# Summary of the Lead Scoring Case Study

1. Overview of the problem
2. Data and its modelling approach
3. Results of the Data Analysis
4. Conclusions

1. Overview of the Problem

   An education company named X Education sells online courses to industry professionals. When those consumers fill out a form with their email address or phone number, they are classified as leads.

   The typical lead to successful sale conversion rate at X education is around 30% which is very poor.

   Therefore, the company wants to identify the most potential leads, also known as 'Hot Leads' with higher chances of conversion thereby focusing more on them and improving the sales.

   The company wants a logistic regression model which can assign a lead score to each of the leads in such a way that a higher lead score means higher conversion chance and lower score means lower chances.

2. Data and its modelling approach

   The dataset has approx. 9000 data points with attributes such as Lead Source, Total Time Spent on the Website, Total Visits, Last Activity, etc., which may or may not help in deciding whether the lead can be converted or not.

   The target variable is column "Converted" containing past information of leads being Converted i.e. (in form of 1) or not (i.e. 0).

   *Modelling Approach:*

   - Data Understanding
   - Data Cleaning – Removed and Imputed missing values, Dropped Imbalanced columns
   - Data Visualization – include plots, charts, etc.

- Data preparation – Created Dummy variables
- Model Building – Fitted a logistic regression model on the dataset
- Model Evaluation – Ran the model on the test data and used Evaluation metrics such as Accuracy, ROC, confusion matrix, etc.

