AMAZON PRODUCT SALES ANALYSIS USING HIVE

A PROJECT REPORT

Submitted by

PONNURI ANIRUDDHA

[Reg No: RA2112704010015]

ABHAY SHAJI

[Reg No: RA2112704010006]
GAURAV SAHA
[Reg No: RA2112704010004]

Under the Guidance of

Dr. K. Sornalakshmi
(Assistant Professor, Department of Data Science and Business Systems)

In partial fulfillment of the Requirements for the Degree of

Masters of Technology (Integrated)

DEPARTMENT OF

DATA SCIENCE AND BUSINESSSYSTEMS

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR-603203

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR-603203

BONAFIDE CERTIFICATE

Certified that this project report titled "**AMAZON PRODUCT SALES AND VISUALIZATION THROUGH TABLEAU"** is the bonafide work of **"Ponnuri Aniruddha [Reg No: RA2112704010015]","Abhay Sahaji [Reg No: RA2112704010006]","Gaurav Saha [Reg No: RA2112704010004]",** carried out the project work under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

Dr. K. Sornalakshmi **GUIDE** Associate Professor Dept. of Data Science and Business Systems	Dr. M Lakshmi **HEAD OF THE DEPARTMENT** Dept. of Data Science and Business Systems
Signature of Internal Examiner	Signature of External Examiner

ABSTRACT

"Amazon Product Sales Analysis Using Hive and Visualization Using Tableau" highlights the use of big data technology and data visualization techniques to analyze and present insights into product sales data obtained from Amazon. The project involves using Apache Hive, a data warehouse infrastructure tool for querying and managing large datasets stored in distributed storage, to process and analyze the data. Furthermore, data visualization tools like Tableau are employed to create interactive dashboards and reports that allow users to explore the data and gain valuable insights. The project provides an opportunity to develop skills in data analysis, data visualization, and big data technologies, which are increasingly important in the field of data science.

Amazon Product Sales Analysis using Hive and Visualization using Tableau is a project that aims to analyze the sales of various products sold on Amazon using the big data tool Hive and visualize the insights obtained using Tableau. The project involves loading and processing the Amazon product sales dataset using Hive, performing exploratory data analysis to gain insights about the data, and generating visualizations using Tableau to effectively communicate the insights obtained. The project aims to provide a comprehensive analysis of the sales data and help identify trends and patterns in product sales, customer behavior, and purchase patterns. The insights obtained from the analysis can be used to improve marketing strategies, optimize pricing, and make better business decisions.

\

ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr T.V.Gopal**, for his invaluable support.

We wish to thank **Dr Revathi Venkataraman**, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department**, Dr M. Lakshmi** Professor, Department of Data Science and Business Systems, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We register our immeasurable thanks to our Faculty Advisor**, Dr K Shantha Kumari, Ph.D.** , Data Science and Business Systems, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr K Sornalakshmi, Ph.D.**, Data Science and Business Systems, SRM Institute of Science and Technology, for providing me with an opportunity to pursue my project under his/her/their mentorship. He/She/They provided me with the freedom and support to explore the research topics of my interest. Her/His/Their passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank the Data Science and Business Systems staff and students, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

PONNURI ANIRUDDHA

ABHAY SHAJI

GAURAV SAHA

TABLE OF CONTENTS

TABLE OF CONTENTS

PROBLEM STATEMENT

The problem statement for the analysis of Amazon product sales using Hive and Tableau is based on the need for businesses to leverage data analytics and visualization tools to gain insights into customer behavior and product trends. With millions of products available on Amazon in various categories, it can be challenging for businesses to identify the most popular products, categories, and sub-categories, and make data-driven decisions to optimize their product offerings, improve customer satisfaction, and increase sales.

The use of tools such as Hive and Tableau can help businesses to process and analyze large volumes of data, and create interactive visualizations and dashboards for better data exploration and communication. Hive is a data warehouse system that enables businesses to store, manage and query large datasets in a distributed manner, while Tableau is a data visualization tool that allows businesses to create interactive dashboards and visualizations for data analysis and decision-making.

The problem statement highlights the need for businesses to use Hive and Tableau to analyze and visualize Amazon product sales data and identify the top-selling products, popular categories and sub-categories, average ratings and prices, and other key metrics for their business success. By doing so, businesses can gain insights into customer behavior and preferences, optimize their product offerings, improve customer satisfaction, and increase sales.

In summary, the problem statement for the analysis of Amazon product sales using Hive and Tableau focuses on the need for businesses to leverage data analytics and visualization tools to gain insights into customer behavior and product trends, and make data-driven decisions to improve their bottom line.

MOTIVATION

Improving Sales Performance: By analyzing the product sales data, Amazon can gain insights into which products are selling well and which ones are not. This can help them make data-driven decisions to improve their sales performance and increase revenue.

Understanding Customer Behavior: Analyzing the customer behavior data can help Amazon understand what customers are buying, when they are buying, and why they are buying. This can help them improve their customer experience, tailor their marketing efforts, and optimize their product offerings.

Identifying Trends and Patterns: Analyzing the sales data over time can help Amazon identify trends and patterns in customer behavior, product performance, and market demand. This can help them stay ahead of the competition and adapt to changing market conditions.

Making Better Business Decisions: By using data-driven insights, Amazon can make better business decisions about product development, pricing, promotions, and inventory management. This can help them optimize their operations and increase profitability.

Enhancing Data Visualization: By using Tableau, Amazon can create visually appealing and interactive dashboards to represent their data in a more meaningful way. This can help stakeholders quickly and easily understand the insights and take action based on them.

SCOPE

The scope of the "Amazon Product Sales Analysis using Hive and Visualization using Tableau" project includes exploring and analyzing the sales data of products on Amazon. The project involves working with a large dataset, which contains information about the products such as their names, categories, sub-categories, prices, ratings, etc.

The project aims to identify the top-selling products, categories, and sub-categories, as well as to gain insights into the purchasing patterns of customers. The analysis will help in understanding the market trends and identifying potential areas for improvement in the sales strategy.

The project also involves visualizing the analyzed data using Tableau, which is a powerful data visualization tool. The visualization will help in presenting the findings in a more effective and easy-to-understand way, which will aid in making informed business decisions.

The scope of the project is not limited to just Amazon; it can be extended to analyze the sales data of other e-commerce platforms as well. The project is relevant for businesses that operate in the e-commerce industry and want to improve their sales strategy by gaining insights into customer behavior and market trends.

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In today's data-driven business world, companies are constantly looking for ways to extract insights from their data to make informed decisions. One of the primary ways to do this is through data analysis, which involves processing, cleaning, and interpreting large amounts of data to identify patterns and trends.

In this context, the analysis of product sales data is of particular interest, as it can provide valuable insights into consumer behavior and market trends. Amazon, being one of the largest online marketplaces in the world, generates a vast amount of sales data that can be used for such analysis.

Hive, a data warehousing system built on top of Hadoop, is a popular tool for analyzing large datasets. It provides a SQL-like interface to query and manage data, making it accessible to users with a background in SQL. Hive also offers the ability to scale to petabytes of data, making it a suitable choice for big data analysis.

In this project, we will analyze Amazon product sales data using Hive and visualize the results using Tableau. The main objective of this project is to extract meaningful insights from the sales data, such as top-performing products, popular categories, and sales trends over time. The use of Hive and Tableau will enable us to handle and visualize large amounts of data, making it easier to draw insights and conclusions.

1.2 HISTORY

The history of Amazon dates back to 1994, when Jeff Bezos founded the company in Seattle, Washington, with the aim of creating an online bookstore. Over time, Amazon expanded to include a wide range of products, including electronics, clothing, and household goods. As the company grew, so did the need for analyzing its vast amount of sales data

Initially, Amazon relied on traditional data analysis tools such as SQL databases and spreadsheets to manage and analyze its sales data. However, as the volume of data grew, these tools proved to be insufficient for handling the scale and complexity of Amazon's sales data. This led to the development of Apache Hadoop, an open-source distributed computing platform that can handle large volumes of data.

Amazon was one of the early adopters of Apache Hadoop, and it became one of the core technologies used by the company for managing and analyzing its sales data. In addition, Amazon developed its own version of Hadoop called Amazon Elastic MapReduce (EMR), which allowed the company to process and analyze its sales data more efficiently.

Hive is a data warehousing tool that was developed on top of Hadoop. It allows users to query and analyze data stored in Hadoop using SQL-like commands. Hive was developed to make it easier for analysts and data scientists to work with Hadoop by providing a familiar SQL interface.

With the advent of big data, the need for scalable and efficient data analysis tools has become increasingly important for businesses. Amazon's use of Hive and Hadoop for analyzing its sales data has become a popular use case for these technologies, and has inspired many other companies to adopt similar technologies for their own data analysis needs.

1.3 BENEFITS OF USING HIVE AND TABLEAU

Benefits of using HIVE:

1. Scalability: Hive is designed to handle large datasets and can process them in a distributed manner. This makes it ideal for processing and analyzing big data.
2. SQL-like language: Hive uses a SQL-like language called HiveQL, which makes it easy for users with SQL knowledge to query and analyze data. This enables users to write complex queries to analyze and explore the data.
3. Data warehouse integration: Hive can be integrated with data warehousing tools such as Hadoop and Spark, which allows users to store and manage large amounts of data in a distributed manner.
4. Flexible data processing: Hive is designed to work with different types of data, including structured and semi-structured data. This means that users can analyze data from a wide range of sources and formats.
5. High-performance processing: Hive uses advanced processing techniques such as query optimization and indexing to improve performance and reduce processing times.

Benefits of using Tablue for visualization:

1. Tablue is a powerful data visualization tool that allows for creating interactive and dynamic dashboards and reports.
2. It supports various types of visualizations, including charts, tables, maps, and more.
3. Tablue provides intuitive drag-and-drop interfaces that make it easy to create andcustomize visualizations.
4. It can connect to various data sources, including MongoDB, and easily integrate with other tools and platforms.

Overall benefits of using HIVE and Tablue for analysis:

1. Scalability: Hive allows users to handle large volumes of data and process it in a distributed manner. This makes it an ideal tool for processing and analyzing big data.
2. SQL-based queries: Hive uses a SQL-like language called HiveQL, which makes it easy for users with SQL knowledge to query and analyze data. This enables users to write complex queries to analyze and explore the data.
3. Data warehouse integration: Hive can be integrated with data warehousing tools such as Hadoop and Spark, which allows users to store and manage large amounts of data in a distributed manner.
4. Flexibility: Hive allows users to work with different types of data, including structured and semi-structured data. This means that users can analyze data from a wide range of sources and formats.
5. Interactive visualizations: Tableau allows users to create interactive visualizations and dashboards, which can help to identify patterns and trends in the data. The visualizations are also easy to share and can be accessed from different devices.
6. Ease of use: Tableau has a user-friendly interface that makes it easy for users to create visualizations and dashboards without the need for programming skills. This means that users can create their own visualizations and explore the data in a more interactive way.
7. Collaboration: Tableau allows users to share their visualizations and dashboards with others, which enables collaboration and knowledge sharing. This can help teams to work together more effectively and make better data-driven decisions.

1.4 WORKING PRINCIPLE

The working principle of Amazon product sales analysis using Hive involves processing and analyzing the sales data of various products sold on Amazon using the Hive data warehousing system. The sales data is first imported into Hive, a distributed data warehousing system built on top of Hadoop, and then transformed and analyzed using HiveQL, a SQL-like language.

The data is queried using HiveQL to obtain insights into the sales trends, popular products, and customer behavior. The queries are designed to aggregate the data by different parameters such as category, sub-category, customer demographics, and time period. The aggregated data is then used to generate reports and visualizations that help in understanding the sales patterns and identifying opportunities for growth.

Once the data is analyzed and insights are obtained, the results can be visualized using various tools such as Tableau, which is a powerful data visualization software. Tableau helps in creating interactive dashboards and visualizations that enable the stakeholders to make informed decisions based on the insights gained from the data.

Overall, the working principle of Amazon product sales analysis using Hive involves importing, transforming, analyzing, and visualizing the sales data using HiveQL and Tableau, to gain insights into the sales patterns and make data-driven decisions for growth.

CHAPTER 2

LITERATURE STUDY

1. "Big Data Analytics on Amazon Web Services (AWS) using Hive and Tableau" by Avinash Prasad, published in the International Journal of Computer Science and Mobile Computing in 2016. This paper discusses the use of Hive and Tableau for big data analytics on AWS, with a focus on Amazon customer review data.
2. "Big Data Analytics using Hive and Tableau" by P. Shanmugapriya and S. Srinivasan, published in the International Journal of Advanced Research in Computer Science and Software Engineering in 2015. This paper provides an overview of Hive and Tableau and their use for big data analytics, with a case study on analyzing social media data.
3. "A Study of Data Visualization Tools: Tableau, QlikView, and MicroStrategy" by M. A. Rizwan and M. A. N. Alias, published in the International Journal of Computer Science and Information Technologies in 2015. This paper compares Tableau, QlikView, and MicroStrategy as data visualization tools, with a focus on their features, strengths, and weaknesses.
4. "Data Warehousing and Business Intelligence for Analyzing Amazon Sales Data using Tableau" by K. Saravanan and R. Kavitha, published in the International Journal of Emerging Technology and Advanced Engineering in 2013. This paper presents a case study on analyzing Amazon sales data using Tableau for data warehousing and business intelligence purposes.
5. "A Comparative Study of Data Visualization Tools for Big Data Analytics" by N. R. Selvarani and V. M. Jeyakumar, published in the International Journal of Advanced Research in Computer Science and Software Engineering in 2016. This paper compares various data visualization tools, including Tableau, for big data analytics, with a focus on their performance, scalability, and ease of use.

CHAPTER 3
DATASET

Amazon Products Sales

This is a Product Sales Dataset scraped from the Amazon website which is download from Kaggle.

- Its product data are separated by 142 categories in csv format, along with the full dataset name **Amazon-Products.csv**.
- Each csv files are consisting of 7 columns and each row has products details accordingly.

