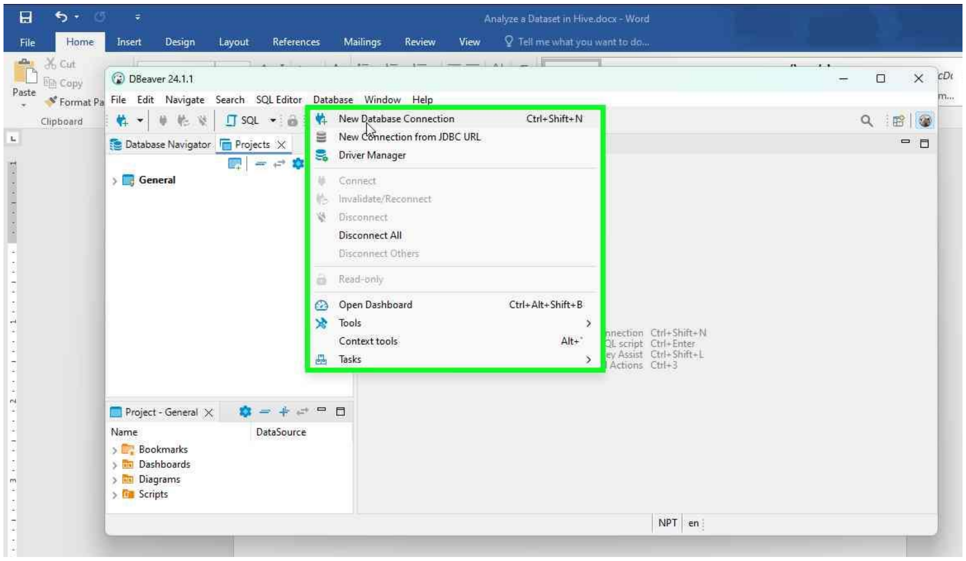
Analyze a Dataset in Hive

***

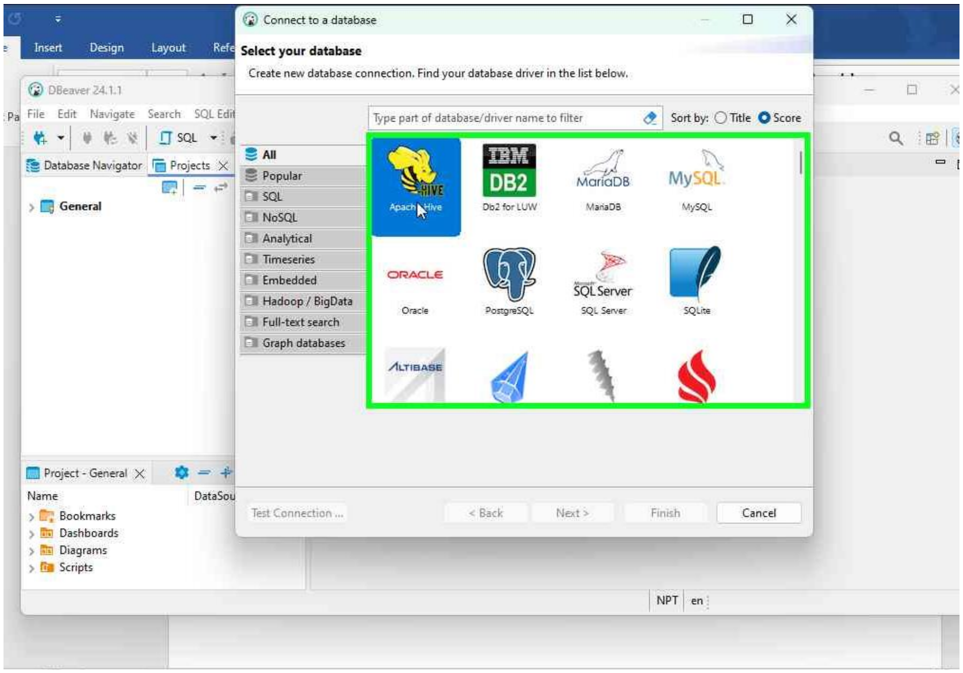
Note: This manual builds up from the previous manual of *Hive Installation*. Please refer to it before proceeding

1. Download DBeaver from site: https://dbeaver.io/download/

2. Open the DBeaver and navigate to Database and click on "New Database Connection"

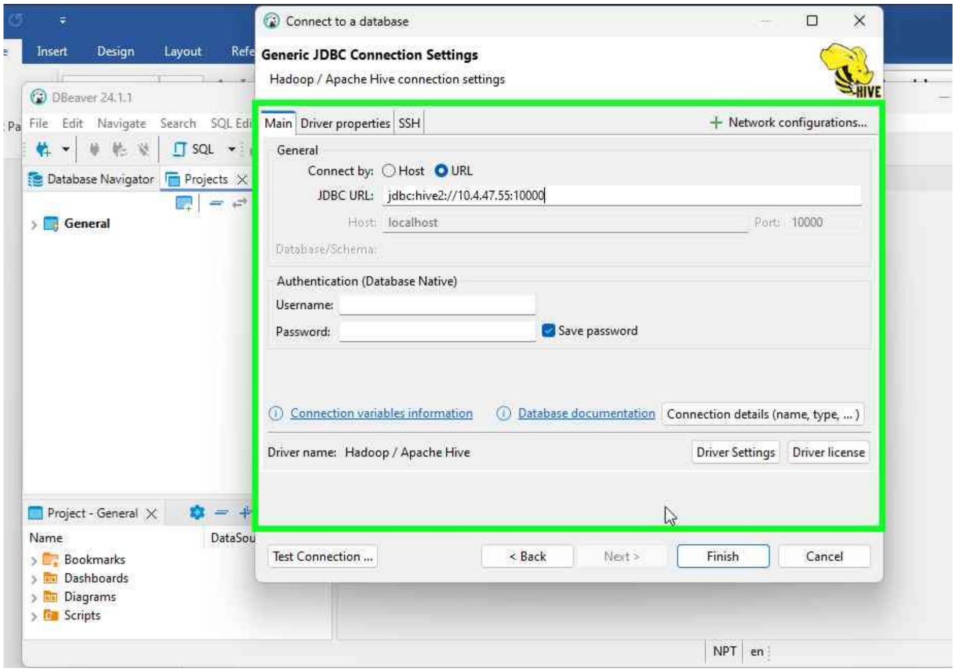


3. Select the Apache Hive as the Database type:



4. Select the "Connect by" option to: URL and enter the URL as shown in the image.

Remember to put the IP of HIVE SERVER instead of the below shown IP 10.4.47.55. The port remains same.



