Assignment5_1

January 3, 2023

###

Work and Pairing together Ovais Saleem (2211-023-KHI-DEG) and Muhammad Humza (2211-018-KHI-DEG)

[1]:
```
from pyspark.sql import SparkSession
from pyspark.sql import SQLContext
from pyspark.sql.types import IntegerType
from pyspark.sql.functions import *
```

[2]:
```
scSpark = SparkSession.builder.appName("Assignment5_1").getOrCreate()
```

[3]:
```
df1 = scSpark.read.csv("data/customers.csv", header=True)
```

[4]:
```
df2 = scSpark.read.csv("data/products.csv", header=True)
```

[5]:
```
df_merged = scSpark.read.csv("data/store_transactions/transactions_*.csv",␣
↪header=True)
```

[6]:
```
df_merged.show()
```

```
+-------+-------------+----------+---------+--------+-------------------+
|StoreId|TransactionId|CustomerId|ProductId|Quantity|    TransactionTime|
+-------+-------------+----------+---------+--------+-------------------+
|      3|          454|        35|        3|       3|2022-12-23 17:36:11|
|      3|          524|        37|        9|      11|2022-12-23 22:02:51|
|      3|          562|         4|        3|       4|2022-12-23 02:51:50|
|      3|          581|        35|       14|      56|2022-12-23 17:05:54|
|      3|          200|        34|       15|      24|2022-12-23 07:15:01|
|      3|          506|        41|       24|      19|2022-12-23 21:26:29|
|      3|          278|         5|        1|       5|2022-12-23 16:41:42|
|      3|          849|        36|       23|      13|2022-12-23 13:22:55|
|      3|          992|        34|        7|       3|2022-12-23 16:47:14|
|      3|          703|        19|        7|      13|2022-12-23 22:36:48|
|      3|          719|        48|       18|      12|2022-12-23 10:11:29|
|      3|          526|        13|       14|       3|2022-12-23 11:57:23|
|      3|          997|        20|        1|      14|2022-12-23 04:02:30|
|      3|          281|        11|       15|      25|2022-12-23 16:07:45|
|      3|          691|        48|       23|       2|2022-12-23 08:12:00|
```

```
|      3|          762|        17|         5|      26|2022-12-23 16:18:27|
|      3|          106|        24|        23|      11|2022-12-23 07:41:50|
|      3|           21|        32|         9|       2|2022-12-23 21:15:10|
|      3|          626|        14|        18|      14|2022-12-23 12:55:02|
|      3|          219|        11|        15|       5|2022-12-23 13:00:17|
+-------+-------------+----------+----------+--------+-------------------+
only showing top 20 rows
```

[7]:
```
df3 = df2.join(df_merged,df2.ProductId ==  df_merged.ProductId, how= "left")
```

[8]:
```
df3 = df2.withColumnRenamed("Name", "Product_Name").join(df_merged,␣
  ↪['ProductId'])
```

