

Dipankar Majumdar
Student no: 11016245

Assignment - Data Transform Activity (20%)

Due: Module 4

Purpose:

The purpose of this assignment is to transform data into an appropriate format for analysis and then extract that data for further problem solving.

Assignment Objective:

For the business problem you selected, you will extract data for further analysis using a series of transformations, including: select, mutate, filter, and arrange.

Instructions:

This activity relates to the Video Games Sales Data:

Click on the link below to download your data:

- [Video Games Sales - Data](#)

In this assignment, you will perform simple transformations on your data (you might need to reload the data in HDInsight and Hive). Do the following steps:

- 1) a) Create a new table for your analysis call "sales_genre".
b) Load the table "sales" into this table.
c) Select these columns: Genre, Global_Sales, Critic_Score.
- 2) Round the data found in the "Global_Sales" column. (HINT: the SQL function to round a number is ROUND(obs)
- 3) Filter the data to only look at those items in the "Critic_Score" column that are greater than 0.
- 4) Order the data in the "Critic_Score" column from highest to lowest. (HINT: Use the DESC query)

This assignment relates to the following Course Learning Requirements:

CLR 3: Recommend which tools and the best use of these tools, to support data analysis, data quality, problem solving, analytics and business decision machining.

CLR 4: Extract and load data to support problem solving and decision-making.

Grading for This Assignment

You will be completing this assignment using the simulation provided.

IMPORTANT: You are free to leave the simulation at any point during the activity and your grade will be calculated based on where you stop. When you leave the simulation, an 8-character code will be generated in the following format: "DA-XX-XXXX".

Be sure to copy this code and submit it to your instructor using the Learning Management System (LMS). The code will confirm where you ended the simulation and the grade earned.

As you complete your tasks, you will notice that your grade is generated right inside the simulation as you progress through the required steps as follows:

Steps required	Grade Earned
1. a) Create a new table for your analysis call "sales_genre". b) Load the table "sales" into this table. c) Select these columns: Genre, Global_Sales, Critic_Score.	50%
2. Round the data found in the "Global_Sales" column. (HINT: the SQL function to round a number is ROUND(obs))	
3. Filter the data to only look at those items in the "Critic_Score" column that are greater than 0.	80%
4. Order the data in the "Critic_Score" column from highest to lowest. (HINT: Use the DESC query)	100%

Q 1a)

```
1 create table sales_genre
2 (Genre string, Global_Sales double, Critic_Score int);
```

Q 1b)&c)

```
1 insert overwrite table sales_genre
2 select Genre, Global_Sales, Critic_Score
3 from sales;
```

Q 2

```
1 select round(global_sales) as RoundedData from sales_genre;
```

83.0

40.0

36.0

33.0

31.0

30.0

30.0

29.0

28.0

28.0

25.0

23.0

23.0

23.0

22.0

Q3

```
1 select * from sales_genre where critic_score > 0;
```

sales_genre.genre	sales_genre.global_sales	sales_genre.critic_score
Sports	82.53	76
Racing	35.52	82
Sports	32.77	80
Platform	29.8	89
Misc	28.92	58
Platform	28.32	87

Q4

```
1 select * from sales_genre order by critic_score desc;
```

sales_genre.genre	sales_genre.global_sales	sales_genre.critic_score
Action	11.01	98
Sports	4.68	98
Fighting	0.34	98
Action	10.5	98
Action	11.01	98
Sports	4.68	98
Fighting	0.34	98
Action	10.5	98
Action	16.27	97
Sports	1.00	97