Connect to a database

Generic JDBC Connection Settings

Hadoop / Apache Hive connection settings

General

Connect by: ○ Host ● URL

JDBC URL: jdbc:hive2://10.4.47.55:10000

Host: localhost Port: 10000

Database/Schema:

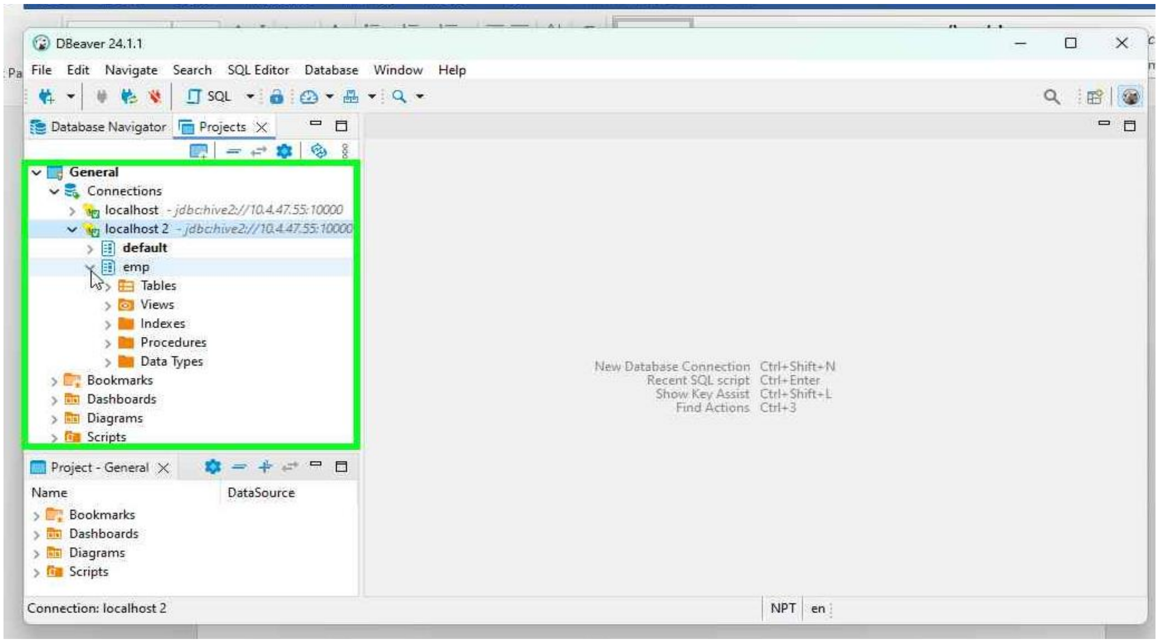
Authentication (Database Native)

Username:

Password: ☑ Save password

Driver name: Hadoop / Apache Hive

5. You should get the Hive database under your connections.

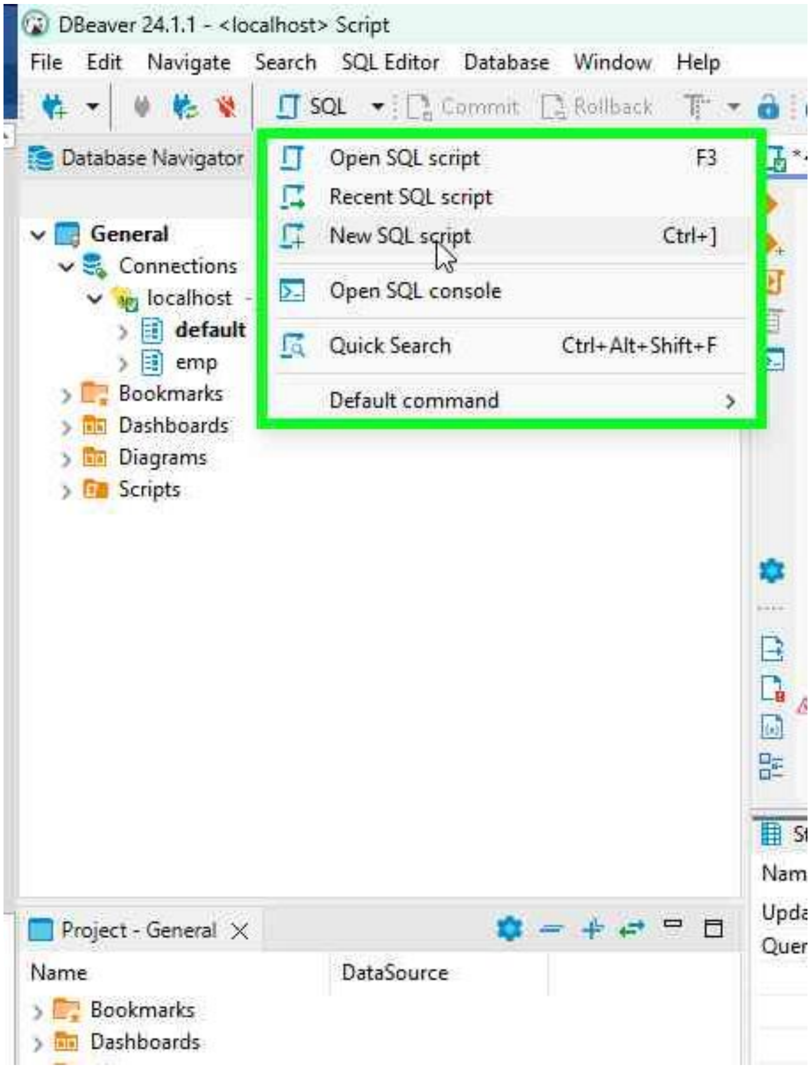


General
- Connections
  - localhost - jdbc:hive2://10.4.47.55:10000
  - localhost 2 - jdbc:hive2://10.4.47.55:10000
    - default
    - emp
      - Tables
      - Views
      - Indexes
      - Procedures
      - Data Types
- Bookmarks
- Dashboards
- Diagrams
- Scripts

