

## Manual de Instrucciones de ejecución del programa Tarea\_2.py

A continuación se describen los pasos para la ejecución del programa Tarea\_2.py. La parte correspondiente a los puntos extras no fue realizada.

### Ejecución del programa principal:

El programa puede ejecutarse con el comando en la imagen de abajo. Nótese que debe indicarse al menos un archivo json al ejecutar el archivo. Para el ejemplo de abajo se incluyen cinco archivos json.

```
bash-5.0# spark-submit Tarea_2.py compras_1.json compras_2.json compras_3.json compras_4.json compras_5.json
```

### Total de Productos:

24/04/18 02:10:30 INFO CodeGenerator: Co

Nombre	Cantidad_Total
Tequila	1
aceite de girasol	1
aceite de oliva	1
aderezo	2
aguacate	4
almendras	1
arroz	2
atun	2
bananos	4
barquillos	6
bolsas	12
brocoli	14
cafe	4
calamares	4
cebollas	1
cereal	6
cervezas	3
chocolate	5
ciruelas	6
coca cola	2

only showing top 20 rows

El archivo csv llamado *'total\_productos'* puede encontrarse en la ubicación indicada en la imagen abajo.

```
24/04/18 02:10:37 INFO ShutdownHookManager: Detecting directory /tmp/spark-a03140c3-94cd-4a4a-80b7-010b2e330294
bash-5.0# ls
Tarea_2.py      __pycache__  compras_2.json  compras_4.json  conftest.py    test_Tarea_2.py  total_productos
__init__.py     compras_1.json  compras_3.json  compras_5.json  metricas        total_cajas
bash-5.0# cd total_productos
bash-5.0# ls
SUCCESS
part-00000-3c445e96-780e-4d5b-918b-b34f08e21e5c-c000.csv
bash-5.0# cat part-00000-3c445e96-780e-4d5b-918b-b34f08e21e5c-c000.csv
Nombre,Cantidad_Total
Tequila,1
aceite de girasol,1
aceite de oliva,1
aderezo,2
aguacate,4
almendras,1
arroz,2
atun,2
bananos,4
barquillos,6
bolsas,12
brocoli,14
```

### Total Cajas:

numero_caja	Total_Vendido
1	230288.0
2	217555.0
3	195033.0
4	192566.0
5	242033.0

El archivo csv llamado *'total\_cajas'* puede encontrarse en la ubicación indicada en la imagen abajo.

```
bash-5.0# ls
Tarea_2.py      __pycache__  compras_2.json  compras_4.json  conftest.py    test_Tarea_2.py  total_productos
__init__.py     compras_1.json  compras_3.json  compras_5.json  metricas        total_cajas
bash-5.0# cd total_cajas
bash-5.0# ls
SUCCESS
part-00000-a0f7ed14-3fc4-4d4c-a39d-23276e5dc230-c000.csv
bash-5.0# cat part-00000-a0f7ed14-3fc4-4d4c-a39d-23276e5dc230-c000.csv
numero_caja,Total_Vendido
1,230288.0
2,217555.0
3,195033.0
4,192566.0
5,242033.0
bash-5.0#
```

## Metricas:

Tipo_de_Metrica	Valor
caja_con_mas_ventas	5
caja_con_menos_ve...	4
percentil_25_por_...	195033.0
percentil_50_por_...	217555.0
percentil_75_por_...	230288.0
producto_mas_vend...	salchichas
producto_de_mayor...	salmon

El archivo csv llamado '*metricas*' puede encontrarse en la ubicación indicada en la imagen abajo.

```
Tarea_2.py      __pycache__  compras_2.json  compras_4.json  conftest.py    test_Tarea_2.py  total_productos
__init__.py    compras_1.json  compras_3.json  compras_5.json  metricas        total_cajas
bash-5.0# cd metricas
bash-5.0# ls
_SUCCESS
bash-5.0# cat part-00000-6414c692-a657-4a4a-a6f1-18cb566e0c7c-c000.csv
part-00000-6414c692-a657-4a4a-a6f1-18cb566e0c7c-c000.csv
Tipo_de_Metrica,Valor
caja_con_mas_ventas,5
caja_con_menos_ventas,4
percentil_25_por_caja,195033.0
percentil_50_por_caja,217555.0
percentil_75_por_caja,230288.0
producto_mas_vendido_por_unidad,salchichas
producto_de_mayor_ingreso,salmon
bash-5.0# |
```

## Pruebas Unitarias:

Para la ejecución de las pruebas unitarias debe correrse el comando indicado en la imagen de abajo:

```
bash-5.0# pytest test_Tarea_2.py
```

Abajo se muestra el resultado de la ejecución del comando anterior

```
test_Tarea_2.py ... [100%]  
===== warnings summary =====  
../../../../usr/lib/python3.7/site-packages/pyspark/context.py:317  
  /usr/lib/python3.7/site-packages/pyspark/context.py:317: FutureWarning: Python 3.7 support is deprecated in Spark 3.4.  
    warnings.warn("Python 3.7 support is deprecated in Spark 3.4.", FutureWarning)  
-- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html  
===== 3 passed, 1 warning in 17.21s =====  
bash-5.0# |
```