[9]:
```
df3.show()
```

```
+---------+------------+--------+---------+-------+------------+----------+---
-----+-------------------+
|ProductId|
Product_Name|Category|UnitPrice|StoreId|TransactionId|CustomerId|Quantity|
TransactionTime|
+---------+------------+--------+---------+-------+------------+----------+---
-----+-------------------+
|        3|  Blue Shorts|  Shorts|   118.88|      3|         454|        35|
3|2022-12-23 17:36:11|
|        9|Green Sandals|   Shoes|   137.53|      3|         524|        37|
11|2022-12-23 22:02:51|
|        3|  Blue Shorts|  Shorts|   118.88|      3|         562|         4|
4|2022-12-23 02:51:50|
|       14|  Red t-shirt|T-Shirts|   121.58|      3|         581|        35|
56|2022-12-23 17:05:54|
|       15|White t-shirt|T-Shirts|   131.13|      3|         200|        34|
24|2022-12-23 07:15:01|
|       24|   Blue Jeans|   Pants|    173.1|      3|         506|        41|
19|2022-12-23 21:26:29|
|        1|   Red Shorts|  Shorts|    89.75|      3|         278|         5|
5|2022-12-23 16:41:42|
|       23| Green Chinos|   Pants|   150.93|      3|         849|        36|
13|2022-12-23 13:22:55|
|        7|White Sandals|   Shoes|   160.96|      3|         992|        34|
3|2022-12-23 16:47:14|
|        7|White Sandals|   Shoes|   160.96|      3|         703|        19|
13|2022-12-23 22:36:48|
|       18|Black t-shirt|T-Shirts|   102.41|      3|         719|        48|
12|2022-12-23 10:11:29|
|       14|  Red t-shirt|T-Shirts|   121.58|      3|         526|        13|
3|2022-12-23 11:57:23|
|        1|   Red Shorts|  Shorts|    89.75|      3|         997|        20|
```

```
14|2022-12-23 04:02:30|
|       15|White t-shirt|T-Shirts|   131.13|      3|          281|        11|
25|2022-12-23 16:07:45|
|       23| Green Chinos|   Pants|   150.93|      3|          691|        48|
2|2022-12-23 08:12:00|
|        5| Black Shorts|  Shorts|    74.58|      3|          762|        17|
26|2022-12-23 16:18:27|
|       23| Green Chinos|   Pants|   150.93|      3|          106|        24|
11|2022-12-23 07:41:50|
|        9|Green Sandals|   Shoes|   137.53|      3|           21|        32|
2|2022-12-23 21:15:10|
|       18|Black t-shirt|T-Shirts|   102.41|      3|          626|        14|
14|2022-12-23 12:55:02|
|       15|White t-shirt|T-Shirts|   131.13|      3|          219|        11|
5|2022-12-23 13:00:17|
+---------+-------------+--------+---------+-------+-------------+----------+---
-----+-------------------+
only showing top 20 rows
```

[10]:
```
df3 = df3.join(df1,["CustomerId"])
```

[11]:
```
df3.show()
```

```
+----------+---------+-------------+--------+---------+-------+-------------+---
-----+------------------+--------------------+--------------------+
|CustomerId|ProductId|
Product_Name|Category|UnitPrice|StoreId|TransactionId|Quantity|
TransactionTime|                Name|               Email|
+----------+---------+-------------+--------+---------+-------+-------------+---
-----+------------------+--------------------+--------------------+
