Project 3

Ana Saavedra | Analiza Asuncion | Avantika Singh | Shilpa Nagendra | Gatlin Rash

**Story –**

We are exploring the global popularity of VIDEO GAMES. We will analyze the gaming industry trends for the last 40 years with the popularity of consoles and specific games. We can visualize the data with the different attributes we have learned in class with an interactive dashboard and visuals with the data provided. The data will be useful to understand the video game industry. Such data can help a video game company's marketing department, who wants to understand the sales generated by a different genre of video games by what consoles. This data can also help make projections about the new games coming in the industry and how likely they are to perform in the industry. One can even understand the trends of the competitor. Our work aims to analyze the consumption of video games worldwide and consider the trends in genres' popularity across time.

**Links –**

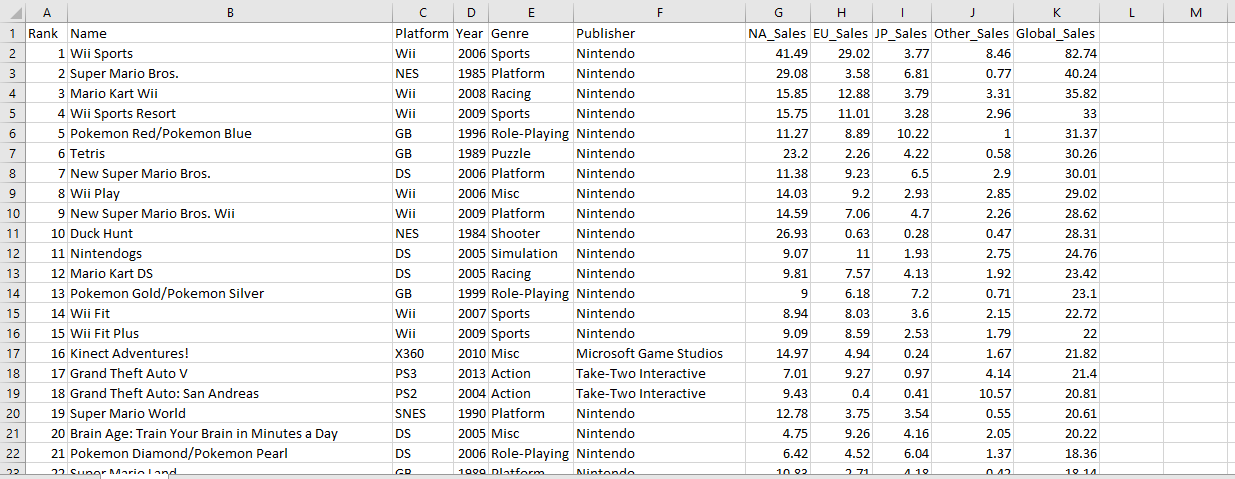
This is where we obtained our data.

