Instituto Tecnológico de Costa Rica

Programa de Ciencia de los Datos - Módulo Big Data

Tarea #1

- Esteban Sáenz Villalobos (**esaenz7@gmail.com**)

- Entrega: 08 de agosto 2021, 23:00.
- Observaciones: Trabajo elaborado desde Google Colab. Ejecutar cada celda de código de forma secuencial.

Instrucciones

```
1- Para cargar el contenedor con los recursos necesarios, ejecute los archivos
build_image.sh y run_image.sh.
2- Para ejecutar el programa principal se debe aplicar el siguiente comando:
    #spark-submit programaestudiante.py ciclista.csv ruta.csv actividad.csv
3- Para ejecutar las pruebas del programa se debe aplicar el siguiente comando:
    #python -m pytest -v
4- Para ejecutar las instrucciones 2 y 3 de forma automática, ejecute el archivo run.sh.
```

Nota1: El código fuente en cada archivo cuenta con comentarios detallados que explican la lógica del programa. Nota2: Se inluye un jupyter notebook de Google Colab con todo el código necesario, como complemento.