|        35|        3|  Blue Shorts|  Shorts|   118.88|      3|          454|
3|2022-12-23 17:36:11|      Dwayne Johnson|dwayne.johnson@gm…|
|        37|        9|Green Sandals|   Shoes|   137.53|      3|          524|
11|2022-12-23 22:02:51|       Brittany Holt|brittany.holt@exa…|
|         4|        3|  Blue Shorts|  Shorts|   118.88|      3|          562|
4|2022-12-23 02:51:50|       Alevtin Paska|alevtin.paska@exa…|
|        35|       14|  Red t-shirt|T-Shirts|   121.58|      3|          581|
56|2022-12-23 17:05:54|      Dwayne Johnson|dwayne.johnson@gm…|
|        34|       15|White t-shirt|T-Shirts|   131.13|      3|          200|
24|2022-12-23 07:15:01|           Avi Shet|avi.shet@example.com|
|        41|       24|   Blue Jeans|   Pants|    173.1|      3|          506|
19|2022-12-23 21:26:29|         Alice Morin|alice.morin@examp…|
|         5|        1|   Red Shorts|  Shorts|    89.75|      3|          278|
5|2022-12-23 16:41:42|      Charlotte Wong|charlotte.wong@ex…|
|        36|       23| Green Chinos|   Pants|   150.93|      3|          849|
13|2022-12-23 13:22:55|     William Nielsen|william.nielsen@e…|
|        34|        7|White Sandals|   Shoes|   160.96|      3|          992|
```

```
3|2022-12-23 16:47:14|            Avi Shet|avi.shet@example.com|
|        19|        7|White Sandals|   Shoes|   160.96|      3|          703|
13|2022-12-23 22:36:48|       Alexia Renaud|alexia.renaud@exa…|
|        48|       18|Black t-shirt|T-Shirts|   102.41|      3|          719|
12|2022-12-23 10:11:29|        Amoli Shenoy|amoli.shenoy@exam…|
|        13|       14|  Red t-shirt|T-Shirts|   121.58|      3|          526|
3|2022-12-23 11:57:23|      Elizabeth Neal|elizabeth.neal@ex…|
|        20|        1|   Red Shorts|  Shorts|    89.75|      3|          997|
14|2022-12-23 04:02:30|         Suzy Gibson|suzy.gibson@examp…|
|        11|       15|White t-shirt|T-Shirts|   131.13|      3|          281|
25|2022-12-23 16:07:45|    Angélique Vennix|angelique.vennix@…|
|        48|       23| Green Chinos|   Pants|   150.93|      3|          691|
2|2022-12-23 08:12:00|        Amoli Shenoy|amoli.shenoy@exam…|
|        17|        5| Black Shorts|  Shorts|    74.58|      3|          762|
26|2022-12-23 16:18:27|Sevastiana Nester…|sevastiana.nester…|
|        24|       23| Green Chinos|   Pants|   150.93|      3|          106|
11|2022-12-23 07:41:50|         Bernd Colin|bernd.colin@examp…|
|        32|        9|Green Sandals|   Shoes|   137.53|      3|           21|
2|2022-12-23 21:15:10|        Ethan Little|ethan.little@exam…|
|        14|       18|Black t-shirt|T-Shirts|   102.41|      3|          626|
14|2022-12-23 12:55:02|      Sylvie Lecomte|sylvie.lecomte@ex…|
|        11|       15|White t-shirt|T-Shirts|   131.13|      3|          219|
5|2022-12-23 13:00:17|    Angélique Vennix|angelique.vennix@…|
+----------+---------+-------------+--------+---------+-------+-------------+---
-----+------------------+--------------------+--------------------+
only showing top 20 rows
```