Connection: localhost 2

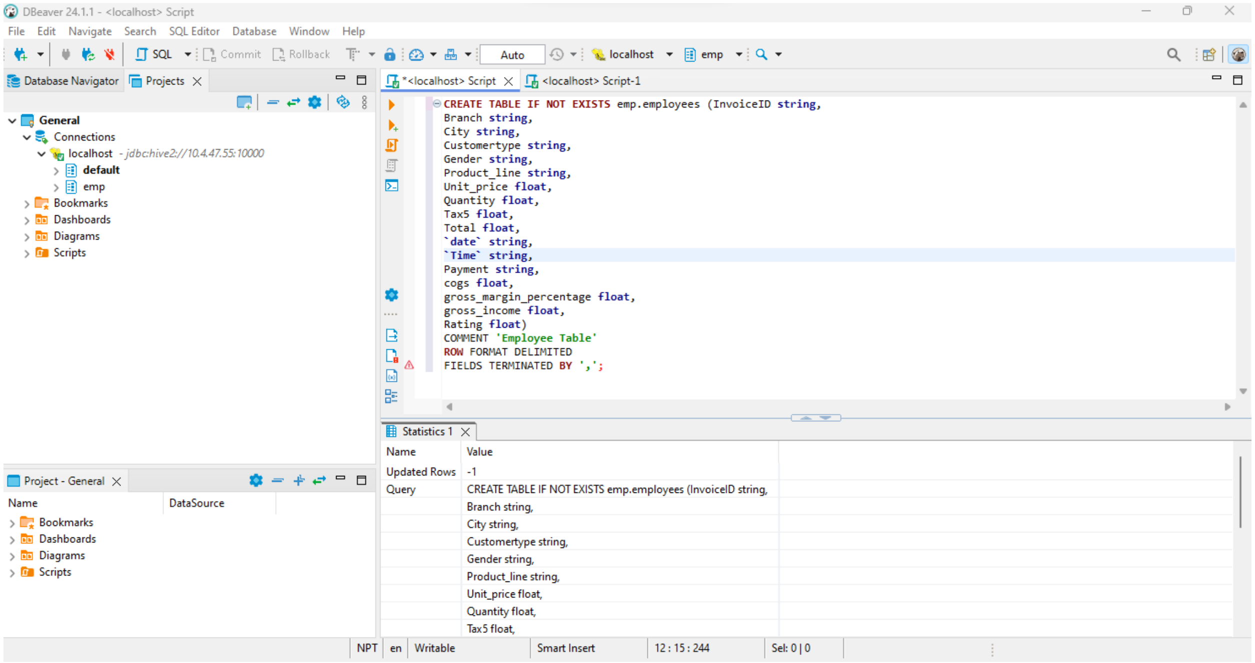
DBeaver 24.1.1
File Edit Navigate Search SQL Editor Database Window Help
SQL
Database Navigator
Projects
New Database Connection Ctrl+Shift+N
Recent SQL script Ctrl+Enter
Show Key Assist Ctrl+Shift+L
Find Actions Ctrl+3
Project - General
Name
DataSource
Bookmarks
Dashboards
Diagrams
Scripts
NPT
en

6. Click on new SQL script to open a SQL window.



7. Create a table using following syntax. hit Ctrl+Enter to execute.

```
CREATE TABLE IF NOT EXISTS emp.employees (InvoiceID string,
Branch string,
City string,
Customertype string,
Gender string,
Product_line string,
Unit_price float,
Quantity float,
Tax5 float,
Total float,
`date` string,
`Time` string,
Payment string,
cogs float,
gross_margin_percentage float,
gross_income float,
Rating float)
COMMENT 'Employee Table'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```



```
CREATE TABLE IF NOT EXISTS emp.employees (InvoiceID string,
Branch string,
City string,
Customertype string,
Gender string,
Product_line string,
Unit_price float,
Quantity float,
Tax5 float,
Total float,
`date` string,
`Time` string,
Payment string,
cogs float,
gross_margin_percentage float,
gross_income float,
Rating float)
COMMENT 'Employee Table'
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```

