From Scikit-Learn: <https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression>

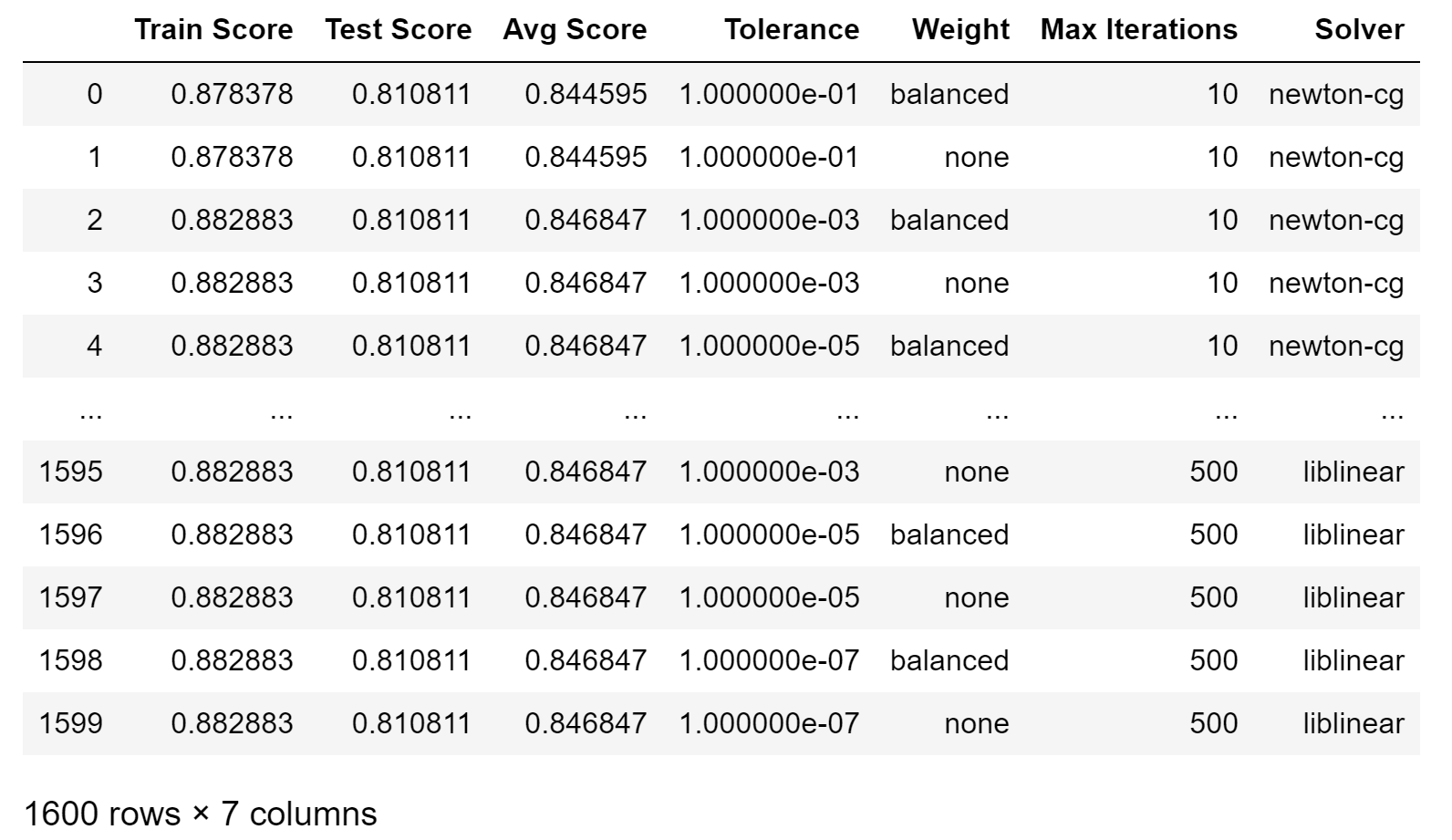
**Logistic Regression**

**Logistic Regression is a linear model used in classification not for regression. It has several alias’ such as logit regression, MaxEnt (maximum-entropy) classification and log-linear classifier. This model technique uses a logistic function to model probabilities that describe possible outcomes of a single trial.**

