

Lead Score Case Study

Report on Technical Flow of Analysis and Business
Recommendations

Table of Content:



Technical Flow

Summarizing the Flow of Model Building & Prediction using Logistic Regression technique



Business Insights from Analysis

Highlighting key insights based on bi-variate analysis of lead demographics



Conclusion

Suggestions and Recommendations for Business

Technical Flow:

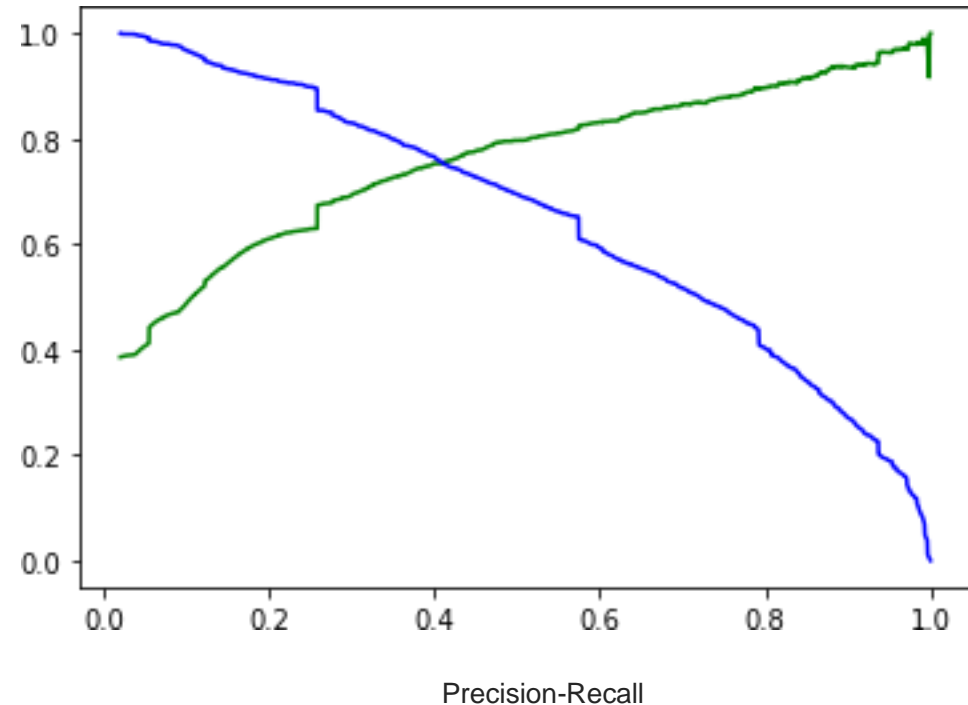
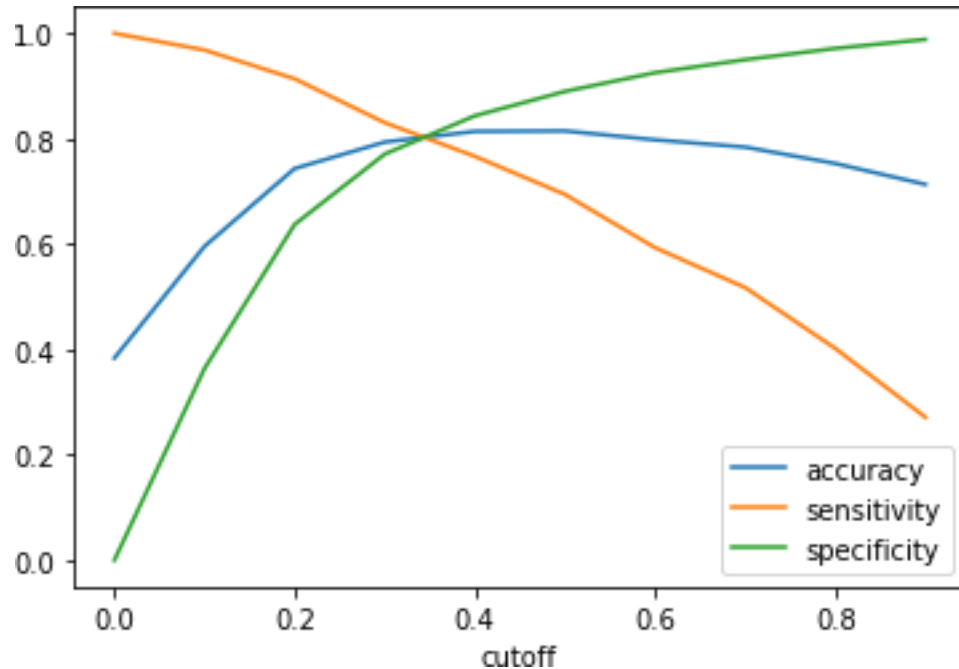
1. **Understanding of Data**– Analyzing the data-type, statistics of numerical attributes were observed for gaining needed context of the data
2. **Data Cleaning** – Identifying missing values and outliers in the data to prevent any imbalance in the Findings/insights generated in further analysis
3. **Data Visualization** –Analysis to highlight the correlation and possible causation between various categorical attributes of a probable lead
4. **Data Preparation** – Modifying the data for utilization in model building by techniques like one-hot encoding
5. **Feature Scaling & Elimination** – Standardizing the features followed by automated feature selection using RFE Methodology for picking impactful features for model building
6. **Model Building** – Iteratively constructing a logistic regression model by analyzing the variance (VIF) & significance (p-value) of selected features
7. **Model Evaluation** – In depth analysis of model accuracy and relevancy using ROC curve, sensitivity & Specificity, precision & recall. Also, identifying the optimal cutoff to identify hot leads efficiently during prediction
8. **Model Prediction** – Implementing the trained model, post accuracy analysis, on the test dataset to make final predictions and assigning lead score
9. **Post Prediction Model evaluation** – Re-checking for accuracy and relevance of the model

