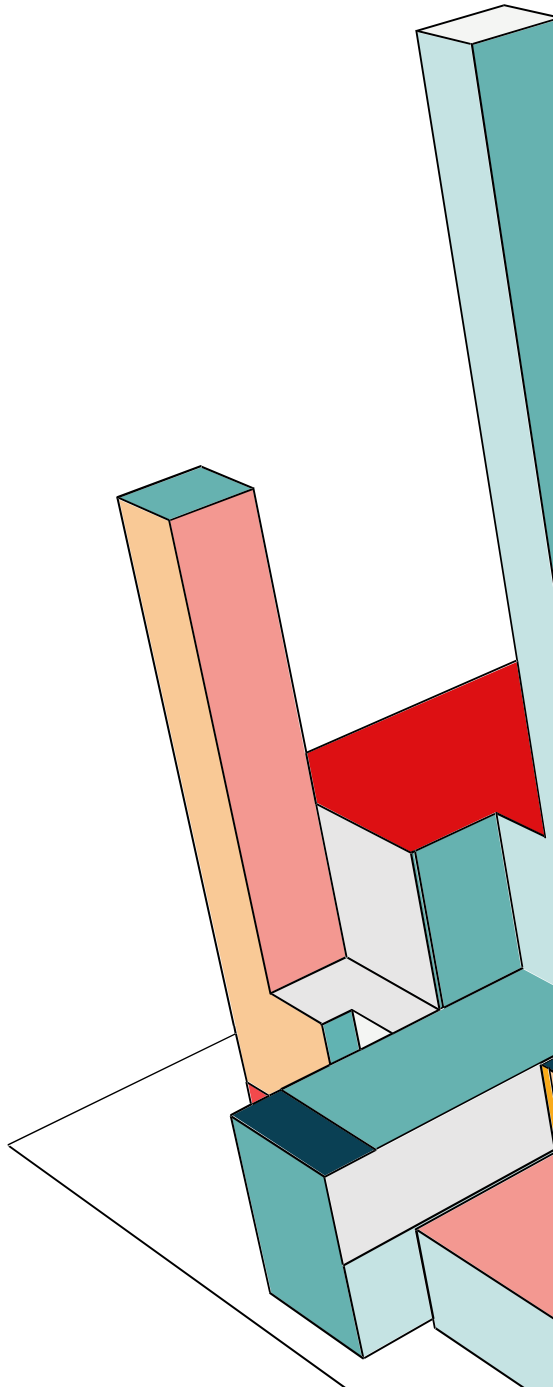






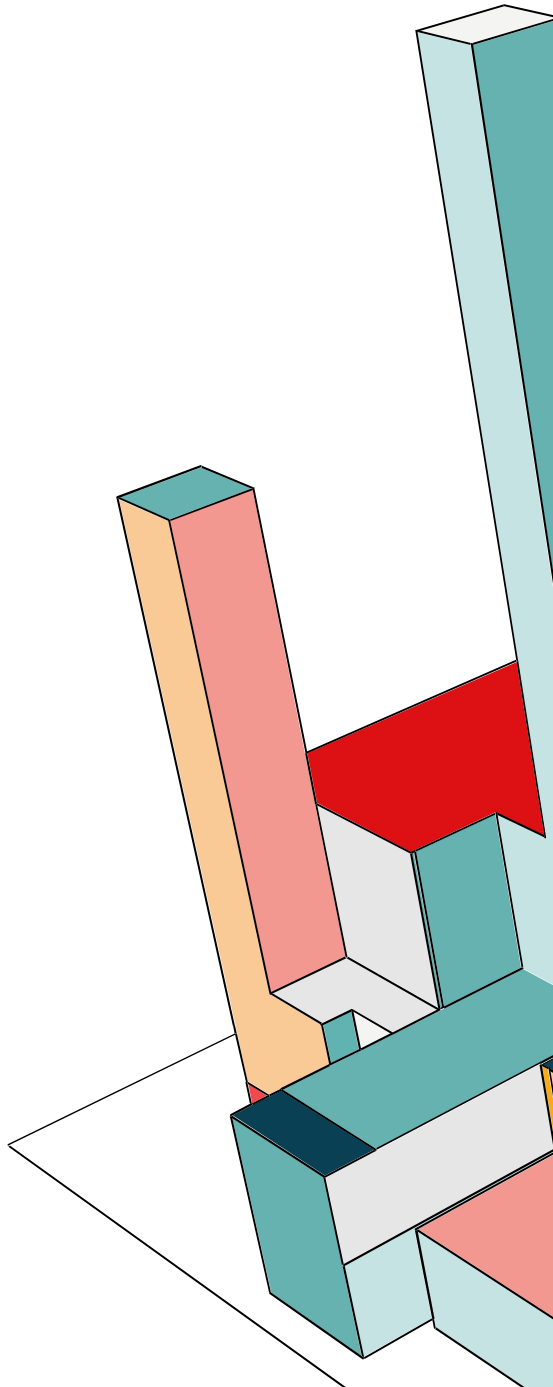
DATA CONVERSION

- Normalized the numeric values
- Created Dummy value for categorical values
- Total rows for analysis 8953
- Total column for analysis 59



MODEL CREATION

- Split train and test dataset from the cleaned dataset
- Tune the model using train dataset using RFE Feature selection from status module.
