**Data Transformation in SQL**

Prepared by:

Name: **Nithin Thomas**

# **CONTENTS**

[1. Assignment Instructions 1](#_Toc84579226)

[2. Answers 2](#_Toc84579227)

[2.1 Provide an example in SQL of how to answer the follow questions: 2](#_Toc84579228)

[2.2 Steps followed 4](#_Toc84579229)

[2.3 Challenges 5](#_Toc84579230)

[3. REFERENCE 6](#_Toc84579231)

# **1. Assignment Instructions**

Using the video game sales in 2016 database that we've been working with in class:

1. Provide an example in SQL of how to answer the follow questions:
2. Was the average of global sales higher before or after 2010 ?
3. Create a new column that labels records before 2010 as 'pre-2010' and after 2010 as 'post-2010'

II. In a short 1 page report, describe how you thought through these 2 questions and any challenges you had to work through.

## **2. Answers**

# **2.1 Provide an example in SQL of how to answer the follow questions:**

1. Was the average of global sales higher before or after 2010?
2. Create a new column that labels records before 2010 as 'pre-2010' and after 2010 as 'post-2010'

**2.1.1 MySQL Code**

-- Data is imported into “videogames” database in “vgsales\_2016” table.

-- Creating " Pre\_or\_POST\_2010" column that labels records before 2010 as 'PRE-2010' and after 2010 as 'POST-2010'.

**SELECT**

\*,

**CASE**

**WHEN** Year\_of\_Release >= 2010 **THEN** 'POST-2010'

**WHEN** Year\_of\_Release < 2010 **THEN** 'PRE-2010'

**END** **AS** Pre\_or\_POST\_2010

**FROM**

videosgames.vgsales\_2016;

-- To feed "POST-2010" or "PRE-2010" into “Pre\_or\_POST\_2010” column based on Year\_of\_Release and to find out average of POST-2010 and Pre-2010 global sales by GROUP function.

