DSC 650

Inman, Gracie

Final Project

06/01/24

Interpreting Sales Data

Interpreting sales data can provide key insights into a business. Having these key insights can help influence business decisions and can help to ensure the longevity of a company. The data that comes from a company can show a lot about the current customer base, the success of current methods, and more. Even a small amount of data can have key insights into the business and the customers.

For the final project, I decided to generate my own data. I did this because I felt as though sales data can often be random depending on industry, company, as well as products. In addition, I felt as though generating data can be a useful skill to practice and showcase. In addition, I felt a lot of the datasets provided an abundance of information which is not always the case. I wanted to show the insights I could make based on minimum data. To do this, I generated the table gensales_data which included basic information regarding sales data for a company. The table included information such as the total sale price of the transaction (between 1 and 1500), the total number of items that were in the transaction (between 1 and 25), and either a 1 or a 2 for method of payment (1 being card and 2 being cash). The data was generated using Spark and uploaded into a hive table. This was done to showcase how the two tools can be used in conjunction.

Using Hive, the data was visualized to confirm the data was uploaded correctly. I checked to make sure there were no missing values and after seeing that the data was uploaded, I
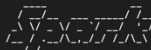
checked the average total sales, and number of products, and the percentage of card and cash transactions. The average total selling price was $787.95 and the average number of items sold was 12.488 which we will say is 13 since you can not have 0.488 of an item. This tells me that it is likely that a significant amount of the transactions contains higher-value items, and it is something that should be explored.

To confirm this, I added a new column and divided the total selling price and the number of items to get an average price per item. Originally, I made a mistake and divided the number of items by the total instead of the total by the number of items. To correct this, I added the columns I wanted to keep to a new table and deleted the original. I renamed the new table back to gensales_data and reperformed the calculation. I created a new column with the correct calculation dividing the total by the number of items. This showed the average spent on each item. I then performed a calculation to calculate the average, average price per item. The average of the average price per item is $122.50. This shows that again it is likely that a significant number of the transactions do contain high-value items. I then calculated the number of transactions where the average price per item was above $100. The number of transactions with a price per item of above $100 is 146. There are 500 transactions in the table which means that 29.2% of the data contains a price per item above $100.

For the type of payment in the transactions, it appears all of the transactions were paid with a card. This is not unusual with online relators. However, if this is not an online realtor, I would recommend that the business has a card-forward POS system that is highly accessible to customers. This will improve customer experience since it appears that the current customer base prefers to pay with a card. However, it is important not to alienate cash payments in case new

customers would like to pay with cash. However, while it should be easy and fast to make cash payments it should be just as easy to make card payments.

