

Dipankar Majumdar  
Student no : 11016245

## `Assignment - Data Aggregation Activity (10%)

**Due:** Module 5

### **Purpose:**

The purpose of this assignment is to aggregate data in a format for analysis and extract those data for further problem solving.

### **Assignment Objective:**

For the business problem you selected, perform the required steps to aggregate data for visualization using a series of summarizations, including: Count and Average.

### **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 aggregate the data for extracting insights (you might need to reload the data in HDInsight and Hive). Do the following steps:

- 1) Create 2 tables. The first table shows the average critic score for Sports games. Store this information in a column called "sports\_critic\_score". The second table shows the average critic score for Shooter games. Store this information in a column called "shooter\_critic\_score".
- 2) Create 3 statistics tables (average, min, max) for the global\_sales for: all games, Sports games, Shooter games. For each table, label the columns as: "average\_global\_sales", "min\_global\_sales", and "max\_global\_sales".
- 3) Create 2 tables containing the global\_sales and the count of games with that global\_sales for: Sports games and Shooter games. For each table, label the columns as: "global\_sales" and "count".

This assignment relates to the following Course Learning Requirements:

CLR 2: Illustrate how current tools are used to address these problems, including the strength and weaknesses of each tool

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

## Grading for This Assignment

Please take a screen shot of each step, showing the result. Also provide the query you used.

Steps required	Grade Earned
1) Create two tables (one for sports games and one for shooter games) displaying the average critic score	50%
2) Create three statistics tables for the global sales; one for all games, one for sports games and one for shooter games	80%
3) Create two tables (one for sports games and one for shooter games) for global sales and the number of games matching that global sales amount	100%

```
1 select AVG(Critic_Score) as sports_critic_score from video_games_sales where Genre = 'Sports';
```

✓ Execute

Save As

Insert UDF ▾

Visual Explain

RESULTS

LOG

VISUAL EXPLAIN

TEZ UI

Filter columns

×

≡

←

→

↗

sports\_critic\_score

71.96817420435511

```
1 select AVG(Critic_Score) as sports_critic_score from video_games_sales where Genre = 'Shooter';
```

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

sports\_critic\_score

70.18114406779661

```
1 select AVG(Global_Sales) as average_global_sales, min(Global_Sales) as min_global_sales, max(Global_Sales) as max_global_sales from video_games_sales
```

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

average\_global\_sales min\_global\_sales max\_global\_sales

0.5335426759974415 0.01 82.53

```
1 select AVG(Global_Sales) as average_global_sales, min(Global_Sales) as min_global_sales, max(Global_Sales) as max_global_sales from video_games_sales where Genre ='Sports'
```

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

x

≡

average_global_sales	min_global_sales	max_global_sales
0.5672913117546791	0.01	82.53

```
1 select AVG(Global_Sales) as average_global_sales, min(Global_Sales) as min_global_sales, max(Global_Sales) as max_global_sales from video_games_sales where Genre ='Shooter'
```

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns

x

≡

average_global_sales	min_global_sales	max_global_sales
0.7958730158729982	0.01	28.31

```
1 select Global_Sales,COUNT(Global_Sales) as count from video_games_sales where Genre='Sports' group by Global_Sales;
```

global_sales	count
0.01	216
0.02	332
0.03	296
0.04	264
0.05	304
0.06	276
0.07	216
0.08	256
0.09	232
0.1	216
0.11	232
0.12	208
0.13	216
0.14	204
0.15	172
0.16	148
0.17	188

---

```
1 select Global_Sales,COUNT(Global_Sales) as count from video_games_sales where Genre='Shooter' group by Global_Sales;
```

0.01	120
0.02	328
0.03	180
0.04	180
0.05	188
0.06	176
0.07	124
0.08	156
0.09	112
0.1	108
0.11	148
0.12	92
0.13	96
0.14	88
0.15	80
0.16	64
0.17	68
0.18	72
0.19	80