Complete the following steps and answer the questions by Oct. 15th. Upload the results on GitHub.

1. Setting Up GitHub Repository
   1. GitHub Repo for our 4-Week STEM program is made. You should have received an email inviting you to it. If you don’t have an access to it, please reach out to me. In the repository, you will be able to find a sample script that we went over during our first meeting and a dataset “SHIPS\_ATL.csv.” Open the csv file and inspect the data and proceed to Task 2.
2. Understanding the SHIPS data
   1. Inspect the dataset and answer the following questions:
      * What is a target variable and feature set?
      * In the data provided, what would be the target variable and the feature set?
3. Notebook Creation and Data Import
   1. Create a new notebook on Jupyter. If you don’t have access to Jupyter you can easily Google search it or refer to the following YouTube video: [LINK](https://www.youtube.com/watch?v=HW29067qVWk&ab_channel=CoreySchafer).
   2. Import pandas and matplotlib libraries (refer the sample script)
   3. Import “Dataset\_SHIPS\_RII\_ATL\_V2.csv” by using “pd.read\_csv()”
      1. You could either specify a file that is saved onto your local computer
      2. Or you could input URL address of the data. In order to refer to the data in our repository, type: “https://raw.githubusercontent.com/kwonkh0424/STEM\_Hurricane/main/Dataset\_SHIPS\_RII\_ATL\_V2.csv?token=AQBXNKZDHVA7YQHOSKQA7ES7QILQK”
4. Data Inspection
   1. Write a script to inspect the first 10 rows of the data and answer the following questions:
      1. What is/are the name(s) of the hurricane(s) shown?
      2. What are the columns shown?
      3. What does “9999” mean?
5. Groupby Function
   1. Let’s start doing some calculations with the data. Use “.groupby().mean()” function to group the data by “ID” and calculate means of the numerical columns. What happened to the data? Are there any columns missing and why do you think it happened?
6. Data Visualization
   1. Let’s create a few visualizations. First, let’s select the first 7 hurricanes by using “.iloc[ ]” function. And then, create a bar chart that compares the average wind speed of those seven hurricanes. (Refer to this [LINK](https://benalexkeen.com/bar-charts-in-matplotlib/) if you want to learn about how to create a bar chart). Add a title, x-axis label, and y-axis label.
   2. Conduct another groupby calculation using “Year” column. Now plot a scatter plot showing average wind speed for each year (Here is the [LINK](https://jakevdp.github.io/PythonDataScienceHandbook/04.02-simple-scatter-plots.html) to a website that explains how you can scatter plot. Is there any trend shown?
7. Do some additional work to find out any trend or interesting statistics about the data.