|  |  |  |
| --- | --- | --- |
| **AIN 380** | **Homework #4** | **DUE:** |

Download the zip file containing the data file named **NFL2023Season.csv**. This data file contains win and loss information for all 32 NFL teams. The data is laid out as indicated by the headers and first few rows as shown below.

A screenshot of a table

Description automatically generated

In Homework #4, you are to code a Python program that loads the data from this file into a Pandas DataFrame. Once the data is loaded, your Python code is to do the following:

1. Print the DataFrame in its entirety.
2. Print the single row for the Green Bay Packers Team data. You should use the row integer number.
3. Print the NFL Team column and winning percentage column (PCT) adjacent to each other.
4. Print the following descriptive stats for the Winning Percentage column:
   1. Max
   2. Min
   3. Mean
   4. Median
   5. Range
   6. Standard deviation
   7. Variance

Make sure that this output is very clearly labeled.

1. Print the following descriptive stats for the *Points For*:
   1. Max
   2. Min
   3. Mean
   4. Median
   5. Range
   6. Standard deviation
   7. Variance

Make sure that this output is very clearly labeled.

1. Plot a histogram for the “Net Pts” column of information. Make sure there are no gaps in the histogram and at least seven bins. Be sure to clearly label it with a title and X- and Y-axis labels. Print the histogram for credit.
2. Which do you think are easier to work with: NumPy arrays or Pandas DataFrames? \_\_\_\_\_\_**I like pandas better.**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. Why? Explain from your personal experience?  
   \_\_\_\_\_**In the BUS280 Class all we used was Pandas DataFrames for the entire semester, so I just feel a lot more comfortable with Pandas since I haven’t used NumPy enough.**\_\_\_\_\_\_\_\_\_\_\_\_\_
4. What was the shape of the histogram obtained in #6 above? \_**Non-Symmetic Monomodal**\_\_\_\_\_  
   What do you think this means for the data? \_\_**This means that we have a big clump of data in the middle, a gap on both sides, and then more irregular peaks on the sides of the histogram, meaning that we have a large group of outliers to go along with the clump in the middle.\_\_\_**
5. Use the web to answer this question: What percentage of ML algorithms are Supervised ML algorithms? \_\_**70%**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
   Cite your reference(s): \_ [**https://www.decypher.com/machine-learning-matters/#:~:text=Popular%20Machine%20Learning%20Methods&text=Two%20popular%20methods%20of%20machine,ranges%20from%2010%20%E2%80%93%2020%20percent**](https://www.decypher.com/machine-learning-matters/#:~:text=Popular%20Machine%20Learning%20Methods&text=Two%20popular%20methods%20of%20machine,ranges%20from%2010%20%E2%80%93%2020%20percent). \_\_\_\_\_\_\_\_

Staple a printout of your Python code and output results to this assignment and submit for credit on the due date. Be sure to also upload your Python code in a zip file together with the **dataset** for credit.