###

Question No. 1

[12]:
```
df_filter = df3.filter(df3.StoreId == 1)
```

[13]:
```
df_filter.show()
```

```
+----------+---------+-------------+----------+---------+-------+-------------
+--------+------------------+--------------------+--------------------+
|CustomerId|ProductId|  Product_Name|
Category|UnitPrice|StoreId|TransactionId|Quantity|    TransactionTime|
Name|               Email|
+----------+---------+-------------+----------+---------+-------+-------------
+--------+------------------+--------------------+--------------------+
|        13|        2|  White Shorts|    Shorts|    89.27|      1|          971|
10|2022-12-23 04:13:05|      Elizabeth Neal|elizabeth.neal@ex…|
|         7|       10|Black Sneakers|     Shoes|   146.41|      1|          605|
5|2022-12-23 09:36:22|          Dominic Lo|dominic.lo@exampl…|
|        37|        2|  White Shorts|    Shorts|    89.27|      1|          567|
8|2022-12-23 19:44:43|       Brittany Holt|brittany.holt@exa…|
```

```
|        38|        5|  Black Shorts|    Shorts|    74.58|      1|          607|
4|2022-12-23 04:36:41|  Filomeno Fernandes|filomeno.fernande…|
|        17|        9| Green Sandals|     Shoes|   137.53|      1|          141|
7|2022-12-23 19:11:29|Sevastiana Nester…|sevastiana.nester…|
|        17|       11|         Watch|Accesories|   179.65|      1|          248|
12|2022-12-23 06:27:58|Sevastiana Nester…|sevastiana.nester…|
|        45|        4|  Green Shorts|    Shorts|   121.43|      1|          726|
13|2022-12-23 14:12:34|   Melissa Patterson|melissa.patterson…|
|         4|        9| Green Sandals|     Shoes|   137.53|      1|          725|
1|2022-12-23 12:15:47|       Alevtin Paska|alevtin.paska@exa…|
|        30|       10|Black Sneakers|     Shoes|   146.41|      1|          232|
9|2022-12-23 01:26:10|    Raymonde Riviere|raymonde.riviere@…|
|        47|        6|   Red Sandals|     Shoes|   138.38|      1|          954|
14|2022-12-23 06:45:59|     Flenn Henderson|flenn.henderson@e…|
|         2|        5|  Black Shorts|    Shorts|    74.58|      1|           38|
3|2022-12-23 10:19:48|        Thies Blümel|thies.blumel@exam…|
|         3|        3|   Blue Shorts|    Shorts|   118.88|      1|          701|
11|2022-12-23 13:22:38|           bhrh|    .aalyzdh@exam…|
|        49|        7| White Sandals|     Shoes|   160.96|      1|          783|
8|2022-12-23 18:00:04|   Jonathan Carrasco|jonathan.carrasco…|
|        23|        8| Blue Sneakers|     Shoes|    111.7|      1|          333|
9|2022-12-23 20:18:44|    Ceyhun Hamzaoğlu|ceyhun.hamzaoglu@…|
|         1|       11|         Watch|Accesories|   179.65|      1|          482|
2|2022-12-23 09:05:36|      Emilia Pedraza|emilia.pedraza@ex…|
|        35|        1|    Red Shorts|    Shorts|    89.75|      1|          286|
12|2022-12-23 01:23:31|      Dwayne Johnson|dwayne.johnson@gm…|
|        43|        5|  Black Shorts|    Shorts|    74.58|      1|          734|
1|2022-12-23 23:58:16|  Lucas Christiansen|lucas.christianse…|
|         1|        3|   Blue Shorts|    Shorts|   118.88|      1|           20|
2|2022-12-23 05:18:30|      Emilia Pedraza|emilia.pedraza@ex…|
|        18|        6|   Red Sandals|     Shoes|   138.38|      1|          203|
10|2022-12-23 23:35:44|          Kiara Brun|kiara.brun@exampl…|
|        30|        5|  Black Shorts|    Shorts|    74.58|      1|          924|
4|2022-12-23 11:35:46|    Raymonde Riviere|raymonde.riviere@…|
+----------+---------+--------------+----------+---------+-------+-------------
+--------+------------------+--------------------+--------------------+
only showing top 20 rows
```

[14]:
```
df_filter = df_filter.withColumn(
    "UnitPrice", df_filter["UnitPrice"].cast(IntegerType())
)
```

[15]:
```
df_filter = df_filter.withColumn(
    "Quantity", df_filter["Quantity"].cast(IntegerType())
)
```

[16]:
```
df_multiply = df_filter.withColumn("Quan*UnitPrice", df_filter.
↪Quantity*df_filter.UnitPrice)
```

