This assignment has been created in order to help you prepare for the course. We would like everyone to complete it before the first live session (Saturday May 14, 2022).

- Read the following. Go in order. You will be asked “What is one thing you learned” regarding each assigned chapter at the end of this assignment. While you read, you can draft your thoughts somewhere that you can easily access when you’re ready to submit.
    - Chapter 1 (“What Is Data Science?”) from *Developing Analytic Talent* by Vincent Granville. We recommend that you download this book as a pdf and read from it locally.
        - The link above should load if you give it a minute (it loads on our end). But if it still does not load for you, please try accessing the book via google drive here.
    - Chapter 2 (“Introduction to NumPy”), Chapter 3 (“Data Manipulation with Pandas”), and Chapter 4 (“Visualization with Matplotlib”) from *Python Data Science Handbook* by Jake VanderPlas.
- Download the dataset named *exam_grades.csv* (attached in the welcome email thread).
- Start a new jupyter notebook in python3.
- Code to complete the following tasks/answer the following questions about the dataset. Go in order. Save your code along the way.
    - Import the pandas library.
    - Load the dataset into a pandas dataframe.
    - How many rows are in the dataset?
    - How many columns are in the dataset?
    - How many unique student IDs are in the dataset?
    - What is the list of unique student IDs in the dataset?
    - How many missing grades are there across all exams in the dataset? Remember that a missing value would be represented as NaN in python.
    - Let’s say: The dataset represents a gradebook, containing students’ exam grades in a particular class. The protocol is that when a student doesn’t make up an exam they missed, their missing grade needs to be replaced with a zero at the end of the term. If you noticed any missing exam grades in the dataset, replace them with zeros, to reflect the class protocol. Make sure to finish this before proceeding.
    - Let's say: You’re interested in looking closer at the exam grades that are considered passing. You could select a subset of the data containing all exam grades greater than or equal to 70. Do this.
    - How many exam grades are passing (i.e. greater than or equal to 70)?
    - Let’s say: You want to know how students performed on average in each exam. You could group the data by exam and get the mean grade per exam. Do this.
    - What is the mean grade for exam 1?
    - What is the mean grade for exam 2?
    - What is the mean grade for exam 3?
    - What is the mean grade for exam 4?
    - What is the mean grade for exam 5?

    - Import the pyplot module in matplotlib's library and give it the alias of plt, to keep things short and sweet.
    - To visualize the distribution of the exam grades, create a histogram.
    - Submit your code via file upload. Do this only after you've done all of the above.
- Fill out this form here to enter your answers to all the questions.