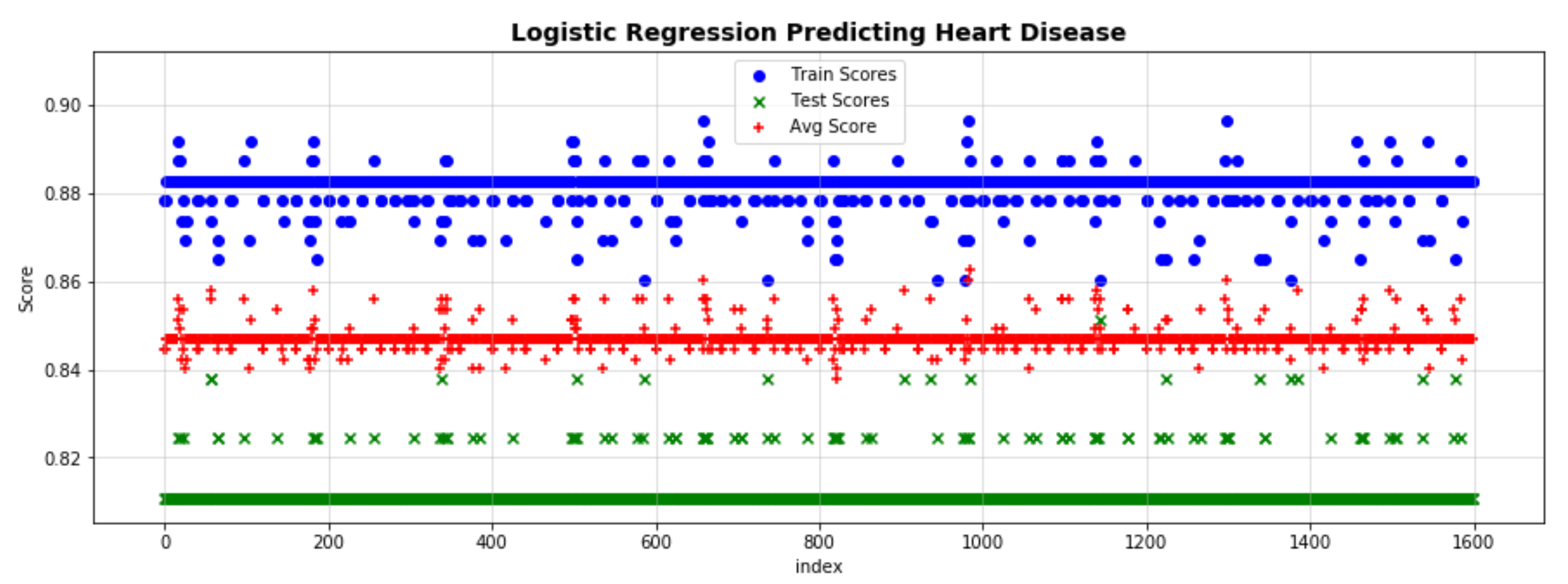
**The scikit learn library implementation of Logistic Regression provides multiple settings to tune a model. Using the Heart Disease data set, four settings were tested as follows using a nested for loops to test all combinations a total of 10 times each:**

* **Tolerances = [0.1, 0.001, 0.00001, 0.0000001]**
* **Weights = ['balanced','none']**
* **Maximum Iterations = [10, 50, 100, 500]**
* **Solvers = ['newton-cg', 'lbfgs', 'sag', 'saga', 'liblinear']**

