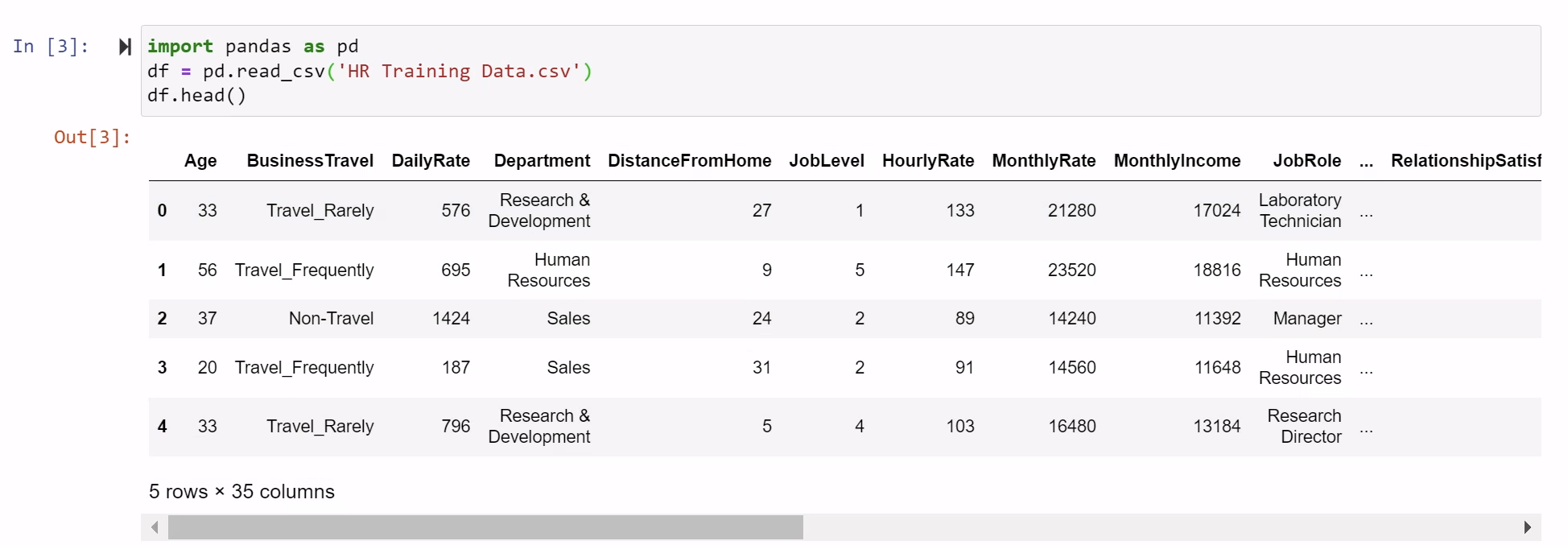
3-2 Assignment: Establishing a Baseline and Hypothesizing Outcomes

DAT-430

Ashley Littles

Southern New Hampshire University

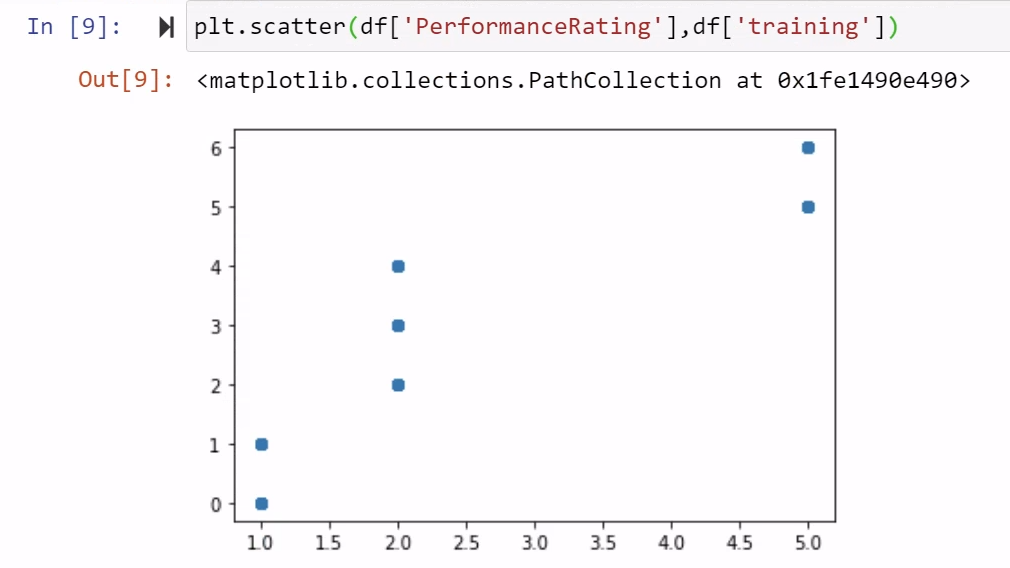
My organization's HR department would like for me to determine if their training has an impact on employee performance or job satisfaction ratings. Overall, they are looking to measure how well their employees are doing with the provided training and with the results the HR department would like to determine if more training would result in much higher ratings in employee performance and job satisfaction.



