Summary

X Education, an esteemed online educational platform catering to professionals in various industries, experiences a consistent influx of enthusiastic visitors to its website daily. Utilizing diverse online channels and search engines, the company actively promotes its array of courses, enticing visitors to explore available educational programs, engage with videos, and submit forms expressing interest. Keen on optimizing its operations, X Education endeavors to pinpoint 'Hot Leads' - those prospects demonstrating the highest likelihood of converting into paying customers. The intent is to refine lead identification, thereby amplifying conversion rates by concentrating sales efforts on these promising leads, streamlining communication strategies, and ultimately bolstering the overall conversion rate.

In pursuit of this objective, a comprehensive study was undertaken to assess leads and assign them scores on a scale of 0 to 100, with 0 denoting a lower probability of conversion and 100 indicating a higher likelihood of conversion. The approach began by gathering and refining data from the company's sources, ensuring enhanced data readability. Subsequently, an exploratory data analysis was conducted, yielding critical insights:

It was revealed that a significant proportion of leads were generated through Google, followed by Direct traffic and Olark Chat. The primary origin of leads stemmed from landing page submissions. Notably, the dataset indicated a dominance of unemployed leads over working professionals, constituting a substantial portion of the leads.

Of the leads within the dataset, approximately 40% were successfully converted into paying customers. Leads sourced from 'Reference' and the 'Welingak Website' exhibited the highest conversion rates, contrasting sharply with leads derived from 'Olark Chat' and 'Direct traffic,' which showcased the lowest conversion rates. Leads specifying their current occupation as working professionals demonstrated a higher likelihood of conversion, while those failing to specify their occupation showed a lower propensity for conversion.

Moreover, leads originating from the Lead Add Form exhibited the highest probability of conversion compared to other lead sources. Additionally, leads with Last Notable Activity labeled as 'Modified' displayed a lower likelihood of conversion, while those marked as 'SMS Sent' or 'Had a Phone Conversation' indicated a higher chance of eventual conversion.

Upon completion of the analysis, a sophisticated Logistic Regression model was constructed to predict lead conversions. The finalized model boasted an accuracy rate of 81% at a threshold of 0.35. To facilitate ongoing adaptability, a Lead Score was assigned to each lead based on its calculated probability of conversion, ranging from 0 to 100. This scoring system offers the company an invaluable tool to prioritize leads, moving beyond predicted values and enabling a more nuanced approach to lead management based on their assigned scores.