

Tarea 2

Jose Piedra Venegas

Se crean 5 archivos .json: Caja_21, Caja22, Caja23, Caja24, Caja25
Estos archivos tienen al menos 10 compras de 1 o varios productos cada una.

1. Construir imagen de Docker (tarea2_jpiedra)

```
...  
sh build_image.sh  
...
```

2. Correr imagen de Docker (tarea2_jpiedra)

```
...  
sh run_image.sh  
...
```

3. Dentro del shell de la imagen (tarea2_jpiedra) ejecutar spark-submit

```
...  
sh run_utilsCajaCompra.sh  
...
```

Se generan 3 carpetas: metricas, totalcajas, totalproductos

Se generan en la carpeta principal: metricas.csv, totalcajas.csv,
totalproductos.csv

4. Las pruebas se ejecutan con

```
```\n\npytest\n\n```\n
```

y son las siguientes:

- test\_CajasTabuladas: prueba la funcion de tabulacion de las compras de una caja, la #21
- test\_obtTotalProducto: toma dos cajas #21 y #22 y suma los productos comunes en las dos
- test\_test\_obtTotalCajas: suma el total de compras (dinero) de cada caja, 21 y 22
- test\_test\_obtMetricas: Prueba las metricas pedidas:
  - test\_single: prueba de promedios ponderados con una sola fila de estudiante
    - caja\_con\_mas\_ventas
    - caja\_con\_menos\_ventas
    - percentil\_25\_por\_caja
    - percentil\_50\_por\_caja
    - percentil\_75\_por\_caja
    - producto\_mas\_vendido\_por\_unidad
    - producto\_de\_mayor\_ingreso