Executive Summary of Lead Scoring Case Study

Team members:

1. Kritisha Panda
2. Uttej Gaikwad

1. **Model Selection**: After evaluating several models, the final model was selected based on specific characteristics.

2. **Statistical Significance**: All variables in the model have p-values less than 0.05, which indicates statistical significance.

3. **Multicollinearity**: The features have very low Variance Inflation Factor (VIF) values, suggesting minimal multicollinearity, which is supported by a heatmap that isn't shown in the image.

4. **Model Accuracy**: The model demonstrates an overall accuracy of 90.56% at a probability threshold of 0.33 on the test dataset.

5. **Performance Metrics:** The model's performance is detailed for different threshold values across various metrics:

- At a 0.50 threshold for the training set, the model has an accuracy of 91.25%, sensitivity of 81.95%, and specificity of 96.90%.

- At a 0.33 threshold for the training set, the model shows slightly lower accuracy (90.32%) and specificity (91.30%) but higher sensitivity (88.70%).

- For the test set at a 0.33 threshold, the model maintains an accuracy of 90.56%, sensitivity of 88.86%, and specificity of 91.63%.

6. **False Positive Rate**: The false positive rate is low, at 0.0310 for the 0.50 threshold and 0.0870 for the 0.33 threshold in the training set, and 0.0837 for the test set at the 0.33 threshold.

7. **Predictive Values**: Positive and negative predictive values are provided for both threshold levels on the training set and the 0.33 threshold on the test set, indicating good predictive power.

8. **Precision and Recall**: Precision and recall are both reported for the 0.33 threshold on both training and test sets, with all values being relatively high.

9. **F1 Score and Cross-Validation Score**: The F1 score and cross-validation score are both reported for the training set at both thresholds, showing robustness in the model's predictive capability.

10. **AUC (Area Under Curve):** The model has a high AUC of 0.9624 for the training set at the 0.50 threshold and 0.9624 for the 0.33 threshold. For the test set at the 0.33 threshold, the AUC is slightly higher at 0.9679, indicating a strong ability to distinguish between classes.

11. **Feature Contribution**: The report identifies certain features that contribute significantly to predicting a successful lead conversion. These features are listed in order of their positive impact:

- Tags indicating a lead has been lost to EINS.
- Tags showing a lead was closed by Horizzon.
- Tags suggesting a lead will revert after reading an email.
- The lead source being the Wellnigak Website.
- The last activity recorded as an SMS being sent.
- The current occupation of the lead being a Working Professional.
- The current occupation of the lead being Unemployed.

12. **Inference**: The probability of a lead converting into a customer increases with the values of the identified features. The model essentially provides a means to prioritize leads based on the likelihood of conversion, which can significantly optimize marketing strategies and resource allocation.

In summary, the chosen model exhibits strong predictive performance with statistically significant features and low multicollinearity, making it a reliable tool for forecasting lead conversion probabilities.