Technical Flow: Optimal probability Cut-off

1. Identifying optimal cut-off:

- Finding the most optimal value of probability threshold, above which a lead should be considered a hot lead and would have chances of converting is identified by analyzing the metrics: Accuracy, Sensitivity, and Specificity

2. The most optimal cut-off value is approximately 0.4 in this case where values of aforementioned metrics is almost in the same range. The same has been further validated through Precision & Recall plot on the right

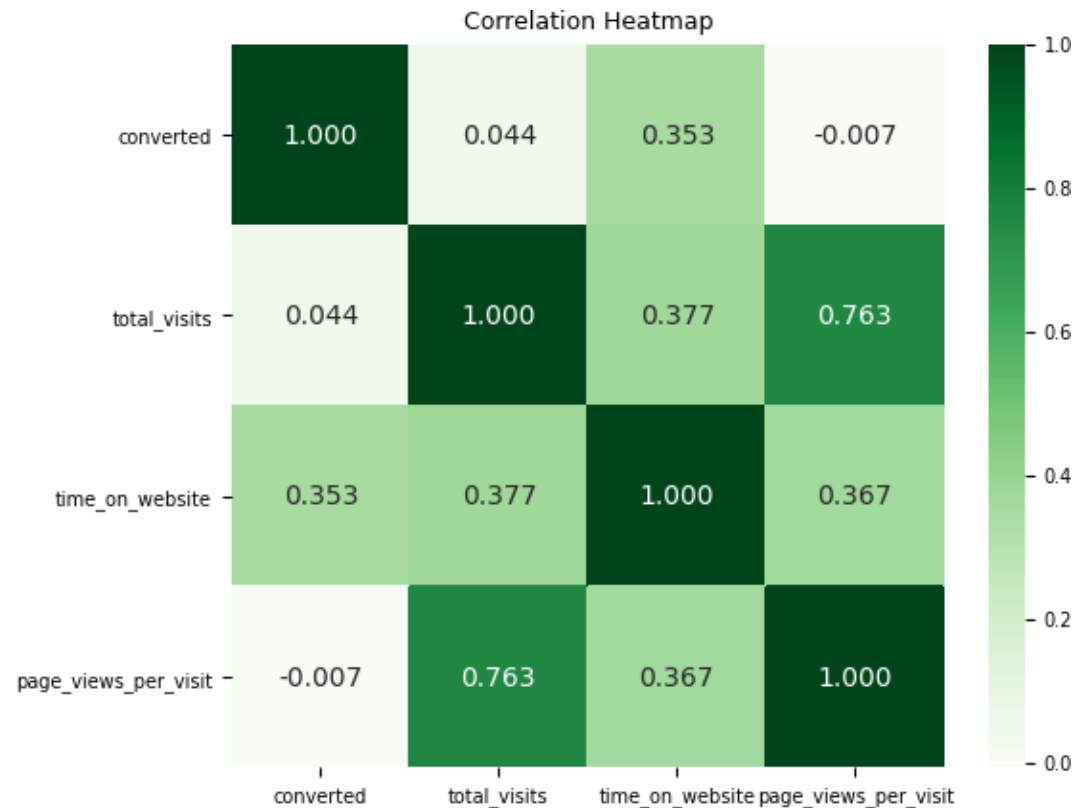


Data Visualization: Findings for Business

1. Correlation across attributes (Numerical Type):

