

Ejercicio 3. Interacción Spark y Python

El objetivo es que el estudiante entienda la interacción del framework Spark y Python, en vista de realizar preprocesamiento y análisis de datos.

Realizar:

Lecturas básicas

<https://spark.apache.org/docs/0.9.2/python-programming-guide.html>

<https://towardsdatascience.com/a-neanderthals-guide-to-apache-spark-in-python-9ef1f156d427>

<https://www.diegocalvo.es/tutorial-de-python-spark/>

Material de soporte

Libro: Data Science with Apache Spark (draft)

<https://george-jen.gitbook.io/data-science-and-apache-spark/>

Realizar los ejercicios de

<https://towardsdatascience.com/six-spark-exercises-to-rule-them-all-242445b24565>

para mayor agilidad, ejecutar los códigos python en línea de comandos (no es necesario hacer los notebooks).

Entregar por correo el link de descarga de un archivo comprimido con lo siguiente:

- códigos python
- reporte con lo siguiente
 - descripción de lo realizado
 - screenshots (archivos jpg, png, etc.) de las pantallas legibles de la ejecución del ejercicio. Nombrarlas de acuerdo al contenido del reporte. Ej.
 - En la Figura 1.1 puede verse la graficación de los datos
 - Entregar el archivo “Figura1.1.jpg”
- fecha de entrega: 14/02/2021 – 23:59