

Lead Score Case Study

Problem Statement

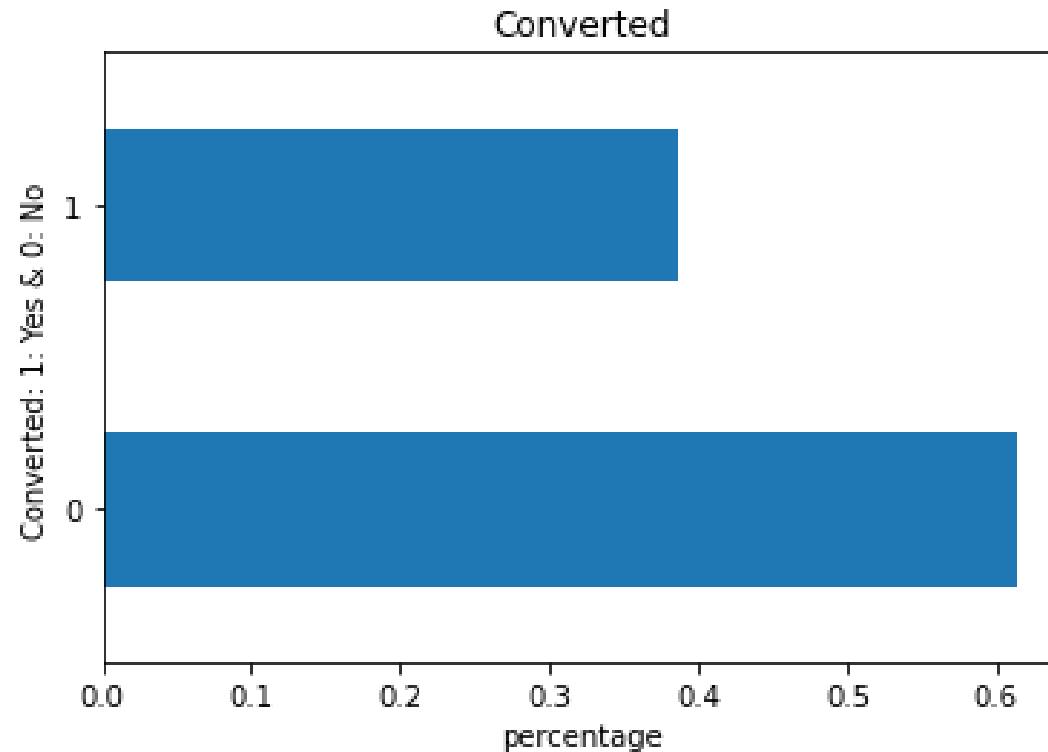
- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the 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.
- The company wants to identify the most promising leads.
- For successful identification, X Education wants to build a ML model.

