

Zihao Li, Jinghao Yan, Brody Vogel, and Ye Zhang

ANLY 501

Project 2

## **README**

1. STOCKS\_FINAL.csv contains the cleaned and treated stock market data from Project 1.

2. SPORTS\_FINAL.csv contains the cleaned and treated sports data from Project 1.

3. Project2.py contains the Python script that reads in the data and goes through the Project 2 analyses. Its annotations explain its workings. Some of the tests have been cut down to samples to lower the runtime; it should complete in 10-15 minutes. The runtime for the full tests would be something like an hour and a half.

*~~~Note that a part of the code entitled “Hypothesis 1” creates a folder on the user’s machine  
~~~and fills it with images of the ROC curves. Comment out if preferred.*

*~~~Also note that the first clustering portion saves multiple images of the clustering plots on  
~~~the user’s machine. Comment out if preferred.*

4. Project2Writeup.pdf is the turn-in report of Project 2. In it, there are explanations for, and results of, everything the Python code does. Also, this contains the interpretation of the results.