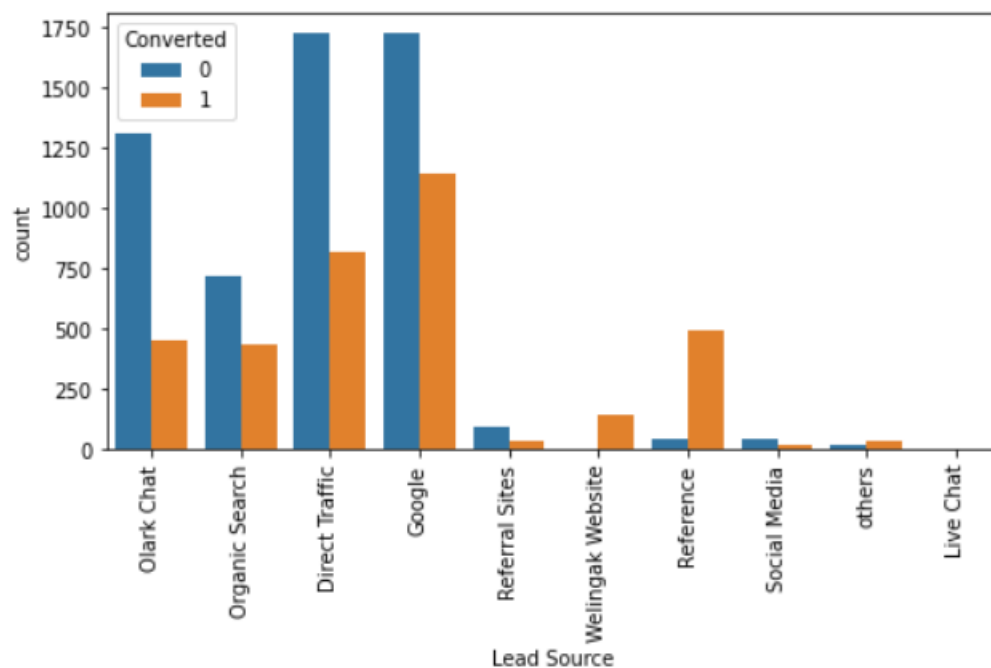3. Results of the Data Analysis

Following are the plots helpful in the analysis:
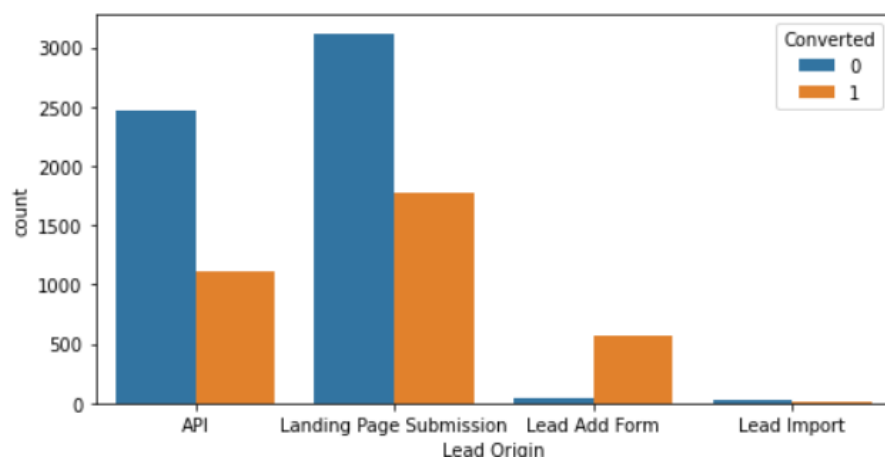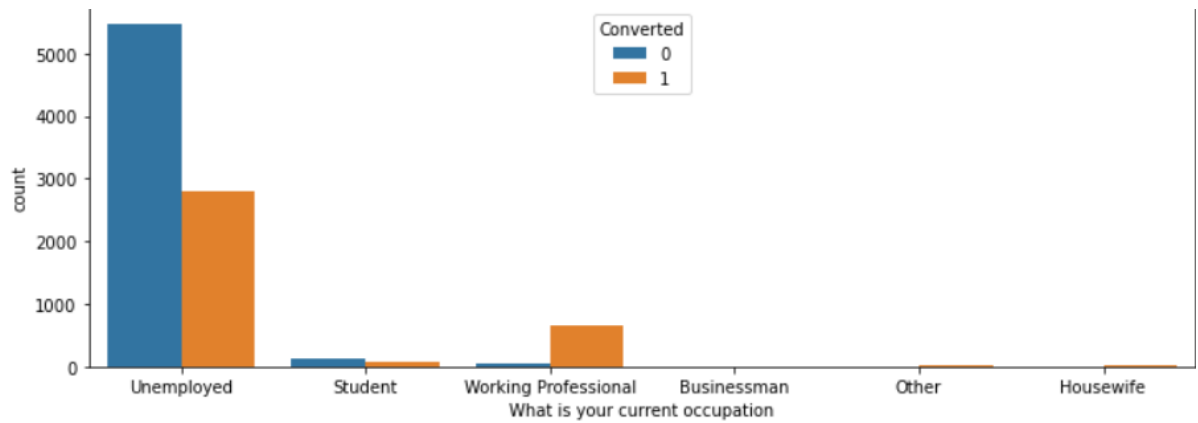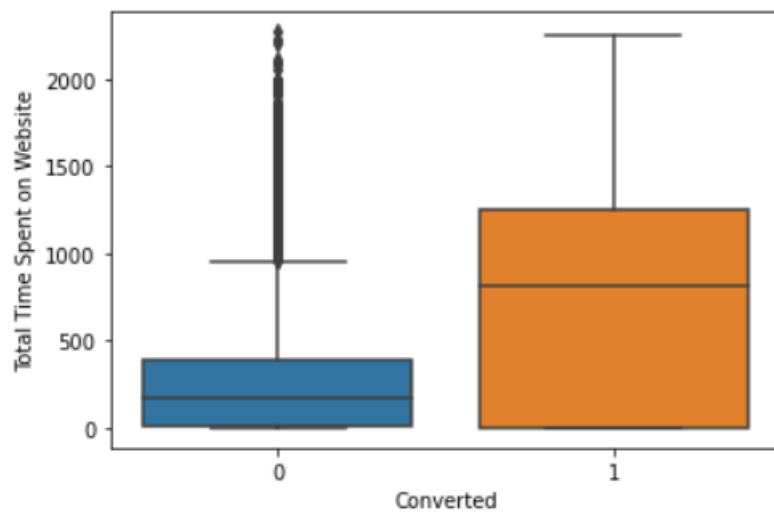