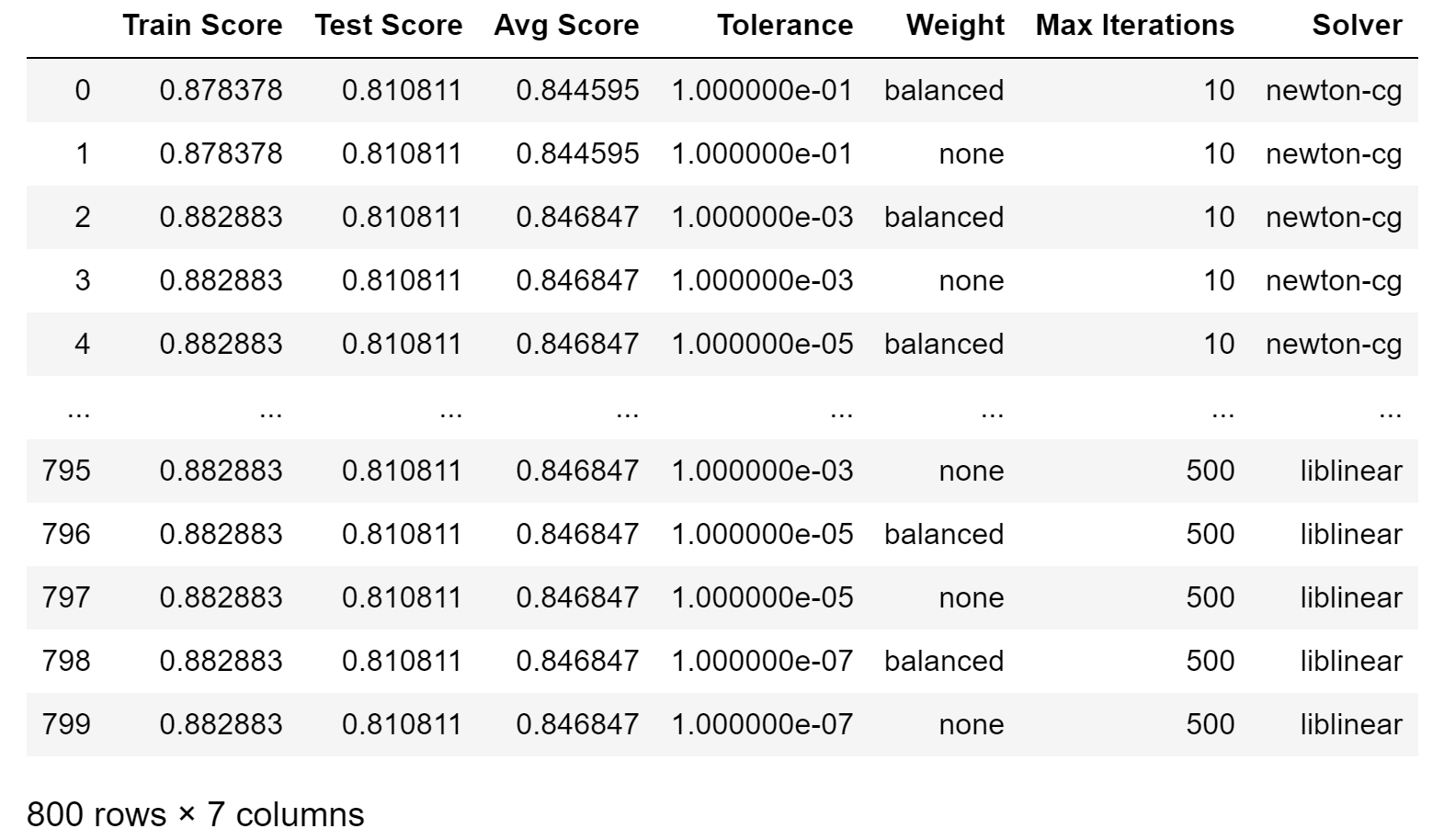
**Results were captured and displayed in a data frame:**

