Name: Chander Parkash
Roll No: DS 032 - 2024-25
Course: Big Data Analytics

- Prepared requirement.txt file with required documentation
- Execute script for requirement.txt for installation of packages
- Prepared dummy data files as below
    - Sports_data.csv
    - Sports_data.xlsx
    - Sports_data.json
- Created new collection with database on mongo db
- Insert random data from file to mongodb
    - If file is not present it'll generate random 1000+400 records in mongo db
- Extract data from files first
- Extract data from mongo then
- Transform them in single manner by removing duplicates with respect to (name, country, sports)
- Loaded data in new database in sports_data collection by removing duplicates as well
- Then executed output file for response

Tools:

- Google Collab
- Python
- .csv
- .xlsx
- .json
- MongoDB (un-structured)