**SELECT**

**AVG**(Global\_Sales) **AS** Average\_Global\_Sales,

**CASE**

**WHEN** Year\_of\_Release >= 2010 **THEN** 'POST-2010'

**WHEN** Year\_of\_Release < 2010 **THEN** 'PRE-2010'

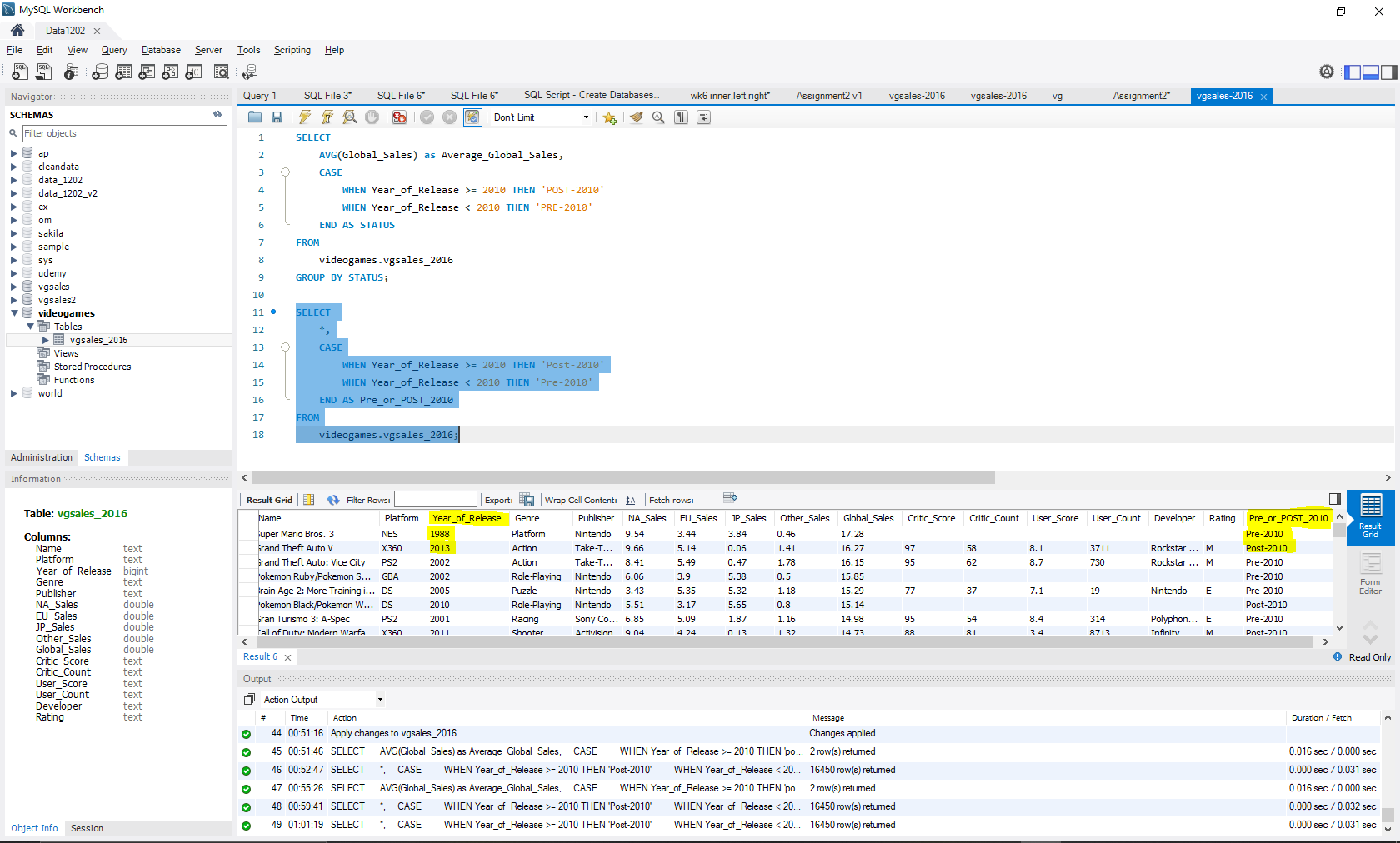
**END** **AS** Pre\_or\_POST\_2010

**FROM**

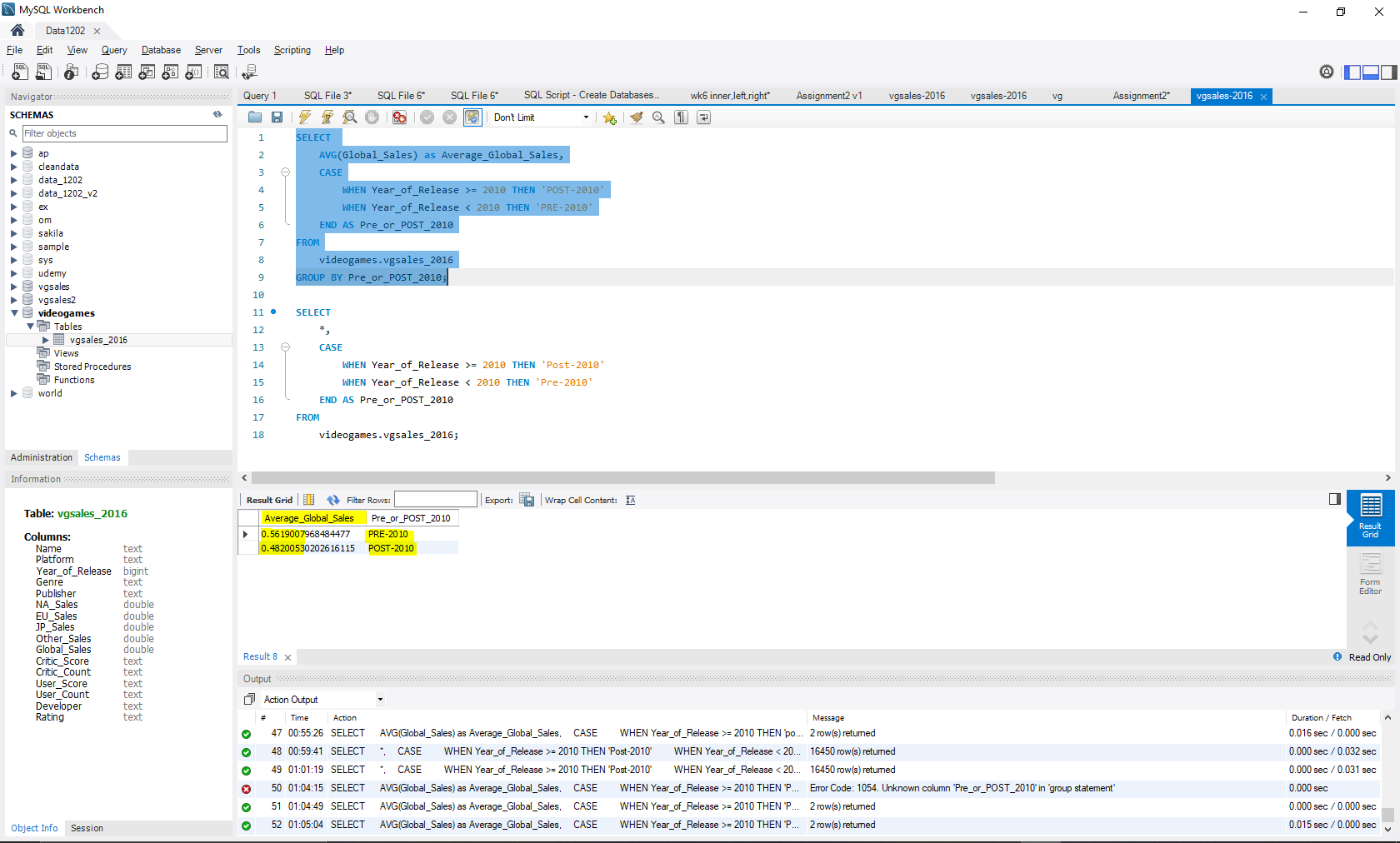
videosgames.vgsales\_2016

**GROUP BY** Pre\_or\_POST\_2010;

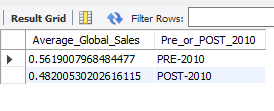
**2.1.2 Screenshot of the Output**



*Figure 1.1 “Pre\_or\_Post\_2010” column for categorizing Pre & Post 2010 dataset*



*Figure 1.2 Average Global Sales Execution*



*Figure 1.3 Average Global Sales Result*

**2.1.3 CONCLUSION**

From the MySQL output (Refer Fig 1.3),

Average Global Sales before 2010 is **0.56** and after 2010 is **0.48**.

**Average global sales are higher before 2010**(i.e., Pre-2010) in the considered dataset.

## **2.2 Steps followed**

* Firstly, I went through the dataset i.e., “vgsales.csv” file to understand the variables and its types.
* Created database and import the dataset into MySQL, also made sure that all the datapoints are imported successfully for correct output.
* Then divided the question into two parts for easy identification of required functions, which are:-