[17]:
```
df_multiply.show()
```

```
+----------+---------+-------------+----------+---------+-------+-------------
+--------+------------------+-------------------+--------------------+--------
------+
|CustomerId|ProductId|  Product_Name|
Category|UnitPrice|StoreId|TransactionId|Quantity|    TransactionTime|
Name|               Email|Quan*UnitPrice|
+----------+---------+-------------+----------+---------+-------+-------------
+--------+------------------+-------------------+--------------------+--------
------+
|        13|        2|  White Shorts|    Shorts|       89|      1|          971|
10|2022-12-23 04:13:05|      Elizabeth Neal|elizabeth.neal@ex…|           890|
|         7|       10|Black Sneakers|     Shoes|      146|      1|          605|
5|2022-12-23 09:36:22|          Dominic Lo|dominic.lo@exampl…|           730|
|        37|        2|  White Shorts|    Shorts|       89|      1|          567|
8|2022-12-23 19:44:43|       Brittany Holt|brittany.holt@exa…|           712|
|        38|        5|  Black Shorts|    Shorts|       74|      1|          607|
4|2022-12-23 04:36:41|  Filomeno Fernandes|filomeno.fernande…|           296|
|        17|        9| Green Sandals|     Shoes|      137|      1|          141|
7|2022-12-23 19:11:29|Sevastiana Nester…|sevastiana.nester…|           959|
|        17|       11|         Watch|Accesories|      179|      1|          248|
12|2022-12-23 06:27:58|Sevastiana Nester…|sevastiana.nester…|          2148|
|        45|        4|  Green Shorts|    Shorts|      121|      1|          726|
13|2022-12-23 14:12:34|   Melissa Patterson|melissa.patterson…|          1573|
|         4|        9| Green Sandals|     Shoes|      137|      1|          725|
1|2022-12-23 12:15:47|       Alevtin Paska|alevtin.paska@exa…|           137|
|        30|       10|Black Sneakers|     Shoes|      146|      1|          232|
9|2022-12-23 01:26:10|    Raymonde Riviere|raymonde.riviere@…|          1314|
|        47|        6|   Red Sandals|     Shoes|      138|      1|          954|
14|2022-12-23 06:45:59|     Flenn Henderson|flenn.henderson@e…|          1932|
|         2|        5|  Black Shorts|    Shorts|       74|      1|           38|
3|2022-12-23 10:19:48|        Thies Blümel|thies.blumel@exam…|           222|
|         3|        3|   Blue Shorts|    Shorts|      118|      1|          701|
11|2022-12-23 13:22:38|           bhrh|    .aalyzdh@exam…|          1298|
|        49|        7| White Sandals|     Shoes|      160|      1|          783|
8|2022-12-23 18:00:04|   Jonathan Carrasco|jonathan.carrasco…|          1280|
|        23|        8| Blue Sneakers|     Shoes|      111|      1|          333|
9|2022-12-23 20:18:44|    Ceyhun Hamzaoğlu|ceyhun.hamzaoglu@…|           999|
|         1|       11|         Watch|Accesories|      179|      1|          482|
2|2022-12-23 09:05:36|      Emilia Pedraza|emilia.pedraza@ex…|           358|
|        35|        1|    Red Shorts|    Shorts|       89|      1|          286|
12|2022-12-23 01:23:31|      Dwayne Johnson|dwayne.johnson@gm…|          1068|
|        43|        5|  Black Shorts|    Shorts|       74|      1|          734|
```

```
1|2022-12-23 23:58:16|  Lucas Christiansen|lucas.christianse…|            74|
|         1|        3|   Blue Shorts|    Shorts|      118|      1|           20|
2|2022-12-23 05:18:30|     Emilia Pedraza|emilia.pedraza@ex…|           236|
|        18|        6|   Red Sandals|     Shoes|      138|      1|          203|
10|2022-12-23 23:35:44|          Kiara Brun|kiara.brun@exampl…|          1380|
|        30|        5|  Black Shorts|    Shorts|       74|      1|          924|
4|2022-12-23 11:35:46|    Raymonde Riviere|raymonde.riviere@…|           296|
+----------+---------+--------------+----------+---------+-------+-------------
+--------+------------------+--------------------+--------------------+--------
------+
only showing top 20 rows
```

[18]:
```
df_group = df_multiply.
  ↪withColumn("TransactionTime",to_date(current_timestamp())).
  ↪groupBy("TransactionTime").sum("Quan*UnitPrice")
```

###

Daily total sales for the store with id 1

[19]:
```
df_group.show()
```

```
+---------------+-------------------+
|TransactionTime|sum(Quan*UnitPrice)|
+---------------+-------------------+
|     2023-01-03|              41070|
+---------------+-------------------+
```