The first step in creating a model of the training's impact on employee performance and job satisfaction is to import the data into the pandas library. To do this we will use the import pandas as pd function and then specify with df = pd.read\_csv(‘HR Training Data.csv’). Use the command df.head() to show 5 rows and 35 columns of the data set.



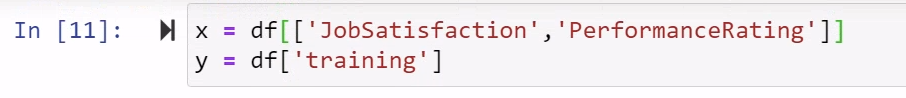
Next, we imported matplotlib.pyplot as plt to plot our values Job Satisfaction, Training, and Performance Rating. Using these commands will help create an inline plot showing any sort of correlation between any of the variables.



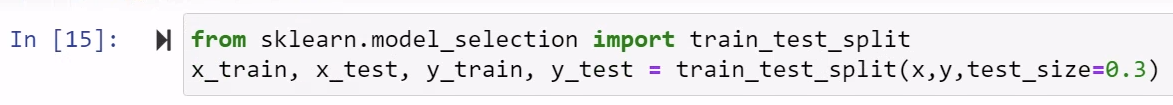
We will utilize the command plt.scatter(df[‘PerformanceRating’],df[‘training’]) to display the graph. This graph compares the Performance Rating to the training with the x-axis being Performance Rating and the y-axis being training. We can see that as the Performance Rating rises so does the training score.



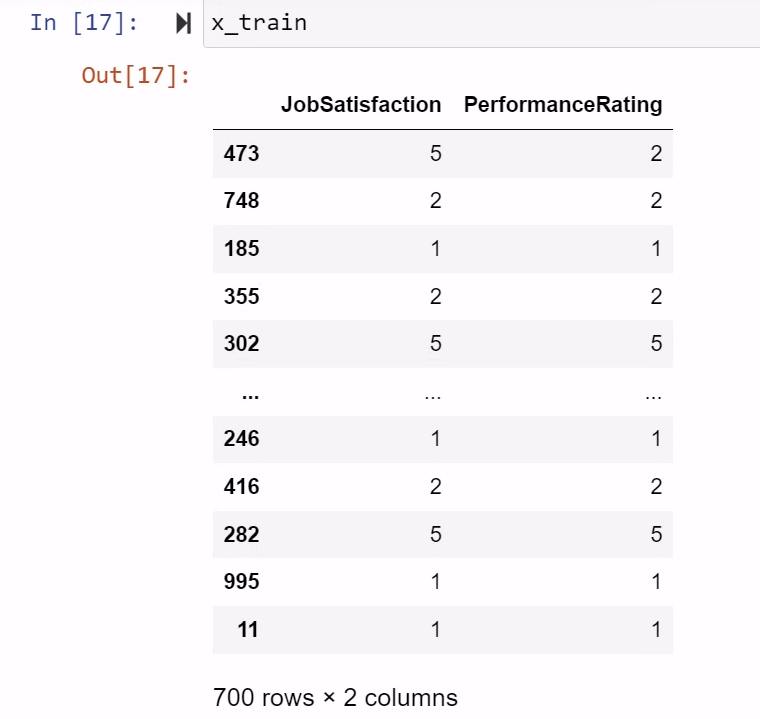
Using the same command, we will plot the Job Satisfaction compared to training. Looking at the plot we can see that as training rises so does the Job Satisfaction. So, we can conclude based off these plots that the more training that is implemented the higher the Performance Ratings and Job Satisfaction ratings are.



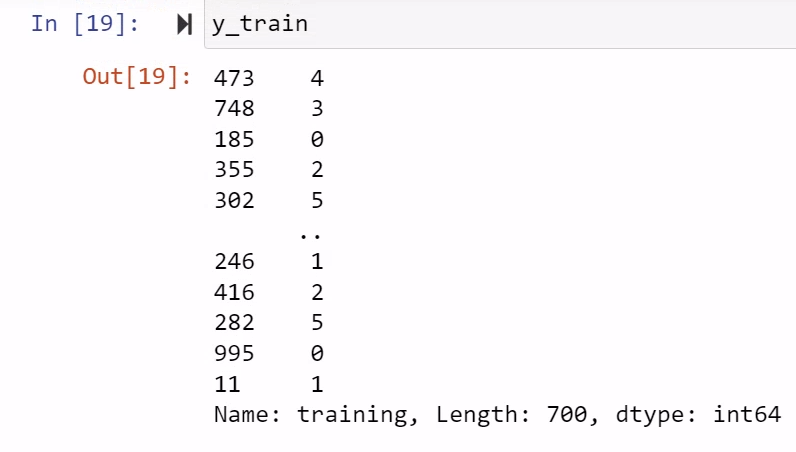
We will continue to define the x and the y using df[[‘JobSatisfactin’,’PerformanceRating’]] and df[‘training’].



