

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large, solid red oval is positioned in the center, containing the main title and subtitle. A dark gray, curved shape is visible on the left side, partially overlapping the red oval.

# Lead Scoring Case Study

By lokesh, ashwin

---

**Contents:** Problem Statement

Data Cleaning

▼ Univariate Analysis

Conclusion

---



## 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:

### ❑ Duplicate Rows :

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

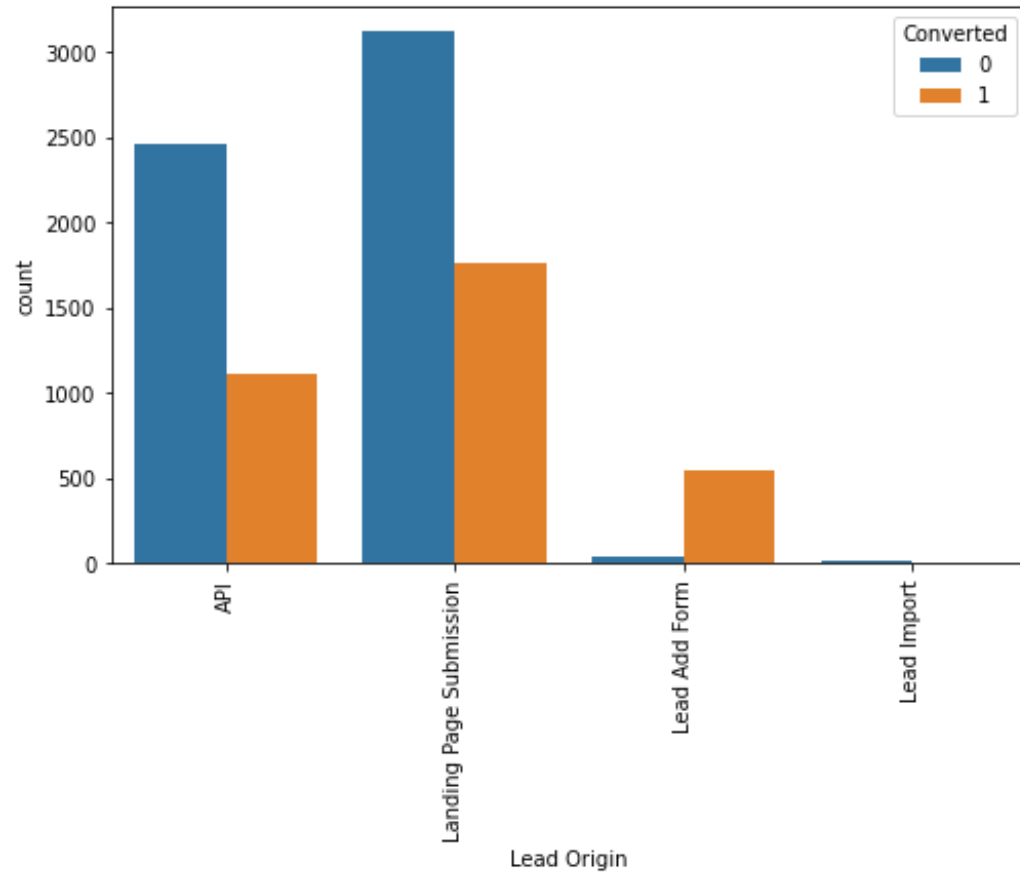
### ❑ Missing Values Treatment :

- 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 :

