






# Lead Scoring CaseStudy

By Lokesh, Maitree, and Manikanta

- 
- 
- Contents:**
- 1.Problem Statement
  - 2.Data Preprocessing steps
  - 3.Visualizations through plots
  - 4.Results From Analysis
  - 5.Conclusion
  6. Recommendations 

## Problem Statement :

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- 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.

## Preprocessing Steps for Dataframe:

■ The following changes have been done for better analysis, visualization. The changes done for the required columns are as below:

❑ Columns with unique entries :**Prospect ID, Lead Number** has been removed.

➤ Checked for duplicate rows and if duplicate rows exists dropped it.

❑ Missing Values Treatment:

➤ If there are variables with greater than 30% missing values, dropped those variables since it wouldn't give much information.

➤ Dropped rows with dropna function.

❑ Outlier Treatment:

➤ There are outliers in **TotalVisits** variable, there is lot of variation from 75% to max. We can't cap these since with domain knowledge we can say there can be persons with 250 views..

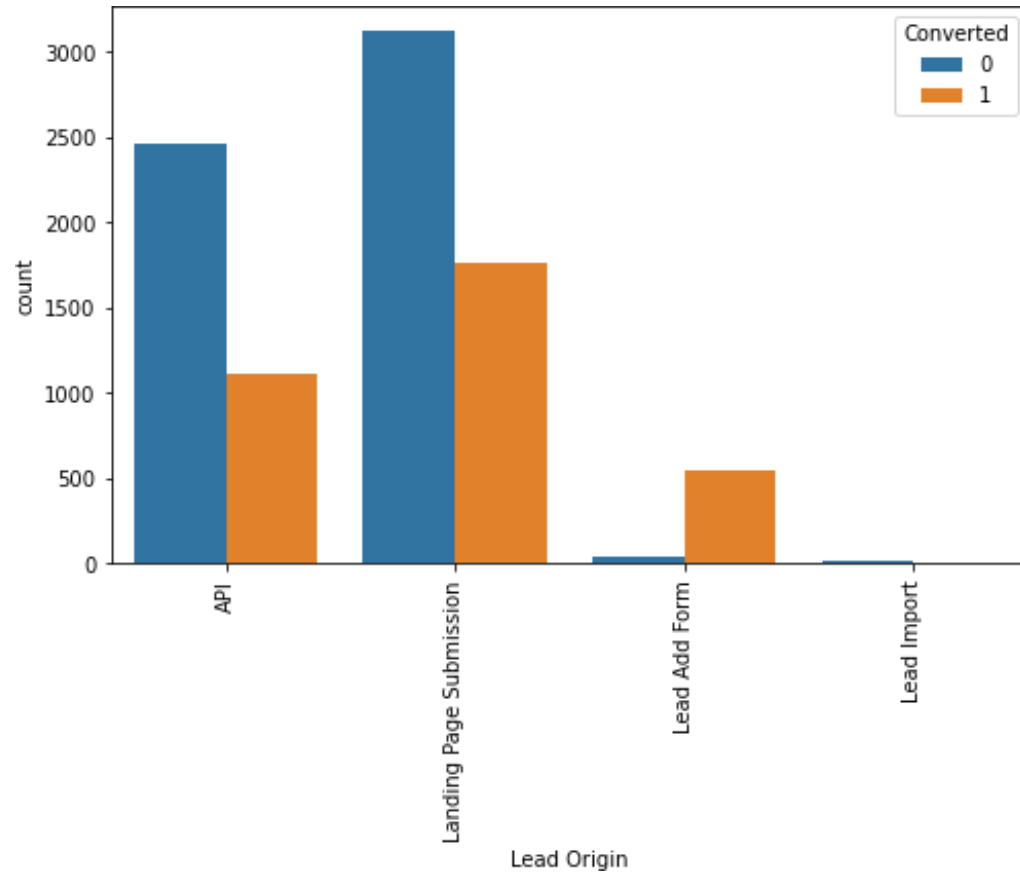
➤ There are outliers in **Page Views Per Visit**, we don't need to deal with these outliers since with the domain knowledge we know there can be views with 50 per visit.