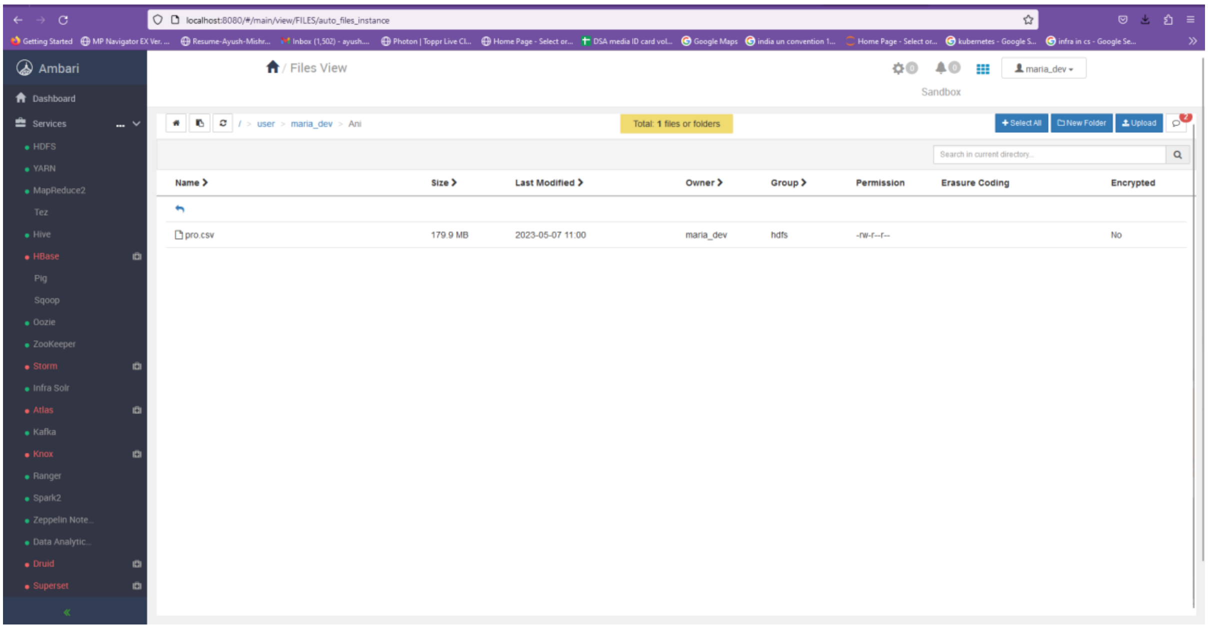
Features

name	**description**
name	The name of the product
main_category	The main category of the product belong
sub_category	The main category of the product belong
ratings	The ratings given by amazon customers of the product
no of ratings	The number of ratings given to this product in amazon shopping
discount_price	The discount prices of the product
actual_price	The actual MRP of the product

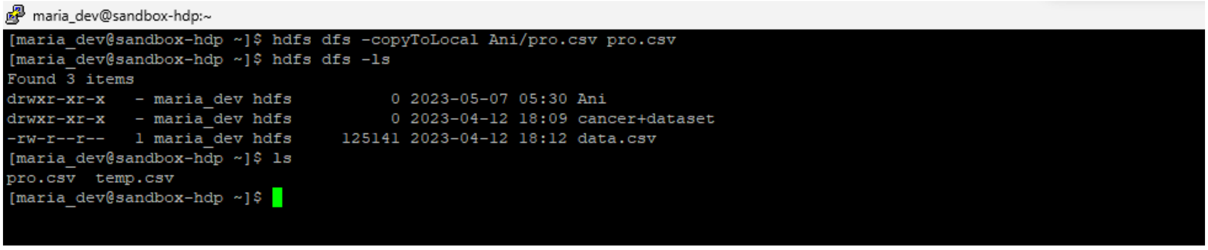
CHAPTER 4
CODE

Importing File into HDFS

Uploading File into HDFS using Apache Ambari

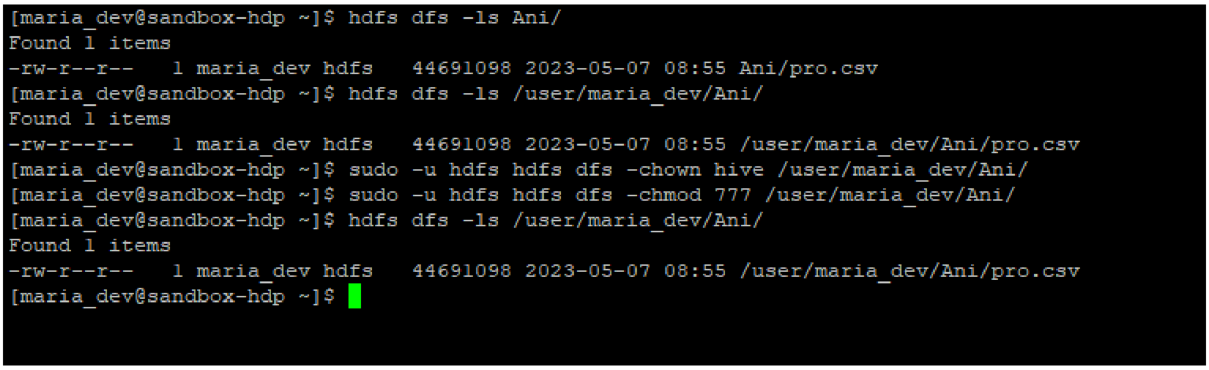


Name	Size	Last Modified	Owner	Group	Permission	Erasure Coding	Encrypted
pro.csv	179.9 MB	2023-05-07 11:00	maria_dev	hdfs	-rw-r--r--		No



```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -copyToLocal Ani/pro.csv pro.csv
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls
Found 3 items
drwxr-xr-x   - maria_dev hdfs          0 2023-05-07 05:30 Ani
drwxr-xr-x   - maria_dev hdfs          0 2023-04-12 18:09 cancer+dataset
-rw-r--r--   1 maria_dev hdfs     125141 2023-04-12 18:12 data.csv
[maria_dev@sandbox-hdp ~]$ ls
pro.csv  temp.csv
[maria_dev@sandbox-hdp ~]$
```

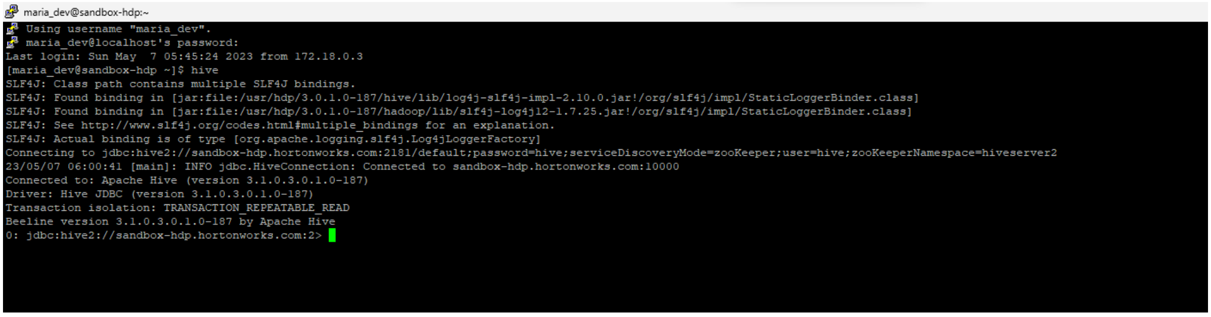
Changing ownership and permissions of data file



```
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls Ani/
Found 1 items
-rw-r--r--   1 maria_dev hdfs   44691098 2023-05-07 08:55 Ani/pro.csv
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev/Ani/
Found 1 items
-rw-r--r--   1 maria_dev hdfs   44691098 2023-05-07 08:55 /user/maria_dev/Ani/pro.csv
[maria_dev@sandbox-hdp ~]$ sudo -u hdfs hdfs dfs -chown hive /user/maria_dev/Ani/
[maria_dev@sandbox-hdp ~]$ sudo -u hdfs hdfs dfs -chmod 777 /user/maria_dev/Ani/
[maria_dev@sandbox-hdp ~]$ hdfs dfs -ls /user/maria_dev/Ani/
Found 1 items
-rw-r--r--   1 maria_dev hdfs   44691098 2023-05-07 08:55 /user/maria_dev/Ani/pro.csv
[maria_dev@sandbox-hdp ~]$
```

Hive Commands

Launching Hive from HortonWorks

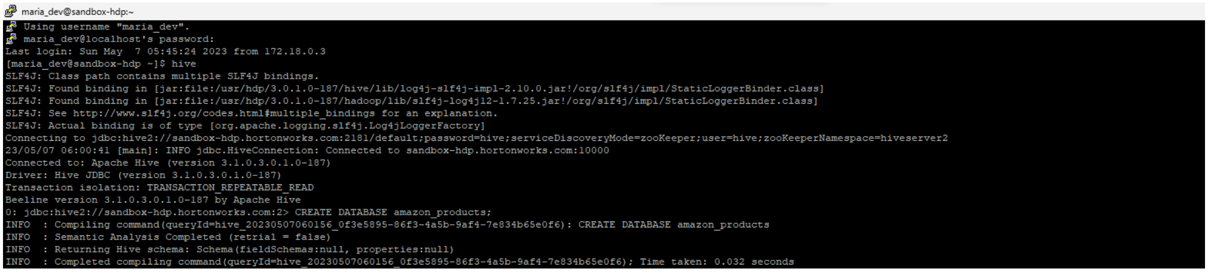


maria_dev@sandbox-hdp:~

```
 Using username "maria_dev".
 maria_dev@localhost's password:
Last login: Sun May  7 05:45:24 2023 from 172.18.0.3
[maria_dev@sandbox-hdp ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hadoop/lib/slf4j-log4j12-1.7.25.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]
Connecting to jdbc:hive2://sandbox-hdp.hortonworks.com:2181/default;password=hive;serviceDiscoveryMode=zooKeeper;user=hive;zooKeeperNamespace=hiveserver2
23/05/07 06:00:41 [main]: INFO jdbc.HiveConnection: Connected to sandbox-hdp.hortonworks.com:10000
Connected to: Apache Hive (version 3.1.0.3.0.1.0-187)
Driver: Hive JDBC (version 3.1.0.3.0.1.0-187)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.0.1.0-187 by Apache Hive
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Creating a Database

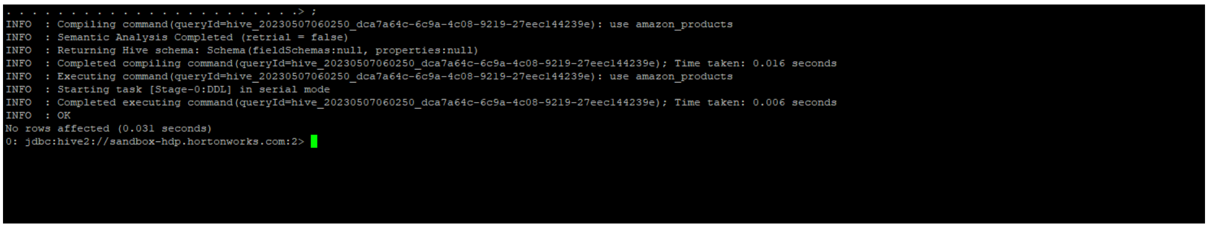
- *create database* amazon_products;



maria_dev@sandbox-hdp:~

```
 Using username "maria_dev".
 maria_dev@localhost's password:
Last login: Sun May  7 05:45:24 2023 from 172.18.0.3
[maria_dev@sandbox-hdp ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hadoop/lib/slf4j-log4j12-1.7.25.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]
Connecting to jdbc:hive2://sandbox-hdp.hortonworks.com:2181/default;password=hive;serviceDiscoveryMode=zooKeeper;user=hive;zooKeeperNamespace=hiveserver2
23/05/07 06:00:41 [main]: INFO jdbc.HiveConnection: Connected to sandbox-hdp.hortonworks.com:10000
Connected to: Apache Hive (version 3.1.0.3.0.1.0-187)
Driver: Hive JDBC (version 3.1.0.3.0.1.0-187)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.0.1.0-187 by Apache Hive
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> CREATE DATABASE amazon_products;
INFO  : Compiling command(queryId=hive_20230507060156_0f3e5895-86f3-4a5b-9af4-7e834b65e0f6): CREATE DATABASE amazon_products
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230507060156_0f3e5895-86f3-4a5b-9af4-7e834b65e0f6); Time taken: 0.032 seconds
```

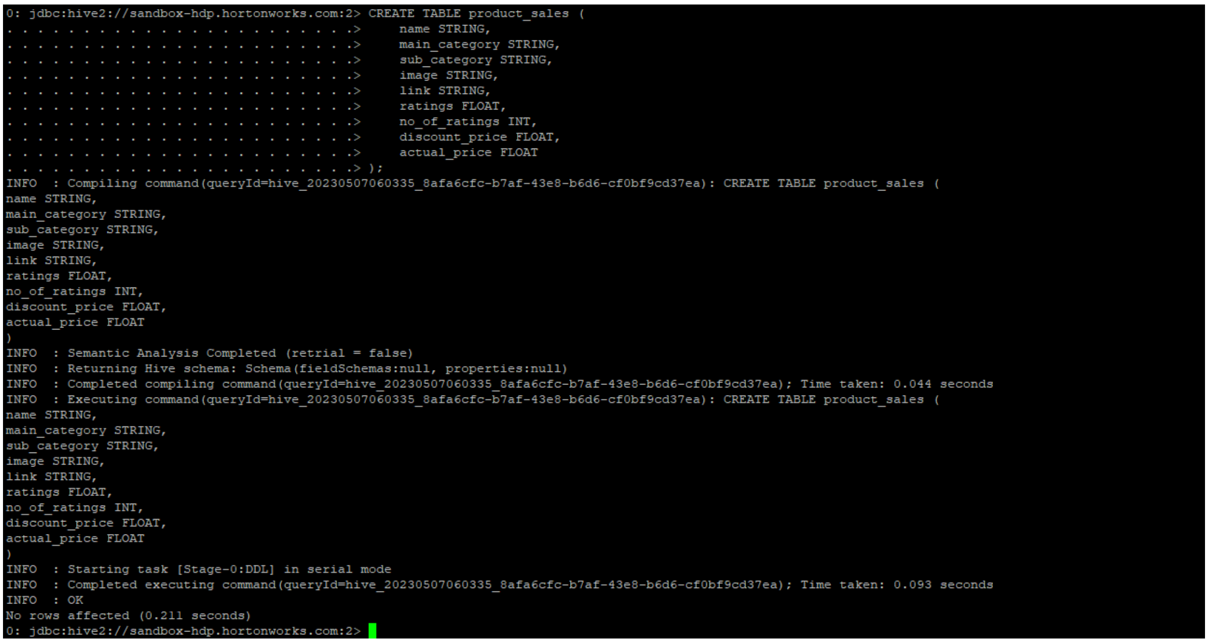
- *use amazon_products*