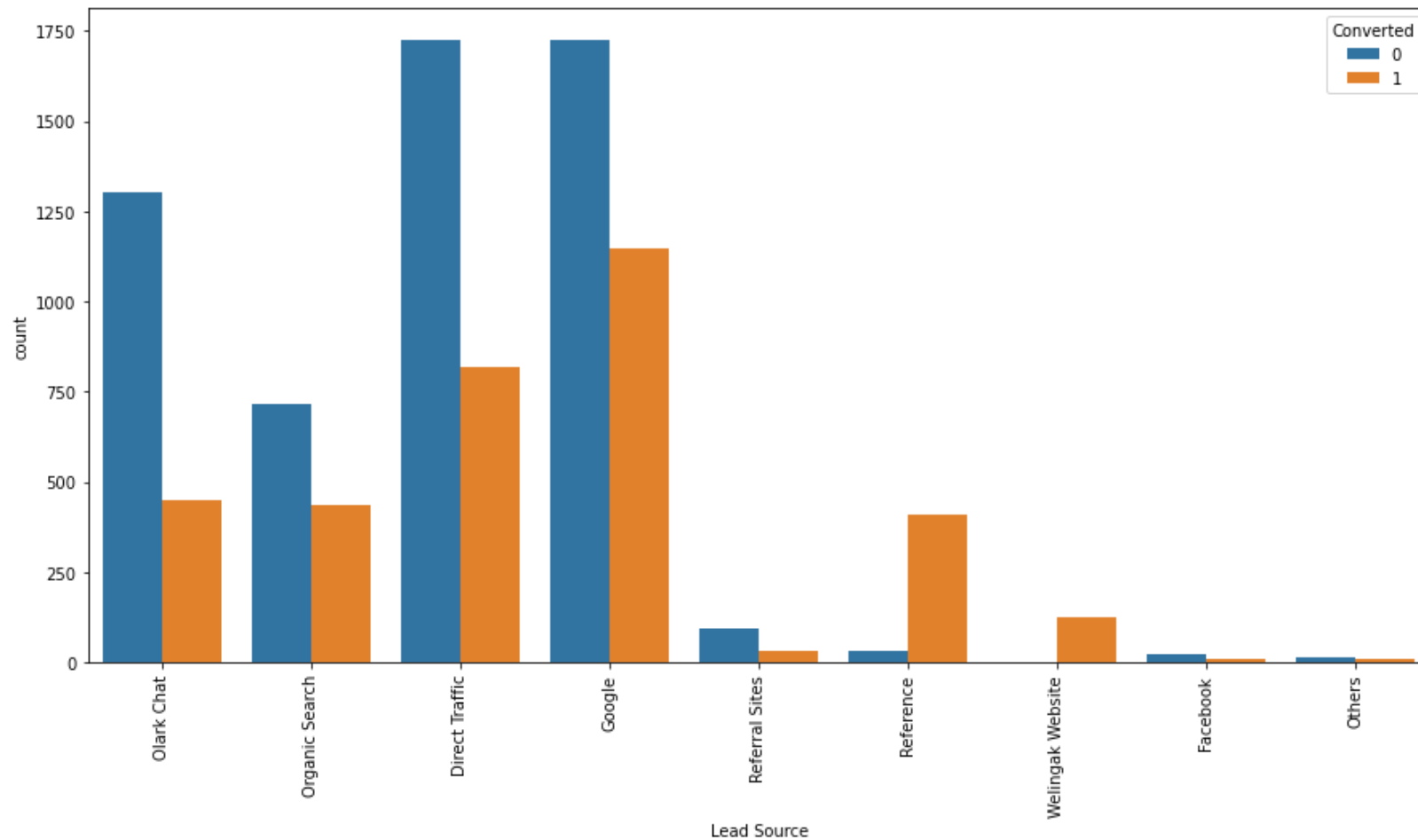
- There are outliers in **TotalVisits** variable, there is lot of variation from 75% to max. We can't cap these since with domain knowledge we can say there can be persons with 250 views..
- There are outliers in **Page Views Per Visit**, we don't need to deal with these outliers since with the domain knowledge we know there can be views with 50 per visit.

