PRE-PREP

The participants at DSR come from various background and skill level when it comes to Data Science and Machine Learning. As we are an advanced level training program, we expect our participants to have fundamental or basic knowledge of Python and Machine Learning before the start of the program. This way we can dive deeper into more important topics during the program.

So here are few resource materials for your reference which you can use to do some pre-preparation before starting the program.

1.   Python

We will cover some advanced Python topics like Unit Testing or OOP, so it is expected from the participants to know the Python basics.

- Books
	- Learning Python, 5th Ed. Mark Lutz. O'Reilly 2013
	- Python Crash Course, 3rd Ed. Eric Matthes. No Starch Press. 2023
- Videos
	- Learn Python, Full Course for Beginners. Freecodecamp.org
	https://www.youtube.com/watch?v=rfscVS0vtbw
	- Python Tutorial - Python Full Course for Beginners. Programming with Mosh
	https://www.youtube.com/watch?v=_uQrJ0TkZlc
- Online courses
	- https://developers.google.com/edu/python?hl=en
	- https://learn.microsoft.com/en-us/training/modules/intro-to-python/
	- https://www.udemy.com/course/pythonforbeginnersintro/
	- https://www.educative.io/courses/learn-python-3-from-scratch
- Check your knowledge here:
	- https://github.com/rachelkberryman/Intro_to_Python_DSR

2.   Numpy and Pandas

You can start taking a look to this important python libraries. You can find some resources here:

- Books
  - Python Data Science Handbook, 2nd Ed. Jake VanderPlas. O'Reilly. 2023
  - Hands-On Data Analysis with NumPy and Pandas. Curtis Miller. Packt. 2018
- Videos
  - Python NumPy Tutorial for Beginners. FreeCodeCamp.Org
  https://www.youtube.com/watch?v=QUT1VHiLmmI
  - NumPy Full Python Course - Data Science Fundamentals. NeuralNine
  https://www.youtube.com/watch?v=4c_mwnYdbhQ
  - Pandas & Python for Data Analysis by Example – Full Course for Beginners. FreeCodeCamp.org
  https://www.youtube.com/watch?v=gtjxAH8uaP0

3.   Machine Learning

For machine learning, we expect students to have:

- Familiarity with random forests & logistic regression, bias & variance, understand the difference between test & train error
- Implemented a machine learning project from scratch, making use of pandas, matplotlib & scikit-learn. The Titanic or MNIST datasets are a good place to start
- Andrew Ng's Stanford Machine Learning is a classic course that is somewhat of a rite of passage for machine learning practitioners