- As observed there is a considerable positive effect of time spent by a lead on the website and the chance of it getting converted

2. Finding can be used by business to increase user engagement on the website by interactive forms, advertisements, promotional campaigns – which can help capture more attributes to the lead and increasing the chance of converting it

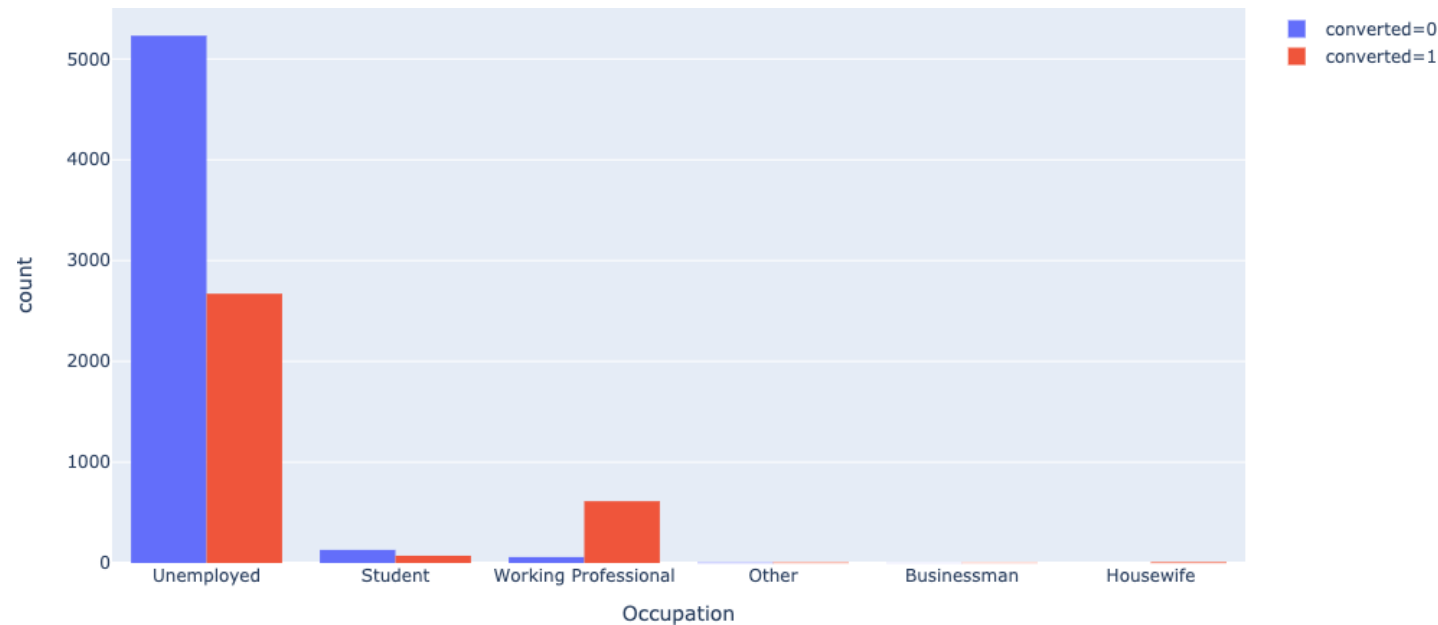


Data Visualization: Findings for Business

1. The Occupation vs Lead conversion:

- As observed, working professionals have the highest conversions, especially when compared from the perspective of total leads. This implies that there is a better chance of converting a lead with working professional background

