

Para utilizar los datos que se generan utilizando el código de “generate_data.py” se debe de utilizar el comando:

- spark-submit generate_data.py

Para correr el programa principal se debe de contar con los archivos con nombre atleta.csv, nadar.csv y correr.csv y estos deben de estar en la carpeta base la cual es “Tarea 1” y utilizar el comando:

- spark-submit programaestudiante.py atleta.csv nadar.csv correr.csv

Para correr los test que se realizaron se debe de correr el comando:

- pytest

En la carpeta principal de “Tarea 1” y este comando correrá los test que se crearon, para ver que realiza cada test en la carpeta “tests” se encuentran los archivos de test con un pequeño comentario sobre lo que está realizando el test.

Los archivos csv intermedios que se generan y los resultados finales que se generan en archivos csv se crearan en la carpeta principal de “Tarea 1”.