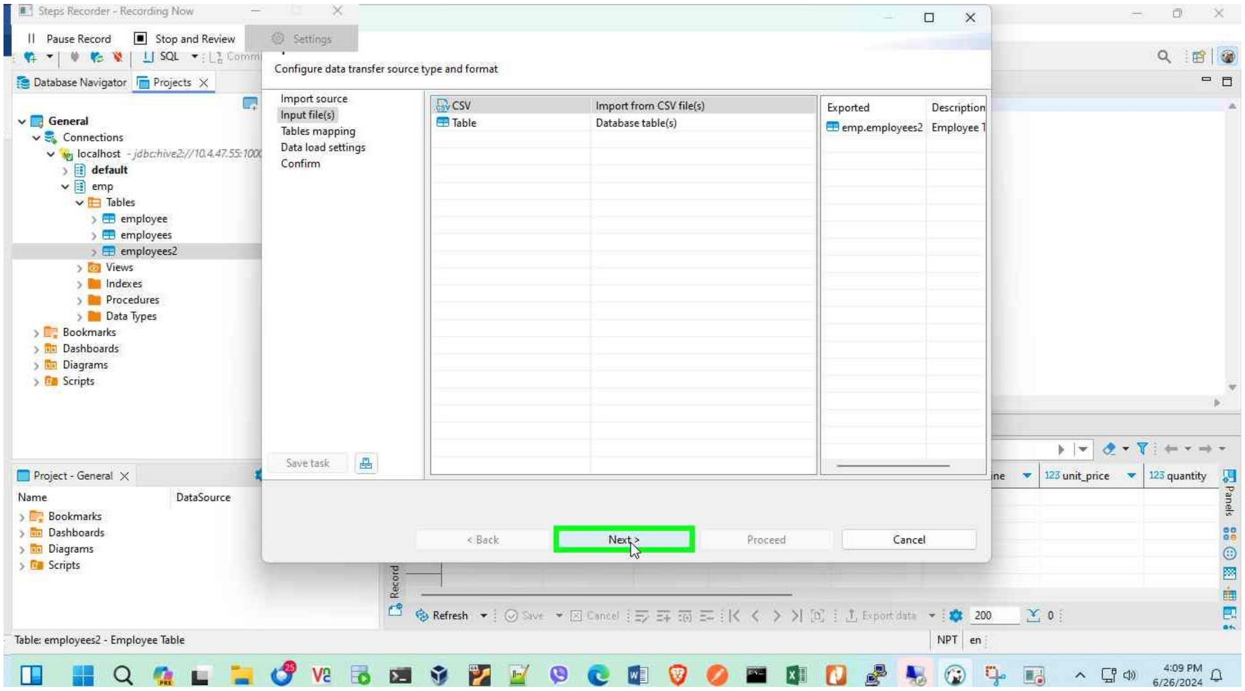
Statistics 1

Name	Value
Updated Rows	-1
Query	CREATE TABLE IF NOT EXISTS emp.employees (InvoiceID string,
	Branch string,
	City string,
	Customertype string,
	Gender string,
	Product_line string,
	Unit_price float,
	Quantity float,
	Tax5 float,

Name	DataSource
Bookmarks	
Dashboards	
Diagrams	
Scripts	

8. Let’s download a CSV file which will be inserted to the table.
https://github.com/sersun/supermarket-sales-analysis/blob/main/supermarket_sales.csv

9. Right click on the database table and select import data.



Configure data transfer source type and format

- Import source
- Input file(s)
- Tables mapping
- Data load settings
- Confirm

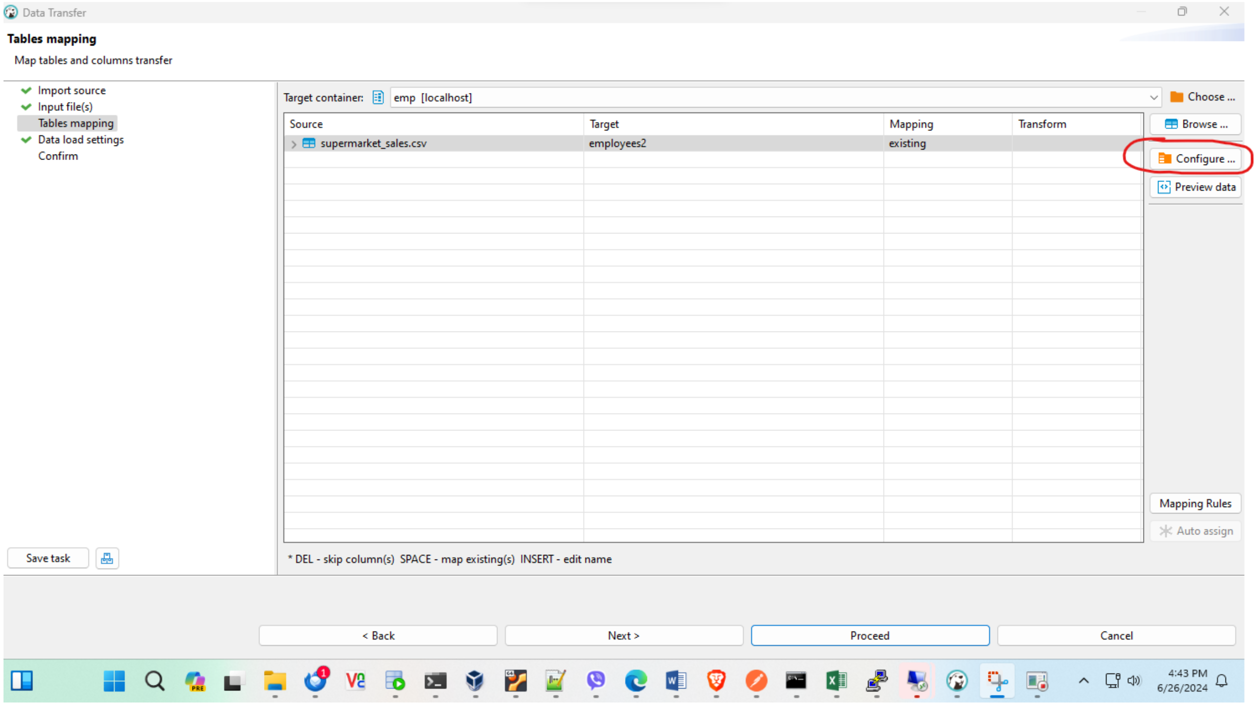
CSV	Import from CSV file(s)
Table	Database table(s)