2. Interestingly, unemployed seems to be the highest number of leads – both successful and unsuccessful, which can be attributed to more unemployed people seeking job opportunities post the course. Although, this set of leads has the lowest chance of converting (approx. 50% only)



Data Visualization: Findings for Business

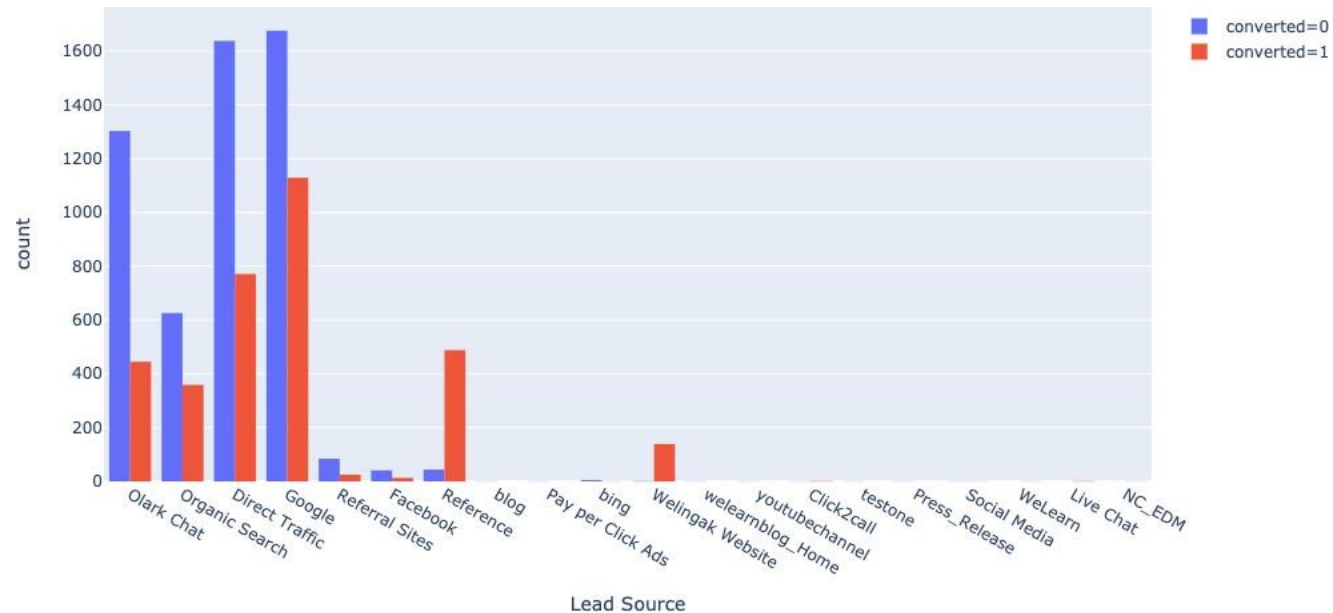
1. Lead vs Lead Conversion:

- As observed, top 5 sources of leads are coming through Google > Direct Traffic > Olark Chat > Organic Search > Reference, in order of total leads

2. Interestingly, it can be seen that with lowest number of leads coming through reference, actually have the most chances of converting which gives a very important insight that “Word of Mouth” is playing a critical role here

3. The focus thus should be to increase the number of leads coming through references and having a better chance at converting these leads

