



Lead Scoring Case Study

By Asha Jyothi Immadisetty



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 improvements listed below have been made to improve analysis and visualization. The necessary columns have undergone the adjustments listed below:

❑ Duplicate Rows :

- checked for duplicate rows and discarded them if any were found.

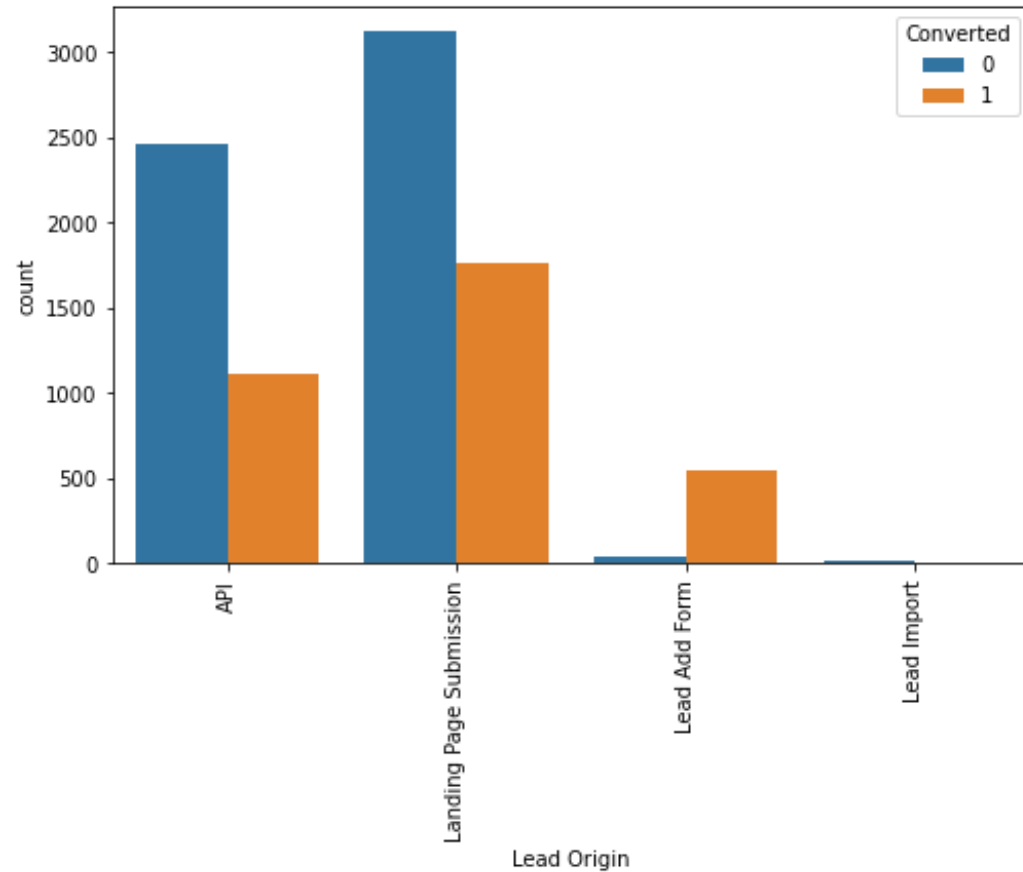
❑ Missing Values Treatment :

- Drop variables that have more than 30% of their missing values if they don't provide much information..
- dropna function rows dropped.

❑ Outlier Treatment :