```
. . . . . . . . . . . . . . . . . . . . . . .> ;
INFO  : Compiling command(queryId=hive_20230507060250_dca7a64c-6c9a-4c08-9219-27eec144239e): use amazon_products
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230507060250_dca7a64c-6c9a-4c08-9219-27eec144239e); Time taken: 0.016 seconds
INFO  : Executing command(queryId=hive_20230507060250_dca7a64c-6c9a-4c08-9219-27eec144239e): use amazon_products
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230507060250_dca7a64c-6c9a-4c08-9219-27eec144239e); Time taken: 0.006 seconds
INFO  : OK
No rows affected (0.031 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Creating Table

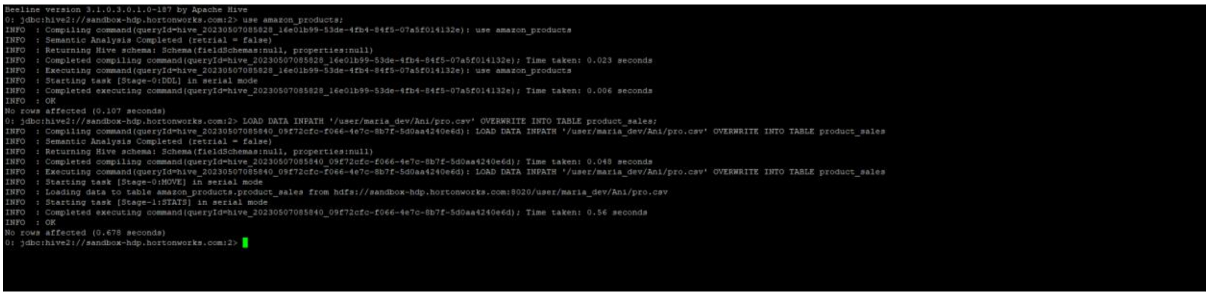
```
CREATE TABLE product_sales (
    name STRING,
    main_category STRING,
    sub_category STRING,
    ratings FLOAT,
    no_of_ratings INT,
    discount_price FLOAT,
    actual_price FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES("skip.header.line.count"="1");
```



```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> CREATE TABLE product_sales (
. . . . . . . . . . . . . . . . . . . . . . .>     name STRING,
. . . . . . . . . . . . . . . . . . . . . . .>     main_category STRING,
. . . . . . . . . . . . . . . . . . . . . . .>     sub_category STRING,
. . . . . . . . . . . . . . . . . . . . . . .>     image STRING,
. . . . . . . . . . . . . . . . . . . . . . .>     link STRING,
. . . . . . . . . . . . . . . . . . . . . . .>     ratings FLOAT,
. . . . . . . . . . . . . . . . . . . . . . .>     no_of_ratings INT,
. . . . . . . . . . . . . . . . . . . . . . .>     discount_price FLOAT,
. . . . . . . . . . . . . . . . . . . . . . .>     actual_price FLOAT
. . . . . . . . . . . . . . . . . . . . . . .> );
INFO  : Compiling command(queryId=hive_20230507060335_8afa6cfc-b7af-43e8-b6d6-cf0bf9cd37ea): CREATE TABLE product_sales (
name STRING,
main_category STRING,
sub_category STRING,
image STRING,
link STRING,
ratings FLOAT,
no_of_ratings INT,
discount_price FLOAT,
actual_price FLOAT
)
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230507060335_8afa6cfc-b7af-43e8-b6d6-cf0bf9cd37ea); Time taken: 0.044 seconds
INFO  : Executing command(queryId=hive_20230507060335_8afa6cfc-b7af-43e8-b6d6-cf0bf9cd37ea): CREATE TABLE product_sales (
name STRING,
main_category STRING,
sub_category STRING,
image STRING,
link STRING,
ratings FLOAT,
no_of_ratings INT,
discount_price FLOAT,
actual_price FLOAT
)
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230507060335_8afa6cfc-b7af-43e8-b6d6-cf0bf9cd37ea); Time taken: 0.093 seconds
INFO  : OK
No rows affected (0.211 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Loading CSV data into HIVE

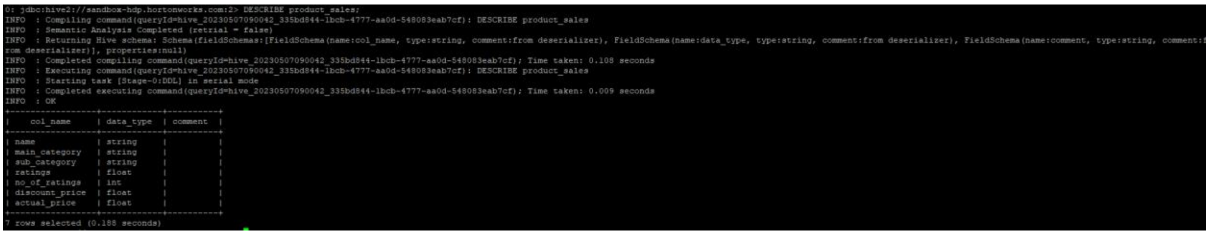
LOAD DATA INPATH '/path/to/csv/file.csv' OVERWRITE INTO TABLE product_sales;



```
Beeline version 3.1.0.3.0.1.0-187 by Apache Hive
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> use amazon_products;
INFO  : Compiling command(queryId=hive_20230507085828_16e01b99-53de-4fb4-84f5-07a5f014132e): use amazon_products
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230507085828_16e01b99-53de-4fb4-84f5-07a5f014132e); Time taken: 0.023 seconds
INFO  : Executing command(queryId=hive_20230507085828_16e01b99-53de-4fb4-84f5-07a5f014132e): use amazon_products
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230507085828_16e01b99-53de-4fb4-84f5-07a5f014132e); Time taken: 0.006 seconds
INFO  : OK
No rows affected (0.107 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> LOAD DATA INPATH '/user/maria_dev/Ani/pro.csv' OVERWRITE INTO TABLE product_sales;
INFO  : Compiling command(queryId=hive_20230507085840_09f72cfc-f066-4e7c-8b7f-5d0aa4240e6d): LOAD DATA INPATH '/user/maria_dev/Ani/pro.csv' OVERWRITE INTO TABLE product_sales
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20230507085840_09f72cfc-f066-4e7c-8b7f-5d0aa4240e6d); Time taken: 0.048 seconds
INFO  : Executing command(queryId=hive_20230507085840_09f72cfc-f066-4e7c-8b7f-5d0aa4240e6d): LOAD DATA INPATH '/user/maria_dev/Ani/pro.csv' OVERWRITE INTO TABLE product_sales
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Loading data to table amazon_products.product_sales from hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/Ani/pro.csv
INFO  : Starting task [Stage-1:STATS] in serial mode
INFO  : Completed executing command(queryId=hive_20230507085840_09f72cfc-f066-4e7c-8b7f-5d0aa4240e6d); Time taken: 0.56 seconds
INFO  : OK
No rows affected (0.678 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Table Schema

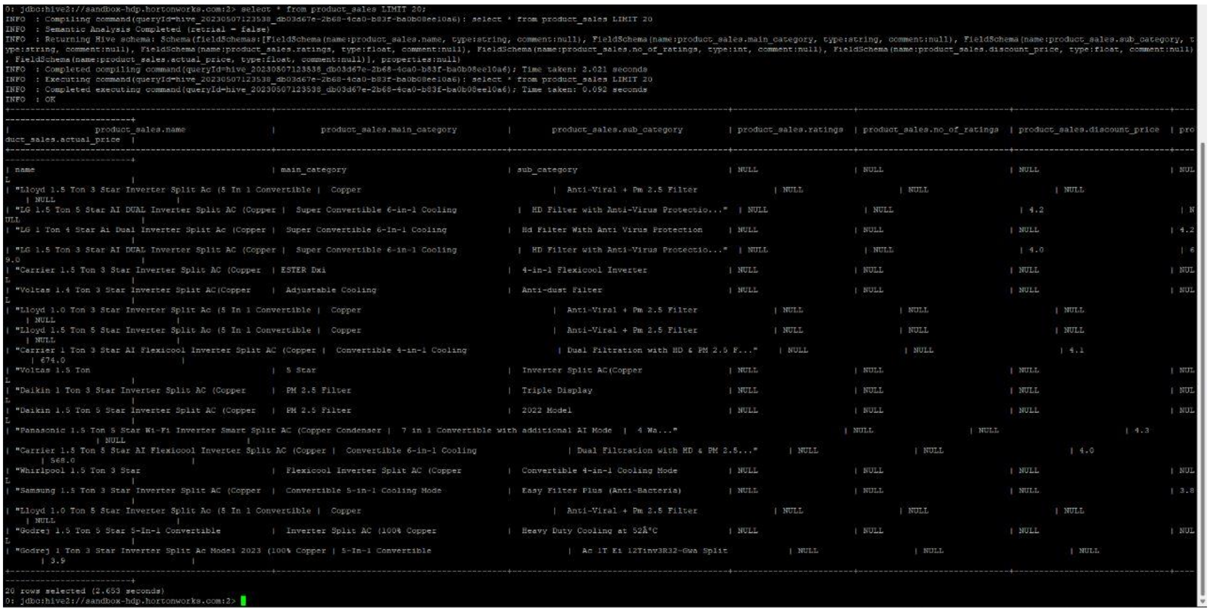
describe product_sales;



```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> DESCRIBE product_sales;
INFO  : Compiling command(queryId=hive_20230507090042_335bd844-1bcb-4777-aa0d-548083eab7cf): DESCRIBE product_sales
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:f
rom deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230507090042_335bd844-1bcb-4777-aa0d-548083eab7cf); Time taken: 0.108 seconds
INFO  : Executing command(queryId=hive_20230507090042_335bd844-1bcb-4777-aa0d-548083eab7cf): DESCRIBE product_sales
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20230507090042_335bd844-1bcb-4777-aa0d-548083eab7cf); Time taken: 0.009 seconds
INFO  : OK
+-----------------+------------+----------+
|    col_name     | data_type  | comment  |
+-----------------+------------+----------+
| name            | string     |          |
| main_category   | string     |          |
| sub_category    | string     |          |
| ratings         | float      |          |
| no_of_ratings   | int        |          |
| discount_price  | float      |          |
| actual_price    | float      |          |
+-----------------+------------+----------+
7 rows selected (0.188 seconds)
```