Fig.1


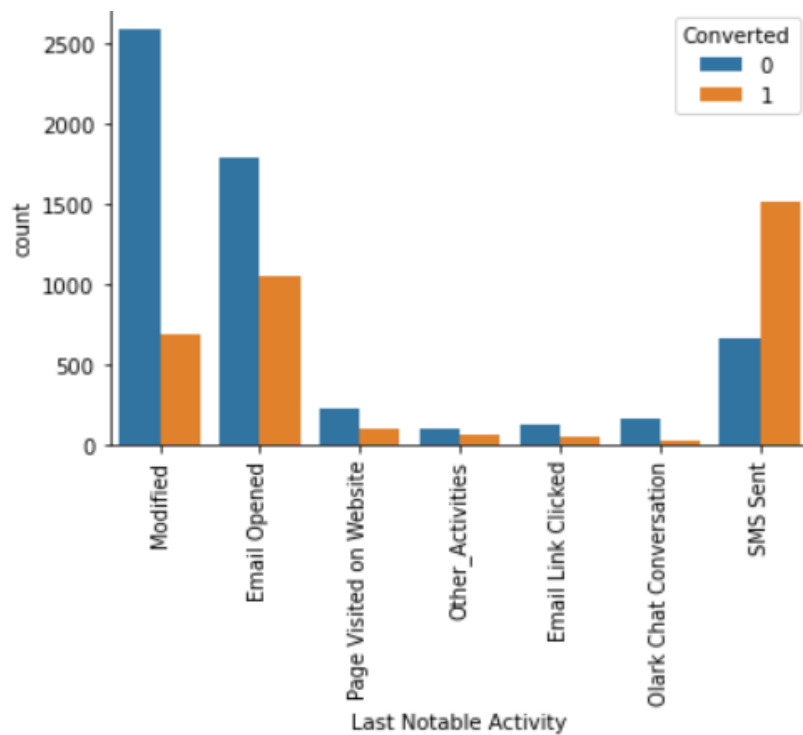
Fig. 2

Fig. 3



Fig. 4



Fig. 5

From the above plots the categories where leads have high rate of conversion can be found as mentioned in the following:

- Fig.1: the leads came through Reference
- Fig. 2: leads with origin as Lead add form
- Fig. 3: leads who are working professionals
- Fig.4: leads spending most time on website
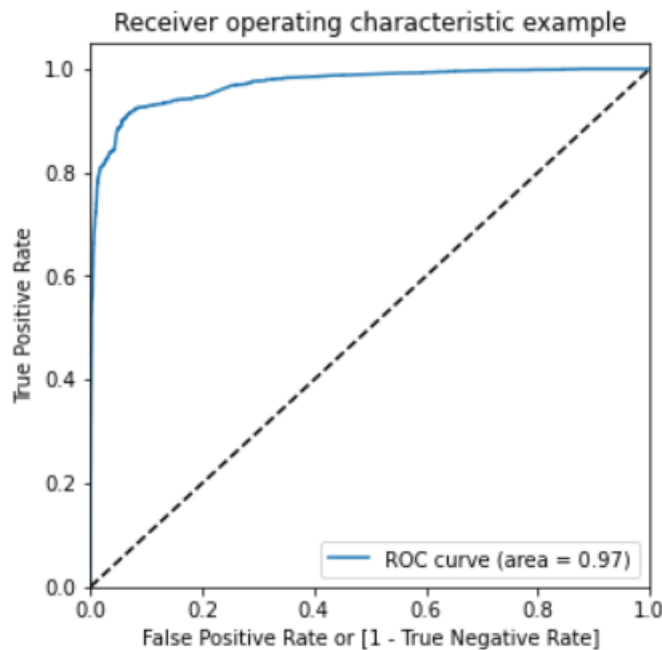- Fig.5: leads whose last notable activity as 'SMS sent'



Fig.6: ROC curve

4. Conclusions

- The accuracy of the model is about 92% and ROC = 0.97.
- Recommendations to the sales team to increase the conversion rate and improve the sales are:

  - spending most time on the Website.

  - coming through Direct Traffic, Referral Sites and Welingak Website as their source.

  - whom Origin is from Lead Add Form.

  - Who have their Last Activity as SMS Sent.

  - Those who have been tagged as Closed by Horizzon, interested in other courses, Lost to EINS, Other Tags, Ringing, Will revert after reading the email.

  - Who have their Last Notable Activity as Modified and Olark Chat Conversation.