All code for my experiments is contained within the notebook logistic_regression_model.ipynb in this repository in the directory logistic_regression_py. It is best run in Google Colab, and the version I used for my experiments is linked here: https://colab.research.google.com/drive/1Z1vLcywoeiBn0diASFUbY2i54MjFTba9?usp=sharing

The notebook is designed to be run sequentially from top to bottom. Before training the model, I applied my entire pipeline to the standardized, merged dataset. This pipeline includes preprocessing steps such as dropping irrelevant columns, applying a custom target transformer, standardizing numerical columns, and performing feature selection. To run the notebook successfully, you'll need to upload the following datasets to your Colab environment: final_merged_training_2020_2022.csv, final_merged_validation_2023.csv, and final_merged_test_2024.csv. Additionally, all the custom transformer classes, helper functions, and code used in the notebook are also available in a standard Python script format called logistic_regression_model.py in the logistic_regression_py directory. To run the logistic regression model locally using these scripts, simply execute python3 logistic_regression_model.py. Note that you'll need to have pandas, scikit-learn, and matplotlib installed in your Python environment.