

Subject: Data Analysis Tools for Analytics – Assignment 2

Instructor: Dr. Kenton White

Name: Priyanka Prasad

Student Id: c0888032

Step 1a: Create a new table for your analysis call “sales_genre”.

The screenshot displays the Hive web interface. At the top, the 'HIVE' logo is on the left, and navigation tabs for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS are on the right. Below the tabs, a 'Worksheet1 *' button with a plus icon is visible. The main area is divided into a 'DATABASE' section on the left, which says 'Select or search database/schema', and a central input field on the right containing 'x default'. Below this, a SQL query is entered in a text area, with line numbers 1 through 14 on the left. The query is:

```
1 CREATE EXTERNAL TABLE sales_genre
2
3 (
4   Name string,
5   Platform string,
6   Year_of_Release int,
7   Genre string,
8   Publisher string,
9   NA_Sales float,
10  EU_Sales float,
11  JP_Sales float,
12  Other_Sales float,
13  Global_Sales float,
14  Critic_Score int,
```

 Below the query area, there are four buttons: 'Execute' (green with a checkmark), 'Save As' (with a floppy disk icon), 'Insert UDF' (with a dropdown arrow), and 'Visual Explain' (with a magnifying glass icon). At the bottom, there is a row of four buttons: 'RESULTS' (dark background with a document icon), 'LOG' (with a list icon), 'VISUAL EXPLAIN' (with a magnifying glass icon), and 'TEZ UI' (with a lightning bolt icon).

```
1 CREATE EXTERNAL TABLE sales_genre
2
3 (
4   Name string,
5   Platform string,
6   Year_of_Release int,
7   Genre string,
8   Publisher string,
9   NA_Sales float,
10  EU_Sales float,
11  JP_Sales float,
12  Other_Sales float,
13  Global_Sales float,
14  Critic_Score int,
```

I created the new table “sales_genre” in the schema.

Step 1b: Load the table “sales” into this table.

✓ Execute

Save As

Insert UDF ▾

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI

Filter columns

×

≡

←

→

↗

sales_genre.name	sales_genre.platform	sales_genre.year_of_release	sales_genre.genre	sales_genre.publisher	sales_genre.na_sales
Wii Sports	Wii	2006	Sports	Nintendo	41.36
Super Mario Bros.	NES	1985	Platform	Nintendo	29.08
Mario Kart Wii	Wii	2008	Racing	Nintendo	15.68
Wii Sports Resort	Wii	2009	Sports	Nintendo	15.61
Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27
Tetris	GB	1989	Puzzle	Nintendo	23.2
New Super Mario Bros.	DS	2006	Platform	Nintendo	11.28
Wii Play	Wii	2006	Misc	Nintendo	13.96
New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.44
Duck Hunt	NES	1984	Shooter	Nintendo	26.93
Nintendogs	DS	2005	Simulation	Nintendo	9.05
Mario Kart DS	DS	2005	Racing	Nintendo	9.71
Pokemon	GB	1999	Role-Playing	Nintendo	9.0

Above screenshot shows the sales file loaded in the table.

Step 1c: Select these columns: Genre, Global_Sales, Critic_Score.

```
1 SELECT Genre, Global_Sales, Critic_Score FROM sales_genre;
```

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns ✕

genre	global_sales	critic_score
Sports	82.53	76
Platform	40.24	null
Racing	35.52	82
Sports	32.77	80
Role-Playing	31.37	null
Puzzle	30.26	null
Platform	29.8	89

In the above screenshot only three variables among all 16 variables been selected.

Step 2: Round the data found in the “Global_Sales” column.

```
1 SELECT Genre, ROUND(Global_Sales), Critic_Score FROM sales_genre;
```

✓ Execute

Save As

Insert UDF ▾

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI

Filter columns ✕

genre	_c1	critic_score
Sports	83.0	76
Platform	40.0	null
Racing	36.0	82
Sports	33.0	80
Role-Playing	31.0	null
Puzzle	30.0	null
Platform	30.0	89

Data from Global_Sales column is been rounded using ROUND function i.e. 82.53 is rounded to 83 and so on.

Step 3: Filter the data to only look at those items in the Critic_Score” column that are greater than 0.

```
1 SELECT Genre, Global_Sales, Critic_Score FROM sales_genre
2 WHERE Critic_Score > 0;
```

Execute

Save As

Insert UDF ▾

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI

Filter columns

genre	global_sales	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
Racing	23.21	91

Step 4: Order the data in the “Critic_Score” column from highest tom lowest.

```
1 SELECT Genre, Global_Sales, Critic_Score FROM sales_genre
2 ORDER BY Critic_Score DESC;
```

✓ Execute Save As Insert UDF ▼ Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns ✕

genre	global_sales	critic_score
Sports	4.68	98
Fighting	0.34	98
Action	11.01	98
Action	10.5	98
Action	12.61	97
Platform	7.51	97
Action	13.1	97

Above figure shows that the Critic_Score column is arranged in descending order i.e. from highest to lowest.