The screenshots below showcase the code used to perform the analysis with breaks between lengthy sections of output.

```
    ____              __
   / __/__  ___ _____/ /__
  _\ \/ _ \/ _ `/ __/  '_/
 /__ / .__/\_,_/_/ /_/\_\   version 3.0.0
    /_/

Using Python version 3.7.10 (default, Mar  2 2021 09:06:08)
SparkSession available as 'spark'.
>>> val spark = SparkSession.builder
  File "<stdin>", line 1
    val spark = SparkSession.builder
              ^
SyntaxError: invalid syntax
>>>    .appName("GenerateSalesData")
  File "<stdin>", line 1
    .appName("GenerateSalesData")
    ^
IndentationError: unexpected indent
>>>    .enableHiveSupport()
  File "<stdin>", line 1
    .enableHiveSupport()
    ^
IndentationError: unexpected indent
>>> spark = SparkSession.builder \
...     .appName("GenerateSalesData") \
...     .enableHiveSupport() \
...     .getOrCreate()
>>> sales_data = spark.range(500).selectExpr(
...     f"CAST((RAND() * 1490 + 10) AS FLOAT) AS amount_spent",
...     f"CAST((RAND() * 24 + 1) AS INT) AS number_of_items",
...     f"CAST((RAND() * 1 + 1) AS INT) AS payment_method"
... )
>>> sales_data.write.mode("overwrite").saveAsTable("gensales_data")
208332 [Thread-4] WARN  org.apache.hadoop.hive.conf.HiveConf  - HiveConf of name hive.strict.managed.tables does not exist
208332 [Thread-4] WARN  org.apache.hadoop.hive.conf.HiveConf  - HiveConf of name hive.create.as.insert.only does not exist
208360 [Thread-4] WARN  org.apache.spark.sql.hive.client.HiveClientImpl  - Detected HiveConf hive.execution.engine is 'tez' and will be reset to 'mr' to disable useless hive logic
220305 [Thread-4] WARN  org.apache.hadoop.hive.ql.session.SessionState  - METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory.
>>> spark.stop()
>>> hive -e "SELECT * FROM gensales_data;"
  File "<stdin>", line 1
    hive -e "SELECT * FROM gensales_data;"
                 ^
SyntaxError: invalid syntax
>>> exit
Use exit() or Ctrl-D (i.e. EOF) to exit
>>> exit()
bash-5.0# huve
bash: huve: command not found
bash-5.0# hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9

Logging initialized using configuration in file:/usr/program/hive/conf/hive-log4j2.properties Async: true
Hive Session ID = a97d2200-7e1b-4d30-9489-c79af6ac915a
2024-06-01 16:49:48,368 INFO  [Tez session start thread] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032
hive> 2024-06-01 16:49:49,387 INFO  [pool-7-thread-1] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032

    > SHOW TABLES LIKE 'gensales_data';
OK
gensales_data
```

```
bash-5.0# hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/program/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/program/tez/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9

Logging initialized using configuration in file:/usr/program/hive/conf/hive-log4j2.properties Async: true
Hive Session ID = a97d2200-7e1b-4d30-9489-c79af6ac915a
2024-06-01 16:49:48,368 INFO  [Tez session start thread] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032
hive> 2024-06-01 16:49:49,387 INFO  [pool-7-thread-1] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032

    > SHOW TABLES LIKE 'gensales_data';
OK
gensales_data
Time taken: 1.315 seconds, Fetched: 1 row(s)
hive> "SELECT * FROM gensales_data;"
    > SELECT * FROM gensales_data;
NoViableAltException(379@[])
        at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1387)
        at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:220)
        at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:74)
        at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:67)
        at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:616)
        at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1826)
        at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1773)
        at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1768)
        at org.apache.hadoop.hive.ql.reexec.ReExecDriver.compileAndRespond(ReExecDriver.java:126)
        at org.apache.hadoop.hive.ql.reexec.ReExecDriver.run(ReExecDriver.java:214)
        at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:239)
        at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:188)
        at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:402)
        at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
        at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
        at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:683)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
FAILED: ParseException line 1:0 cannot recognize input near '"SELECT * FROM gensales_data;"' 'SELECT' '*'
hive> SELECT * FROM gensales_data;
OK
298.76892       5       1
47.37617        20      1
1442.7307       6       1
449.16257       6       1
871.92334       10      1
1246.7609       4       1
1090.3108       3       1
649.4959        3       1
142.94347       7       1
838.36456       22      1
1328.1687       19      1
210.18256       10      1
817.43195       15      1
213.78856       19      1
1493.339        24      1
668.3966        21      1
1017.68256      16      1
1372.7251       1       1
426.18167       16      1
```

```
1248.2864       5       1
620.4092        15      1
1400.1814       12      1
1345.389        20      1
472.94693       9       1
Time taken: 2.805 seconds, Fetched: 500 row(s)
hive> SELECT
    > SUM(CASE WHEN amount_spent IS NULL THEN 1 ELSE 0 END) AS null_amount_spent,
    > SUM(CASE WHEN number_of_items IS NULL THEN 1 ELSE 0 END) AS null_number_of_items,
    > SUM(CASE WHEN payment_method IS NULL THEN 1 ELSE 0 END) AS null_payment_method
    > FROM
    > gensales_data;
2024-06-01 16:52:18,375 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorReduceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.
ql.plan.VectorReduceSinkInfo@758ebec4
Query ID = root_20240601165217_65293c30-c00e-41c7-8adc-5af038397966