Viewing Data using HiveQL

*Select * from product_sales Limit 20*

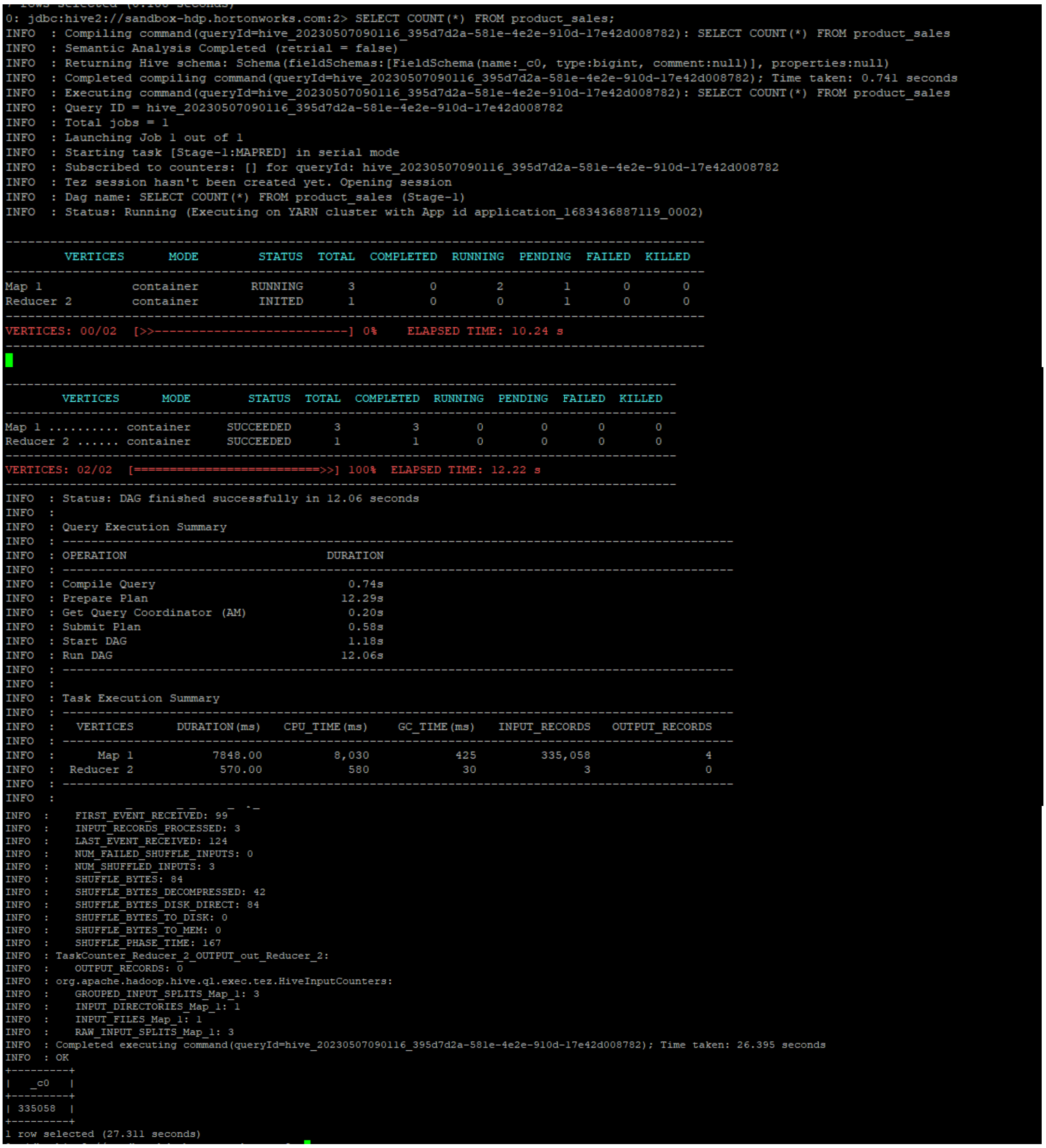


```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> select * from product_sales LIMIT 20;
INFO  : Compiling command(queryId=hive_20230507123538_db03d67e-2b68-4ca0-b83f-ba0b08ee10a6): select * from product_sales LIMIT 20
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:product_sales.name, type:string, comment:null), FieldSchema(name:product_sales.main_category, type:string, comment:null), FieldSchema(name:product_sales.sub_category, t
ype:string, comment:null), FieldSchema(name:product_sales.ratings, type:float, comment:null), FieldSchema(name:product_sales.no_of_ratings, type:int, comment:null), FieldSchema(name:product_sales.discount_price, type:float, comment:null)
, FieldSchema(name:product_sales.actual_price, type:float, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230507123538_db03d67e-2b68-4ca0-b83f-ba0b08ee10a6); Time taken: 2.021 seconds
INFO  : Executing command(queryId=hive_20230507123538_db03d67e-2b68-4ca0-b83f-ba0b08ee10a6): select * from product_sales LIMIT 20
INFO  : Completed executing command(queryId=hive_20230507123538_db03d67e-2b68-4ca0-b83f-ba0b08ee10a6); Time taken: 0.092 seconds
INFO  : OK
+-----------------------------------------------------+-----------------------------------------------+-----------------------------------------+-------------------------+-------------------------------+-------------------------------+----
------------------------+
|                 product_sales.name                  |          product_sales.main_category          |         product_sales.sub_category      | product_sales.ratings   | product_sales.no_of_ratings   | product_sales.discount_price  | pro
duct_sales.actual_price  |
+-----------------------------------------------------+-----------------------------------------------+-----------------------------------------+-------------------------+-------------------------------+-------------------------------+----
------------------------+
| name                                                | main_category                                 | sub_category                            | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Lloyd 1.5 Ton 3 Star Inverter Split Ac (5 In 1 Convertible |  Copper                                        |  Anti-Viral + Pm 2.5 Filter              | NULL                    | NULL                          | NULL
    | NULL                   |
| "LG 1.5 Ton 5 Star AI DUAL Inverter Split AC (Copper |  Super Convertible 6-in-1 Cooling          |  HD Filter with Anti-Virus Protectio..."  | NULL                    | NULL                          | 4.2                           | N
ULL                        |
| "LG 1 Ton 4 Star Ai Dual Inverter Split Ac (Copper |  Super Convertible 6-In-1 Cooling             |  Hd Filter With Anti Virus Protection     | NULL                    | NULL                          | NULL                          | 4.2
                         |
| "LG 1.5 Ton 3 Star AI DUAL Inverter Split AC (Copper |  Super Convertible 6-in-1 Cooling          |  HD Filter with Anti-Virus Protectio..."  | NULL                    | NULL                          | 4.0                           | 6
9.0                        |
| "Carrier 1.5 Ton 3 Star Inverter Split AC (Copper  | ESTER Dxi                                     |  4-in-1 Flexicool Inverter               | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Voltas 1.4 Ton 3 Star Inverter Split AC(Copper    |  Adjustable Cooling                           |  Anti-dust Filter                        | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Lloyd 1.0 Ton 3 Star Inverter Split Ac (5 In 1 Convertible |  Copper                                        |  Anti-Viral + Pm 2.5 Filter              | NULL                    | NULL                          | NULL
    | NULL                   |
| "Lloyd 1.5 Ton 5 Star Inverter Split Ac (5 In 1 Convertible |  Copper                                        |  Anti-Viral + Pm 2.5 Filter              | NULL                    | NULL                          | NULL
    | NULL                   |
| "Carrier 1 Ton 3 Star AI Flexicool Inverter Split AC (Copper |  Convertible 4-in-1 Cooling                   | Dual Filtration with HD & PM 2.5 F..."    | NULL                    | NULL                          | 4.1
     | 674.0                  |
| "Voltas 1.5 Ton                                    |  5 Star                                       |  Inverter Split AC(Copper                | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Daikin 1 Ton 3 Star Inverter Split AC (Copper     |  PM 2.5 Filter                                |  Triple Display                          | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Daikin 1.5 Ton 5 Star Inverter Split AC (Copper   |  PM 2.5 Filter                                |  2022 Model                              | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Panasonic 1.5 Ton 5 Star Wi-Fi Inverter Smart Split AC (Copper Condenser |  7 in 1 Convertible with additional AI Mode  |  4 Wa..."                    | NULL                    | NULL                          | 4.3
                 | NULL                          |
| "Carrier 1.5 Ton 5 Star AI Flexicool Inverter Split AC (Copper |  Convertible 6-in-1 Cooling                 | Dual Filtration with HD & PM 2.5..."      | NULL                    | NULL                          | 4.0
      | 568.0                  |
| "Whirlpool 1.5 Ton 3 Star                          |  Flexicool Inverter Split AC (Copper          |  Convertible 4-in-1 Cooling Mode         | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Samsung 1.5 Ton 3 Star Inverter Split AC (Copper  |  Convertible 5-in-1 Cooling Mode              |  Easy Filter Plus (Anti-Bacteria)        | NULL                    | NULL                          | NULL                          | 3.8
                         |
| "Lloyd 1.0 Ton 5 Star Inverter Split Ac (5 In 1 Convertible |  Copper                                        |  Anti-Viral + Pm 2.5 Filter              | NULL                    | NULL                          | NULL
    | NULL                   |
| "Godrej 1.5 Ton 5 Star 5-In-1 Convertible          |  Inverter Split AC (100% Copper               |  Heavy Duty Cooling at 52Â°C             | NULL                    | NULL                          | NULL                          | NUL
L                        |
| "Godrej 1 Ton 3 Star Inverter Split Ac Model 2023 (100% Copper | 5-In-1 Convertible                         |  Ac 1T Ei 12Tinv3R32-Gwa Split          | NULL                    | NULL                          | NULL
      | 3.9                    |
+-----------------------------------------------------+-----------------------------------------------+-----------------------------------------+-------------------------+-------------------------------+-------------------------------+----
------------------------+
20 rows selected (2.653 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Data Analysis

Viewing Number of rows

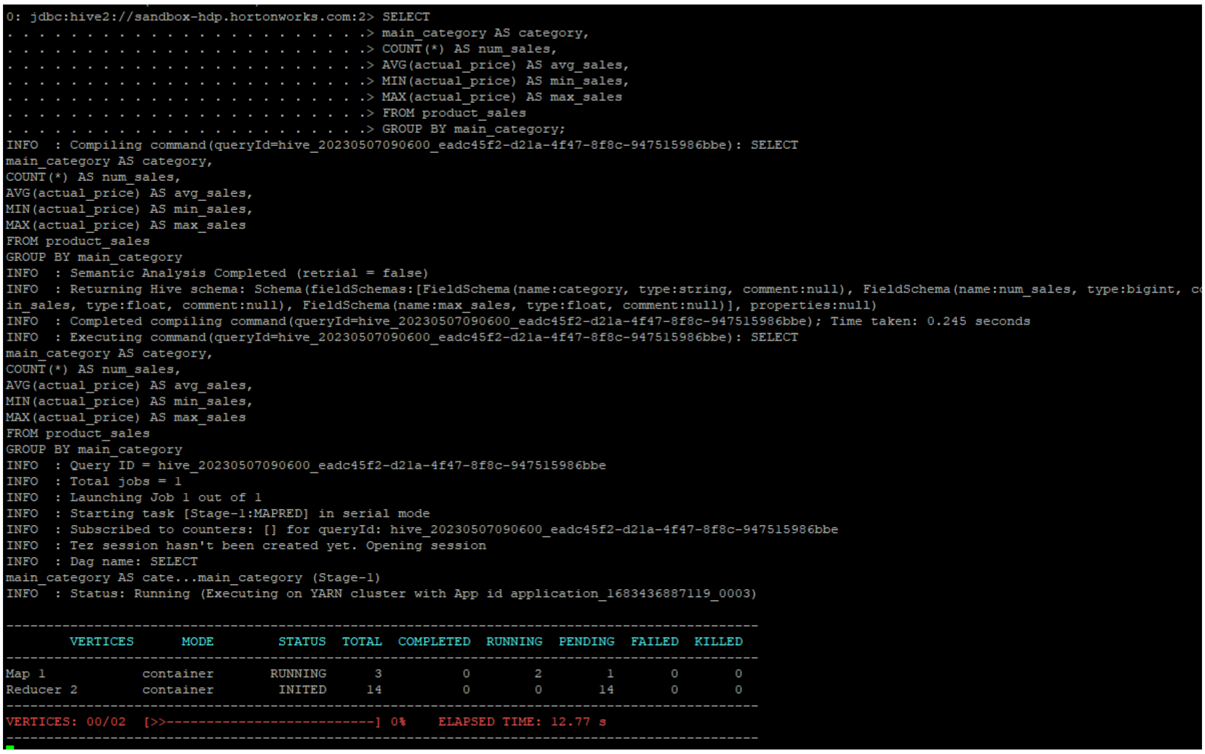
Select count(*) from product_sales;



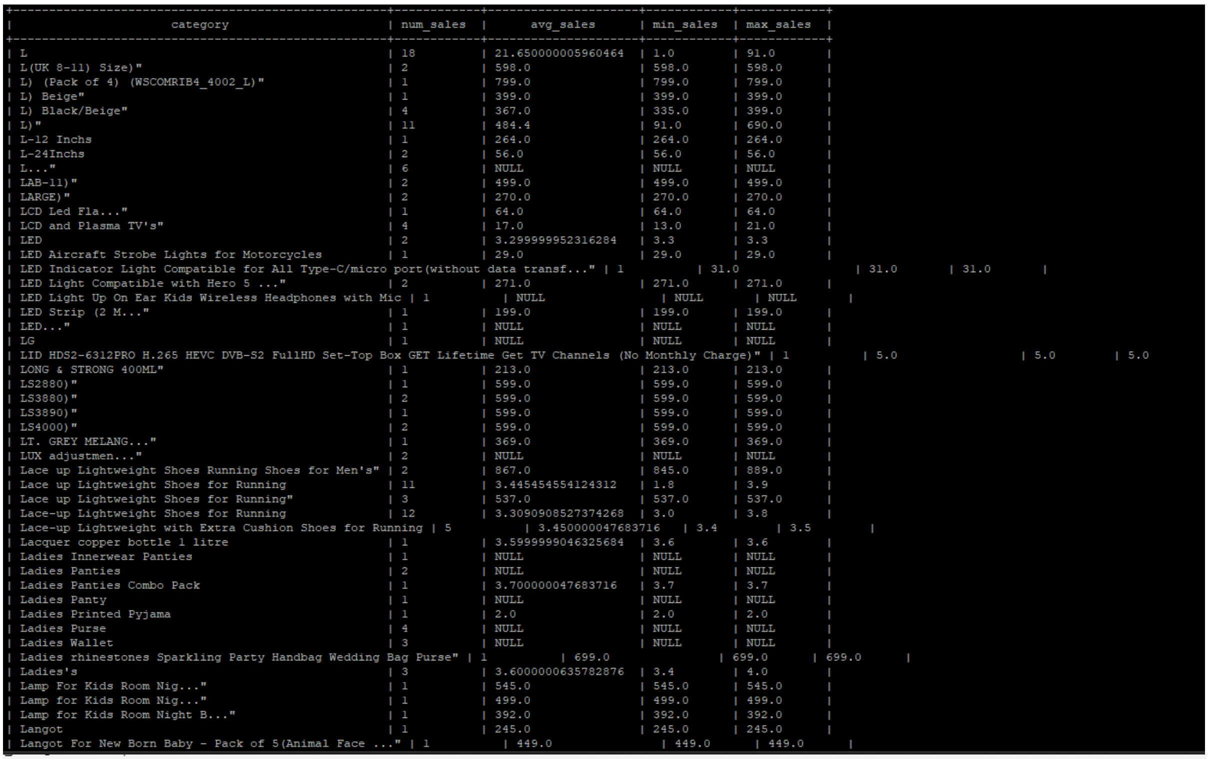
```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> SELECT COUNT(*) FROM product_sales;
INFO  : Compiling command(queryId=hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782): SELECT COUNT(*) FROM product_sales
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:_c0, type:bigint, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782); Time taken: 0.741 seconds
INFO  : Executing command(queryId=hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782): SELECT COUNT(*) FROM product_sales
INFO  : Query ID = hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782
INFO  : Tez session hasn't been created yet. Opening session
INFO  : Dag name: SELECT COUNT(*) FROM product_sales (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1683436887119_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1            container       RUNNING      3          0        2        1       0       0
Reducer 2        container        INITED      1          0        0        1       0       0
----------------------------------------------------------------------------------------------
VERTICES: 00/02  [>>--------------------------] 0%    ELAPSED TIME: 10.24 s
----------------------------------------------------------------------------------------------