- ❑ There are few columns with 'select' as an entry. 'Select' represent the option is not chosen by the customer, hence can be replaced with nan values
- ❑ Data Imbalance: **Search ,Newspaper Article ,X Education Forums ,Newspaper, Digital Advertisement ,Through Recommendations** these columns had data imbalance so have been dropped
- ❑ Imputation of missing values: Nan values have been imputed with mean or mode wherever necessary



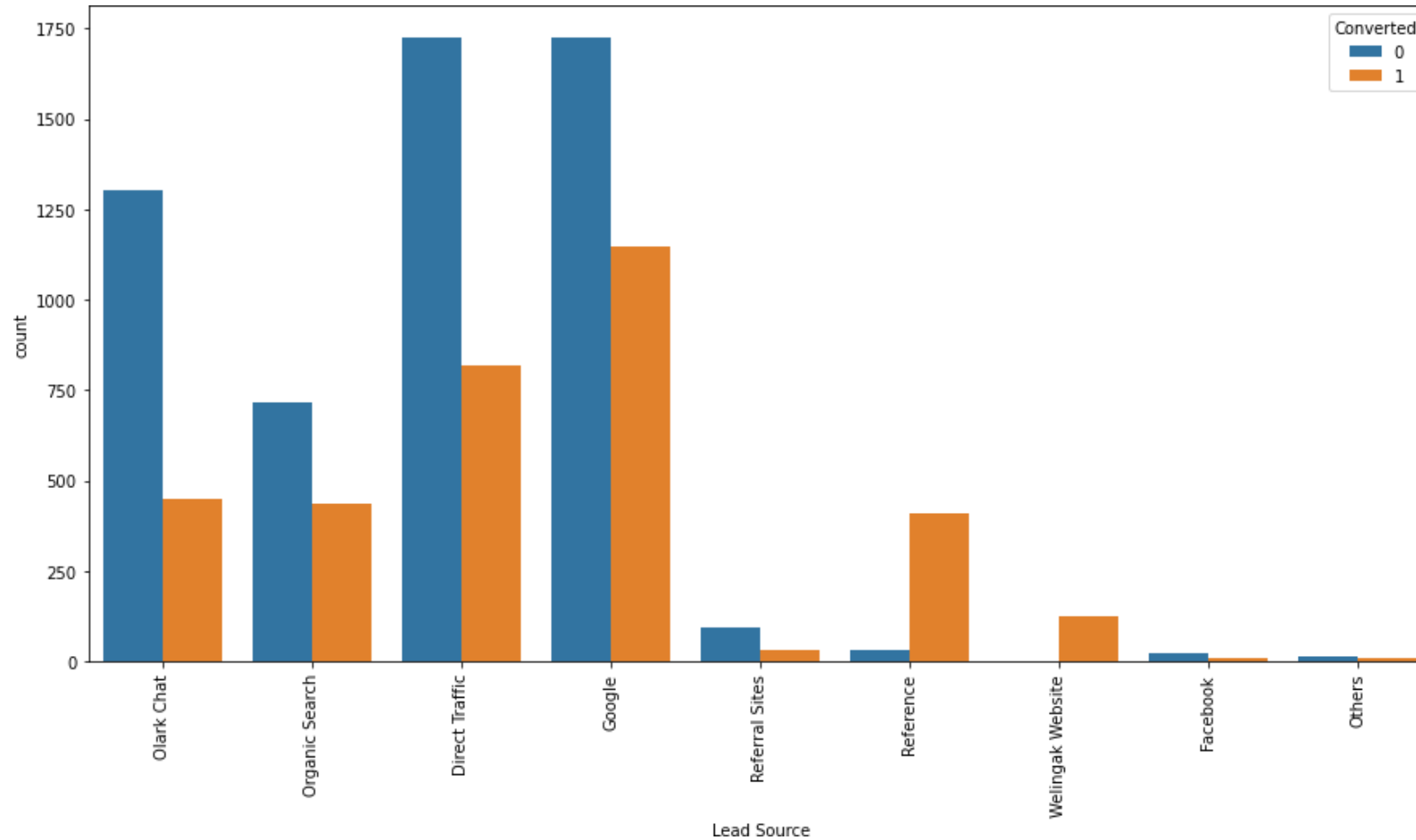
# Visualizations through Plots:

## Lead Origin



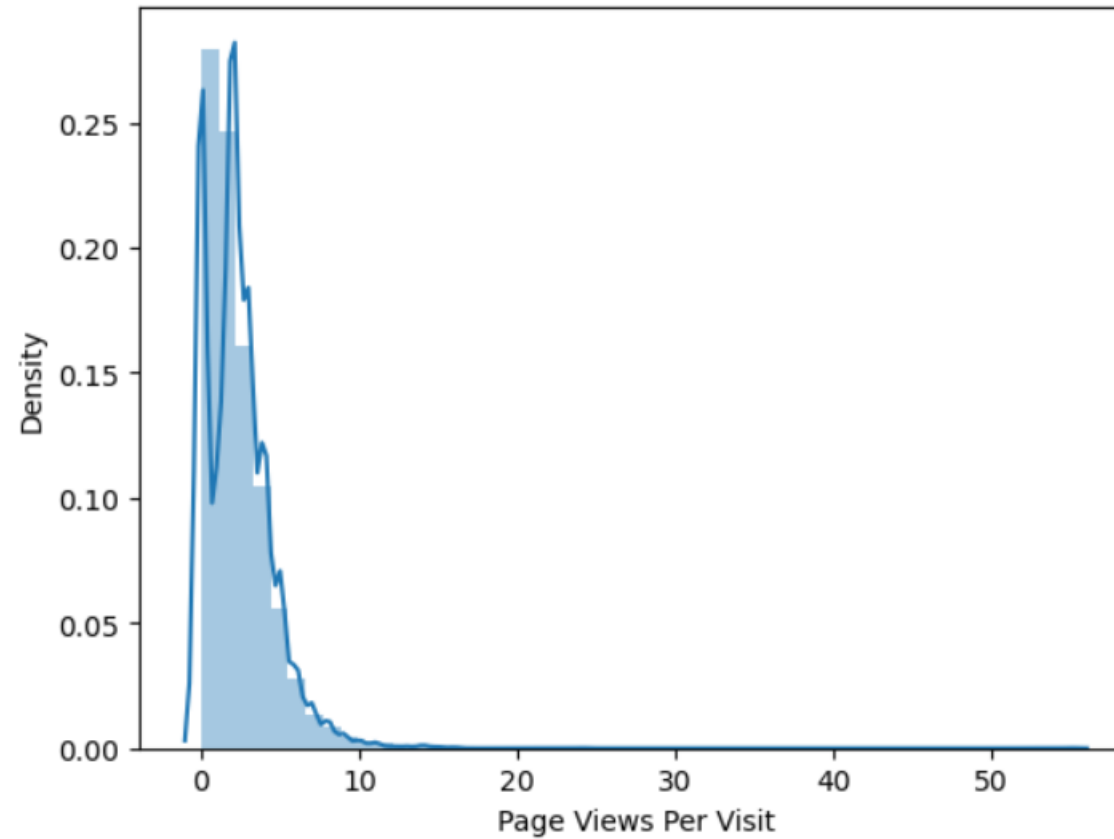
- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.

## Lead Source:



- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.