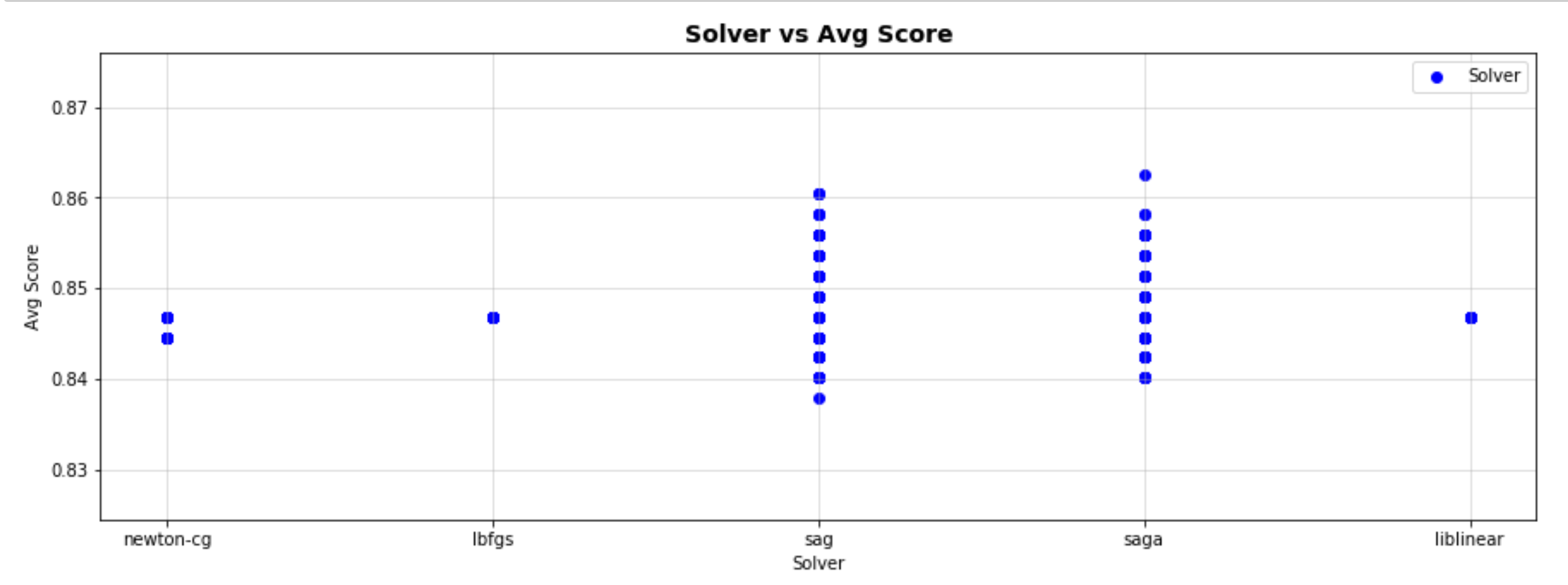


**All scores were plotted in run order:**

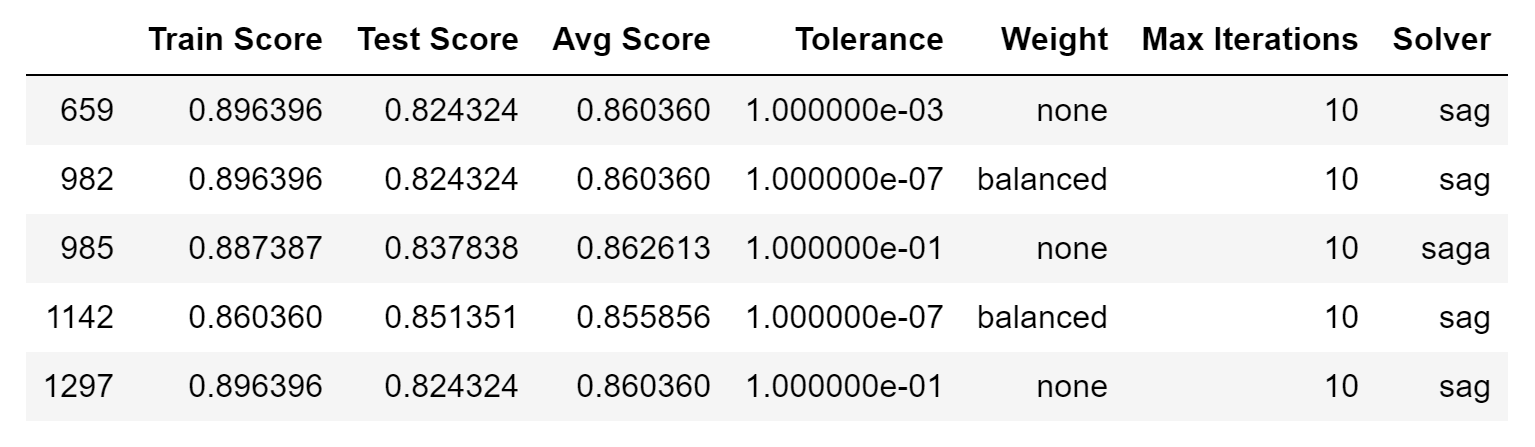


**Results were grouped by the four variables and plotted vs the Average Score:**

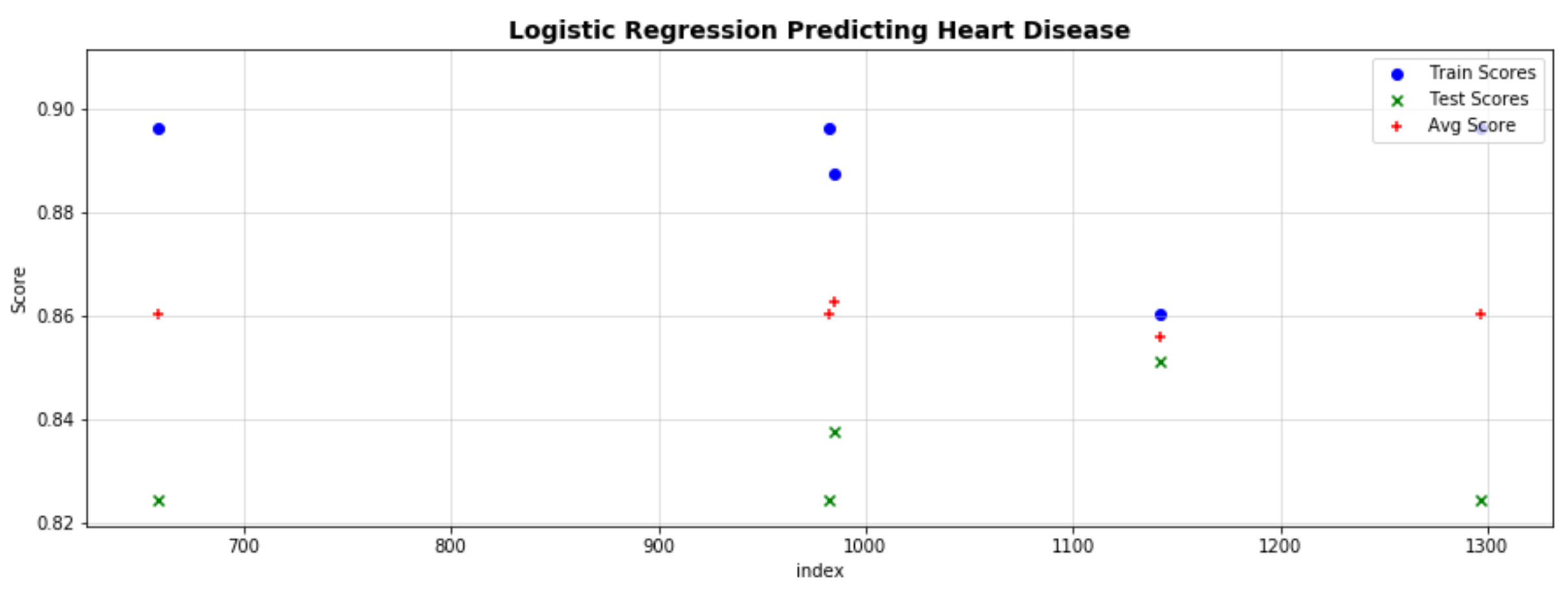




**Finally, the max Avg, Train and Test scores were calculated and the data set was reduced to only rows that contain at least one these max values:**



**These values were then plotted by their index value.**



**Based on the results, sample 1142 was chosen as it has the most consistent scores for all three and the highest for the Train set.**

**The following model was then selected:**

* **Tolerance = 0.0000001**
* **Weight = 'balanced'**
* **Max Iterations = 10**
* **Solver = 'sag'**

**Finally, the test data was run with the new model. The results were broken into ‘same’ (correct) and ‘differences’ (incorrect) populations. They were then plotted vs each variable feature. The mean (ANOVA, Kruskal) and the variation (Student t) of the two populations were checked to ensure they were not different. If they had been, we may have had a problem / opportunity with our model. A sample of the analysis is shown below for max heart rate:**