###

Question No. 2

[20]:
```
df_filter2 = df3.filter(df3.StoreId == 2)
```

[21]:
```
df_filter2.show()
```

```
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+
|CustomerId|ProductId| Product_Name|
Category|UnitPrice|StoreId|TransactionId|Quantity|    TransactionTime|
Name|               Email|
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 18:49:45|        Thies Blümel|thies.blumel@exam…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 13:19:51|        Thies Blümel|thies.blumel@exam…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
```

```
2|2022-12-23 22:39:21|        Thies Blümel|thies.blumel@exam…|
|        14|       21|   Red Chinos|     Pants|   134.42|      2|          514|
5|2022-12-23 00:24:15|      Sylvie Lecomte|sylvie.lecomte@ex…|
|        44|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          363|
2|2022-12-23 10:46:04|        Dobrik Svida|dobrik.svida@exam…|
|        47|       19| Green jacket|   Jackets|   223.69|      2|          773|
6|2022-12-23 22:18:53|     Flenn Henderson|flenn.henderson@e…|
|        39|       15|White t-shirt|  T-Shirts|   131.13|      2|          822|
6|2022-12-23 02:39:02|    Gládis das Neves|gladis.dasneves@e…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 15:34:54|        Thies Blümel|thies.blumel@exam…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 08:35:19|        Thies Blümel|thies.blumel@exam…|
|        42|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          227|
5|2022-12-23 19:58:57|     Sofia Jørgensen|sofia.jorgensen@e…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 02:23:41|        Thies Blümel|thies.blumel@exam…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 21:49:12|        Thies Blümel|thies.blumel@exam…|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 18:30:51|        Thies Blümel|thies.blumel@exam…|
|        42|       14|  Red t-shirt|  T-Shirts|   121.58|      2|          372|
3|2022-12-23 19:25:27|     Sofia Jørgensen|sofia.jorgensen@e…|
|        17|       25|  Black Jeans|     Pants|   129.72|      2|          713|
3|2022-12-23 19:27:14|Sevastiana Nester…|sevastiana.nester…|
|         1|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          846|
4|2022-12-23 05:09:24|      Emilia Pedraza|emilia.pedraza@ex…|
|        17|       13|     Earrings|Accesories|    185.9|      2|          969|
4|2022-12-23 03:22:26|Sevastiana Nester…|sevastiana.nester…|
|        28|       17|Green t-shirt|  T-Shirts|   130.13|      2|          694|
5|2022-12-23 21:56:11|         Efe Tazegül|efe.tazegul@examp…|
|        28|       18|Black t-shirt|  T-Shirts|   102.41|      2|          269|
1|2022-12-23 21:34:17|         Efe Tazegül|efe.tazegul@examp…|
|        12|       19| Green jacket|   Jackets|   223.69|      2|           67|
2|2022-12-23 16:38:56|           Eric King|eric.king@example…|
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+
only showing top 20 rows
```

[22]:
```
df_mean = df_filter2.withColumn("Quan*UnitPrice", df_filter2.
↪Quantity*df_filter2.UnitPrice)
```

[23]:
```
df_mean.show()
```

```
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+----------
--------+
```

```
|CustomerId|ProductId| Product_Name|
Category|UnitPrice|StoreId|TransactionId|Quantity|    TransactionTime|
Name|               Email|    Quan*UnitPrice|
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+----------
--------+
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 18:49:45|        Thies Blümel|thies.blumel@exam…|
178.54|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 13:19:51|        Thies Blümel|thies.blumel@exam…|
178.54|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 22:39:21|        Thies Blümel|thies.blumel@exam…|
178.54|
|        14|       21|   Red Chinos|     Pants|   134.42|      2|          514|
5|2022-12-23 00:24:15|      Sylvie Lecomte|sylvie.lecomte@ex…|
672.0999999999999|
|        44|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          363|
2|2022-12-23 10:46:04|        Dobrik Svida|dobrik.svida@exam…|
281.36|
|        47|       19| Green jacket|   Jackets|   223.69|      2|          773|
6|2022-12-23 22:18:53|     Flenn
Henderson|flenn.henderson@e…|1342.1399999999999|
|        39|       15|White t-shirt|  T-Shirts|   131.13|      2|          822|
6|2022-12-23 02:39:02|    Gládis das Neves|gladis.dasneves@e…|
786.78|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 15:34:54|        Thies Blümel|thies.blumel@exam…|
178.54|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 08:35:19|        Thies Blümel|thies.blumel@exam…|
178.54|
|        42|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          227|
5|2022-12-23 19:58:57|     Sofia Jørgensen|sofia.jorgensen@e…|
703.4000000000001|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 02:23:41|        Thies Blümel|thies.blumel@exam…|
178.54|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 21:49:12|        Thies Blümel|thies.blumel@exam…|
178.54|
|         2|        2| White Shorts|    Shorts|    89.27|      2|            2|
2|2022-12-23 18:30:51|        Thies Blümel|thies.blumel@exam…|
178.54|
|        42|       14|  Red t-shirt|  T-Shirts|   121.58|      2|          372|
3|2022-12-23 19:25:27|     Sofia Jørgensen|sofia.jorgensen@e…|
364.74|
```

```
|        17|       25|  Black Jeans|     Pants|   129.72|      2|          713|
3|2022-12-23 19:27:14|Sevastiana
Nester…|sevastiana.nester…|389.15999999999997|
|         1|       16| Blue t-shirt|  T-Shirts|   140.68|      2|          846|
4|2022-12-23 05:09:24|      Emilia Pedraza|emilia.pedraza@ex…|
562.72|
|        17|       13|     Earrings|Accesories|    185.9|      2|          969|
4|2022-12-23 03:22:26|Sevastiana Nester…|sevastiana.nester…|
743.6|
|        28|       17|Green t-shirt|  T-Shirts|   130.13|      2|          694|
5|2022-12-23 21:56:11|         Efe Tazegül|efe.tazegul@examp…|
650.65|
|        28|       18|Black t-shirt|  T-Shirts|   102.41|      2|          269|
1|2022-12-23 21:34:17|         Efe Tazegül|efe.tazegul@examp…|
102.41|
|        12|       19| Green jacket|   Jackets|   223.69|      2|           67|
2|2022-12-23 16:38:56|           Eric King|eric.king@example…|
447.38|
+----------+---------+-------------+----------+---------+-------+-------------+-
-------+------------------+--------------------+--------------------+----------
--------+
only showing top 20 rows
```

