**Steps to run Bigdata (Final project) code:**

* With a sentiment dataset and word cloud visualizations - This can be executed as normal “. ipynb” file named "**Data603\_Final\_Project\_Amazon\_Data\_Gundu\_Kavali\_wordclouds. ipynb**"
* Second file is required to have **MongoDb** connection as the data is retrieved from **MongoDb** in the notebook. This notebook will also need Apache Spark as it is coded in Apache Spark named "Data603\_Final\_Project\_Amazon\_Data\_Gundu\_Kavali.ipynb"
* Initially with csv file "**Amazon\_Products.csv**", create database: "**AmazonReviews**" with Collection: "**Products**"
* later, through code, cleaned dataset with columns with proper datatypes for modelling is created then stored in database with csv file "**Amazon\_2022.csv**"(which is created in the path in which this jupyter notebook is running).
* Later, with this file database is created in mongoDB with database: "**Amazon\_2022**" and collection: "**Review\_ml**", as it is created twice because most of the data needs ETL operations on columns as it contained several special characters which are not feasible for working on ml models.
* with csv file "**DatafinitiElectronicsProductData.csv**", third database is created, in which database: "**Amazon\_2018**" and collection: "**reviews**".
* Later csv files "**DatafinitiElectronicsProductData.csv**", "**Amazon\_Products.csv**" and "**test.csv**" ("**train.csv**") stored in the path as per the code.