LEAD SCORE CASE STUDY

CONTENT:

- ## Technical Flow of Analysis

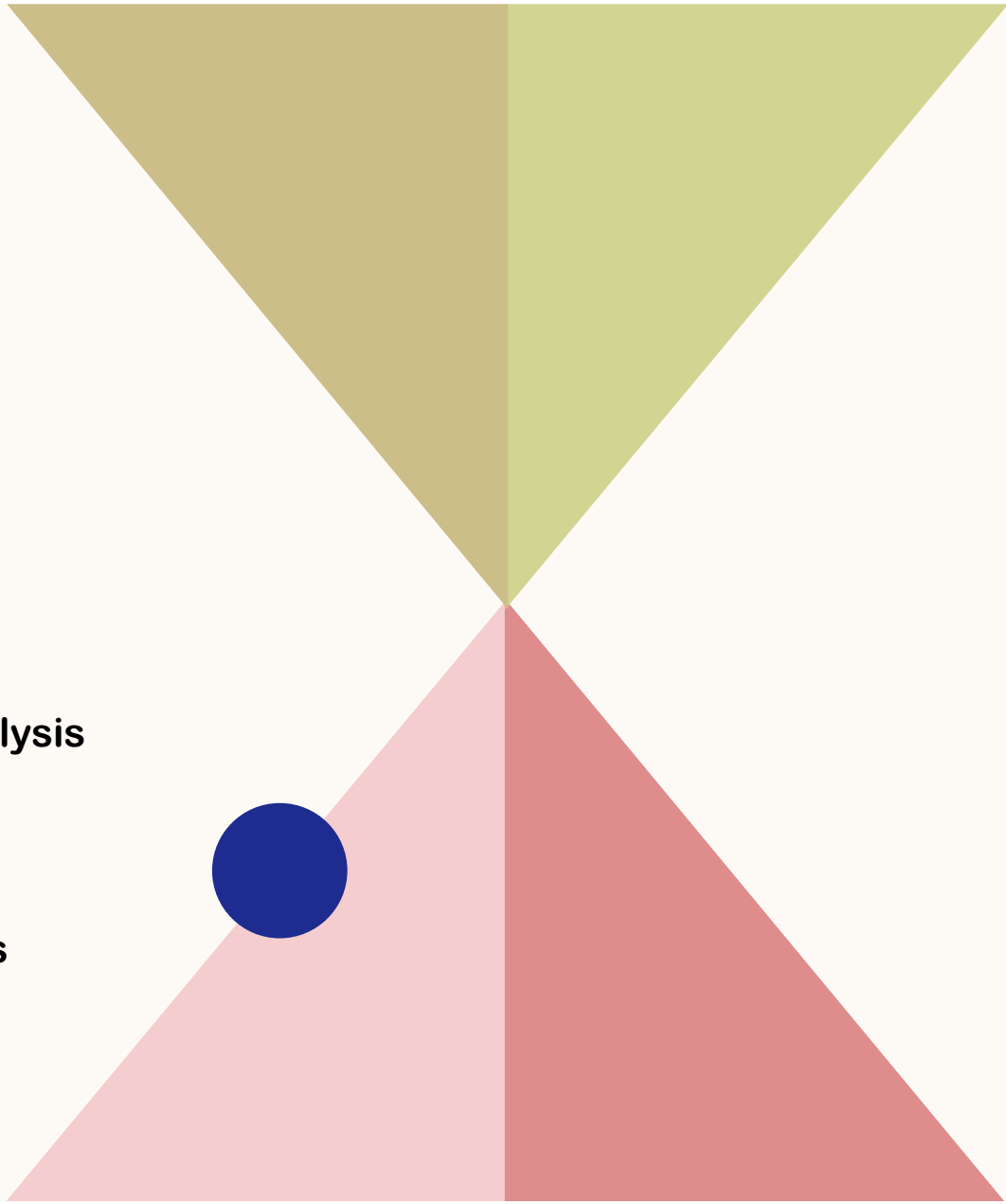
  Summarizing the Flow of Model Building & Prediction using Logistic Regression technique.