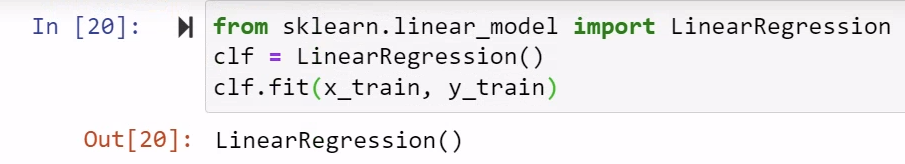
Then we will import and test the training model.

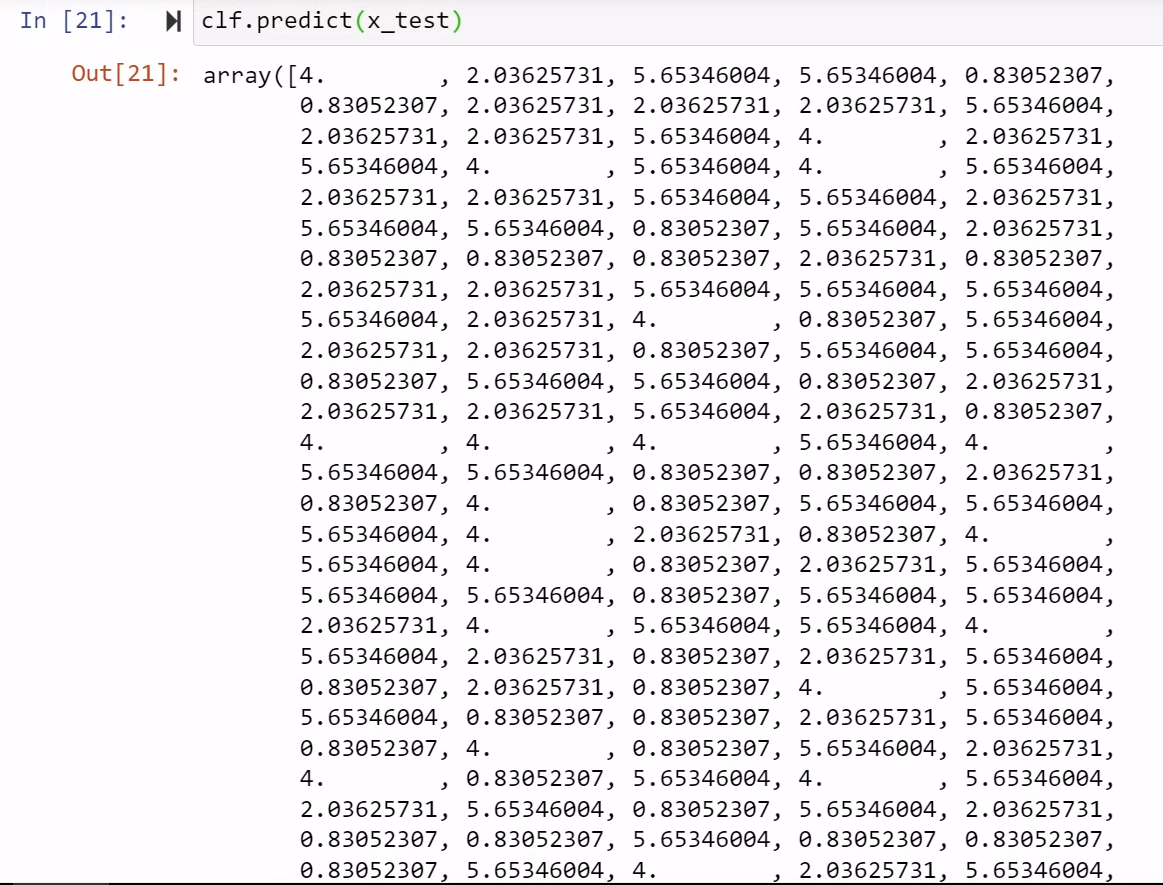


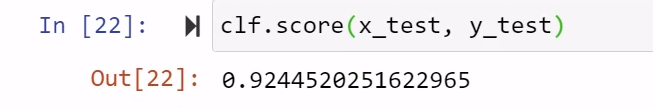
For the x axis we will get the above results from JobSatisfaction and PerformanceRating.



For the y-axis we will get the above results for the training.







Through creating a linear model to test our values with the above commands we can see that the output of the accuracy of the model is 0.924 so 92% accurate. We are able to confirm that there is a correlation between training and employee performance and job satisfaction ratings. More training would most certainly result in higher ratings in employee performance and job satisfaction.

References

Gupta, T. (Aug 23, 2019) *“Data Preprocessing in Python for Machine Learning with working code example…”* towardsdatascience <https://towardsdatascience.com/data-preprocessing-in-python-b52b652e37d5>

Pandas (n.d.) *“pandas.DataFrame.plot.scatter”* <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.scatter.html>