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      3          3        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.22 s
----------------------------------------------------------------------------------------------
INFO  : Status: DAG finished successfully in 12.06 seconds
INFO  :
INFO  : Query Execution Summary
INFO  : ----------------------------------------------------------------------------------------------
INFO  : OPERATION                            DURATION
INFO  : ----------------------------------------------------------------------------------------------
INFO  : Compile Query                           0.74s
INFO  : Prepare Plan                           12.29s
INFO  : Get Query Coordinator (AM)              0.20s
INFO  : Submit Plan                             0.58s
INFO  : Start DAG                               1.18s
INFO  : Run DAG                                12.06s
INFO  : ----------------------------------------------------------------------------------------------
INFO  :
INFO  : Task Execution Summary
INFO  : ----------------------------------------------------------------------------------------------
INFO  :   VERTICES      DURATION(ms)   CPU_TIME(ms)    GC_TIME(ms)   INPUT_RECORDS   OUTPUT_RECORDS
INFO  : ----------------------------------------------------------------------------------------------
INFO  :      Map 1           7848.00          8,030            425         335,058                4
INFO  :  Reducer 2            570.00            580             30               3                0
INFO  : ----------------------------------------------------------------------------------------------
INFO  :
INFO  :    FIRST_EVENT_RECEIVED: 99
INFO  :    INPUT_RECORDS_PROCESSED: 3
INFO  :    LAST_EVENT_RECEIVED: 124
INFO  :    NUM_FAILED_SHUFFLE_INPUTS: 0
INFO  :    NUM_SHUFFLED_INPUTS: 3
INFO  :    SHUFFLE_BYTES: 84
INFO  :    SHUFFLE_BYTES_DECOMPRESSED: 42
INFO  :    SHUFFLE_BYTES_DISK_DIRECT: 84
INFO  :    SHUFFLE_BYTES_TO_DISK: 0
INFO  :    SHUFFLE_BYTES_TO_MEM: 0
INFO  :    SHUFFLE_PHASE_TIME: 167
INFO  : TaskCounter_Reducer_2_OUTPUT_out_Reducer_2:
INFO  :    OUTPUT_RECORDS: 0
INFO  : org.apache.hadoop.hive.ql.exec.tez.HiveInputCounters:
INFO  :    GROUPED_INPUT_SPLITS_Map_1: 3
INFO  :    INPUT_DIRECTORIES_Map_1: 1
INFO  :    INPUT_FILES_Map_1: 1
INFO  :    RAW_INPUT_SPLITS_Map_1: 3
INFO  : Completed executing command(queryId=hive_20230507090116_395d7d2a-581e-4e2e-910d-17e42d008782); Time taken: 26.395 seconds
INFO  : OK
+----------+
|   _c0    |
+----------+
| 335058   |
+----------+
1 row selected (27.311 seconds)
```

Viewing Min ,Max, Avg sales of each department grouped by categories

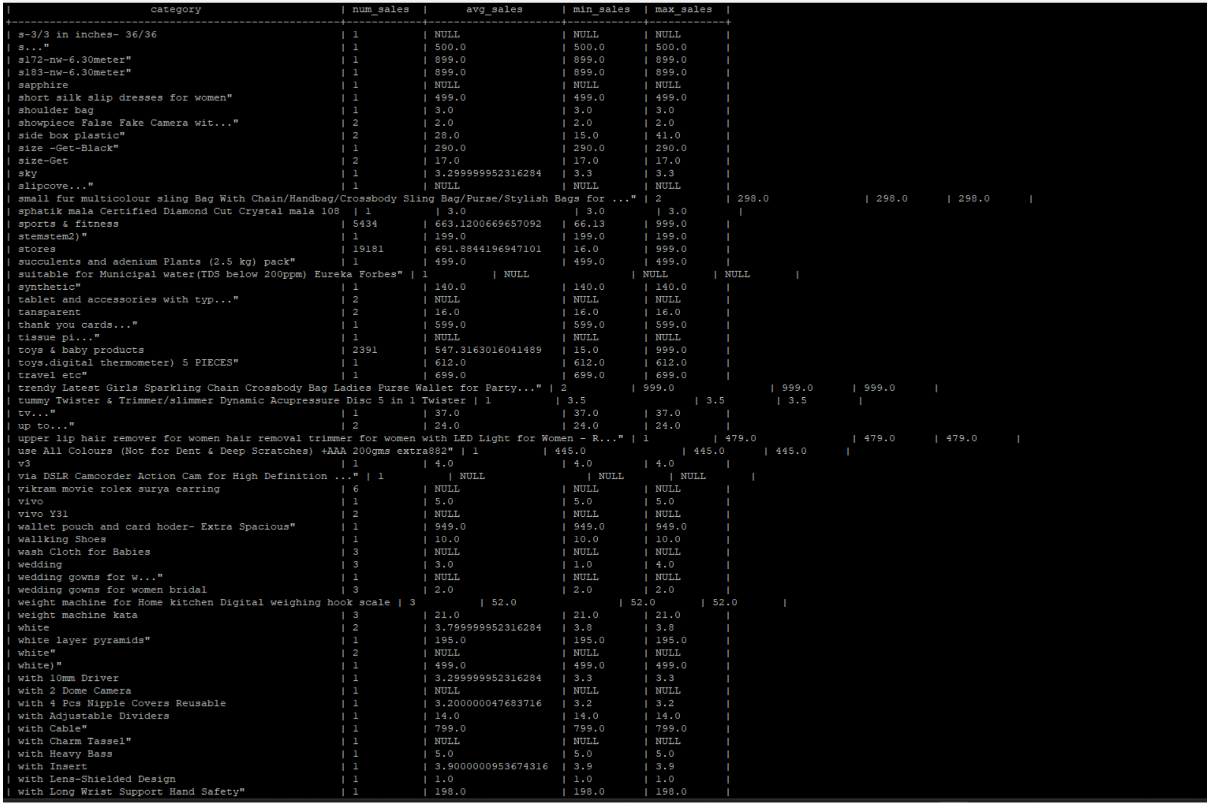
```
SELECT
  category,
  COUNT(*) AS num_sales,
  AVG(sales) AS avg_sales,
  MIN(sales) AS min_sales,
  MAX(sales) AS max_sales
FROM product_sales
GROUP BY category;
```



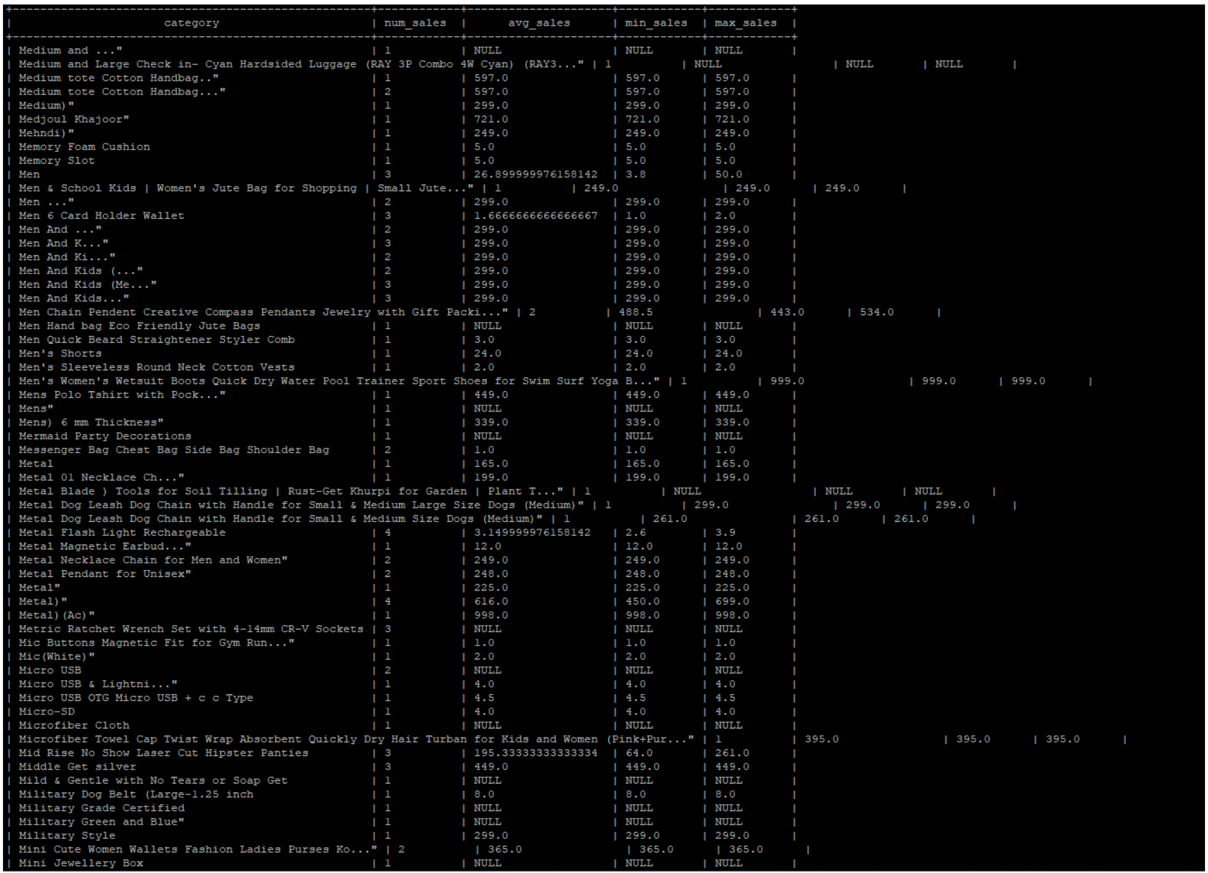
```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> SELECT
. . . . . . . . . . . . . . . . . . . . . . .> main_category AS category,
. . . . . . . . . . . . . . . . . . . . . . .> COUNT(*) AS num_sales,
. . . . . . . . . . . . . . . . . . . . . . .> AVG(actual_price) AS avg_sales,
. . . . . . . . . . . . . . . . . . . . . . .> MIN(actual_price) AS min_sales,
. . . . . . . . . . . . . . . . . . . . . . .> MAX(actual_price) AS max_sales
. . . . . . . . . . . . . . . . . . . . . . .> FROM product_sales
. . . . . . . . . . . . . . . . . . . . . . .> GROUP BY main_category;
INFO  : Compiling command(queryId=hive_20230507090600_eadc45f2-d21a-4f47-8f8c-947515986bbe): SELECT
main_category AS category,
COUNT(*) AS num_sales,
AVG(actual_price) AS avg_sales,
MIN(actual_price) AS min_sales,
MAX(actual_price) AS max_sales
FROM product_sales
GROUP BY main_category
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:category, type:string, comment:null), FieldSchema(name:num_sales, type:bigint, co
in_sales, type:float, comment:null), FieldSchema(name:max_sales, type:float, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230507090600_eadc45f2-d21a-4f47-8f8c-947515986bbe); Time taken: 0.245 seconds
INFO  : Executing command(queryId=hive_20230507090600_eadc45f2-d21a-4f47-8f8c-947515986bbe): SELECT
main_category AS category,
COUNT(*) AS num_sales,
AVG(actual_price) AS avg_sales,
MIN(actual_price) AS min_sales,
MAX(actual_price) AS max_sales
FROM product_sales
GROUP BY main_category
INFO  : Query ID = hive_20230507090600_eadc45f2-d21a-4f47-8f8c-947515986bbe
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230507090600_eadc45f2-d21a-4f47-8f8c-947515986bbe
INFO  : Tez session hasn't been created yet. Opening session
INFO  : Dag name: SELECT
main_category AS cate...main_category (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1683436887119_0003)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1            container       RUNNING      3          0        2        1       0       0
Reducer 2        container        INITED     14          0        0       14       0       0
----------------------------------------------------------------------------------------------
VERTICES: 00/02  [>>--------------------------] 0%    ELAPSED TIME: 12.77 s
----------------------------------------------------------------------------------------------
```



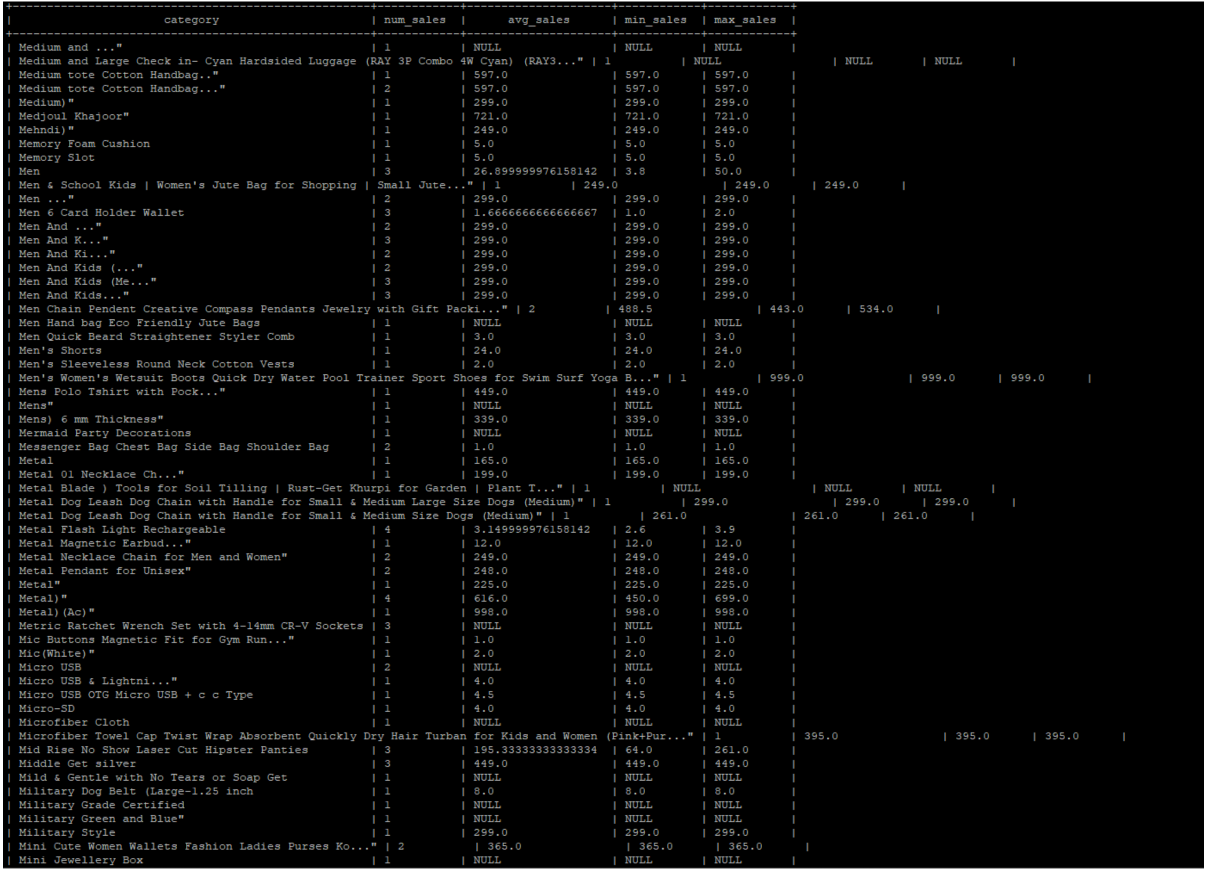
```
+----------------------------------------------------+------------+---------------------+-----------+-----------+
|                      category                      | num_sales  |      avg_sales      | min_sales | max_sales |
+----------------------------------------------------+------------+---------------------+-----------+-----------+
| L                                                  | 18         | 21.650000005960464  | 1.0       | 91.0      |
| L(UK 8-11) Size)"                                  | 2          | 598.0               | 598.0     | 598.0     |
| L) (Pack of 4) (WSCOMRIB4_4002_L)"                 | 1          | 799.0               | 799.0     | 799.0     |
| L) Beige"                                          | 1          | 399.0               | 399.0     | 399.0     |
| L) Black/Beige"                                    | 4          | 367.0               | 335.0     | 399.0     |
| L)"                                                | 11         | 484.4               | 91.0      | 690.0     |
| L-12 Inchs                                         | 1          | 264.0               | 264.0     | 264.0     |
| L-24Inchs                                          | 2          | 56.0                | 56.0      | 56.0      |
| L..."                                              | 6          | NULL                | NULL      | NULL      |
| LAB-11)"                                           | 2          | 499.0               | 499.0     | 499.0     |
| LARGE)"                                            | 2          | 270.0               | 270.0     | 270.0     |
| LCD Led Fla..."                                    | 1          | 64.0                | 64.0      | 64.0      |
| LCD and Plasma TV's"                               | 4          | 17.0                | 13.0      | 21.0      |
| LED                                                | 2          | 3.299999952316284   | 3.3       | 3.3       |
| LED Aircraft Strobe Lights for Motorcycles         | 1          | 29.0                | 29.0      | 29.0      |
| LED Indicator Light Compatible for All Type-C/micro port(without data transf..." | 1          | 31.0                | 31.0      | 31.0      |
| LED Light Compatible with Hero 5 ..."              | 2          | 271.0               | 271.0     | 271.0     |
| LED Light Up On Ear Kids Wireless Headphones with Mic | 1          | NULL                | NULL      | NULL      |
| LED Strip (2 M..."                                 | 1          | 199.0               | 199.0     | 199.0     |
| LED..."                                            | 1          | NULL                | NULL      | NULL      |
| LG                                                 | 1          | NULL                | NULL      | NULL      |
| LID HDS2-6312PRO H.265 HEVC DVB-S2 FullHD Set-Top Box GET Lifetime Get TV Channels (No Monthly Charge)" | 1          | 5.0                 | 5.0       | 5.0       |
| LONG & STRONG 400ML"                               | 1          | 213.0               | 213.0     | 213.0     |
| LS2880)"                                           | 1          | 599.0               | 599.0     | 599.0     |
| LS3880)"                                           | 2          | 599.0               | 599.0     | 599.0     |
| LS3890)"                                           | 1          | 599.0               | 599.0     | 599.0     |
| LS4000)"                                           | 2          | 599.0               | 599.0     | 599.0     |
| LT. GREY MELANG..."                                | 1          | 369.0               | 369.0     | 369.0     |
| LUX adjustmen..."                                  | 2          | NULL                | NULL      | NULL      |
| Lace up Lightweight Shoes Running Shoes for Men's" | 2          | 867.0               | 845.0     | 889.0     |
| Lace up Lightweight Shoes for Running              | 11         | 3.445454554124312   | 1.8       | 3.9       |
| Lace up Lightweight Shoes for Running"             | 3          | 537.0               | 537.0     | 537.0     |
| Lace-up Lightweight Shoes for Running              | 12         | 3.3090908527374268  | 3.0       | 3.8       |
| Lace-up Lightweight with Extra Cushion Shoes for Running | 5          | 3.450000047683716   | 3.4       | 3.5       |
| Lacquer copper bottle 1 litre                      | 1          | 3.5999999046325684  | 3.6       | 3.6       |
| Ladies Innerwear Panties                           | 1          | NULL                | NULL      | NULL      |
| Ladies Panties                                     | 2          | NULL                | NULL      | NULL      |
| Ladies Panties Combo Pack                          | 1          | 3.700000047683716   | 3.7       | 3.7       |
| Ladies Panty                                       | 1          | NULL                | NULL      | NULL      |
| Ladies Printed Pyjama                              | 1          | 2.0                 | 2.0       | 2.0       |
| Ladies Purse                                       | 4          | NULL                | NULL      | NULL      |
| Ladies Wallet                                      | 3          | NULL                | NULL      | NULL      |
| Ladies rhinestones Sparkling Party Handbag Wedding Bag Purse" | 1          | 699.0               | 699.0     | 699.0     |
| Ladies's                                           | 3          | 3.6000000635782876  | 3.4       | 4.0       |
| Lamp For Kids Room Nig..."                         | 1          | 545.0               | 545.0     | 545.0     |
| Lamp for Kids Room Nig..."                         | 1          | 499.0               | 499.0     | 499.0     |
| Lamp for Kids Room Night B..."                     | 1          | 392.0               | 392.0     | 392.0     |
| Langot                                             | 1          | 245.0               | 245.0     | 245.0     |
| Langot For New Born Baby - Pack of 5(Animal Face ..." | 1          | 449.0               | 449.0     | 449.0     |
```