- ## Business Insights from Analysis

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

- ## Conclusion

  Suggestions and Recommendations for Business



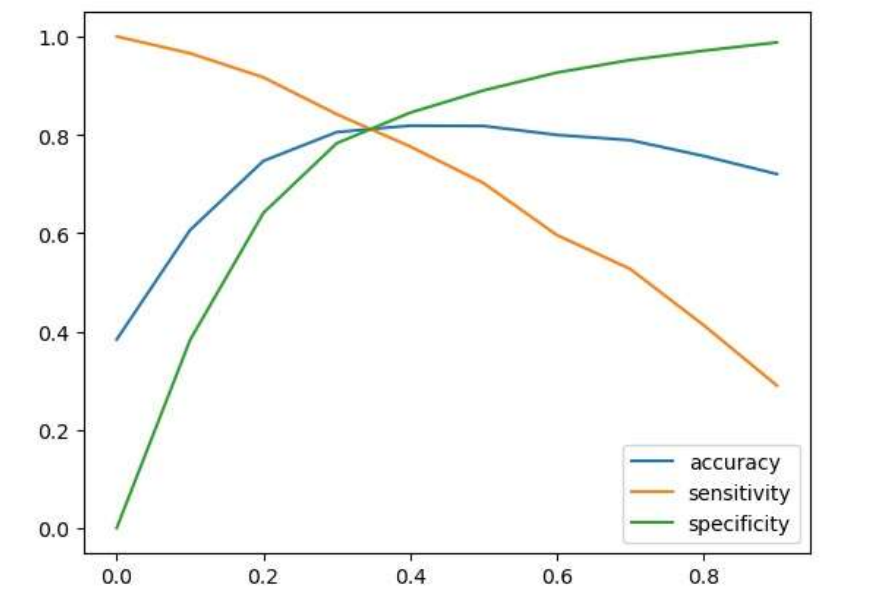
TECHNICAL FLOW:

- **Data Understanding** – Statistics of numerical attributes were observed during the data type analysis to provide the necessary context for the data.
- **Data Cleaning** – To avoid any imbalance in the conclusions or insights drawn from subsequent analysis, we identified missing values and outliers in the data.
- **Data Visualization** – A thorough examination to show the connections and potential causes between several categorical characteristics of a likely lead.
- **Data Preparation** – Modifying the data for use in model construction using methods like one-hot encoding.
- **Feature Scaling & Elimination** – Standardizing the features and then automatically choosing the most important ones for model creation using the RFE technique.
- **Model Building** – Building a logistic regression model iteratively while examining the variance (VIF) and significance (p-value) of specific features
- **Model Evaluation** – Thorough evaluation of model relevancy and accuracy utilizing the ROC curve, sensitivity and specificity, recall, and precision. Choosing the ideal cutoff to effectively discover hot leads during forecast.
- **Model Prediction** – Putting the trained model to use on the test dataset after accuracy analysis to create final predictions and assign lead score.
- **Post Prediction Model evaluation** – Verifying the model's precision and applicability once again.

