**Data Collection and Preparation:**

**Data Collection:**

Kaggle, a renowned website for sharing and accessing datasets, provided me with the dataset for my research. This dataset, titled "The Golden Age of Video Games | SQL Project | EDA," is rich in video game information. It includes information about game titles, genres, platforms, sales numbers, player demographics, and review ratings. After downloading the dataset from Kaggle, I discovered that it was in a useable format, basically a structured file, such as a CSV or Excel spreadsheet. This first structure made it easy to integrate into my SQL project.

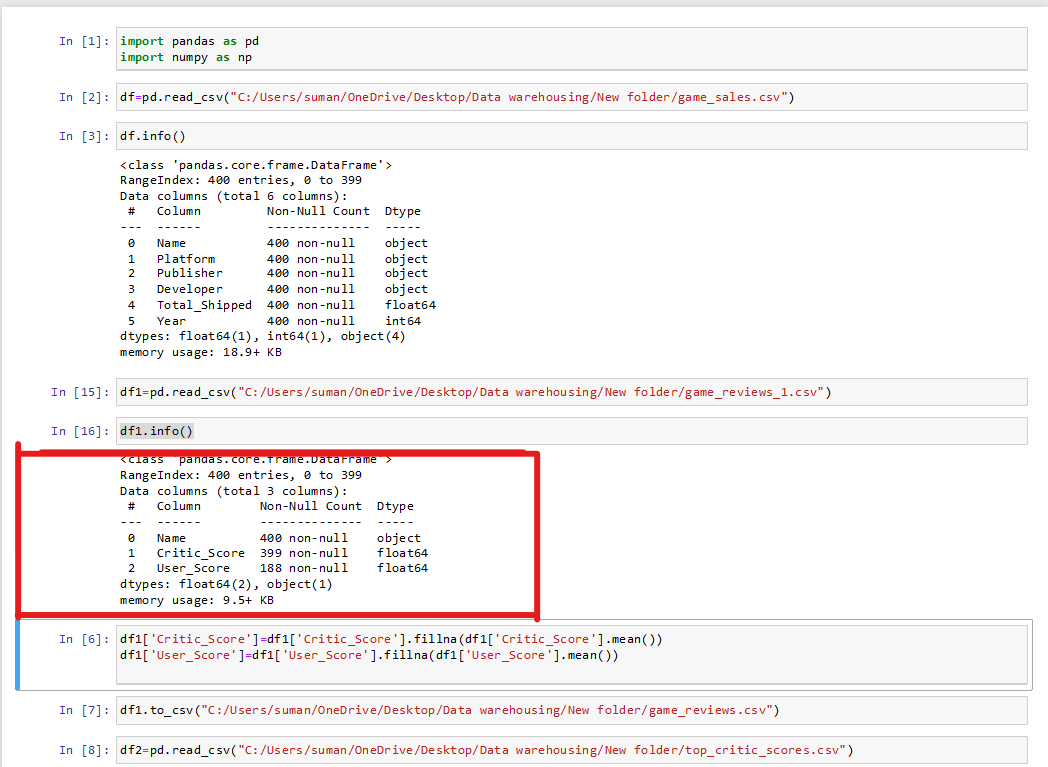
Data Source : -<https://www.kaggle.com/code/ahmedabdallah1/the-golden-age-of-video-games-sql-project-eda/input>

**Data preparation:**

In the process of preparing the dataset, I first performed data cleaning to ensure its quality and accuracy. This involved addressing issues like missing values, correcting data types, and identifying and dealing with any outliers or anomalies.

During my data cleaning process in Jupyter Notebook, I focused on enhancing the dataset's quality and completeness, particularly in the "reviews" field; none of the other tables had missing values. The columns "game," "user\_score," and "critic\_score" in this table held critical information. I fixed missing values in the "user\_score" and "critic\_score" columns to ensure the dataset was ready for analysis. Missing data might jeopardize the validity of analytical results; therefore I used imputation techniques to fill in the gaps. I attempted to improve the dataset's usability for further studies by imputing missing values in the "user\_score" and "critic\_score" columns, ensuring that no crucial information was neglected or removed. In my study, this data cleansing phase was critical for getting robust and dependable findings.