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1717258132785_0016)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container      SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container      SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 9.29 s
----------------------------------------------------------------------------------------------
OK
0       0       0
Time taken: 13.726 seconds, Fetched: 1 row(s)
hive> SELECT
    > AVG(amount_spent) AS avg_amount_spent,
    > AVG(number_of_items) AS avg_number_of_items,
    > (COUNT(CASE WHEN payment_method = 1 THEN 1 END) / COUNT(*)) * 100 AS percentage_payment_method_1,
    > (COUNT(CASE WHEN payment_method = 2 THEN 1 END) / COUNT(*)) * 100 AS percentage_payment_method_2,
    > number_of_items / amount_spent AS items_per_amount
    > FROM
    > gensales_data;
FAILED: SemanticException [Error 10025]: Line 6:2 Expression not in GROUP BY key 'amount_spent'
hive> SELECT
    > AVG(amount_spent),
    > AVG(number_of_items),
    > (COUNT(CASE WHEN payment_method = 1 THEN 1 END) / COUNT(*)) * 100 AS percentage_payment_method_1,
    > (COUNT(CASE WHEN payment_method = 2 THEN 1 END) / COUNT(*)) * 100 AS percentage_payment_method_2,
    > AVG(number_of_items / amount_spent)
    > FROM
    > gensales_data;
2024-06-01 16:55:56,735 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorReduceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.
ql.plan.VectorReduceSinkInfo@24e31f06
Query ID = root_20240601165556_e9b80fd1-42eb-4577-ab51-9edc4ee406d8
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1717258132785_0016)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container      SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container      SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 9.39 s
----------------------------------------------------------------------------------------------
OK
787.9480350189209       12.488  100.0   0.0     0.04505173621108595
Time taken: 10.583 seconds, Fetched: 1 row(s)
```

```
hive> SELECT
    >    amount_spent,
    >    number_of_items,
    >    payment_method,
    >    number_of_items / amount_spent AS items_per_amount
    > FROM
    >    gensales_data;
OK
298.76892      5      1      0.01673534176501472
47.37617       20     1      0.42215152746204130
1442.7307      6      1      0.00415878025357797710
449.16257      6      1      0.013358192420668916
871.92334      10     1      0.0114688981737683680
1246.7609      4      1      0.003208313730942437
1090.3108      3      1      0.0027515090419361057
649.4959       3      1      0.004618966726092135
142.94347      7      1      0.04897040897879076
838.36456      22     1      0.02624156717882113
1328.1687      19     1      0.01430541164178605210
210.18256      10     1      0.04757768762100237
817.43195      15     1      0.018350151443256284
213.78856      19     1      0.08887285686582688
1493.339       24     1      0.01607136770193616
668.3966       21     1      0.03141847190350479
1017.68256     16     1      0.01572199494161797410
1372.7251      1      1      7.28477975457264E-4
426.18167      16     1      0.03754267506883757
1438.9019      12     1      0.00833969318643401810
527.0639       16     1      0.03035685024981428
1070.8181      10     1      0.009338654116634227
1112.9349      12     1      0.01078230146812117
696.6449       2      1      0.0028709031061440377
1421.6904      17     1      0.011957596144004956
164.21014      23     1      0.14006442862617310
1259.8485      1      1      7.937462254179207E-4
834.54504      13     1      0.015577349712056867
582.46014      1      1      0.0017168556685420674
917.66504      2      1      0.0021794444757786886
215.92921      24     1      0.11114753535351965
517.02917      8      1      0.01547301648387647110
1335.2513      2      1      0.0014978453388875166
407.26382      3      1      0.00736623245130223
1329.8596      9      1      0.00676763161349010110
886.7329       4      1      0.004510941179904066
1139.4746      22     1      0.01930714367744180310
54.417088      7      1      0.12863606478266243
916.2749       14     1      0.0152792573104307910
560.2799       14     1      0.02498751038440999810
1443.821       21     1      0.01454473881916322110
471.61407      4      1      0.00848151107976329610
1100.4382      13     1      0.01181347541096354110
391.12668      24     1      0.0613611940102863210
976.12195      12     1      0.01229354592590582410
1419.3025      23     1      0.0162051431306950780
945.9229       8      1      0.008457348789692296
726.5861       22     1      0.03027858553321561
1495.4683      18     1      0.012036363766966542
374.85995      7      1      0.018673640408189550
1018.75476     5      1      0.0049079525246058760
518.9565       6      1      0.0115616630852059910
1418.077       3      1      0.0021155409362249972
1167.406       4      1      0.0034264000527009773
566.3094       10     1      0.0176581921929968
1481.4717      14     1      0.0094506252523607028
268.8987       17     1      0.06322083086064863
```

```
hive> SELECT
    >    AVG(items_per_amount) AS avg_items_per_amount
    > FROM
    >    (
    >       SELECT
    >          amount_spent,
    >          number_of_items,
    >          payment_method,
    >          number_of_items / amount_spent AS items_per_amount
    >       FROM
    >          gensales_data
    >    ) subquery;
2024-06-01 16:58:56,785 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorReduceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.
ql.plan.VectorReduceSinkInfo@38991781
Query ID = root_20240601165856_9adbe52c-af29-42c9-a399-ef1b263c3eba
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1717258132785_0016)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container      SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container      SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.37 s
----------------------------------------------------------------------------------------------
OK
0.04505173621108595
Time taken: 9.271 seconds, Fetched: 1 row(s)