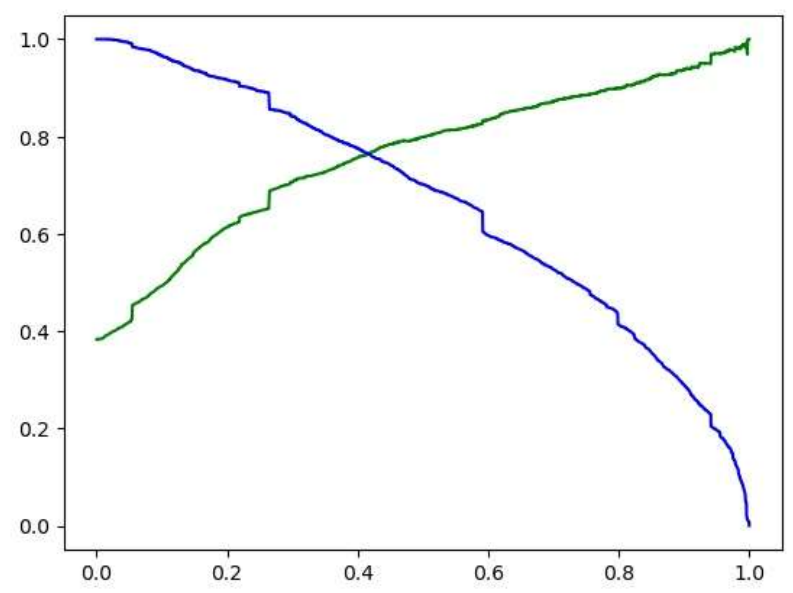
TECHNICAL FLOW: OPTIMAL PROBABILITY CUT-OFF

1) Determining the ideal cut-off: - By studying the metrics, it is possible to determine the probability threshold at which a lead should be regarded as hot and has a good possibility of converting: Sensitivity, specificity, and accuracy.

2) In this situation, where the values of the aforementioned indicators are almost in the same range, the most ideal cut-off value is roughly 0.4. The Precision & Recall plot on the right has further verified this.



Cut-off

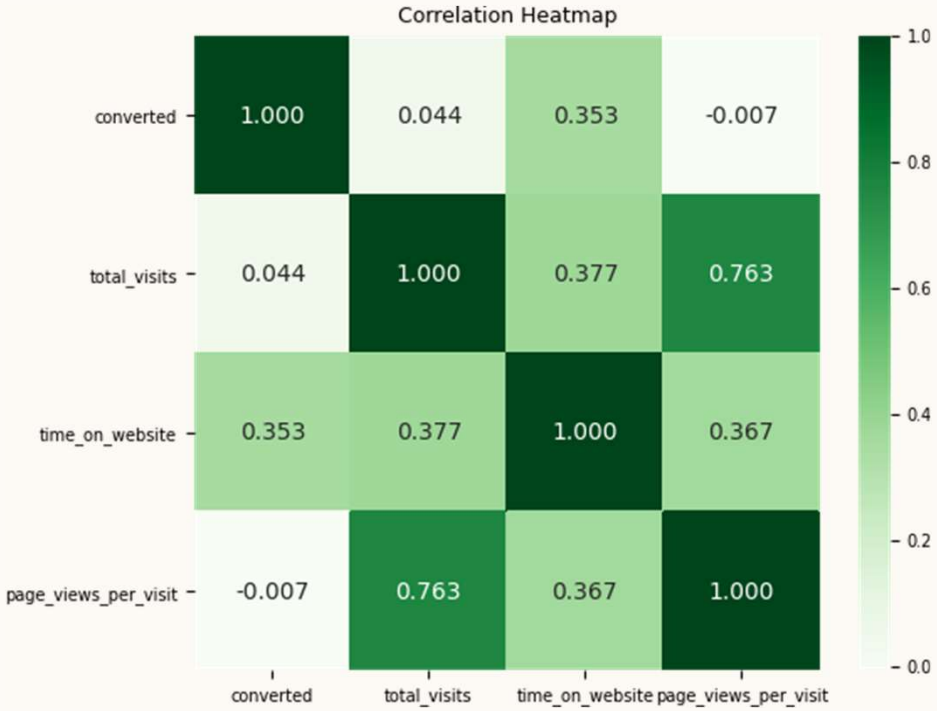


Precision-Recall

DATA VISUALIZATION: BUSINESS FINDINGS

1) Correlation across Attributes (Numerical Type): It has been determined that a lead's time spent on a website significantly increases the likelihood that the lead will convert.