Exported	Description
emp.employees2	Employee T

Name	DataSource
Bookmarks	
Dashboards	
Diagrams	
Scripts	

Table: employees2 - Employee Table

10. Under the “Table Mapping” option, click on configure.



Tables mapping

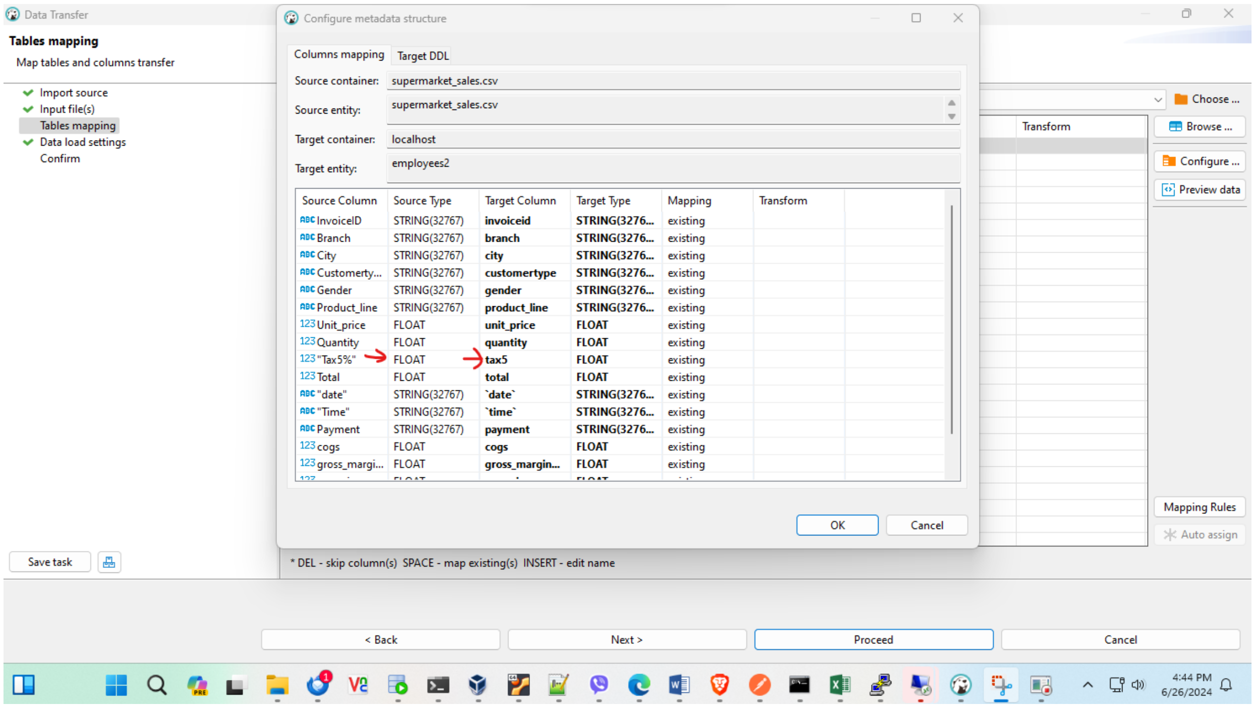
Map tables and columns transfer

Target container: emp [localhost]

Source	Target	Mapping	Transform
supermarket_sales.csv	employees2	existing	

* DEL - skip column(s) SPACE - map existing(s) INSERT - edit name

11. Make sure the special characters such as % are removed in the target Column.



Tables mapping

Map tables and columns transfer

Source container: supermarket_sales.csv

Source entity: supermarket_sales.csv

Target container: localhost

Target entity: employees2

Source Column	Source Type	Target Column	Target Type	Mapping	Transform
InvoiceID	STRING(32767)	**invoiceid**	**STRING(3276...**	existing	
Branch	STRING(32767)	**branch**	**STRING(3276...**	existing	
City	STRING(32767)	**city**	**STRING(3276...**	existing	
Customerty...	STRING(32767)	**customertype**	**STRING(3276...**	existing	
Gender	STRING(32767)	**gender**	**STRING(3276...**	existing	
Product_line	STRING(32767)	**product_line**	**STRING(3276...**	existing	
Unit_price	FLOAT	**unit_price**	**FLOAT**	existing	
Quantity	FLOAT	**quantity**	**FLOAT**	existing	
"Tax5%"	FLOAT	**tax5**	**FLOAT**	existing	
Total	FLOAT	**total**	**FLOAT**	existing	
"date"	STRING(32767)	**`date`**	**STRING(3276...**	existing	
"Time"	STRING(32767)	**`time`**	**STRING(3276...**	existing	
Payment	STRING(32767)	**payment**	**STRING(3276...**	existing	
cogs	FLOAT	**cogs**	**FLOAT**	existing	
gross_margi...	FLOAT	**gross_margin...**	**FLOAT**	existing	

* DEL - skip column(s) SPACE - map existing(s) INSERT - edit name

12. Click OK, Next, Next and finally Proceed. This will take some time to complete as the data is being transferred from host computer to Virtual Box VM to Hive DB.

13. To do this fast, we can download the CSV file in our LINUX machine first. And then load it to HIVE. Remember you should either import from DBeaver or do this step 13. Don't do both.

a. First open the hive shell in CLI.

cd $HIVE_HOME

bin/beeline -u jdbc:hive2://10.4.47.55:10000 hadoop

b. Load data into the HIVE.

LOAD DATA LOCAL INPATH '/home/hadoop/data.csv' INTO TABLE emp.employees;

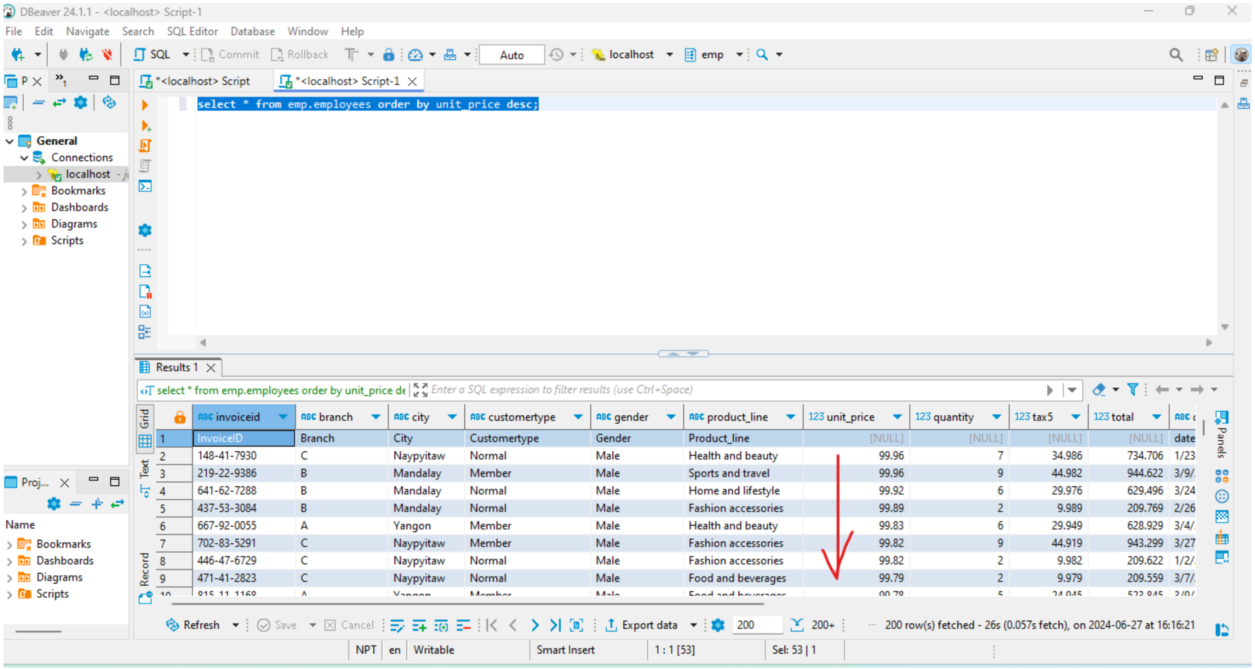
in above command put the appropriate location of your CSV file. in my case, the I am uploading the file data.csv which is in location /home/hadoop.