hive> CREATE TABLE gensales_data_new AS
    > SELECT
    >    amount_spent,
    >    number_of_items,
    >    payment_method
    > FROM
    >    gensales_data;
Query ID = root_20240601165957_30d57394-fdcb-49df-b871-ac0584bf482b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1717258132785_0016)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container      SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 7.79 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://master:9000/usr/hive/warehouse/gensales_data_new
OK
Time taken: 8.899 seconds
hive> DROP TABLE gensales_data;
OK
Time taken: 0.132 seconds
hive> ALTER TABLE gensales_data_new RENAME TO gensales_data;
OK
Time taken: 0.118 seconds
hive> SELECT
    >    amount_spent,
    >    number_of_items,
    >    payment_method,
    >    amount_spent / number_of_items AS amount_per_item
    > FROM
    >    gensales_data;
```

```
hive> SELECT
    >   amount_spent,
    >   number_of_items,
    >   payment_method,
    >   amount_spent / number_of_items AS amount_per_item
    > FROM
    >   gensales_data;
OK
298.76892      5     1     59.7537841796875
47.37617       20    1     2.368808555603027
1442.7307      6     1     240.45511881510416
449.16257      6     1     74.86042785644531
871.92334      10    1     87.192333984375
1246.7609      4     1     311.6902160644531
1090.3108      3     1     363.4369303385417
649.4959       3     1     216.4986368815104
142.94347      7     1     20.420495169503347
838.36456      22    1     38.10748013583097
1328.1687      19    1     69.90361585115132
210.18256      10    1     21.018255615234374
817.43195      15    1     54.495463053385414
213.78856      19    1     11.252029418945312
1493.339       24    1     62.22245788574219
668.3966       21    1     31.828409830729168
1017.68256     16    1     63.605159759521484
1372.7251      1     1     1372.72509765625
426.18167      16    1     26.636354446411133
1438.9019      12    1     119.90848795572917
527.0639       16    1     32.94149398803711
1070.8181      10    1     107.0818115234375
1112.9349      12    1     92.7445780436198
696.6449       2     1     348.32244873046875
1421.6904      17    1     83.62884880514706
164.21014      23    1     7.139571480129076
1259.8485      1     1     1259.8485107421875
834.54504      13    1     64.19577261117789
582.46014      1     1     582.4601440429688
917.66504      2     1     458.83261953125
215.92921      24    1     8.997050603230795
517.0291       8     1     64.6286392211914
1335.2513      2     1     667.6256713867188
407.26382      3     1     135.75460815429688
1329.8596      9     1     147.76217990451389
886.7329       4     1     221.6832275390625
1139.4746      22    1     51.79430042613637
54.417088      7     1     7.77386965070452
916.2749       14    1     65.44820731026786
560.2799       14    1     40.019993373325896
1443.821       21    1     68.75338309151786
471.61407      4     1     117.90351867675781
1100.4382      13    1     84.6490948016827
391.12668      24    1     16.296944936116535
976.12195      12    1     81.34349568684895
1419.3025      23    1     61.708803923233695
945.9229       8     1     118.24036407470703
726.5861       22    1     33.026641845703125
1495.4683      18    1     83.08157009548611
374.85995      7     1     53.551422119140625
1018.75476     5     1     203.7509521484375
518.9565       6     1     86.4927469889323
1418.077       3     1     472.6923421223958
1167.406       4     1     291.85150146484375
566.3094       10    1     56.63093872070313
1481.4717      14    1     105.81940569196429
268.8987       17    1     15.817571303423714
```
```
416.27634      11    1     37.84330333362926
175.71806      11    1     15.974369395862926
70.54707       16    1     4.409192085266113
1490.0887      23    1     64.78646717900816
763.60693      8     1     95.45086669921875
579.4782       16    1     36.21738815307617
1175.5818      22    1     53.43553577769887
44.673042      21    1     2.1272877284458707
1248.2864      5     1     249.65727539625
620.4092       15    1     41.36061197916667
1400.1814      12    1     116.68178304036458
1345.389       20    1     67.26945190429687
472.94693      9     1     52.549658881293404
Time taken: 0.145 seconds, Fetched: 500 row(s)
hive> SELECT
    >   AVG(amount_spent / number_of_items) AS avg_amount_per_item
    > FROM
    >   gensales_data;
2024-06-01 17:03:24,194 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorReduceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.
ql.plan.VectorReduceSinkInfo@f69e35a
Query ID = root_20240601170324_e0c2c7ec-944e-4886-a4ab-f810eb553755
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1717258132785_0016)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 ......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 7.45 s
----------------------------------------------------------------------------------------------
OK
122.50352575013109
Time taken: 8.283 seconds, Fetched: 1 row(s)
hive> SELECT
    >   COUNT(*) AS num_transactions_above_100
    > FROM
    >   gensales_data
    > WHERE
    >   amount_spent / number_of_items > 100;
2024-06-01 17:55:57,116 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] reducesink.VectorReduceSinkEmptyKeyOperator: VectorReduceSinkEmptyKeyOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.
ql.plan.VectorReduceSinkInfo@5c94d4b8
Query ID = root_20240601175556_087400d5-6748-4832-8b6e-f127a2e719b5
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
2024-06-01 17:55:57,289 INFO  [d3e855fb-ef1a-4adc-b5fc-4bae01f1d0b9 main] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1717258132785_0018)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 ......... container     SUCCEEDED      2          2        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 8.30 s
----------------------------------------------------------------------------------------------
OK
146
Time taken: 16.775 seconds, Fetched: 1 row(s)
hive>
```