## Univariate Analysis : Lead Origin



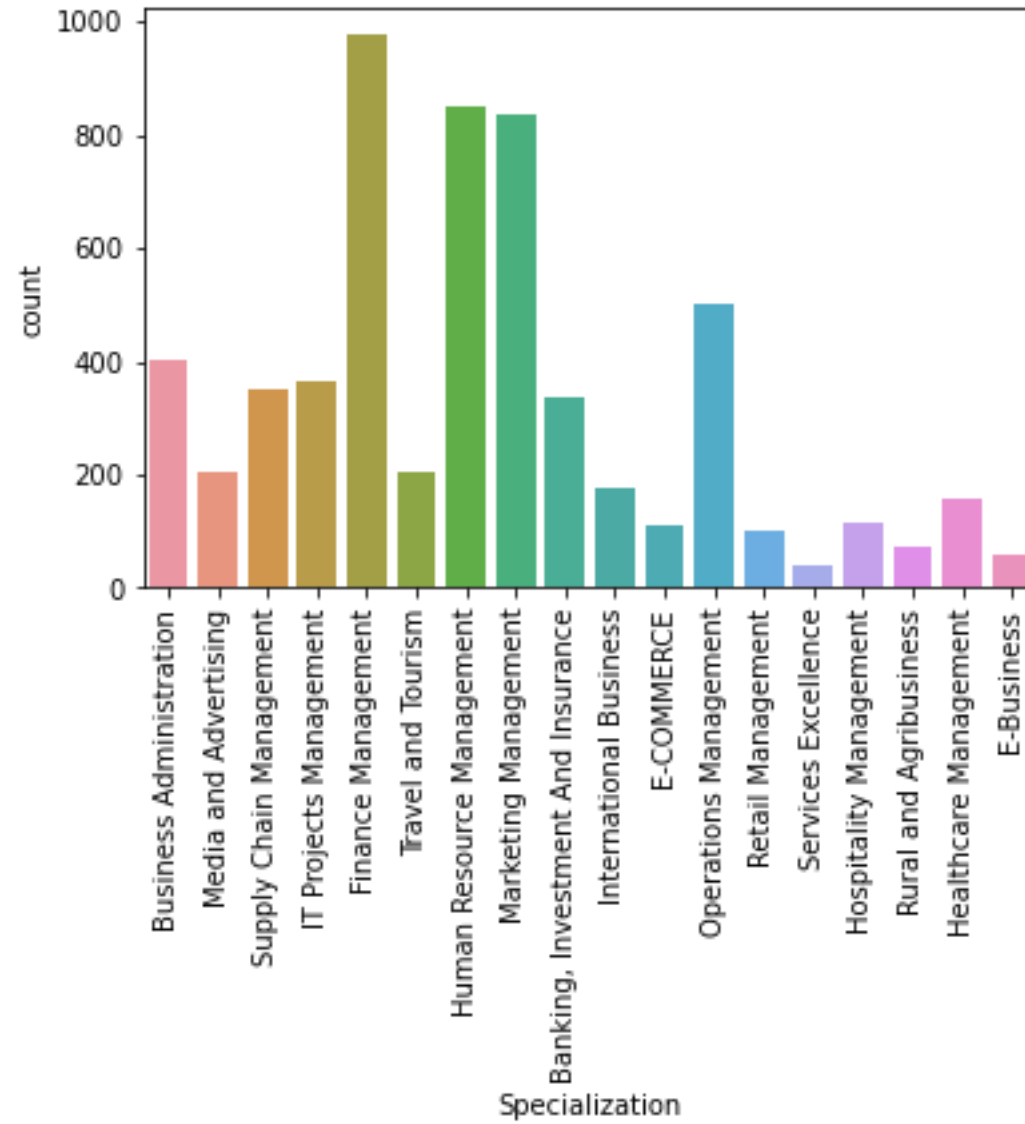
- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.

### Lead Source :



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

## Specialization



The bargraph clearly depicts specialization of Finance Management are high.

# Conclusion :

## Insights

Tags and Lead Source and What is your current occupation and Last Notable activity are the top 4 variables which contribute most towards the probability of a lead getting converted

Country and What matters most to you in choosing a course and Last Notable Activity are the 3 variables which should be focused for the most in order to increase the probability of lead conversion.

## Accuracy

At 0.4 threshold, accuracy is high, which is 90%

Specificity is 91%

Sensitivity is 88%