###

Mean sales for the store with id 2

[24]:
```
df_mean.select(mean ("Quan*UnitPrice")).show()
```

```
+------------------+
|avg(Quan*UnitPrice)|
+------------------+
|  513.4598039215689|
+------------------+
```

###

Question No. 3

[25]:
```
df3 = df3.withColumn(
    "Quantity", df3["Quantity"].cast(IntegerType())
)
```

###

Email of the client who spent the most when summing up purchases from all of the stores

[26]:
```
df3.groupBy("CustomerId","Name","Email").sum("Quantity").
  ↪orderBy('sum(Quantity)', ascending=False).show(10)
```

```
+----------+--------------------+--------------------+-------------+
|CustomerId|                Name|               Email|sum(Quantity)|
+----------+--------------------+--------------------+-------------+
|        35|      Dwayne Johnson|dwayne.johnson@gm…|           93|
|        17|Sevastiana Nester…|sevastiana.nester…|           65|
|         2|        Thies Blümel|thies.blumel@exam…|           54|
|        11|    Angélique Vennix|angelique.vennix@…|           40|
|        34|            Avi Shet|avi.shet@example.com|           38|
|        39|    Gládis das Neves|gladis.dasneves@e…|           37|
|        44|        Dobrik Svida|dobrik.svida@exam…|           34|
|         7|          Dominic Lo|dominic.lo@exampl…|           34|
|        20|         Suzy Gibson|suzy.gibson@examp…|           33|
|        41|         Alice Morin|alice.morin@examp…|           32|
+----------+--------------------+--------------------+-------------+
only showing top 10 rows
```

###

Question No. 4

###

5 products are most frequently bought across all stores

[27]:
```
df3.groupBy("ProductId","Product_Name").sum("Quantity").
  ↪orderBy('sum(Quantity)', ascending=False).show(5)
```

```
+---------+------------+-------------+
|ProductId| Product_Name|sum(Quantity)|
+---------+------------+-------------+
|       14|  Red t-shirt|           82|
|       24|   Blue Jeans|           77|
|       15|White t-shirt|           76|
|        5| Black Shorts|           75|
|       19| Green jacket|           74|
+---------+------------+-------------+
only showing top 5 rows
```

###

Work and Pairing together Ovais Saleem (2211-023-KHI-DEG) and Muhammad Humza (2211-018-KHI-DEG)

[ ]:
```

```