**February 27, 2024**

Creating a GitHub repository and testing Git Bash to push files from a local folder.

**February 28, 2024**

Starting reading about the assignment topic ***“combining Advanced data analytics (specifically Neural Networks) and Big Data Storage & Processing”.*** I’ve found interesting books to enrich my knowledge about this subject:

1. Mastering Apache Spark 2.x (Second Edition) by Romeo Kienzler. Packt.
2. Advanced Analytics with Spark by Sandy Ryza, Uri Laserson, Sean Owen and Josh Wills. O’Reilly.
3. Deep Learning with Python by Francois Chollet. Manning.
4. Large Scale Machine Learning with Spark by Rezaul Karim and Mahedi Kaysar. Packt.
5. Learning Spark by Holden Karau, Andy Konwinski, Patrick Wendell and Matei Zaharia. O’Reilly.
6. Mastering Machine Learning with Python in Six Steps, by Manohar Swamynathan. Apress.

I’ve installed Apache Spark 3.5.1 on my Windows 10 source:  
<https://www.youtube.com/watch?v=0F4fokX5MPQ>

To open Apache on http://desktop-ik25uha:4040: C:\Program Files\spark-3.5.1-bin-hadoop3\bin>spark-shell

I had to install Java Development Kit aka JDK. When you install Apache Spark or any other Java-based software, you need to have JDK installed on your system to ensure that the necessary tools and libraries are available for development and execution. Source:  
<https://www.youtube.com/watch?v=-O4QVijnA7Y>

The next step is to push records into Apache Spark and link a Jupyter Notebook to it, stablish the connection and pull records into a pandas dataframe.

Videos to connect Jupyter Notebooks to Apache:

Part 1:

https://www.youtube.com/watch?v=XvbEADU0IPU

Part 2:

https://www.youtube.com/watch?v=e\_QoFQjZwqc