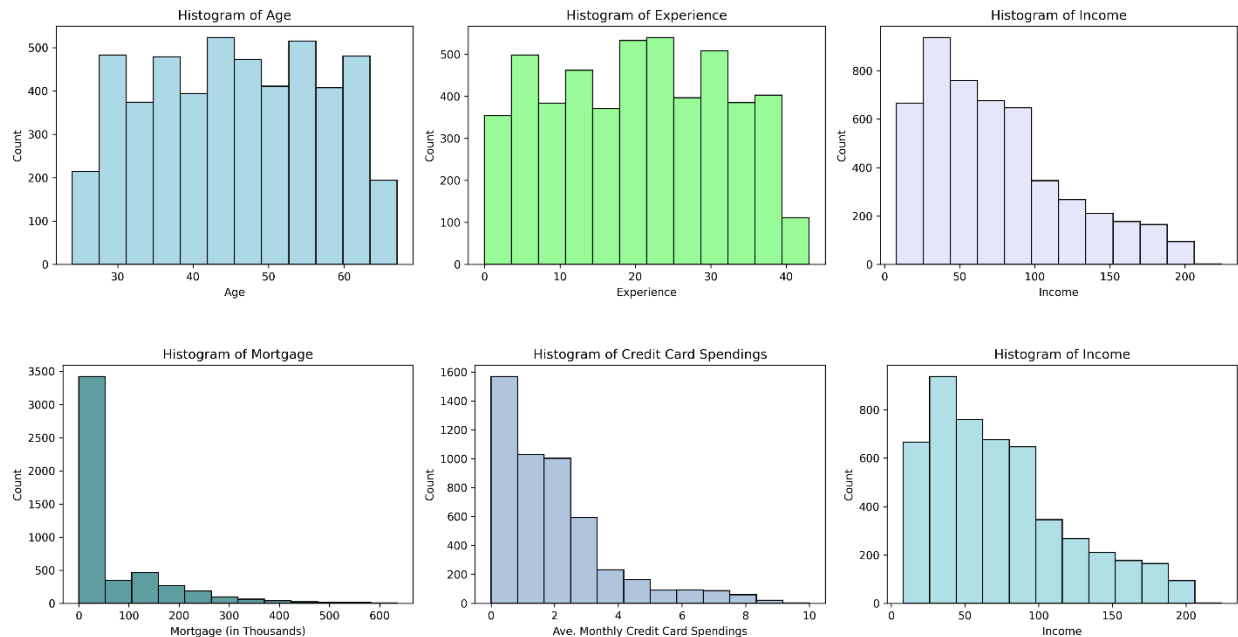


## Cleaning:

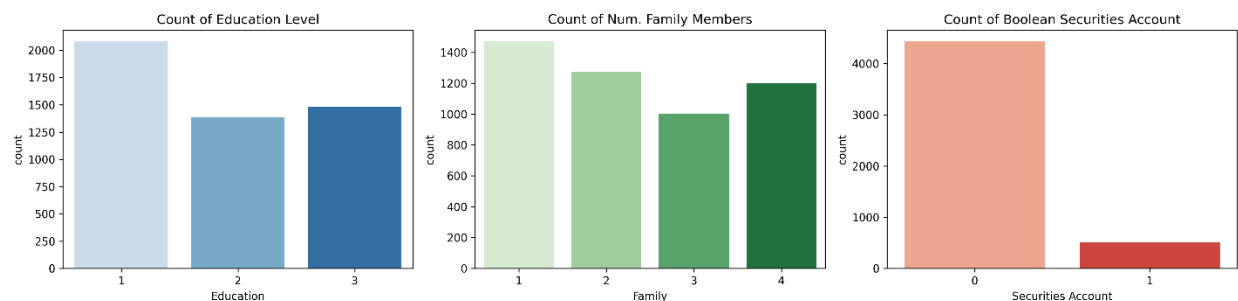
Cleaning the data took a lot less work for this Neural Network project than the Random Forest project. The data was already all numerical and had column headers. I found that there were data points with negative experience, which made no sense, so I removed all lines with negative experience. Likewise, there were zip codes with only four digits, so I removed any zip code below 10000. This only removed 53 values out of the 5000 data points, which should not significantly impact our predictions.

