**Willy Esquivel-Lopez**

**(wesquive - G01127937)**

**(miner2 user-name): Gopher123**

**HW2 Report**

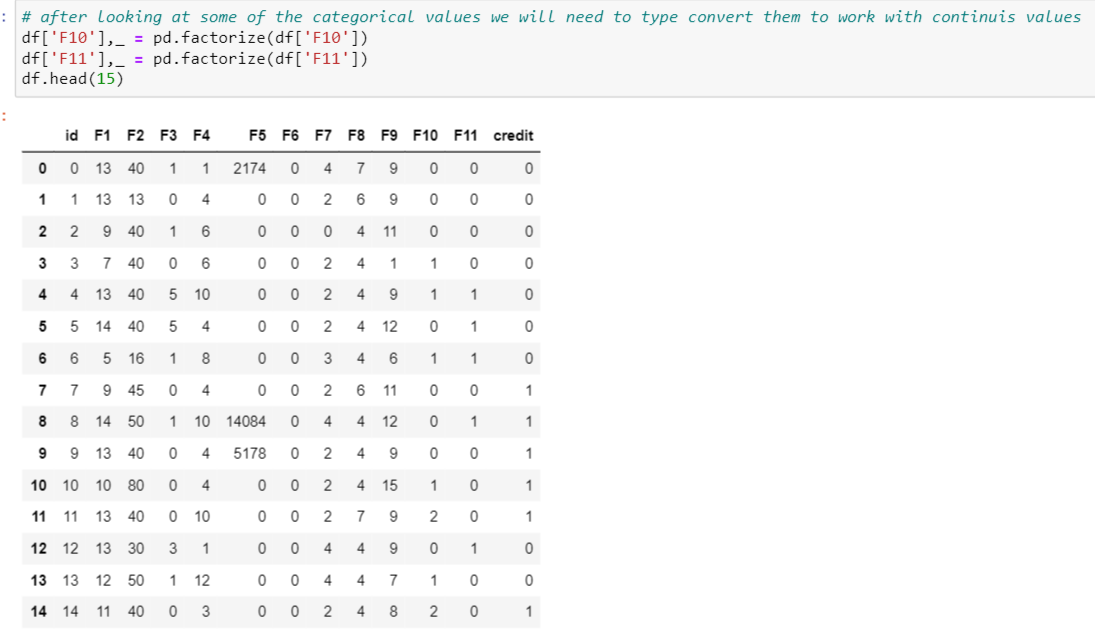
**My Approach:**

For This assignment I approached by looking at the data and seeing what I can do to decide the best model to use for classification. I saw a combination of categorical and continuous data, thus between SVMs and Decision trees, decision trees made more sense as it would be faster to compute, it would be less complex, and would give me more control over the classification.



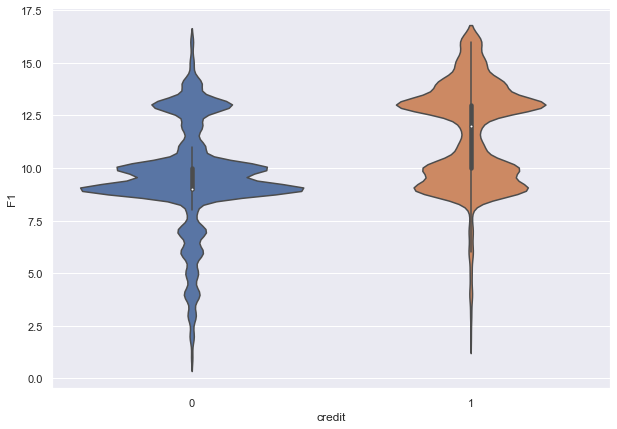
Figure - Train data unchanged

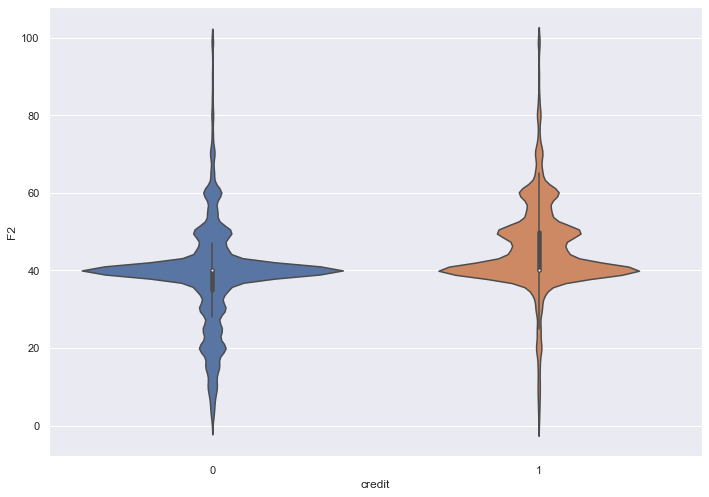
I had to modify two columns to continuous variable to simplify the tree

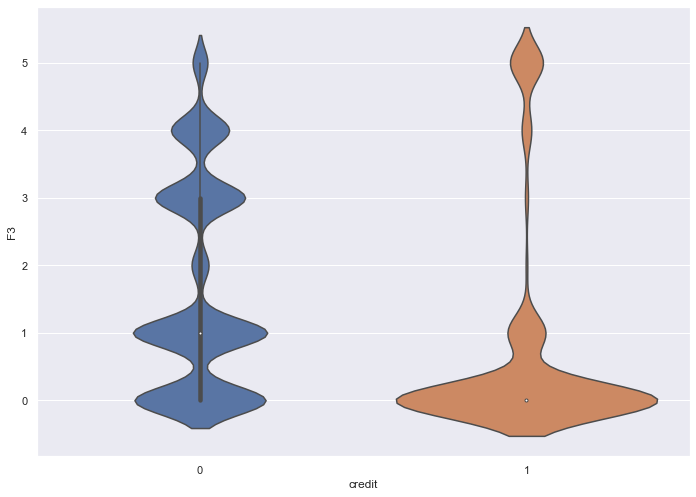


**Analysis:**

To Gather a clearer and broader picture of the distribution in the data I plotted violin plots for each feature against my target value (credit).



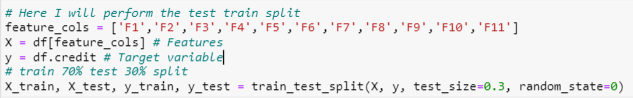


 etc.

F3 had no label but appeared to be crucial to the outcome of the decision tree.

**The Training and Testing:**

I Played it safe and used a 70-30 split for my models. Labeled the features and my target variables like so:



I decided to test the Gini Index on using a decision tree with max depth of 3, 5, and 7. All of which appeared to do fine but would not break on 85% accuracy. This could be due to a