Lead Score Case Study Summary

Steps taken:

1. Loading and Analysing the Data

2. Data Manipulation and Cleaning

Dropped the columns which had more than 40% of null values. Imputing null values with mode or 'Others' where applicable. Outliers were identified and treated.

3. EDA

4. Dummy Variable Creation

Dummy variables were created for categorical columns.

5. Train-Test Split

Data was split into train set and test set by 70% and 30% respectively.

6. Scaling the Data

StandardScaler was used to scale numerical variables.

7. Model Building

We started with RFE while selecting top 20 features. Later features with high p-values and high VIF values were removed.

8. Model Evaluation

Model was evaluated using confusion matrix. And after finding optimal cut-off point (using ROC curve) accuracy was 92%, sensitivity was 91.69% and specificity was 92.22%.

9. Precision and Recall

Precision was 87.86% and Recall was 91.69%.

10. Predictions on Test Set

According to the model built, test data set gave accuracy of 92.10%, sensitivity of 90.89% and specificity of 92.84%.

11. Conclusions

	coef
const	-0.4709
Total Time Spent on Website	1.0425
Lead Source_Direct Traffic	-1.6601
Lead Source_Google	-1.0360
Lead Source_Organic Search	-1.1346
Lead Source_Welingak Website	4.9715
Last Activity_Olark Chat Conversation	-1.0254
Last Activity_SMS Sent	2.0650
Tags_Closed by Horizzon	7.5370
Tags_Interested in other courses	-2.1876
Tags_Lost to EINS	5.9659
Tags_Other	-2.2780
Tags_Ringing	-3.3331
Tags_Will revert after reading the email	4.7111
Last Notable Activity_Modified	-1.4729

These are the features that affect the chances of a lead converting or not the most, company should focus on leads that are closed by Horrizon, leads that are sourced from Welingak Website, leads should be contacted via Email more compared to SMS as Email has a higher coefficient for conversion.