**DATA SOURCE & LOAD**

**Name**: Video Game Sales With Ratings

**Provide**r: Kaggle.com

**URL**: <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>

**Date**: July 31, 2022

**Size**: 488 kB

**Download Instructions:**

1. Click the URL provided above in this README document OR click the URL shortcut provided in the “data” folder called, “Dataset -- Video Game Sales With Ratings.url”
2. The Kaggle website will load in the web browser and you should see the title, “Video game sales with ratings”.
3. Click the download button found at the top of the screen to download the dataset.

**Important:**

When this analysis was performed last, the data set was 488 kB in size.

**Dataset Column Content**

The following references the Kaggle dataset website. URL provided above.

**Name** Name of the game

**Platform** Console on which the game is running

**Year\_of\_Release** Year of the game released

**Genre** Type of game

**Publisher** Company that released the game

**NASales** North American Sales

**EUSales** European Sales

**JPSales** Japanese Sales

**OtherSales** Other Geographical Sales

**Global\_Sales** Global Sales

**Critic\_score**  Aggregate score compiled by Metacritic staff

**Criticcount**  The number of critics used in coming up with the Critic score

**User\_score**  Score by Metacritic's subscribers

**Usercount**  Number of users who gave the user score

**Developer** Party responsible for creating the game

**Rating**  The [ESRB](https://www.esrb.org/) ratings

**Tools Used:**

* Azure HDInsight Cluster
* Azure file storage
  + Resource group: SNeudorf
  + Location: Canada Central
  + Storage account: sneudorfhdistorage
* Apache Hive 2.0

**Data Load Process Performed:**

1. Created a container called, “sales”. Created a folder within the sales container called, “Sales” and uploaded the CSV file called, “Video\_Games\_Sales\_as\_at\_22\_Dec\_2016.csv”. This was done using the MS Azure Storage Explorer.

1. Created a schema and created a table called, “sales” and pulled in all the data from the “Video Game Sales With Ratings” CSV data file.

**Scripts Used:**

Below were the scripts used for loading the data into Hive for the items listed in the ‘Data Load Process Performed’ section above:

2.

CREATE EXTERNAL TABLE sales

(

Name string,

Platform string,

Year\_of\_Release int,

Genre string,

Publisher string,

NA\_Sales string,

EU\_Sales float,

JP\_Sales float,

Other\_Sales float,

Global\_Sales float,

Critic\_Score float,

Critic\_Count int,

User\_Score float,

User\_Count int,

Developer string,

Rating string

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED AS TEXTFILE LOCATION '/Sales'

TBLPROPERTIES("skip.header.line.count"="1");