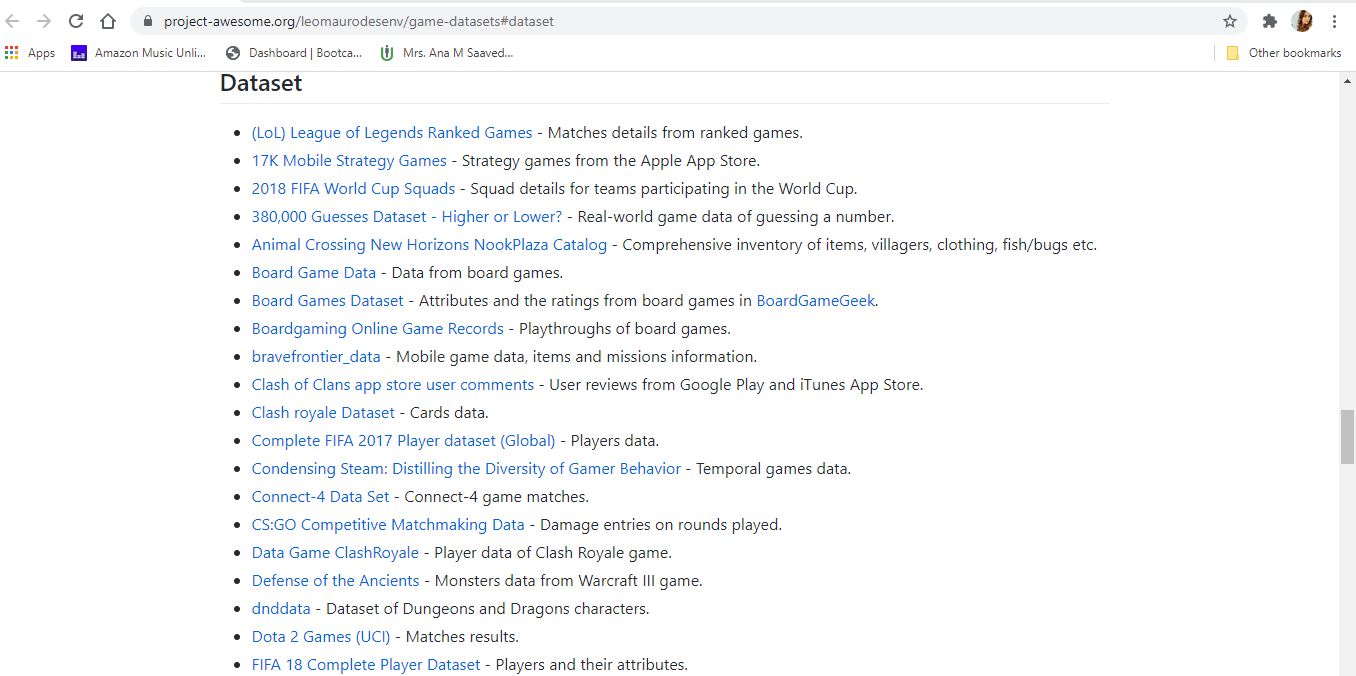
<https://www.kaggle.com/gregorut/videogamesales>

* Rank - Ranking of overall sales
* Name - The games name
* Platform - Platform of the game's release (i.e., PC, PS4, etc.)
* Year - Year of the game's release
* Genre - Genre of the game
* Publisher - Publisher of the game
* NA\_Sales - Sales in North America (in millions)
* EU\_Sales - Sales in Europe (in millions)
* JP\_Sales - Sales in Japan (in millions)
* Other\_Sales - Sales in the rest of the world (in millions)
* Global\_Sales - Total worldwide sales.

<https://project-awesome.org/leomaurodesenv/game-datasets#dataset>

**Screenshots –**

The following is the snippet of the dataset we found to execute our project. ****

****

**AWS and Cloud Services –**

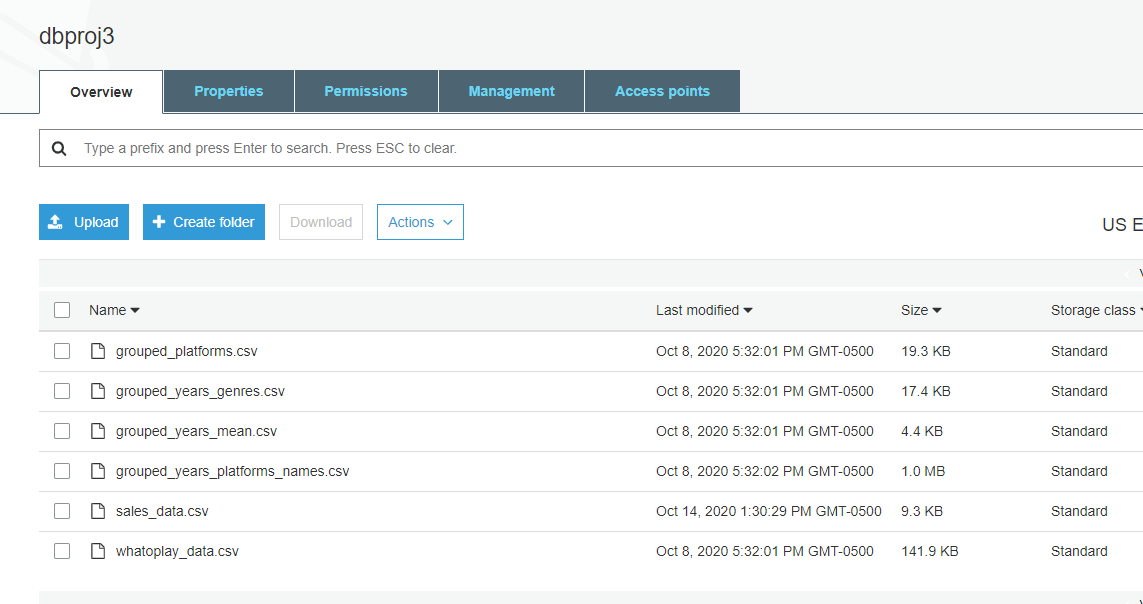
This section illustrates how we utilized a few services available in AWS:

* S3 Bucket
* RDS
* EC2
* Quicksight

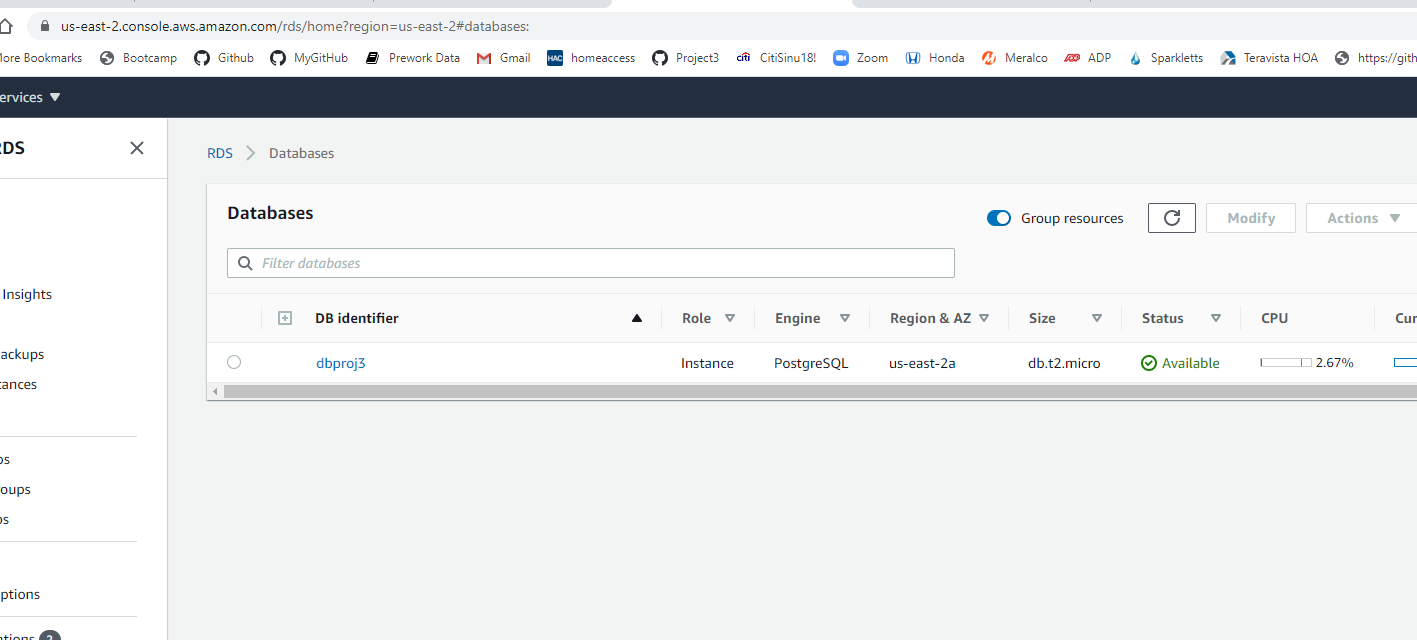
Other tools used:

* 5. Google Collaborator
* 6. Python
* 7. Postgres

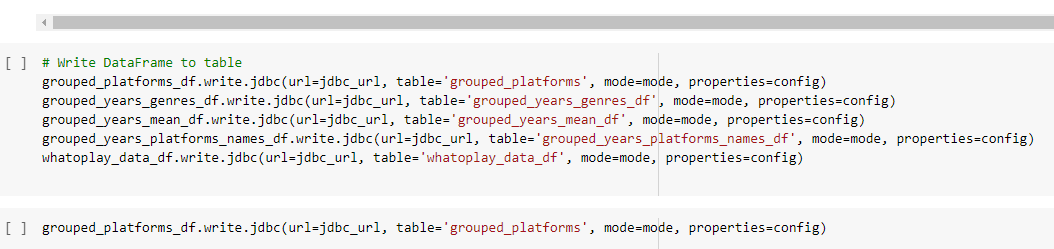
dbproj3 bucket with CSV files:



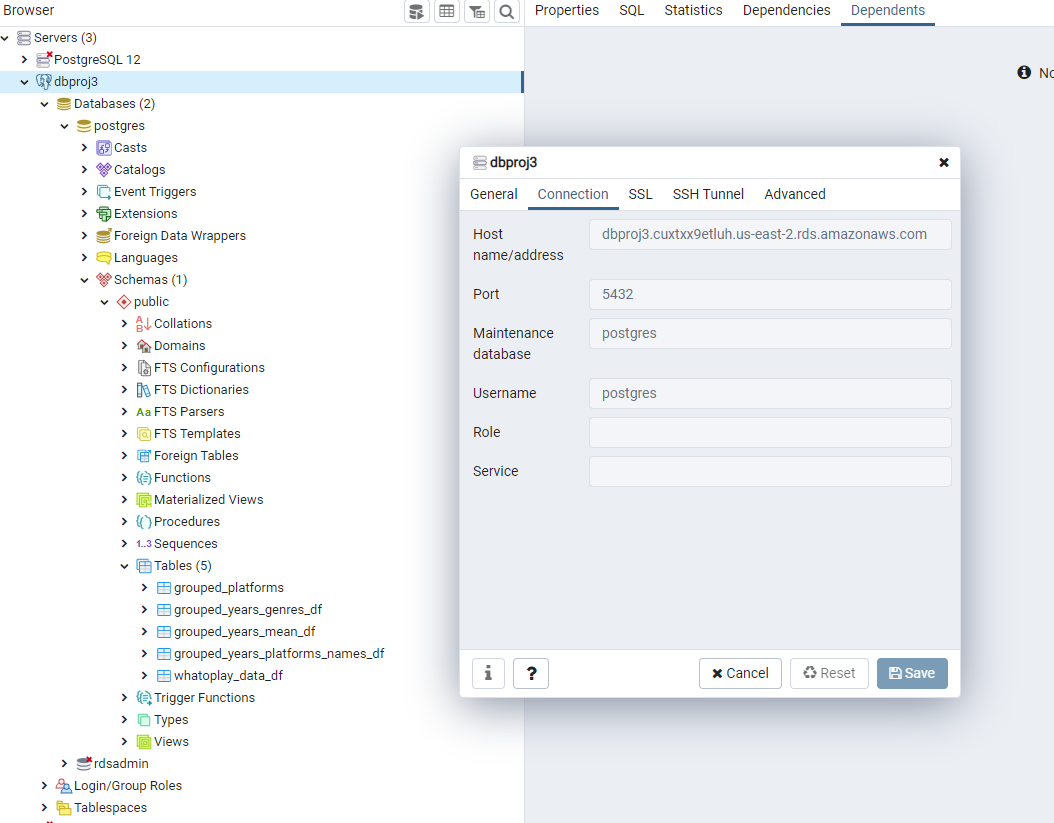
RDS database:



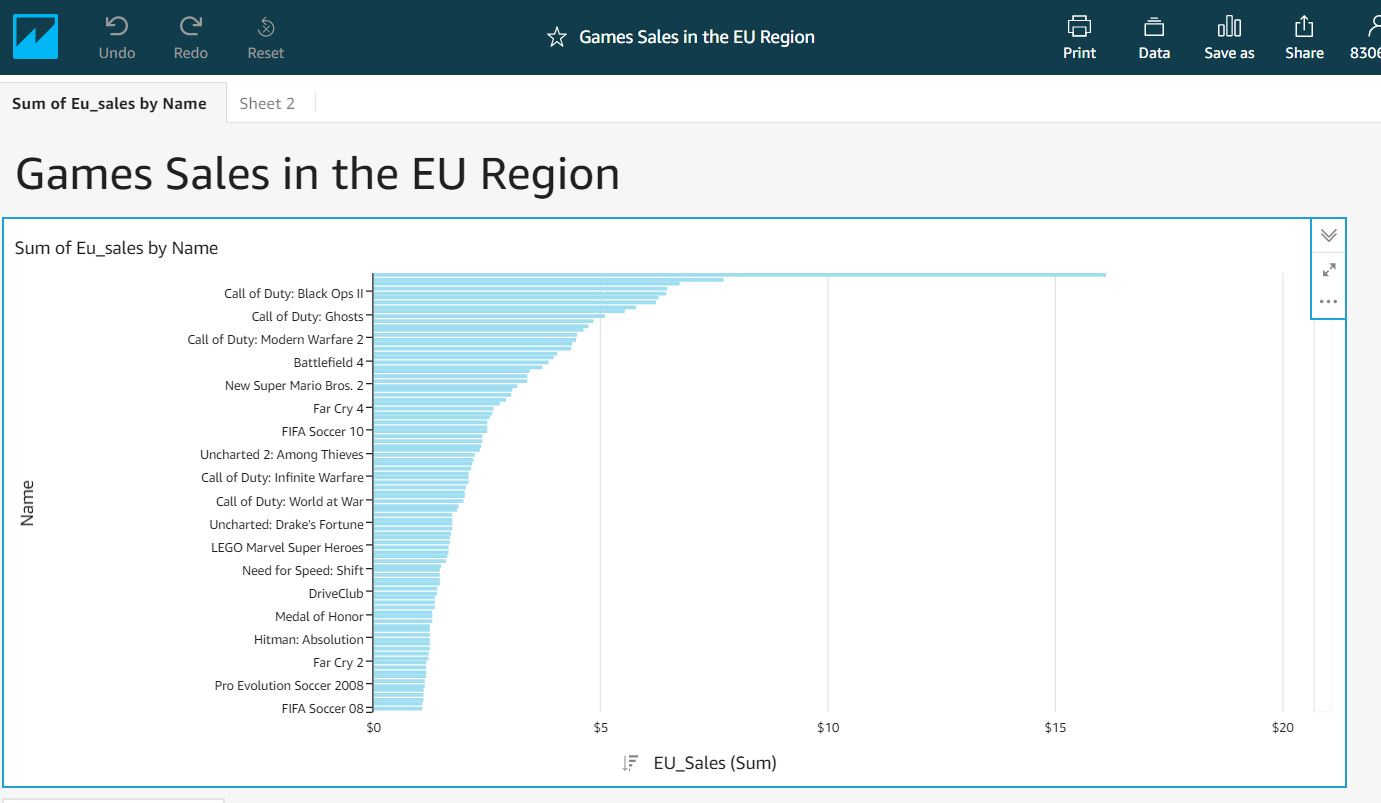
Python file used to write the data frames to Postgres using Google Collaborator:



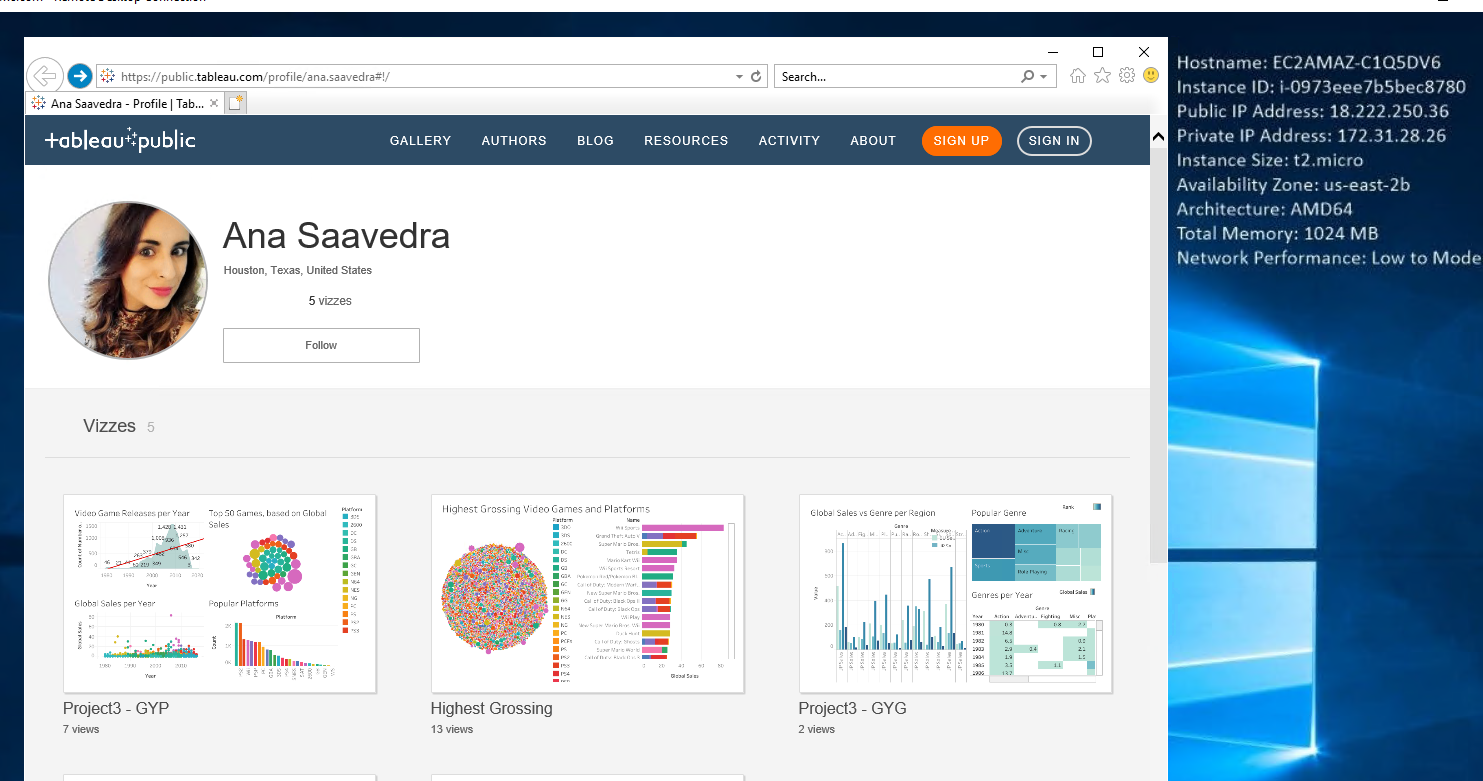
RDS dbproj3 and Postgres SQL Query Editor:



Quicksight Analytics using RDS as a data source:



EC2 showing Tableau graphs:



Challenge: No Admin rights to upgrade the IE browser or install Google Chrome. Unable to access the Quicksight website. Very slow.

**Tableau Visualizations and Dashboards –**

Thru Tableau software, we were able to show the data in new ways with different perspectives. Considering that our data sets are so large, this is a great way to summarize and display the data for the general population to understand. Tableau makes the different charts very easy to understand, and they are also interactive. The tooltips are great to hover over a marker and show the related information. We were able to adapt to the trend line and create a forecast for future sales.

**Machine Learning –**

The Video Game data had three scores in the form of the rating score, which was Play Score, Game Score, and Critic Score, respectively. To describe each of them, the Play Score is the score obtained from the Player Ratings on various games, Game Score is the rating given to each game by the developer of the game, and

the Critic Score is obtained from the critic rating for the games. To study the relevance of these scores, we ran some Machine Learning Algorithms on our data (the analysis was done on North America Sales since it is the biggest market in sales). The following are the conclusions drawn.

**Final design –**

We designed a multi-page website to display different attributes of the project. The visualizations and the Machine Learning outcomes are divided into separate pages for viewing ease. We also show the two sets of raw data that were used to mitigate to the data used for each data source.

**GitHub Repository –**

This is the link to our GitHub repository.

https://github.com/Anabn357/dbproj3