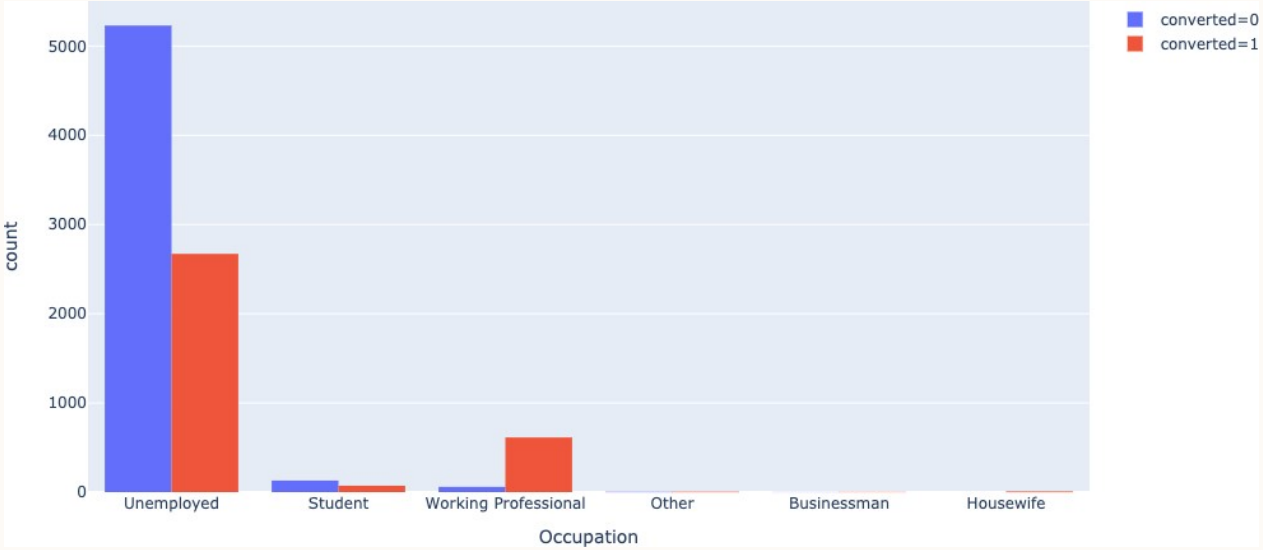
2) This discovery can be applied by businesses to boost user engagement on the website through interactive forms, adverts, and promotional efforts, which can help capture more qualities to the lead and raise the likelihood that it will be converted.



DATA VISUALIZATION: BUSINESS FINDINGS

1) Lead conversion vs Occupation: Working professionals are found to have the highest conversion rates, especially when looked at in terms of total leads. This suggests that a lead with professional experience has a higher likelihood of being converted.

2) It's interesting to note that leads from unemployed people seem to be more frequent than leads from employed people, both successful and unsuccessful, which may be due to more unemployed people looking for work after the course. Although this particular group of leads has the lowest conversion rate.



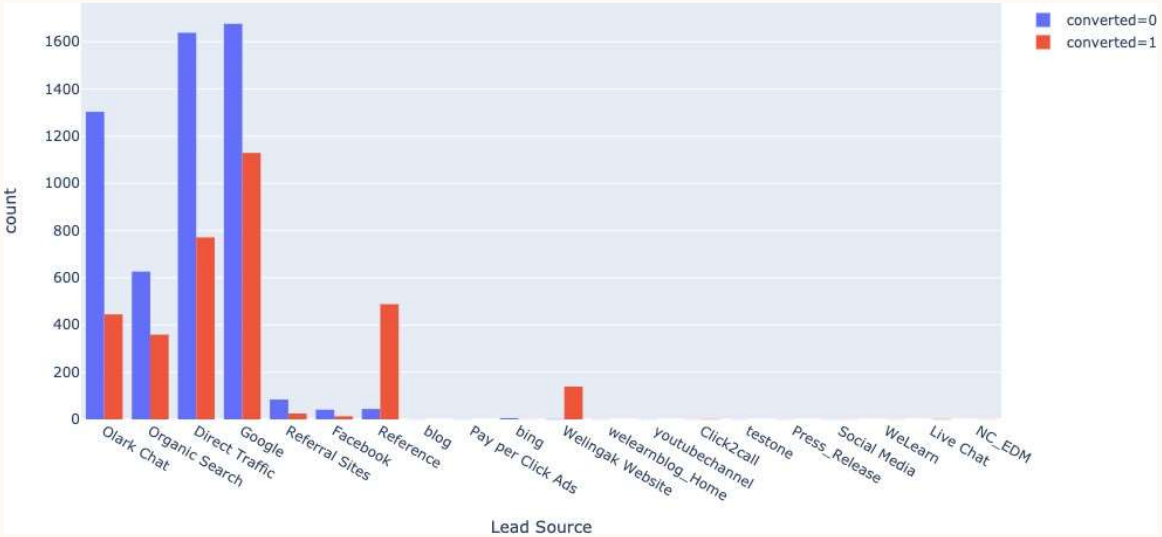
DATA VISUALIZATION: BUSINESS FINDINGS

1) Source of Lead vs. Lead Conversion: As seen, the top 5 lead sources are Google, Direct Traffic, Olark Chat, Organic Search, and Reference, in that order.

