* Correlation, heatmap
* Unfortunately, a lot of the demographic variables seem very strongly related.
* Performed train test split, excluded ‘employed’ variable from test data set
* Doing an 80-20 split
* y\_train only contains only 1 colums with many rows
* using solver=’liblinear’ in our logistic regression model

Classification Report:

* Precision and Recall averages around 50%. Simply means that it can be improved

K-fold cross validation:

* Confusion matrix : diagonal down: shows correct predictions
* Diagonal up shows incorrect predictions
* Precision score: tells us what our accuracy rate of predictions is

Prediction:

* Ok so currently, this model is predicting everything as not employed.
* Its prediction accuracy is very low, below 40%. Which is bad. We want correct predictions to average above 70%.