Now you can continue over to DBeaver.

14. Examples:

a. Sort by unit_price

```
select * from emp.employees order by unit_price desc;
```



DBeaver 24.1.1 - <localhost> Script-1
File Edit Navigate Search SQL Editor Database Window Help
SQL Commit Rollback Auto localhost emp
*<localhost> Script
*<localhost> Script-1
General
Connections
localhost
Bookmarks
Dashboards
Diagrams
Scripts
Proj...
Name
Bookmarks
Dashboards
Diagrams
Scripts
Results 1
Enter a SQL expression to filter results (use Ctrl+Space)
Grid
Text
Record
Panels
Refresh Save Cancel Export data
200
200+
NPT en Writable Smart Insert 1 : 1 [53] Sel: 53 | 1

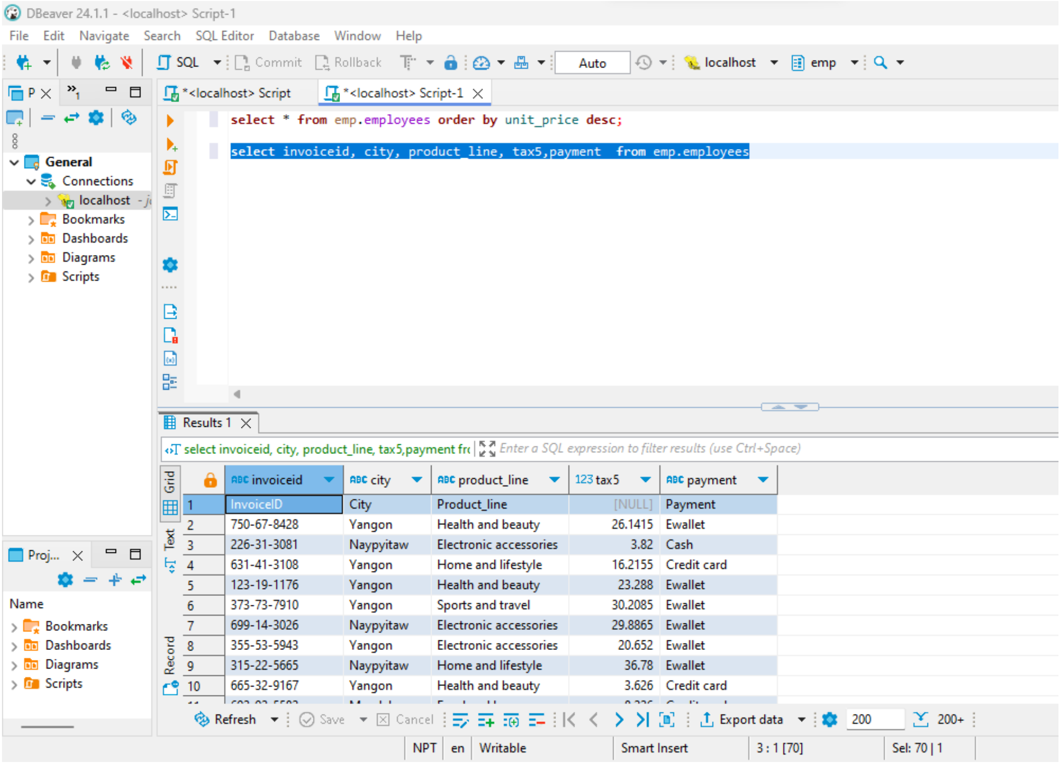
```
select * from emp.employees order by unit_price desc;
```

	invoiceid	branch	city	customertype	gender	product_line	unit_price	quantity	tax5	total	d
1	InvoiceID	Branch	City	Customertype	Gender	Product_line	[NULL]	[NULL]	[NULL]	[NULL]	date
2	148-41-7930	C	Naypyitaw	Normal	Male	Health and beauty	99.96	7	34.986	734.706	1/23
3	219-22-9386	B	Mandalay	Member	Male	Sports and travel	99.96	9	44.982	944.622	3/9/
4	641-62-7288	B	Mandalay	Normal	Male	Home and lifestyle	99.92	6	29.976	629.496	3/24
5	437-53-3084	B	Mandalay	Normal	Male	Fashion accessories	99.89	2	9.989	209.769	2/26
6	667-92-0055	A	Yangon	Member	Male	Health and beauty	99.83	6	29.949	628.929	3/4/
7	702-83-5291	C	Naypyitaw	Member	Male	Fashion accessories	99.82	9	44.919	943.299	3/27
8	446-47-6729	C	Naypyitaw	Normal	Male	Fashion accessories	99.82	2	9.982	209.622	1/2/
9	471-41-2823	C	Naypyitaw	Normal	Male	Food and beverages	99.79	2	9.979	209.559	3/7/
10	815-11-1168	A	Yangon	Member	Male	Food and beverages	99.78	5	24.945	523.845	3/9/

200 row(s) fetched - 26s (0.057s fetch), on 2024-06-27 at 16:16:21

b. Selecting only invoiceid, city, product_line, tax5,payment columns

```
select invoiceid, city, product_line, tax5,payment  from emp.employees;
```