- Drop the column has high P value and coefficient values
- Test final model against train set and test set to analysis the precision of the model.



MODEL RESULTS

	precision	recall	f1-score	support
0	0.89	0.95	0.92	3881
1	0.91	0.82	0.86	2386
accuracy			0.90	6267
macro avg	0.90	0.88	0.89	6267
weighted avg	0.90	0.90	0.90	6267

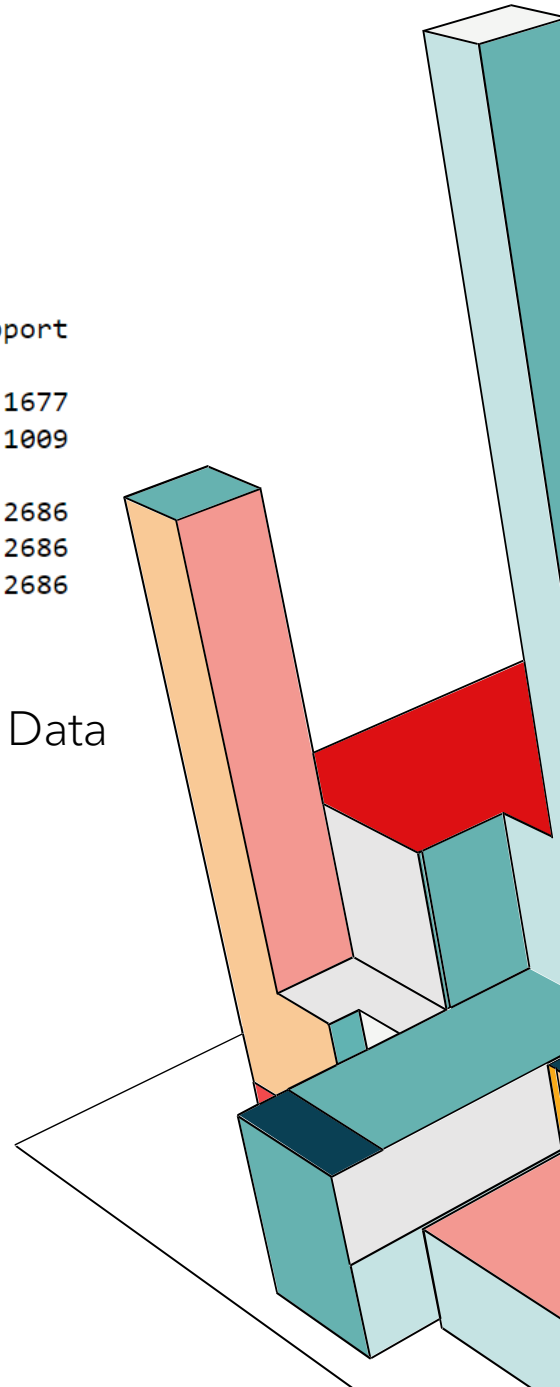
After running the model on the Train Data these are the figures we obtain:

