Regression Project
Due Monday June 3rd

Description of Dataset

Big Company Inc is concerned about employee turnover. They want to keep their data scientists happy (to prevent them from leaving). The below data-set is of data scientists who filled out employee satisfaction surveys and data on weather or not they remained at the company three months after the survey.

The variables in the dataset are:

Stay: Weather or not the data Scientist is there three months later. Pay : Monthly Salary in Dollars Estimated Happiness: This comes from a complicated model reviewing both the employees reported happiness and there comments on what they'd like to see changed. Performance: Results of managers performance review.

Your task is to fit a logistic regression model to predict weather an employee remains with the company or not.

1. Show your data exploration step and show your reasoning [20]

2. Discuss how you selected or engineered variables. Discuss why you did things. Discuss each interation [25]

3. Fit the model using a training set. Use a validation set to tune the hyperparameters/features of your model. Give the final accuracy of your tuned model on the test set (show code) [20]

4. Discuss the model and whether it makes sense in terms of the variables [20]

5. Does the data match the intuition that you gained by looking at various scatter plots [15]