1. How to grouping the data into Pre and Post 2010

By using **CASE WHEN THEN** function to check whether the “**Year\_of\_Release**” is before or after 2010 and assigned it to “Pre\_or\_POST\_2010” column using AS function.

1. How to find the average for pre and post 2010 data points

By using **AVG** function and **GROUP BY** function to **“Pre\_or\_POST\_2010”** column, the average **Global\_Sales** of pre and post 2010 was found.

## **2.3 Challenges**

* Identification of proper functions for execution i.e., CASE WHEN THEN, GROUP BY and AS functions
* Syntax for using CASE WHEN along with GROUP BY and AS function to find the average Global\_Sales.
* I also tried using IF function, which also gave the same result

SELECT

AVG(Global\_Sales),

IF(Year\_of\_Release > 2010, 'post-2010', 'pre-2010') AS Pre\_or\_POST\_2010

FROM

videogames.vgsales\_2016

GROUP BY Pre\_or\_POST\_2010;

* Tried to represent data in the below format but failed,

|  |  |  |
| --- | --- | --- |
|  | **PRE-2010** | **POST- 2010** |
| **Average Global Sales** | 0.56 | 0.48 |

* Data import – importing vgsales dataset into the SQL table was taking a lot of time.

# **3. REFERENCE**

* <https://stackoverflow.com/>
* <https://www.w3schools.com/>
* DC Home >> Contents