Approach

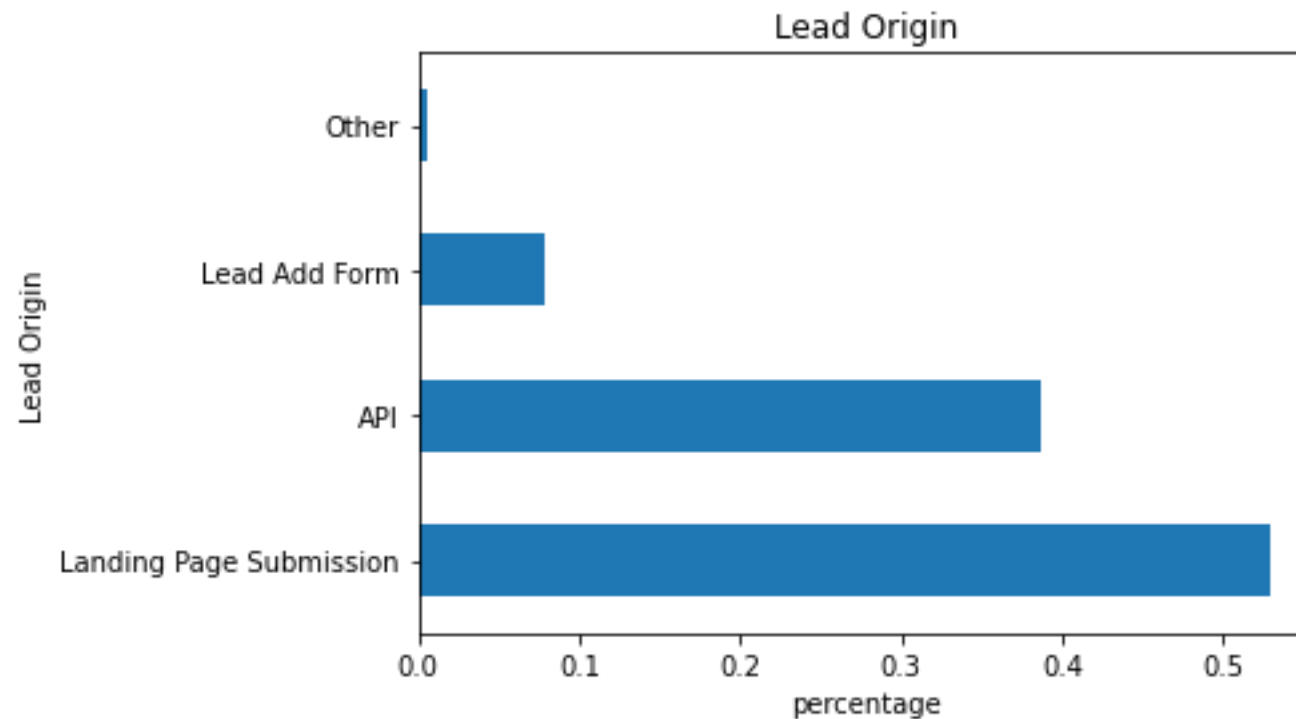
- Data Preparation :
 - Checked data description and shape.
 - Checked and handled NAN and missing values.
 - Dropped columns with large amount of missing values.
 - Imputation of values if required.
 - Checked and handle outliers.
- EDA :
- Univariate Analysis of data.
- Bivariate Analysis of data.
- Feature Scaling and Dummy Encoding of data.
- Logistic Regression for classification.

Exploratory Data Analysis

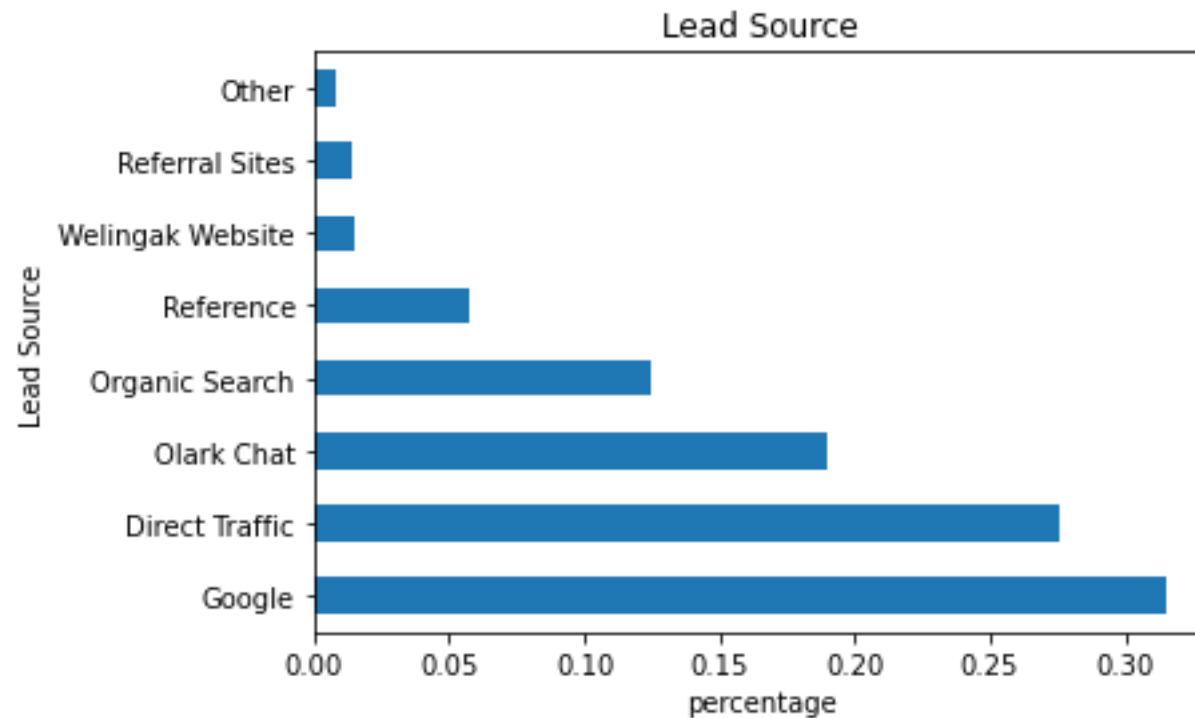
Univariate Analysis of target variable - Converted