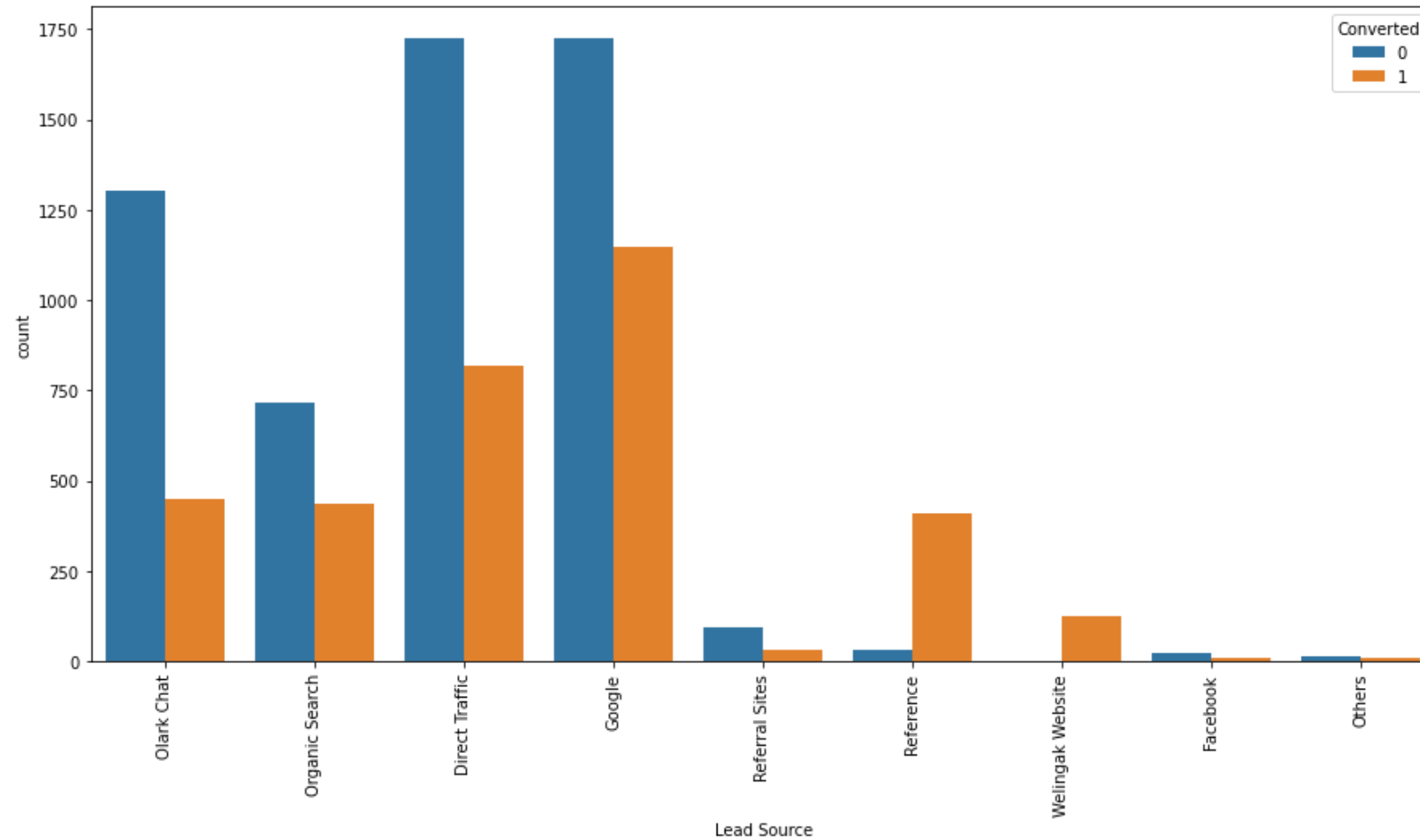
- There are outliers in **TotalVisits** variable, there is lot of variation from 75% to max. We can't cap these since with domain knowledge we can say there can be persons with 250 views..
- There are outliers in **Page Views Per Visit**, we don't need to deal with these outliers since with the domain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

