Welcome!

This is a simple exercise, it is only intended to test your knowledge of Scientific Python, your approach to data science, and your ability to explain what you've done to someone else.

It also shows if you are able to learn new tools and technologies, so we specify which ones you should use. If you already know all of them - good for you :)

It is an open exercise. You are given one dataset to explore.
We expect to see how you think, perform exploratory data analysis, present the data as you see it and discover meaningful behaviors worth mentioning.

You are free to focus on parts of the dataset itself and highlight only what you think is interesting.

Write clear and comprehensive code, and make sure you structure your analysis into easily understood steps.

Use https://ipython.org and http://jupyter.org as your data science environment.

Use http://pandas.pydata.org to load and manipulate the data.

Try and incorporate http://scikit-learn.org as part of your code.

All frameworks have nice examples on homepages, so if you didn't use any of them before, you should be able to look at examples and figure out how to do it.

Your end result should be a jupyter notebook.
Upload your solution to a public github repository where we can see it. It should contain clear instructions about how to launch the code you wrote.

Send us the link to the repository when you're done.