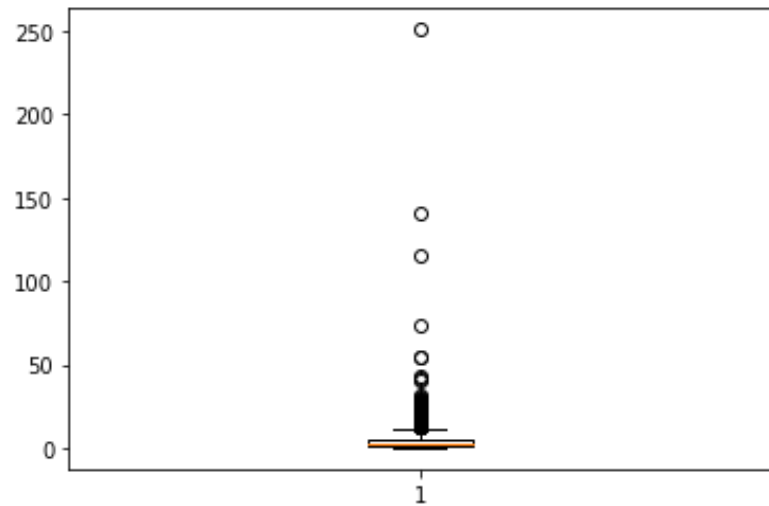
Univariate Analysis of Lead Origin



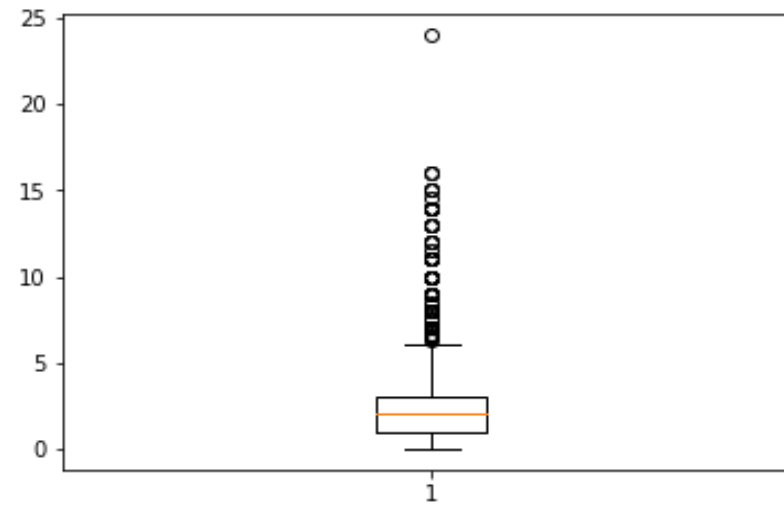
Univariate Analysis of Lead Source



Outlier Analysis



TotalVisits



Page Views Per Visit

EDA Summary

- Around 40% of the leads converted.
- In "Lead Origin", "Landing Page Submission" is more than 50%, the second highest is through "API".
- In "Lead Source", "Google" is the highest through which leads search on, second and third highest is "Direct traffic" & "Olark Chat" respectively.
- More than 80% leads opted for "Email contact" & around 95% opted for "call contact".
- In "Last Activity", around 40% opened email and around 30% SMS Sent.
- Around 70% of leads are from country "India", near 25% country are "unknown" and rest 5% are others.
- In "Specialization" more than 35% data is "unknown", near 10% are from "Finance Management".