Univariate Analysis : Lead Origin



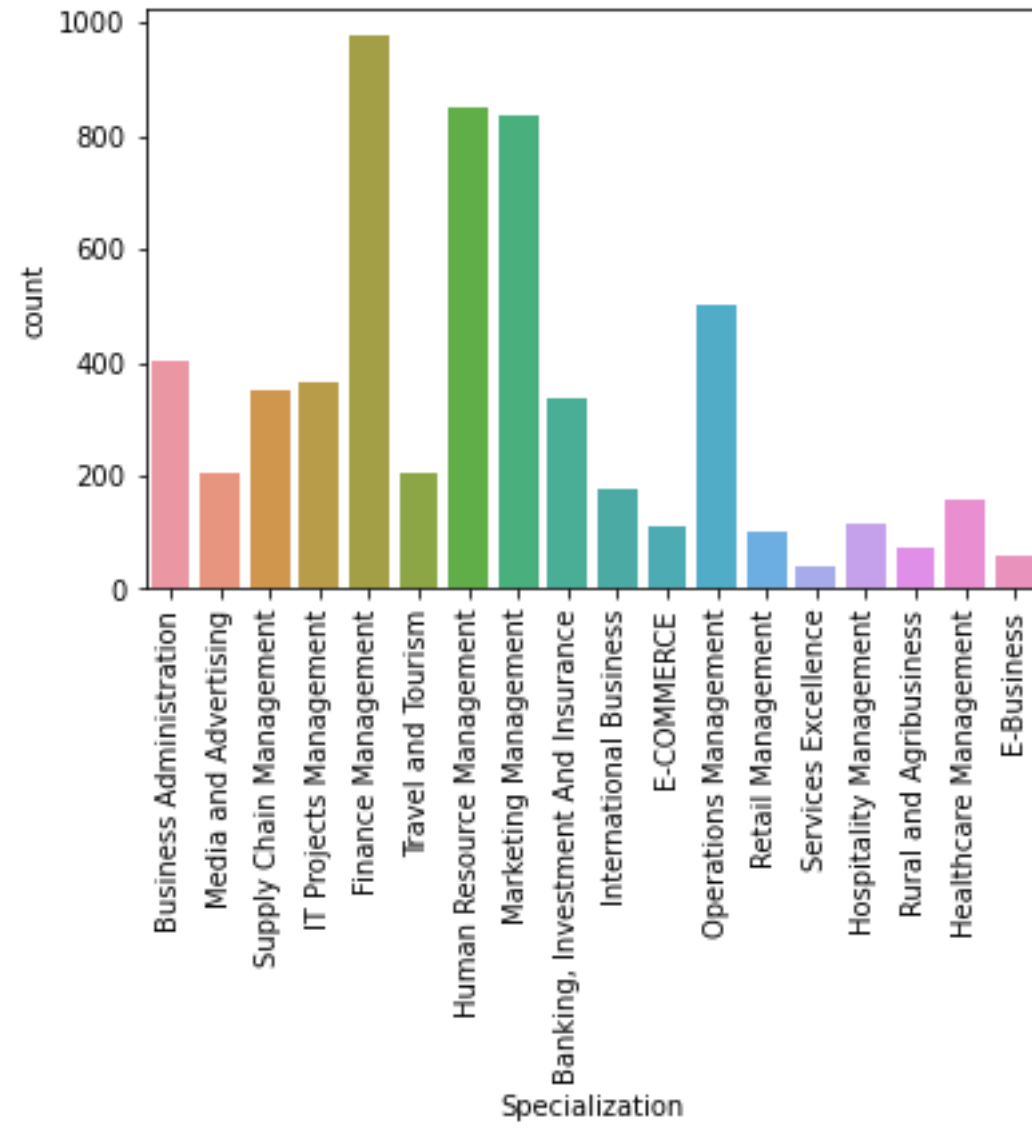
- Although API and landing page submissions have a 30–35% conversion rate, they nonetheless generate a sizable number of leads.
- Despite having a lead count of only a few hundred, the lead add form has a conversion rate of over 90%.
- Lead imports are quite rare.

Lead Source :



- The greatest amount of leads are generated by Google and Direct traffic.
- Reference leads and leads generated from the Welingak website convert at a high rate.

Specialization



Conclusion :

Insights It was found that the variables that mattered the most in the potential leads are (In descending order) :

The total time spend on the Website.

- * Total number of visits.
- * When the lead source was:
 - * Google
 - * Direct traffic
 - * Organic search
 - * Welingak website
- * When the last activity was:
 - * SMS
 - * Olark chat conversation
- * When the lead origin is Lead add format.
- * When their current occupation is as a working professional.

Accuracy

At 0.4 threshold, accuracy is high, which is 89.20%

Specificity is 89.56%

Sensitivity is 88.66%