|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Analysis:   |  | | --- | | After looking at the dataset and considering the two model types, I believe that the logistic regression model | | will perform better for this dataset. Both Random Forest and Logistic Regression models are good for classification | | problems, but I believe with over 80 variables to consider, the Logistic model will win out. Based on my research, | | Random Forest is recommended for simpler classification problems. | |
|  |
| Evaluating the Models - without scaling the data, the Random Forest model performed better on the test data. Perhaps my understanding |
| of what a "simple" classification problem is is lacking, or the lack of scaling the data influenced the results. |
|  |
| Results w/ Scaling - Once I scaled the data, the logistic regression became the better classifier. It had an accuracy score of .66. |
| The Random Forest model decreased in accuracy from .64 to .5. From what I understand, this is to be expected as Random Forest modeling |
| doesn't rely on scaled data to make accurate calculations, while logistic regression models do. |