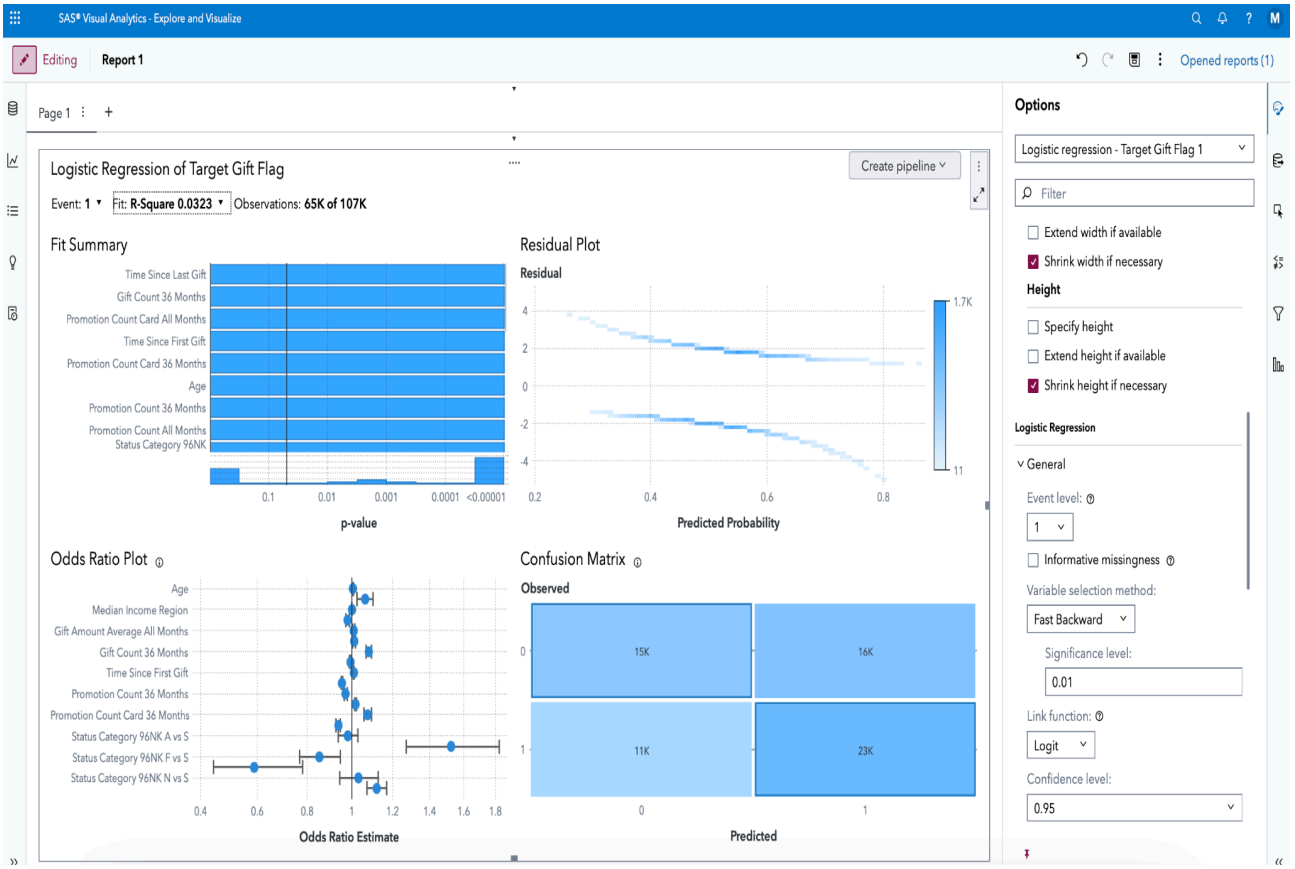
National University of Science and Technology Politehnica Bucharest
Advanced Analytics for Business
Predictive Modeling

Name: Merve Pakcan Tufenk

Practice 1. Logistic Regression



Questions:

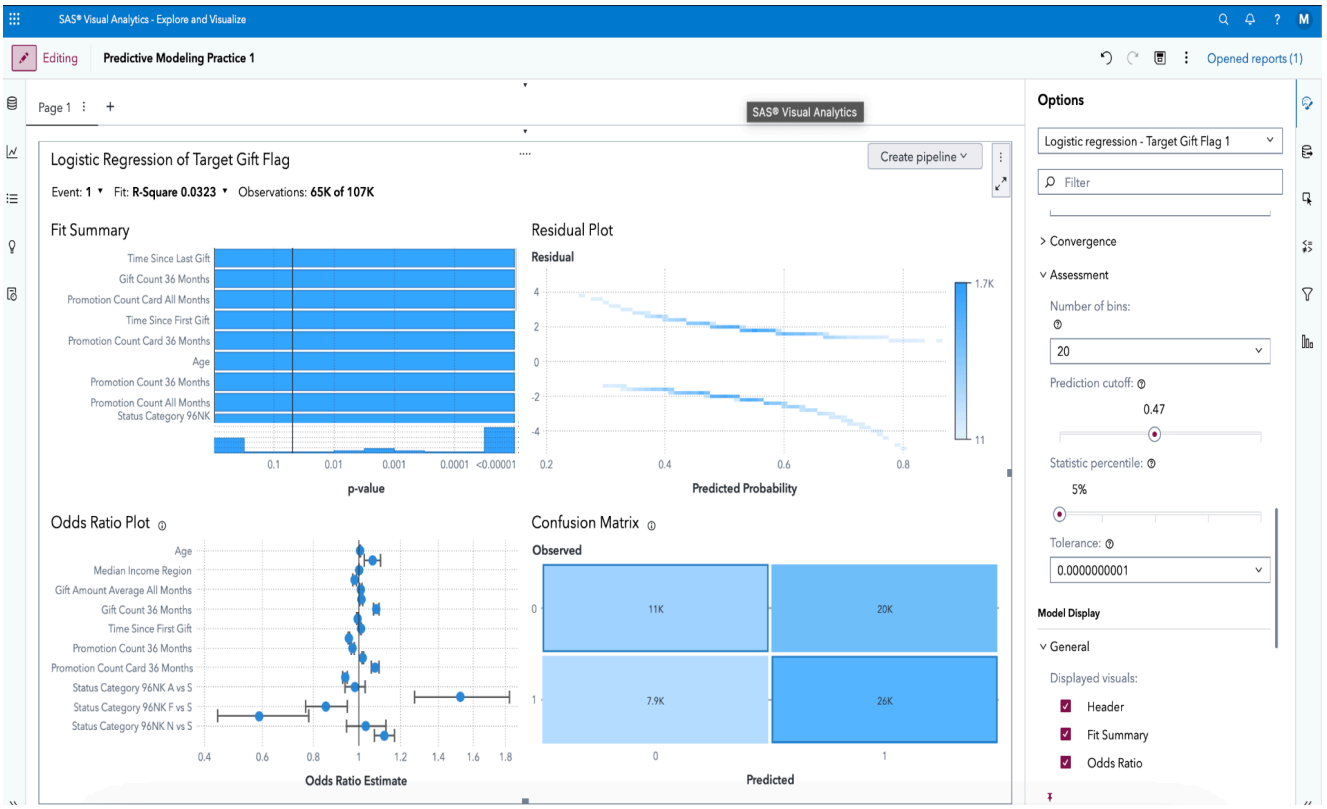
a. What is the values of the **R-square** statistics? In a short sentence asses the quality of the model based on the R-square.

   Since the R-square is 0.0323, the model explains approximately 3.2% of the variability in the target variable. This shows that its prediction performance is quite low and may not be enough to trust for donation predictions.

a. Examine the **Fit Summary** panel. How many of the 23 input variables are not included in this model?
   As it seen from the Fit Summary panel, 9 out of 23 input variables were included in the model. Therefore, 14 variables were excluded during the Fast Backward selection process because they were not statistically significant.

b. In this model, are any of the insignificant variables that are not included classification effects?
   Two classification effects(Gender and Home Owner) were not included in the final model, indicating they were statistically insignificant.

c. Explain the **Assement charts** and indetify the best cutoff treshold. Explain how the adjustment of the cutoff treshold affects the model.
Based on my review of the assessment charts and confusion matrix analysis, I determined that the best cutoff threshold is 0.47. At this point, the model offers a solid balance between sensitivity and specificity, successfully identifying most donors while keeping false positives at a reasonable level. As I evaluated different thresholds between 0.45 and 0.50, I noticed that lower cutoffs like 0.45 resulted in very high sensitivity (82.4%) but poor specificity, leading to too many false positives. In contrast, higher thresholds such as 0.49 and 0.50 improved specificity, but sensitivity dropped significantly, which means the model missed more actual donors. After comparing all these values, I found that 0.47 is the most balanced and practical option. This clearly shows how adjusting the cutoff affects the trade-off between detecting donors and minimizing incorrect predictions.



2. Compute the **Sensivity** and the **Specifity** of the designed model (*! Hint – look an the Confusion Matrix. !)*

   To compute the sensitivity and specificity of the model, I used the values from the confusion matrix at the selected cutoff threshold of 0.47:

- **True Positives (TP):** 26K
- **False Negatives (FN):** 7.9K
- **True Negatives (TN):** 11K
- **False Positives (FP):** 20K

Sensitivity= TP/(TP+FN) = 26/(26+7.9)= 26/33.9= approximately 76%

Specificity=TN/(TN+FP) = 11/(11+20)= 11/31= approximately 35%