Accuracy : 90.00%
Sensitivity : 81.60%
Specificity : 95.33%

	precision	recall	f1-score	support
0	0.95	0.87	0.91	1677
1	0.81	0.93	0.87	1009
accuracy			0.89	2686
macro avg	0.88	0.90	0.89	2686
weighted avg	0.90	0.89	0.89	2686

After running the model on the Test Data these are the figures we obtain:

Accuracy : 89.00%
Sensitivity : 92.96%
Specificity : 86.88%

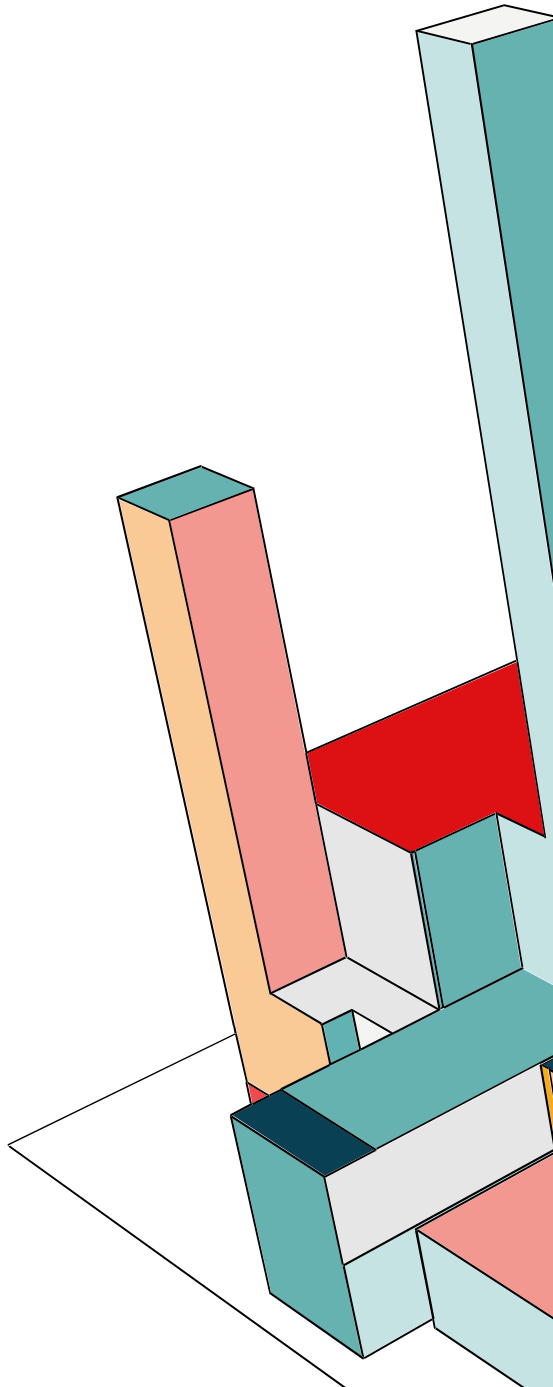


CONCLUSION

We were able to achieve **90% accuracy average on both train and test set**. With the use of this model, X Education will be able to better plan their company plans and develop by concentrating more on the elements that are most important to them.

It was discovered that the following factors, in descending order, were the most important to the potential buyers:

- Welingak website
- Google
- Direct traffic
- Organic search



THANK YOU

KRITHIKA CHANDRASEKARAN