**OVERVIEW**

**Business Problem**

The customer is looking to identify which video game genre (sport or shooter) is more popular and generates the highest revenue. The customer is trying to determine which genre to invest in next.

**Objectives**

To identify the most popular video game genre (sport or shooter) that has the highest sales and ratings.

To note the global sales of each game, and the volume sold, as well as compare the critic ratings between the sport and shooter games.

**Process Used**

* Loaded the ‘Video game sales with ratings’ CSV data file into HDInsight using the Azure Data Storage Tool, and created a Data Schema using Apache Hive
* Transformed the data for further analysis regarding genre, global sales, and critic scores
* Observed the average critic scores for sports and shooter games
* Observed the average/min/max global sales values for sports games and shooter games
* Observed the total global sales and count for each game
* Created dashboards for the average critic scores for sports games and shooter games

**Insights**

Sports games were found to be the most popular video game genre sold generating the most sales over the shooter video game genre. However, video game critic ratings for these two genres are split down the middle. Shooter video games had about a 2% higher critic rating over sport video games.

To conclude, it would be more profitable for the customer to invest in the sports video game genre, as it has the highest RIO regarding global sales.

**Directory structure**

**Folders**

* **data** *Contains information about the dataset, URL, and how to download the dataset, tools and the data load process and scripts that were used.*
* **munge** *Contains information about the data transformations performed, tools and scripts that were used.*
* **scr** *Contains information about the data aggregations performed, tools and scripts that were used.*
* **reports** *Contains an Excel dashboard used to visualize the data for analysis and presentation purposes.*