2) It's interesting to note that the leads with the lowest referral volume actually have the highest conversion rates, which provides a crucial insight into the importance of "Word of Mouth" in this situation.

3) The goal should be to increase the quantity of leads generated through referrals and improve the likelihood that these leads will be converted.

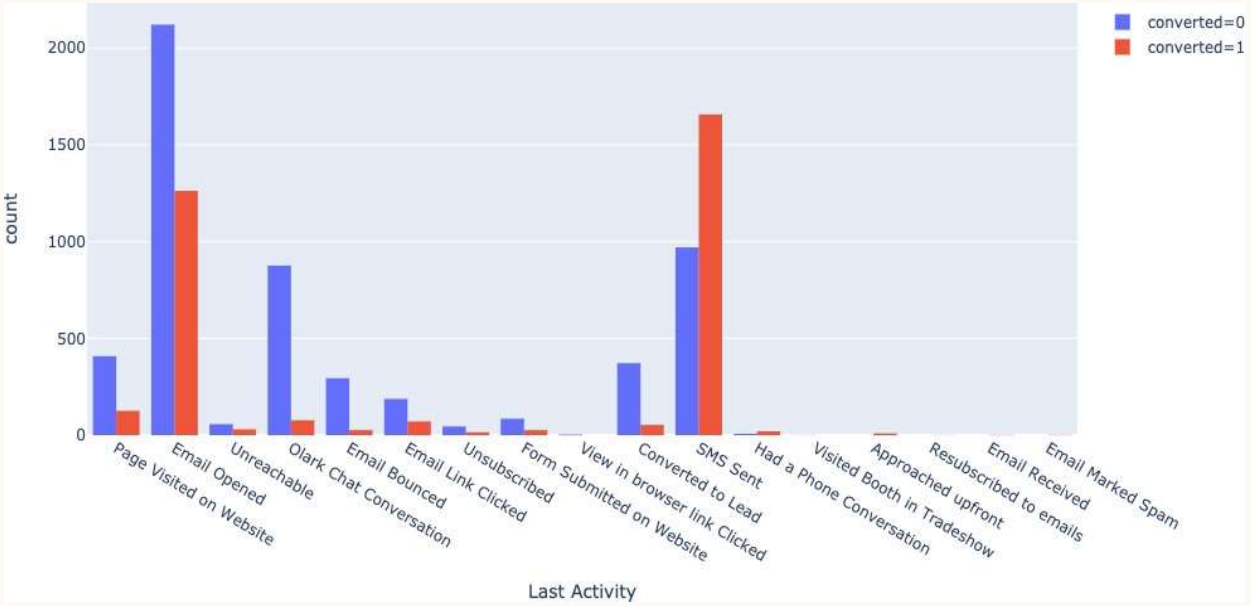
4) There are numerous ways to do this, for as by implementing referral programs and using alumni endorsements, among others.



DATA VISUALIZATION: BUSINESS FINDINGS

1) Last activity type vs Lead Conversion: Based on the most recent action taken by the user in response to a contact from a company representative, it is found that SMS are producing the best outcomes, with the largest percentage of leads that are actually converted.

2) The company can enhance SMS contact with potential leads and possibly add personalization to it to better grab their attention. By focusing efforts and resources on the most effective connection pathway, this may assist enhance conversion rates. .



CONCLUSION: SUMMARIZING THE FINDINGS & RECOMMENDATIONS

Relative Feature Importance:- To highlight the most impactful features as identified and used by the Logistic Regression model.

Business Recommendations: As observed from Bi-variate analysis & now from the below plot, it can be understood that –

- ❑ Lead conversion is positively impacted by time spent on a website.
- ❑ SMS usage is facilitating lead conversion.
- ❑ Because non-working occupations have a negative effect, leads with a working background can be more effectively targeted.
- ❑ Welingak website leads should be prioritized because they have a higher possibility of converting.