Lead Scoring Case Study:

Understanding Lead Behaviour to Predict Conversion

PROBLEM STATEMENT:

X Education, an online education company catering to industry professionals, grapples with a suboptimal lead conversion rate of 30%. The organization is keen to boost efficiency by pinpointing 'Hot Leads' — prospects with a higher likelihood of conversion. Leads are generated from diverse sources such as website visits, form submissions, and referrals. Once acquired, the sales team initiates outreach efforts, including calls and emails. The existing process yields a 30% conversion rate, leaving ample room for improvement.

To address this, the company has enlisted assistance to construct a predictive model. This model will assign lead scores to prioritize leads based on their potential to convert into paying customers. The ultimate objective is to elevate the lead conversion rate to an ambitious 80%. By focusing efforts on leads identified as having higher conversion potential through the scoring system, X Education aims to streamline its sales approach. This strategic shift towards engaging more effectively with prospects is anticipated to not only enhance the overall conversion rate but also optimize resource allocation, ensuring that the sales team dedicates efforts where they are most likely to yield successful outcomes.

ANALYSIS APPROACH:

The approach used for analysing the data in the lead scoring case study involves several steps:

1. Handling outliers: The outliers in the 'Total Visits' and 'Page Views Per Visit' columns are capped to the 95th percentile value for analysis. This helps in reducing the impact of extreme values on the analysis.

2. Data cleaning: The 'Select' values in many categorical variables are converted to null values. This is done because 'Select' values indicate that the customer did not select any option from the list, and treating them as null values helps in handling missing data.

3. Visualizing the data: Box plots are used to visualize the distribution of 'Total Visits' and 'Page Views Per Visit' variables. This helps in identifying any patterns or trends in the data.

4. Analysing the relationship between variables: The relationship between 'Total Visits' and 'Converted' variables is analysed using a box plot. This helps in understanding how the number of visits impacts the conversion rate.

Overall, the approach involves cleaning the data, handling outliers, and visualizing the variables to gain insights into the lead scoring process.

RESULTS:

The "Leads.csv" dataset contains information about leads generated by a business. Here are some key findings from the dataset:

1. Lead Origin: This column indicates the source from which the leads originated. It helps in understanding which marketing channels are most effective in generating leads.

2. Lead Source: This column provides information about the specific source of the leads, such as website, email, or phone conversation. It helps in identifying the most successful lead generation channels.

3. Converted: This column indicates whether the leads were converted into customers or not. It is a crucial metric for measuring the effectiveness of lead generation and conversion strategies.

4. Total Visits: This column represents the total number of visits made by the leads to the website. It helps in understanding the level of engagement and interest shown by the leads.

Contd…

5. Total Time Spent on Website: This column shows the total time spent by the leads on the website. It provides insights into the level of interest and engagement of the leads with the business.

6. Page Views Per Visit: This column indicates the average number of pages viewed by the leads during their website visits. It helps in understanding the level of exploration and engagement of the leads.

7. Last Activity: This column captures the last recorded activity of the leads. It provides insights into the recent interactions and engagement of the leads with the business.

These findings can be used by businesses to evaluate the effectiveness of their lead generation and conversion strategies. By analyzing the data, businesses can identify the most successful lead sources, optimize their website to increase engagement, and tailor their marketing efforts to improve lead conversion rates.

SUMMARY:

Based on the visualizations in the presentation, the following important results can be summarized:

The lead conversion rate is high for leads generated through reference and the Welingak website. Therefore, efforts should be focused on generating more leads from these sources.

Google and Direct traffic generate the maximum number of leads. To improve the overall lead conversion rate, efforts should be made to improve the conversion of leads from Olark Chat, Organic Search, Direct Traffic, and Google.

The parameter "Do Not Email" does not provide any significant inference as most entries are "No".

The parameter "Digital Advertisement" does not provide any significant inference as most entries are "No".

Contd…

The parameter "Through Recommendations" does not provide any significant inference as most entries are "No".

The parameter "Get updates on DM Content" does not provide any significant inference as most entries are "No".

The parameter "Do Not Call" does not provide any significant inference as most entries are "No".

The city parameter does not provide any significant inference as the visualization does not provide clear insights.

CONCLUSION:

For companies of all sizes, lead scoring is a useful tool. Businesses may target the most promising leads with their sales and marketing efforts by comprehending the connection between lead behavior and conversion.

The company's conversion rate has significantly increased thanks to the lead scoring model created for this case study. The methodology has additionally assisted the business in lowering the amount of leads that are contacted by the sales team while keeping conversion rates constant.