4. There can be multiple ways to increase this by rolling out referral programs, alumni testimonials, etc

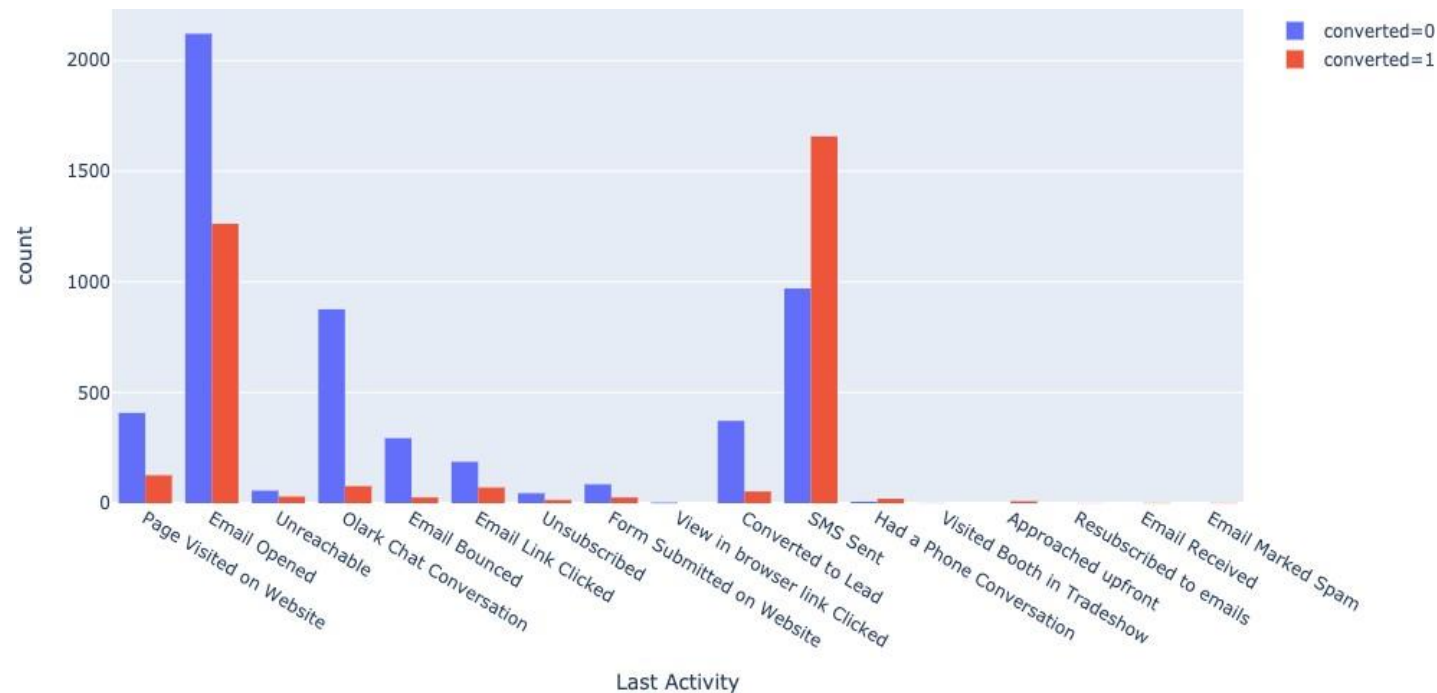


Data Visualization: Findings for Business

1. Last activity type vs Lead Conversion:

- based on the last activity performed by the user in response to the contact made by company's representatives, it is observed that SMS are showing the best results, where the percentage of converted leads is the highest.

2. The business can improve and increase interaction through SMS to potential leads and possibly add personalisation to it, to capture attention more effectively. This can potentially help increase conversion rates by investing efforts & resources in the most impactful connection pathway.



Conclusion: Summarizing the Findings & Recommendations

Relative Feature Importance:

- To highlight the most impactful features as identified and used by the Logistic

Regressionmodel

Business Recommendations: As observed

from analysis & now from the below

highest coefficient, it can be understood that –

1. Time spent on website has a positive impact on lead conversion
2. SMS activity is enabling lead conversion
3. Occupation other than working, are having a negative impact, hence leads with working background can be targeted more

4. Leads from Is_not_act_Had a Phone Conversation, lead_org_Lead Add Form ,

Welingak website have a higher chance of converting and should be focused upon