```
|                   category                    | num_sales  |      avg_sales      | min_sales | max_sales |
+-----------------------------------------------+------------+---------------------+-----------+-----------+
| s-3/3 in inches- 36/36                        | 1          | NULL                | NULL      | NULL      |
| s..."                                         | 1          | 500.0               | 500.0     | 500.0     |
| s172-nw-6.30meter"                            | 1          | 899.0               | 899.0     | 899.0     |
| s183-nw-6.30meter"                            | 1          | 899.0               | 899.0     | 899.0     |
| sapphire                                      | 1          | NULL                | NULL      | NULL      |
| short silk slip dresses for women"            | 1          | 499.0               | 499.0     | 499.0     |
| shoulder bag                                  | 1          | 3.0                 | 3.0       | 3.0       |
| showpiece False Fake Camera wit..."           | 2          | 2.0                 | 2.0       | 2.0       |
| side box plastic"                             | 2          | 28.0                | 15.0      | 41.0      |
| size -Get-Black"                              | 1          | 290.0               | 290.0     | 290.0     |
| size-Get                                      | 2          | 17.0                | 17.0      | 17.0      |
| sky                                           | 1          | 3.299999952316284   | 3.3       | 3.3       |
| slipcove..."                                  | 1          | NULL                | NULL      | NULL      |
| small fur multicolour sling Bag With Chain/Handbag/Crossbody Sling Bag/Purse/Stylish Bags for ..." | 2          | 298.0               | 298.0     | 298.0     |
| sphatik mala Certified Diamond Cut Crystal mala 108  | 1          | 3.0                 | 3.0       | 3.0       |
| sports & fitness                              | 5434       | 663.1200669657092   | 66.13     | 999.0     |
| stemstem2)"                                   | 1          | 199.0               | 199.0     | 199.0     |
| stores                                        | 19181      | 691.8844196947101   | 16.0      | 999.0     |
| succulents and adenium Plants (2.5 kg) pack"  | 1          | 499.0               | 499.0     | 499.0     |
| suitable for Municipal water(TDS below 200ppm) Eureka Forbes" | 1          | NULL                | NULL      | NULL      |
| synthetic"                                    | 1          | 140.0               | 140.0     | 140.0     |
| tablet and accessories with typ..."           | 2          | NULL                | NULL      | NULL      |
| tansparent                                    | 2          | 16.0                | 16.0      | 16.0      |
| thank you cards..."                           | 1          | 599.0               | 599.0     | 599.0     |
| tissue pi..."                                 | 1          | NULL                | NULL      | NULL      |
| toys & baby products                          | 2391       | 547.3163016041489   | 15.0      | 999.0     |
| toys.digital thermometer) 5 PIECES"           | 1          | 612.0               | 612.0     | 612.0     |
| travel etc"                                   | 1          | 699.0               | 699.0     | 699.0     |
| trendy Latest Girls Sparkling Chain Crossbody Bag Ladies Purse Wallet for Party..." | 2          | 999.0               | 999.0     | 999.0     |
| tummy Twister & Trimmer/slimmer Dynamic Acupressure Disc 5 in 1 Twister | 1          | 3.5                 | 3.5       | 3.5       |
| tv..."                                        | 1          | 37.0                | 37.0      | 37.0      |
| up to..."                                     | 2          | 24.0                | 24.0      | 24.0      |
| upper lip hair remover for women hair removal trimmer for women with LED Light for Women - R..." | 1          | 479.0               | 479.0     | 479.0     |
| use All Colours (Not for Dent & Deep Scratches) +AAA 200gms extra882" | 1          | 445.0               | 445.0     | 445.0     |
| v3                                            | 1          | 4.0                 | 4.0       | 4.0       |
| via DSLR Camcorder Action Cam for High Definition ..." | 1          | NULL                | NULL      | NULL      |
| vikram movie rolex surya earring              | 6          | NULL                | NULL      | NULL      |
| vivo                                          | 1          | 5.0                 | 5.0       | 5.0       |
| vivo Y31                                      | 2          | NULL                | NULL      | NULL      |
| wallet pouch and card hoder- Extra Spacious"  | 1          | 949.0               | 949.0     | 949.0     |
| wallking Shoes                                | 1          | 10.0                | 10.0      | 10.0      |
| wash Cloth for Babies                         | 3          | NULL                | NULL      | NULL      |
| wedding                                       | 3          | 3.0                 | 1.0       | 4.0       |
| wedding gowns for w..."                       | 1          | NULL                | NULL      | NULL      |
| wedding gowns for women bridal                | 3          | 2.0                 | 2.0       | 2.0       |
| weight machine for Home kitchen Digital weighing hook scale | 3          | 52.0                | 52.0      | 52.0      |
| weight machine kata                           | 3          | 21.0                | 21.0      | 21.0      |
| white                                         | 2          | 3.799999952316284   | 3.8       | 3.8       |
| white layer pyramids"                         | 1          | 195.0               | 195.0     | 195.0     |
| white"                                        | 2          | NULL                | NULL      | NULL      |
| white)"                                       | 1          | 499.0               | 499.0     | 499.0     |
| with 10mm Driver                              | 1          | 3.299999952316284   | 3.3       | 3.3       |
| with 2 Dome Camera                            | 1          | NULL                | NULL      | NULL      |
| with 4 Pcs Nipple Covers Reusable             | 1          | 3.200000047683716   | 3.2       | 3.2       |
| with Adjustable Dividers                      | 1          | 14.0                | 14.0      | 14.0      |
| with Cable"                                   | 1          | 799.0               | 799.0     | 799.0     |
| with Charm Tassel"                            | 1          | NULL                | NULL      | NULL      |
| with Heavy Bass                               | 1          | 5.0                 | 5.0       | 5.0       |
| with Insert                                   | 1          | 3.9000000953674316  | 3.9       | 3.9       |
| with Lens-Shielded Design                     | 1          | 1.0                 | 1.0       | 1.0       |
| with Long Wrist Support Hand Safety"          | 1          | 198.0               | 198.0     | 198.0     |
```



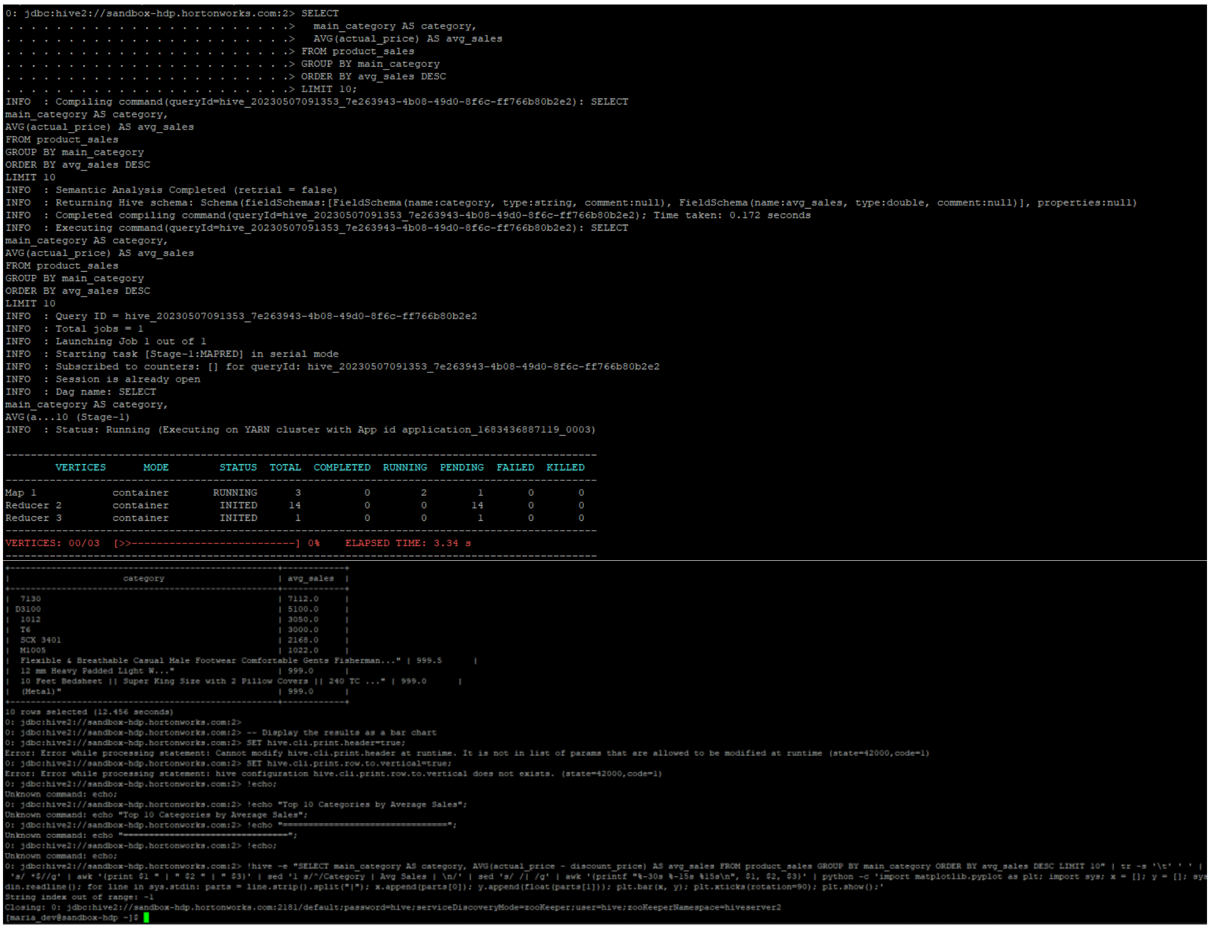
```
+----------------------------------------------------+------------+---------------------+------------+------------+
|                      category                      | num_sales  |      avg_sales      | min_sales  | max_sales  |
+----------------------------------------------------+------------+---------------------+------------+------------+
| Medium and ..."                                    | 1          | NULL                | NULL       | NULL       |
| Medium and Large Check in- Cyan Hardsided Luggage (RAY 3P Combo 4W Cyan) (RAY3..." | 1          | NULL                | NULL       | NULL       |
| Medium tote Cotton Handbag.."                      | 1          | 597.0               | 597.0      | 597.0      |
| Medium tote Cotton Handbag..."                     | 2          | 597.0               | 597.0      | 597.0      |
| Medium)"                                           | 1          | 299.0               | 299.0      | 299.0      |
| Medjoul Khajoor"                                   | 1          | 721.0               | 721.0      | 721.0      |
| Mehndi)"                                           | 1          | 249.0               | 249.0      | 249.0      |
| Memory Foam Cushion                                | 1          | 5.0                 | 5.0        | 5.0        |
| Memory Slot                                        | 1          | 5.0                 | 5.0        | 5.0        |
| Men                                                | 3          | 26.899999976158142  | 3.8        | 50.0       |
| Men & School Kids | Women's Jute Bag for Shopping | Small Jute..." | 1          | 249.0               | 249.0      | 249.0      |
| Men ..."                                           | 2          | 299.0               | 299.0      | 299.0      |
| Men 6 Card Holder Wallet                           | 3          | 1.6666666666666667  | 1.0        | 2.0        |
| Men And ..."                                       | 2          | 299.0               | 299.0      | 299.0      |
| Men And K..."                                      | 3          | 299.0               | 299.0      | 299.0      |
| Men And Ki..."                                     | 2          | 299.0               | 299.0      | 299.0      |
| Men And Kids (..."                                 | 2          | 299.0               | 299.0      | 299.0      |
| Men And Kids (Me..."                               | 3          | 299.0               | 299.0      | 299.0      |
| Men And Kids..."                                   | 3          | 299.0               | 299.0      | 299.0      |
| Men Chain Pendent Creative Compass Pendants Jewelry with Gift Packi..." | 2          | 488.5               | 443.0      | 534.0      |
| Men Hand bag Eco Friendly Jute Bags                | 1          | NULL                | NULL       | NULL       |
| Men Quick Beard Straightener Styler Comb           | 1          | 3.0                 | 3.0        | 3.0        |
| Men's Shorts                                       | 1          | 24.0                | 24.0       | 24.0       |
| Men's Sleeveless Round Neck Cotton Vests           | 1          | 2.0                 | 2.0        | 2.0        |
| Men's Women's Wetsuit Boots Quick Dry Water Pool Trainer Sport Shoes for Swim Surf Yoga B..." | 1          | 999.0               | 999.0      | 999.0      |
| Mens Polo Tshirt with Pock..."                     | 1          | 449.0               | 449.0      | 449.0      |
| Mens"                                              | 1          | NULL                | NULL       | NULL       |
| Mens) 6 mm Thickness"                              | 1          | 339.0               | 339.0      | 339.0      |
| Mermaid Party Decorations                          | 1          | NULL                | NULL       | NULL       |
| Messenger Bag Chest Bag Side Bag Shoulder Bag      | 2          | 1.0                 | 1.0        | 1.0        |
| Metal                                              | 1          | 165.0               | 165.0      | 165.0      |
| Metal 01 Necklace Ch..."                           | 1          | 199.0               | 199.0      | 199.0      |
| Metal Blade ) Tools for Soil Tilling | Rust-Get Khurpi for Garden | Plant T..." | 1          | NULL                | NULL       | NULL       |
| Metal Dog Leash Dog Chain with Handle for Small & Medium Large Size Dogs (Medium)" | 1          | 299.0               | 299.0      | 299.0      |
| Metal Dog Leash Dog Chain with Handle for Small & Medium Size Dogs (Medium)" | 1          | 261.0               | 261.0      | 261.0      |
| Metal Flash Light Rechargeable                     | 4          | 3.149999976158142   | 2.6        | 3.9        |
| Metal Magnetic Earbud..."                          | 1          | 12.0                | 12.0       | 12.0       |
| Metal Necklace Chain for Men and Women"            | 2          | 249.0               | 249.0      | 249.0      |
| Metal Pendant for Unisex"                          | 2          | 248.0               | 248.0      | 248.0      |
| Metal"                                             | 1          | 225.0               | 225.0      | 225.0      |
| Metal)"                                            | 4          | 616.0               | 450.0      | 699.0      |
| Metal)(Ac)"                                        | 1          | 998.0               | 998.0      | 998.0      |
| Metric Ratchet Wrench Set with 4-14mm CR-V Sockets | 3          | NULL                | NULL       | NULL       |
| Mic Buttons Magnetic Fit for Gym Run..."           | 1          | 1.0                 | 1.0        | 1.0        |
| Mic(White)"                                        | 1          | 2.0                 | 2.0        | 2.0        |
| Micro USB                                          | 2          | NULL                | NULL       | NULL       |
| Micro USB & Lightni..."                            | 1          | 4.0                 | 4.0        | 4.0        |
| Micro USB OTG Micro USB + c c Type                 | 1          | 4.5                 | 4.5        | 4.5        |
| Micro-SD                                           | 1          | 4.0                 | 4.0        | 4.0        |
| Microfiber Cloth                                   | 1          | NULL                | NULL       | NULL       |
| Microfiber Towel Cap Twist Wrap Absorbent Quickly Dry Hair Turban for Kids and Women (Pink+Pur..." | 1          | 395.0               | 395.0      | 395.0      |
| Mid Rise No Show Laser Cut Hipster Panties         | 3          | 195.33333333333334  | 64.0       | 261.0      |
| Middle Get silver                                  | 3          | 449.0               | 449.0      | 449.0      |
| Mild & Gentle with No Tears or Soap Get            | 1          | NULL                | NULL       | NULL       |
| Military Dog Belt (Large-1.25 inch                 | 1          | 8.0                 | 8.0        | 8.0        |
| Military Grade Certified                           | 1          | NULL                | NULL       | NULL       |
| Military Green and Blue"                           | 1          | NULL                | NULL       | NULL       |
| Military Style                                     | 1          | 299.0               | 299.0      | 299.0      |
| Mini Cute Women Wallets Fashion Ladies Purses Ko..." | 2          | 365.0               | 365.0      | 365.0      |
| Mini Jewellery Box                                 | 1          | NULL                | NULL       | NULL       |
```