## Page views Per Visit



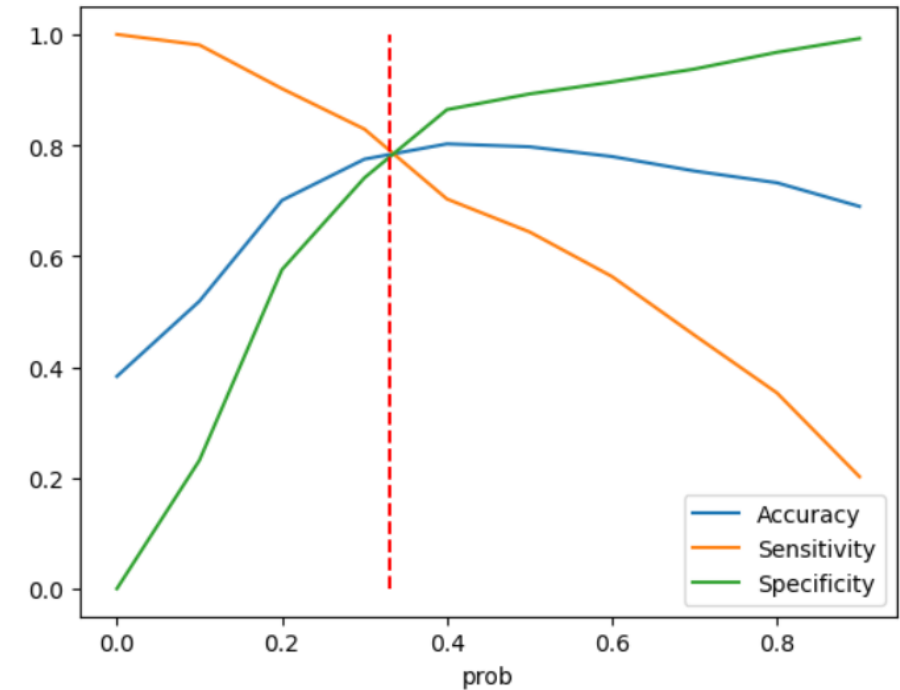
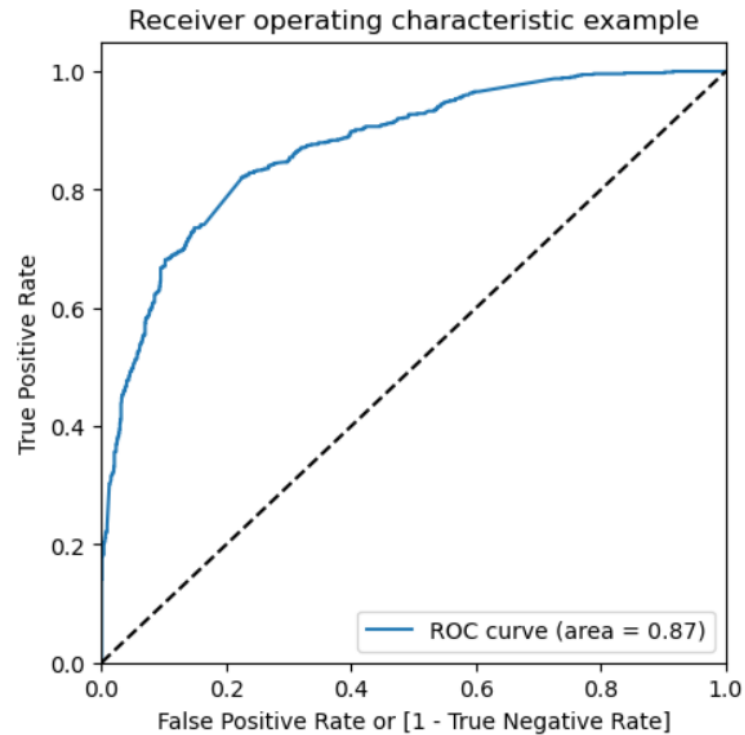
- We see two peaks.
- We see outliers which goes beyond 50 page view per visit
- The average lies somewhere between 0 to 10



## Results from our analysis

Accuracy	= 0.8014
Sensitivity	= 0.7366
Specificity	= 0.8436
False Positive Rate	= 0.1564
Precision	= 0.7542
Recall	= 0.7366
Negative Predictive Value	= 0.831

# ROC Curve



- Finding optimal cutoff point:
- Optimal cutoff probability is that probability where we get balanced sensitivity and specificity.
- From the graph it is visible that the optimal cutoff is 0.33

## **Conclusions**

- **Lead Add Form, current occup\_Working Professional** and **Lead Source\_Welingak Website** are effective factors that contribute to a good conversion rate.
- **Working professionals** and **Unemployed customers** and **Students** tend to have higher conversion rates.
- Referral leads generated by old customers have a significantly higher conversion rate
- Google and Direct Traffic are channels that are showing promising conversion rates.
- Leads whose 'Last Activity' is 'SMS Sent' or 'Email Opened' tend to have a higher conversion rate.
- The 'Others' specialization category is the most common among customers followed by Finance Management, HR Management and Marketing Management

# Recommendations

- 1. Lead Source\_Welingak Website 2. Lead Add Form 3. Current occupation \_Working Professional , these features have high rate of conversion.
- Working Professional should be aggressively targeted as they have high chances of conversion and have better financial situation to pay the service.
- Increasing the frequency of media usage such as Google ads or email campaigns save time and increase the conversion rate.
- Analyzing behavior of customer who spend more time on website can help improving the user experience and increase conversion rate.