DBeaver 24.1.1 - <localhost> Script-1
File Edit Navigate Search SQL Editor Database Window Help
SQL Commit Rollback Auto localhost emp
*<localhost> Script
*<localhost> Script-1
General
Connections
localhost
Bookmarks
Dashboards
Diagrams
Scripts
Proj...
Name
Bookmarks
Dashboards
Diagrams
Scripts
Results 1
Grid
Text
Record
Refresh Save Cancel Export data 200 200+
NPT en Writable Smart Insert 3 : 1 [70] Sel: 70 | 1

```
select * from emp.employees order by unit_price desc;

select invoiceid, city, product_line, tax5,payment  from emp.employees
```

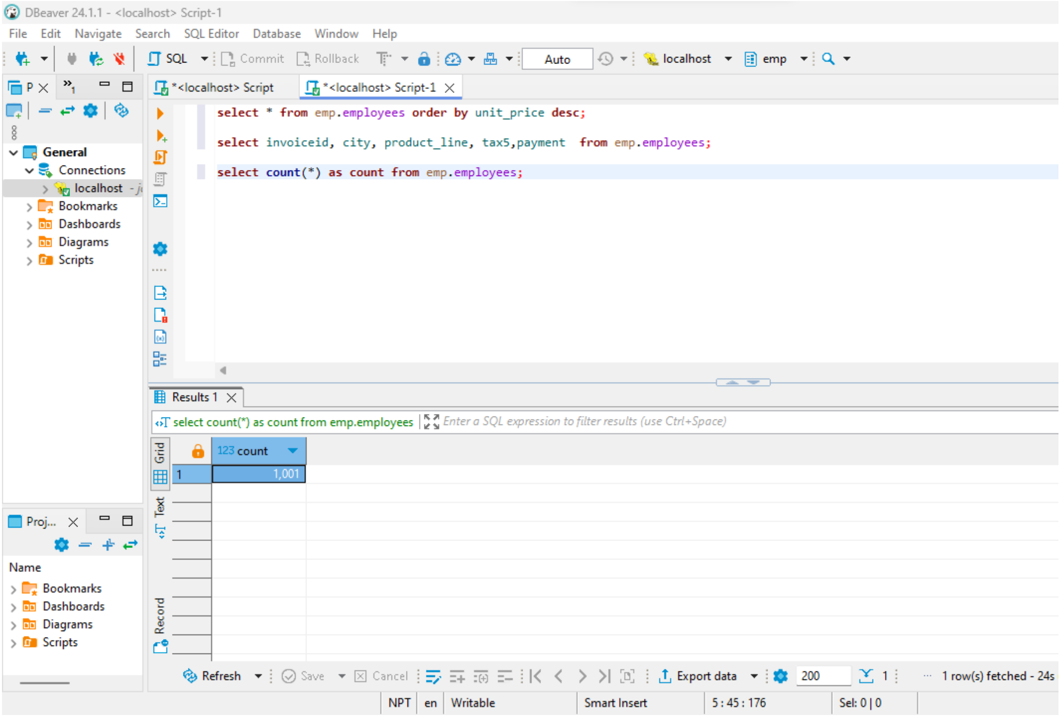
select invoiceid, city, product_line, tax5,payment fro

Enter a SQL expression to filter results (use Ctrl+Space)

	invoiceid	city	product_line	tax5	payment
1	InvoiceID	City	Product_line	[NULL]	Payment
2	750-67-8428	Yangon	Health and beauty	26.1415	Ewallet
3	226-31-3081	Naypyitaw	Electronic accessories	3.82	Cash
4	631-41-3108	Yangon	Home and lifestyle	16.2155	Credit card
5	123-19-1176	Yangon	Health and beauty	23.288	Ewallet
6	373-73-7910	Yangon	Sports and travel	30.2085	Ewallet
7	699-14-3026	Naypyitaw	Electronic accessories	29.8865	Ewallet
8	355-53-5943	Yangon	Electronic accessories	20.652	Ewallet
9	315-22-5665	Naypyitaw	Home and lifestyle	36.78	Ewallet
10	665-32-9167	Yangon	Health and beauty	3.626	Credit card

c. Count the total entries

```
select count(*) from emp.employees;
```



DBeaver 24.1.1 - <localhost> Script-1
File Edit Navigate Search SQL Editor Database Window Help
SQL Commit Rollback Auto localhost emp
*<localhost> Script
*<localhost> Script-1
General
Connections
localhost
Bookmarks
Dashboards
Diagrams
Scripts
Proj...
Name
Bookmarks
Dashboards
Diagrams
Scripts
Results 1
Grid
Text
Record
Refresh Save Cancel Export data 200 1
NPT en Writable Smart Insert 5 : 45 : 176 Sel: 0 | 0

```
select * from emp.employees order by unit_price desc;

select invoiceid, city, product_line, tax5,payment  from emp.employees;

select count(*) as count from emp.employees;
```

select count(*) as count from emp.employees

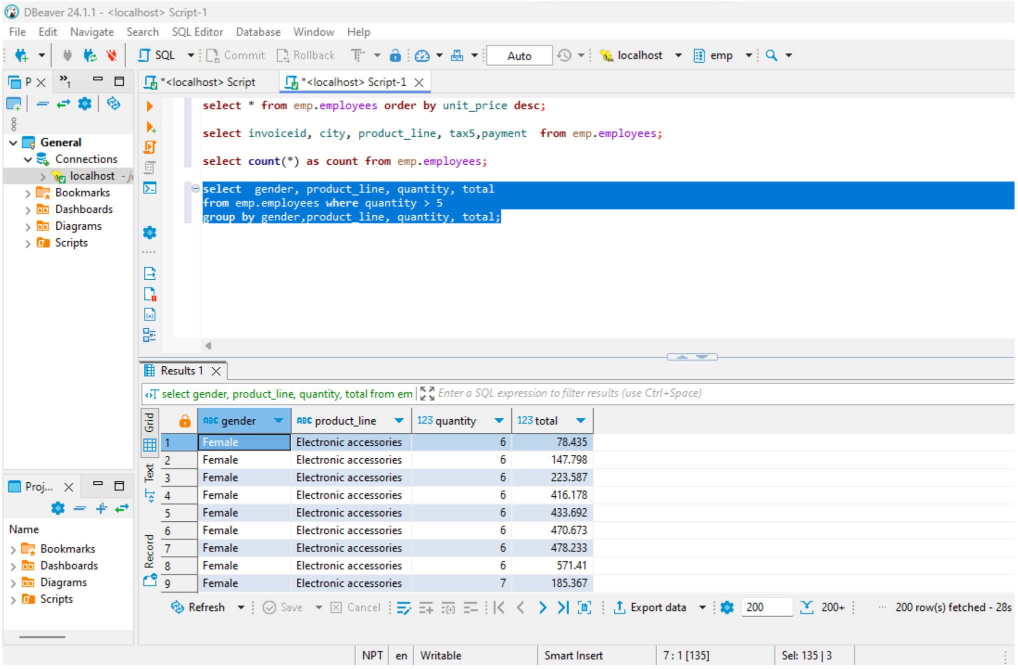
Enter a SQL expression to filter results (use Ctrl+Space)