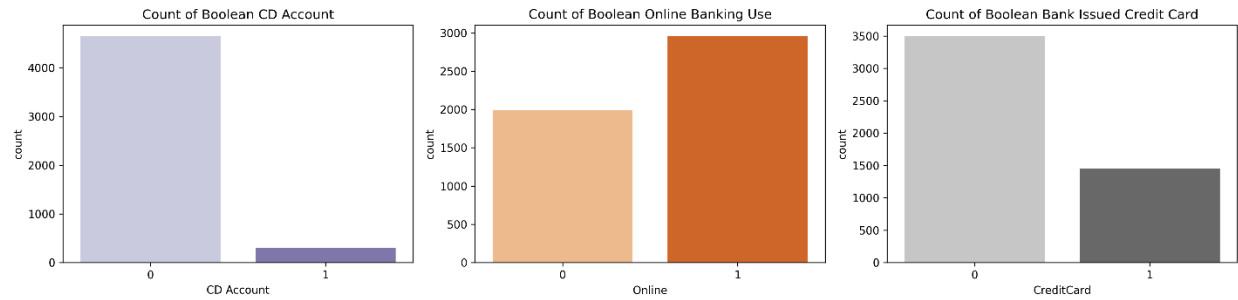
## Looking at counts:

After cleaning and initial analysis of the data, I looked at the counts of each variable. For continuous variables, I used histograms with 12 bins to get an idea of how the variable is distributed. For categorical and binary variables, I used count plots. These graphs are below, starting with the 6 histograms.

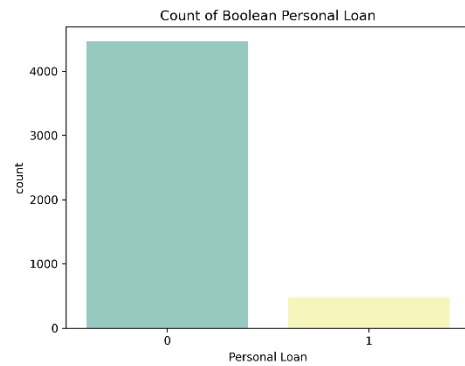


Next are the count plots of the categorical and Boolean values.





Lastly, is the response.



### Model Accuracy, Confusion Matrix, Precision, Recall, F1:

The final model's test accuracy was ~.971. The confusion matrix is as follows,

879	10
18	83

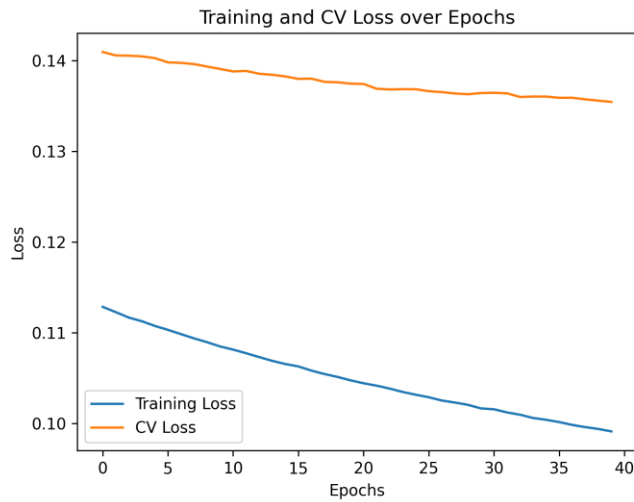
From this confusion matrix, we find the precision, recall, and F1 score to be:

Precision: 0.8924731182795699

Recall: 0.8217821782178217

F1 Score: 0.8556701030927835

## Training and CV loss:



## Summary:

The model's accuracy was very close to being 100%, so not much room for improvement. The most challenging part of this project was parsing the data and deciding on model parameters. This part of the project was easier than the first because the data was all already numeric and didn't have any outliers that I could find. Also, we have made a lot of neural networks for this class, so making the network was pretty simple.