```
+----------------------------------------------------+------------+---------------------+------------+------------+
|                      category                      | num_sales  |      avg_sales      | min_sales  | max_sales  |
+----------------------------------------------------+------------+---------------------+------------+------------+
| Medium and ..."                                    | 1          | NULL                | NULL       | NULL       |
| Medium and Large Check in- Cyan Hardsided Luggage (RAY 3P Combo 4W Cyan) (RAY3..." | 1          | NULL                | NULL       | NULL       |
| Medium tote Cotton Handbag.."                      | 1          | 597.0               | 597.0      | 597.0      |
| Medium tote Cotton Handbag..."                     | 2          | 597.0               | 597.0      | 597.0      |
| Medium)"                                           | 1          | 299.0               | 299.0      | 299.0      |
| Medjoul Khajoor"                                   | 1          | 721.0               | 721.0      | 721.0      |
| Mehndi)"                                           | 1          | 249.0               | 249.0      | 249.0      |
| Memory Foam Cushion                                | 1          | 5.0                 | 5.0        | 5.0        |
| Memory Slot                                        | 1          | 5.0                 | 5.0        | 5.0        |
| Men                                                | 3          | 26.899999976158142  | 3.8        | 50.0       |
| Men & School Kids | Women's Jute Bag for Shopping | Small Jute..." | 1          | 249.0               | 249.0      | 249.0      |
| Men ..."                                           | 2          | 299.0               | 299.0      | 299.0      |
| Men 6 Card Holder Wallet                           | 3          | 1.6666666666666667  | 1.0        | 2.0        |
| Men And ..."                                       | 2          | 299.0               | 299.0      | 299.0      |
| Men And K..."                                      | 3          | 299.0               | 299.0      | 299.0      |
| Men And Ki..."                                     | 2          | 299.0               | 299.0      | 299.0      |
| Men And Kids (..."                                 | 2          | 299.0               | 299.0      | 299.0      |
| Men And Kids (Me..."                               | 3          | 299.0               | 299.0      | 299.0      |
| Men And Kids..."                                   | 3          | 299.0               | 299.0      | 299.0      |
| Men Chain Pendent Creative Compass Pendants Jewelry with Gift Packi..." | 2          | 488.5               | 443.0      | 534.0      |
| Men Hand bag Eco Friendly Jute Bags                | 1          | NULL                | NULL       | NULL       |
| Men Quick Beard Straightener Styler Comb           | 1          | 3.0                 | 3.0        | 3.0        |
| Men's Shorts                                       | 1          | 24.0                | 24.0       | 24.0       |
| Men's Sleeveless Round Neck Cotton Vests           | 1          | 2.0                 | 2.0        | 2.0        |
| Men's Women's Wetsuit Boots Quick Dry Water Pool Trainer Sport Shoes for Swim Surf Yoga B..." | 1          | 999.0               | 999.0      | 999.0      |
| Mens Polo Tshirt with Pock..."                     | 1          | 449.0               | 449.0      | 449.0      |
| Mens"                                              | 1          | NULL                | NULL       | NULL       |
| Mens) 6 mm Thickness"                              | 1          | 339.0               | 339.0      | 339.0      |
| Mermaid Party Decorations                          | 1          | NULL                | NULL       | NULL       |
| Messenger Bag Chest Bag Side Bag Shoulder Bag      | 2          | 1.0                 | 1.0        | 1.0        |
| Metal                                              | 1          | 165.0               | 165.0      | 165.0      |
| Metal 01 Necklace Ch..."                           | 1          | 199.0               | 199.0      | 199.0      |
| Metal Blade ) Tools for Soil Tilling | Rust-Get Khurpi for Garden | Plant T..." | 1          | NULL                | NULL       | NULL       |
| Metal Dog Leash Dog Chain with Handle for Small & Medium Large Size Dogs (Medium)" | 1          | 299.0               | 299.0      | 299.0      |
| Metal Dog Leash Dog Chain with Handle for Small & Medium Size Dogs (Medium)" | 1          | 261.0               | 261.0      | 261.0      |
| Metal Flash Light Rechargeable                     | 4          | 3.149999976158142   | 2.6        | 3.9        |
| Metal Magnetic Earbud..."                          | 1          | 12.0                | 12.0       | 12.0       |
| Metal Necklace Chain for Men and Women"            | 2          | 249.0               | 249.0      | 249.0      |
| Metal Pendant for Unisex"                          | 2          | 248.0               | 248.0      | 248.0      |
| Metal"                                             | 1          | 225.0               | 225.0      | 225.0      |
| Metal)"                                            | 4          | 616.0               | 450.0      | 699.0      |
| Metal)(Ac)"                                        | 1          | 998.0               | 998.0      | 998.0      |
| Metric Ratchet Wrench Set with 4-14mm CR-V Sockets | 3          | NULL                | NULL       | NULL       |
| Mic Buttons Magnetic Fit for Gym Run..."           | 1          | 1.0                 | 1.0        | 1.0        |
| Mic(White)"                                        | 1          | 2.0                 | 2.0        | 2.0        |
| Micro USB                                          | 2          | NULL                | NULL       | NULL       |
| Micro USB & Lightni..."                            | 1          | 4.0                 | 4.0        | 4.0        |
| Micro USB OTG Micro USB + c c Type                 | 1          | 4.5                 | 4.5        | 4.5        |
| Micro-SD                                           | 1          | 4.0                 | 4.0        | 4.0        |
| Microfiber Cloth                                   | 1          | NULL                | NULL       | NULL       |
| Microfiber Towel Cap Twist Wrap Absorbent Quickly Dry Hair Turban for Kids and Women (Pink+Pur..." | 1          | 395.0               | 395.0      | 395.0      |
| Mid Rise No Show Laser Cut Hipster Panties         | 3          | 195.33333333333334  | 64.0       | 261.0      |
| Middle Get silver                                  | 3          | 449.0               | 449.0      | 449.0      |
| Mild & Gentle with No Tears or Soap Get            | 1          | NULL                | NULL       | NULL       |
| Military Dog Belt (Large-1.25 inch                 | 1          | 8.0                 | 8.0        | 8.0        |
| Military Grade Certified                           | 1          | NULL                | NULL       | NULL       |
| Military Green and Blue"                           | 1          | NULL                | NULL       | NULL       |
| Military Style                                     | 1          | 299.0               | 299.0      | 299.0      |
| Mini Cute Women Wallets Fashion Ladies Purses Ko..." | 2          | 365.0               | 365.0      | 365.0      |
| Mini Jewellery Box                                 | 1          | NULL                | NULL       | NULL       |
```

Calculating Top 10 selling Categories

SELECT
  main_category AS category,
  AVG(actual_price) AS avg_sales
FROM product_sales
GROUP BY main_category
ORDER BY avg_sales DESC
LIMIT 10;

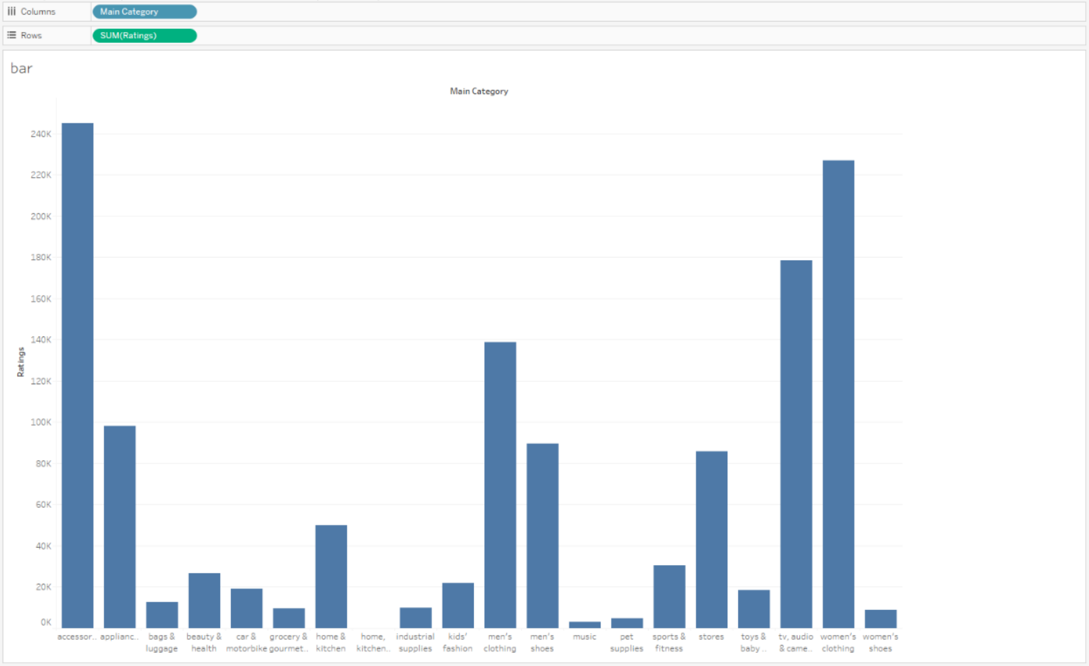