	count
1	1,001

1 row(s) fetched - 24s

d. Group by gender, product_line, quantity, total

```sql
select  gender, product_line, quantity, total
from emp.employees where quantity > 5
group by gender,product_line, quantity, total;
```


DBeaver 24.1.1 - <localhost> Script-1
File Edit Navigate Search SQL Editor Database Window Help
SQL Commit Rollback Auto localhost emp
*<localhost> Script
*<localhost> Script-1
General
Connections
localhost
Bookmarks
Dashboards
Diagrams
Scripts
Proj...
Name
Results 1
Grid
Text
Record
Refresh Save Cancel Export data 200 200+
NPT en Writable Smart Insert 7 : 1 [135] Sel: 135 | 3

```sql
select * from emp.employees order by unit_price desc;

select invoiceid, city, product_line, tax5,payment  from emp.employees;

select count(*) as count from emp.employees;

select  gender, product_line, quantity, total
from emp.employees where quantity > 5
group by gender,product_line, quantity, total;
```

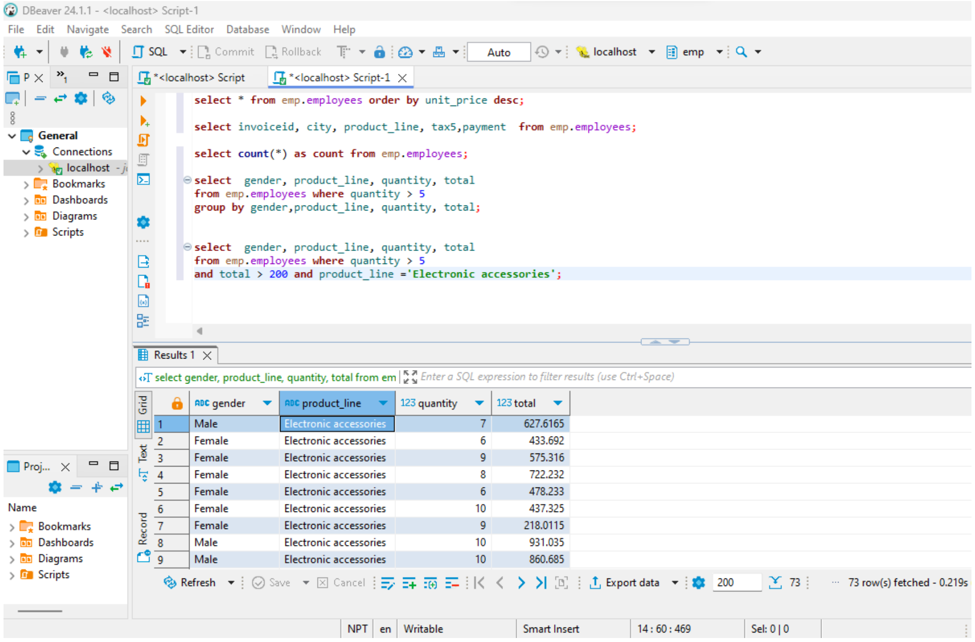
select gender, product_line, quantity, total from em | Enter a SQL expression to filter results (use Ctrl+Space)

	gender	product_line	quantity	total
1	Female	Electronic accessories	6	78.435
2	Female	Electronic accessories	6	147.798
3	Female	Electronic accessories	6	223.587
4	Female	Electronic accessories	6	416.178
5	Female	Electronic accessories	6	433.692
6	Female	Electronic accessories	6	470.673
7	Female	Electronic accessories	6	478.233
8	Female	Electronic accessories	6	571.41
9	Female	Electronic accessories	7	185.367

200 row(s) fetched - 28s

e. Select using multiple criteria’s

```sql
select  gender, product_line, quantity, total
from emp.employees where quantity > 5
and total > 200 and product_line ='Electronic accessories';
```


DBeaver 24.1.1 - <localhost> Script-1
File Edit Navigate Search SQL Editor Database Window Help
SQL Commit Rollback Auto localhost emp
*<localhost> Script
*<localhost> Script-1
General
Connections
localhost
Bookmarks
Dashboards
Diagrams
Scripts
Proj...
Name
Results 1
Grid
Text
Record
Refresh Save Cancel Export data 200 73
NPT en Writable Smart Insert 14 : 60 : 469 Sel: 0 | 0

```sql
select * from emp.employees order by unit_price desc;

select invoiceid, city, product_line, tax5,payment  from emp.employees;

select count(*) as count from emp.employees;

select  gender, product_line, quantity, total
from emp.employees where quantity > 5
group by gender,product_line, quantity, total;


select  gender, product_line, quantity, total
from emp.employees where quantity > 5
and total > 200 and product_line ='Electronic accessories';
```

select gender, product_line, quantity, total from em | Enter a SQL expression to filter results (use Ctrl+Space)

	gender	product_line	quantity	total
1	Male	Electronic accessories	7	627.6165
2	Female	Electronic accessories	6	433.692
3	Female	Electronic accessories	9	575.316
4	Female	Electronic accessories	8	722.232
5	Female	Electronic accessories	6	478.233
6	Female	Electronic accessories	10	437.325
7	Female	Electronic accessories	9	218.0115
8	Male	Electronic accessories	10	931.035
9	Male	Electronic accessories	10	860.685

73 row(s) fetched - 0.219s

*** END OF MANUAL***