EDA Summary

- Near 60% of leads are unemployed, around 25% are unknown, around 10% are working professional and near 5% are student.
- Everyone in this dataset opted for "no" to this variable "I agree to pay the amount through cheque".
- Near 70% of leads choosing this course for "Better Career Prospects".
- API and Landing Page Submission have 30-35% conversion rate.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.

Model Building

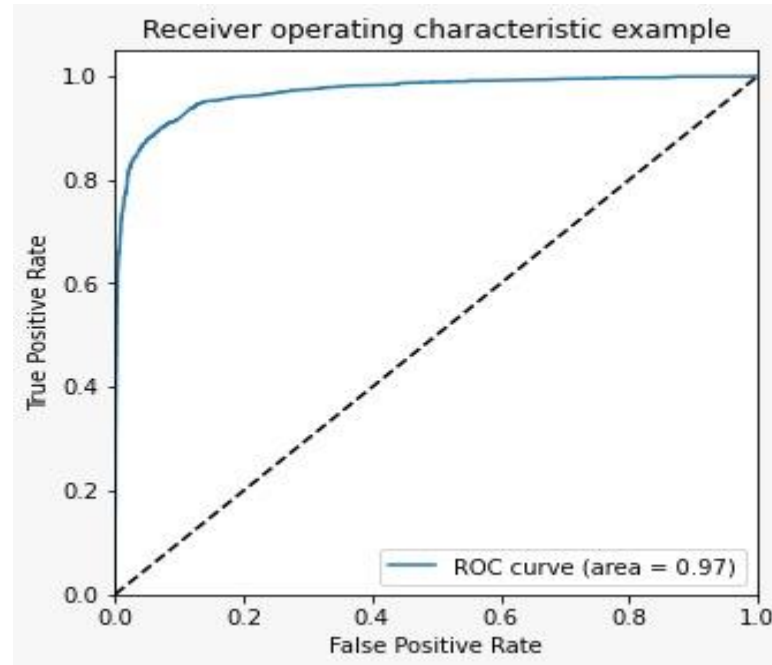
- Data split into train and test sets by 70:30 ratio.
- RFE and P-value for Feature Selection.
- Final Logistic Regression model with 16 variable after feature selection with VIF less than 5.

	Features	VIF
4	Country_unknown	1.81
3	Last Activity_SMS Sent	1.56
12	Tags_Will revert after reading the email	1.53
14	Last Notable Activity_Modified	1.53
7	What is your current occupation_unknown	1.45
1	Total Time Spent on Website	1.42
9	Tags_Closed by Horizon	1.13
0	Do Not Email	1.11
11	Tags_Ringing	1.11
2	Lead Source_Welingak Website	1.08
15	Last Notable Activity_Olark Chat Conversation	1.06
8	Tags_Busy	1.05
10	Tags_Lost to EINS	1.04
6	Specialization_Travel and Tourism	1.03
13	Tags_switched off	1.03
5	Specialization_Hospitality Management	1.01

Performance of the model

Performance of the final model with 16 columns having VIF less than 5.

- Sensitivity = 0.86
- Specificity = 0.96
- Precision Score = 0.92
- Recall score = 0.86
- ROC curve with area under = 0.97



Performance of the model on Test data

- Performance of the model on Test data:
- Accuracy score = 0.93
- Sensitivity = 0.895
- Specificity = 0.955
- The model predicted the metrics of test set under good agreement with the train set.

Conclusion

- People spending more time on the website are more likely to be converted.
- People headed through Google, Direct traffic and organic search are more likely to be converted.
- People with current occupation as working professionals are most likely to be converted among others.
- People with Last Activity as SMS and Chat conversation are more likely to be converted.
- Keeping in mind the above hot leads, X Education can focus on those potential buyers and convert them.