```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> SELECT
. . . . . . . . . . . . . . . . . . . . . . . .>   main_category AS category,
. . . . . . . . . . . . . . . . . . . . . . . .>   AVG(actual_price) AS avg_sales
. . . . . . . . . . . . . . . . . . . . . . . .> FROM product_sales
. . . . . . . . . . . . . . . . . . . . . . . .> GROUP BY main_category
. . . . . . . . . . . . . . . . . . . . . . . .> ORDER BY avg_sales DESC
. . . . . . . . . . . . . . . . . . . . . . . .> LIMIT 10;
INFO  : Compiling command(queryId=hive_20230507091353_7e263943-4b08-49d0-8f6c-ff766b80b2e2): SELECT
main_category AS category,
AVG(actual_price) AS avg_sales
FROM product_sales
GROUP BY main_category
ORDER BY avg_sales DESC
LIMIT 10
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:category, type:string, comment:null), FieldSchema(name:avg_sales, type:double, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230507091353_7e263943-4b08-49d0-8f6c-ff766b80b2e2); Time taken: 0.172 seconds
INFO  : Executing command(queryId=hive_20230507091353_7e263943-4b08-49d0-8f6c-ff766b80b2e2): SELECT
main_category AS category,
AVG(actual_price) AS avg_sales
FROM product_sales
GROUP BY main_category
ORDER BY avg_sales DESC
LIMIT 10
INFO  : Query ID = hive_20230507091353_7e263943-4b08-49d0-8f6c-ff766b80b2e2
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Subscribed to counters: [] for queryId: hive_20230507091353_7e263943-4b08-49d0-8f6c-ff766b80b2e2
INFO  : Session is already open
INFO  : Dag name: SELECT
main_category AS category,
AVG(a...10 (Stage-1)
INFO  : Status: Running (Executing on YARN cluster with App id application_1683436887119_0003)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1            container       RUNNING      3          0        2        1       0       0
Reducer 2        container        INITED     14          0        0       14       0       0
Reducer 3        container        INITED      1          0        0        1       0       0
----------------------------------------------------------------------------------------------
VERTICES: 00/03  [>>--------------------------] 0%    ELAPSED TIME: 3.34 s
----------------------------------------------------------------------------------------------
+----------------------------------------------------+------------+
|                      category                      | avg_sales  |
+----------------------------------------------------+------------+
|  7130                                              | 7112.0     |
| D3100                                              | 5100.0     |
|  1012                                              | 3050.0     |
|  T6                                                | 3000.0     |
|  SCX 3401                                          | 2168.0     |
|  M1005                                             | 1022.0     |
|  Flexible & Breathable Casual Male Footwear Comfortable Gents Fisherman..." | 999.5      |
|  12 mm Heavy Padded Light W..."                    | 999.0      |
|  10 Feet Bedsheet || Super King Size with 2 Pillow Covers || 240 TC ..." | 999.0      |
|  (Metal)"                                          | 999.0      |
+----------------------------------------------------+------------+
10 rows selected (12.456 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> -- Display the results as a bar chart
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> SET hive.cli.print.header=true;
Error: Error while processing statement: Cannot modify hive.cli.print.header at runtime. It is not in list of params that are allowed to be modified at runtime (state=42000,code=1)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> SET hive.cli.print.row.to.vertical=true;
Error: Error while processing statement: hive configuration hive.cli.print.row.to.vertical does not exists. (state=42000,code=1)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> !echo;
Unknown command: echo;
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> !echo "Top 10 Categories by Average Sales";
Unknown command: echo "Top 10 Categories by Average Sales";
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> !echo "================================";
Unknown command: echo "================================";
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> !echo;
Unknown command: echo;
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> !hive -e "SELECT main_category AS category, AVG(actual_price - discount_price) AS avg_sales FROM product_sales GROUP BY main_category ORDER BY avg_sales DESC LIMIT 10" | tr -s '\t' ' ' |
 's/ *$//g' | awk '{print $1 " | " $2 " | " $3}' | sed '1 s/^/Category | Avg Sales | \n/' | sed 's/ /| /g' | awk '{printf "%-30s %-15s %15s\n", $1, $2, $3}' | python -c 'import matplotlib.pyplot as plt; import sys; x = []; y = []; sys
din.readline(); for line in sys.stdin: parts = line.strip().split("|"); x.append(parts[0]); y.append(float(parts[1])); plt.bar(x, y); plt.xticks(rotation=90); plt.show();'
String index out of range: -1
Closing: 0: jdbc:hive2://sandbox-hdp.hortonworks.com:2181/default;password=hive;serviceDiscoveryMode=zooKeeper;user=hive;zooKeeperNamespace=hiveserver2
[maria_dev@sandbox-hdp ~]$
```

CHAPTER 5
VISUALIZATION

ALL GENRES SORTED ACCORDING TO THE SUM OF THE ENERGY:

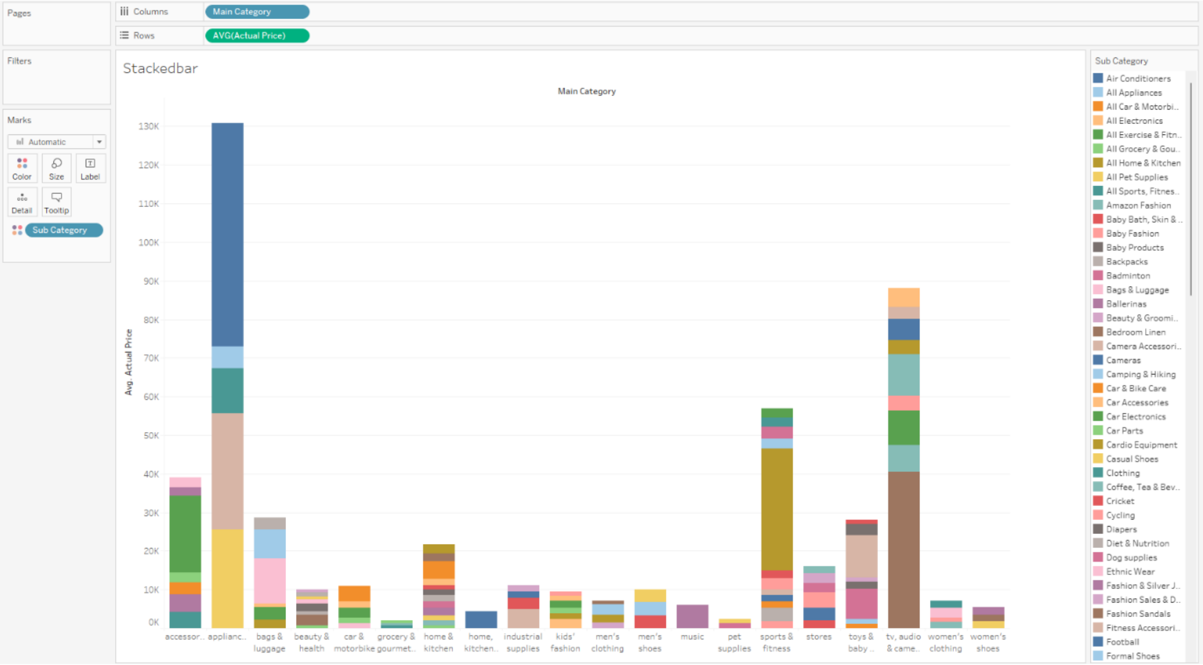
Bar Chart



Columns: Main Category

Rows: SUM(Ratings)

Stacked Bar Chart

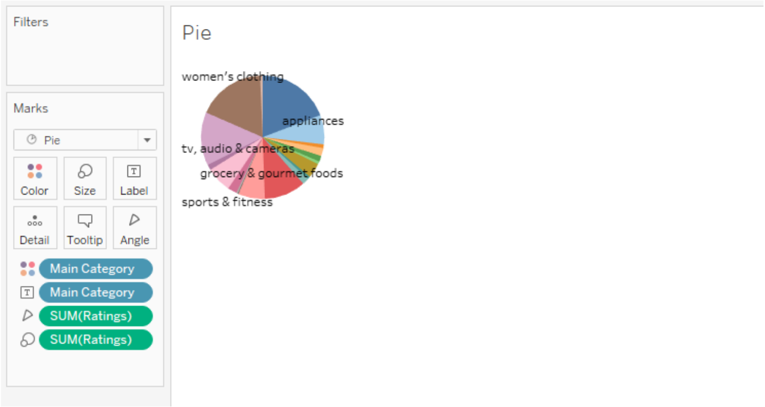


Columns: Main Category

Rows: AVG(Actual Price)

Marks: Automatic; Sub Category (Color)

PIE Chart



Filters

Marks

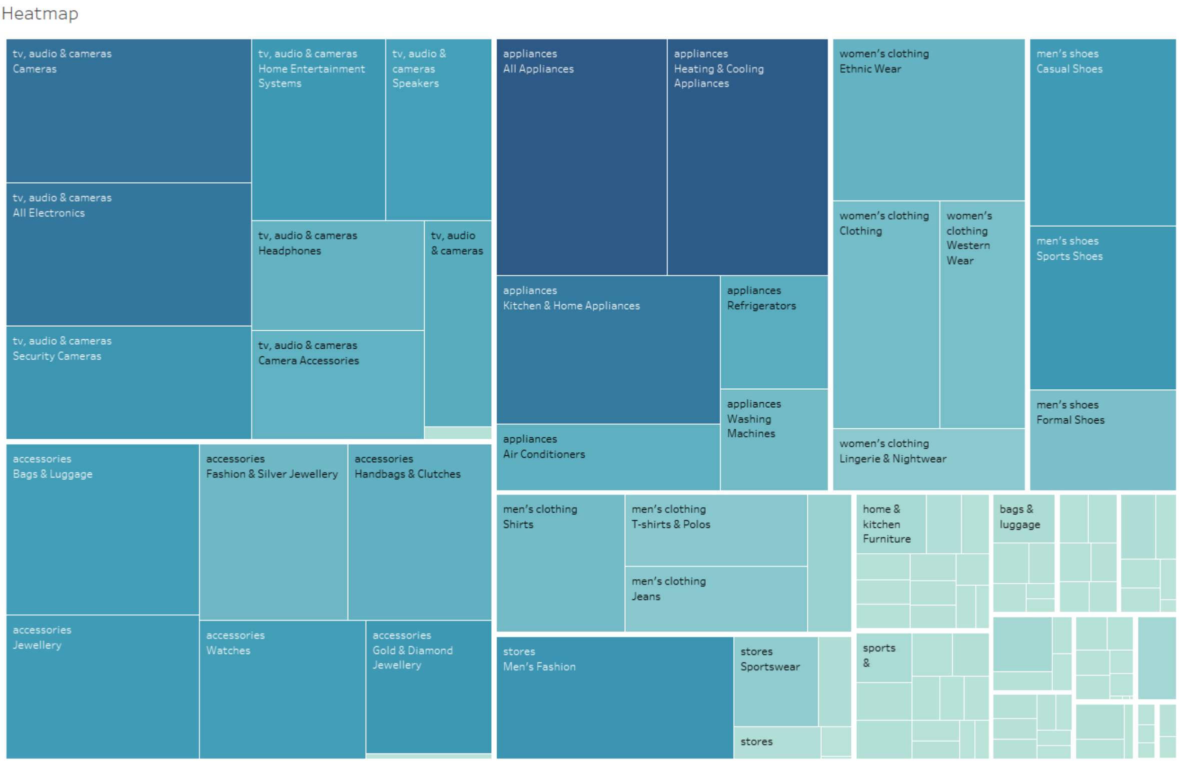
Pie

Color Size Label

Detail Tooltip Angle

Main Category
Main Category
SUM(Ratings)
SUM(Ratings)

Heatmap



CHAPTER 7
CONCLUSION

In conclusion, the analysis of Amazon product sales data using Hive and Tableau has provided valuable insights into customer behavior and product trends. The use of Hive has allowed us to process and analyze large volumes of data in a distributed manner, while the use of Tableau has enabled us to create interactive visualizations and dashboards for better data exploration and communication. Through this analysis, we have identified the top-selling products in each category, the average rating and number of ratings for each product, as well as the discount and actual prices of each product. We have also identified the most popular product categories and sub-categories, as well as the average rating and number of ratings for each category. The insights gained from this analysis can help businesses to make data-driven decisions in areas such as product pricing, marketing, and inventory management. For example, businesses can use this information to optimize their product offerings, improve customer satisfaction, and increase sales.

Overall, the use of Hive and Tableau for analyzing and visualizing Amazon product sales data has provided a powerful tool for businesses to gain insights into customer behavior and product trends, and make data-driven decisions to improve their bottom line.

REFERENCES

1. "Analyzing Amazon Sales Data with Hive and Tableau" by Thomas Kachura, published in the Journal of Big Data. (2021)

2. "Big Data Analytics using Hadoop, Hive and Tableau for Sales Data of an E-commerce Company" by Ramya M, published in the International Journal of Advanced Research in Computer Science and Software Engineering. (2021)

3. "Analysis of Amazon Product Reviews Using Hadoop and Tableau" by Sai Vamsi Maddipati and Anil Kumar Pasupuleti, published in the International Journal of Emerging Trends & Technology in Computer Science. (2020)

4. "Exploratory Data Analysis of Amazon Product Reviews Using Hive and Tableau" by Vinay Kumar and Praveen Kumar, published in the International Journal of Computer Science and Mobile Computing. (2020)

5. "Big Data Analytics for E-commerce Sales Data using Hive and Tableau" by Sanjay Bhat, published in the International Journal of Advanced Trends in Computer Science and Engineering. (2019)