

Lead Scoring – Summary

Submitted By

Kakani Venkata Hari

Himanshu Shekhar

Hrithik Kumar

The analysis began with an inspection of the '**Leads**' dataset, carefully interpreting variables based on the 'Leads Data Dictionary'. Initial data cleaning involved converting '**Select**' values to null, as these defaults from drop-down menus needed adjustment. Missing data, exceeding **40%**, were either imputed (mean for numeric, mode for categorical) or dropped to ensure data integrity. Post-cleaning, indexes were reset for clarity.

Subsequent exploratory analysis revealed a lead conversion rate of 38.54%, identifying India and Mumbai as leaders in total leads and conversions. 'Finance Management' emerged as the top specialization for both metrics, while Google proved the most fruitful lead source. Notably, '**unemployed**' individuals represented the largest group among leads and conversions.

Correlation analysis highlighted significant associations: conversions correlated positively with total website visits and time spent on the site, but inversely with page views per visit. With categorical values transformed into dummies, the dataset was split into 70% training and 30% test sets, then numerically scaled via **StandardScaler**.

Model development ensued: the initial iteration included all variables, though many displayed high p-values, signaling limited significance. Refinement via Recursive Feature Elimination (RFE) in the second model improved feature selection. The final model employed features chosen by RFE, with those showing high p-values excluded.

Using this final model, predictions on the training set yielded an **80.46% accuracy**. A confusion matrix provided further insights, revealing a **sensitivity** of 80%, **specificity** of 80.71%, and recall of 75%. A precision-recall tradeoff analysis and ROC curve informed the selection of 0.345 as the optimal **cutoff** for predictions.

Recommendations drawn from the analysis are as follows:

- Prioritize leads with high 'Lead Scores' to enhance conversion rates.
- Allocate resources towards Google marketing, given its high conversion rates.
- Foster referrals from converted leads through incentivization.
- Expand marketing efforts beyond Mumbai to other cities with substantial lead potential.
- Target 'unemployed' prospects and those specializing in 'Finance Management'.
- Adjust marketing strategies away from students due to lower conversion rates.

In summary, the analysis underscores actionable insights derived from rigorous data exploration, feature selection, and model refinement. These recommendations aim to optimize conversion rates by focusing efforts on high-potential leads and refining marketing strategies accordingly.