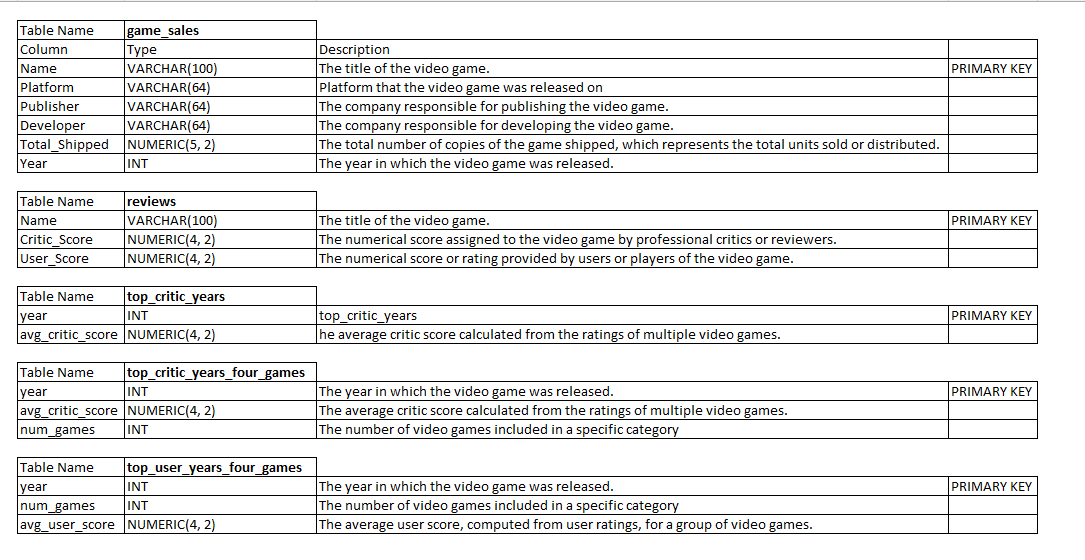
To ensure a clear understanding of the dataset, I created a data dictionary or metadata, documenting the meaning of each column and units of measurement. Finally, I formatted and exported the cleaned and prepared dataset in a manner compatible with my SQL project, including any schema design to represent the data effectively. This meticulous data collection and preparation process laid the foundation for meaningful exploratory data analysis and SQL-based insights into the golden age of video games.

I ran a comprehensive check on the tables to ensure there are no missing values.

A screenshot of a computer program

Description automatically generated

**Data Dictionary:**



**Database Design:**

**ERD Diagram**

The Entity-Relationship Diagram (ERD) was a critical stage in my data warehousing project, which began with a dataset collected from Kaggle and offered insights on the Golden Age of Video Games. I converted this rich dataset into a structured database using painstaking design and execution within Oracle SQL Developer. The ERD serves as the architectural plan for the database, displaying the relationships between the many tables built for this project. These tables, which include "game\_sales," "game\_reviews," "top\_critic\_score," "top\_critic\_scores\_more\_than\_four\_games," and "top\_user\_scores\_more\_than\_four\_games," serve as the foundation of my data warehouse project. The ERD acts as a guiding framework, encapsulating not only the links between these tables but also highlighting the main entities, traits, and connections required for complete data analysis.

A computer screen shot of a computer screen

Description automatically generated

**Data Cubes:**

In data warehousing and business intelligence, data cubes are an essential component of multidimensional data modeling. In your project titled "The Golden Age of Video Games," you created data cubes to aid in multidimensional analysis.

**1. Game Sales Data Cube Query:**

This query returns total game sales grouped out by platform, developer, and year. It enables you to find the best-selling games in various combinations of these parameters.

A screenshot of a computer

Description automatically generated

**2.Game Review Ratings Data Cube Query:**

This query computes the average critic and user scores for games. It aids in the identification of games with the highest average rating from both reviewers and users.

A screenshot of a computer

Description automatically generated

**3.** **Top Critic Scores Data Cube Query:**

This query identifies the year with the average critic scores and the num\_games they were available on.

**A screenshot of a computer

Description automatically generated**

**Creating table**

In order to gain significant insights from a Kaggle dataset, I began on a complete database design journey. The Golden Age of Video Games dataset offered a treasure mine of information on game sales, reviews, and critical ratings. My database design approach started with the development of an organized schema for effective data storage and retrieval. Using Oracle SQL Developer, I painstakingly created tables to house this abundance of data, laying the groundwork for my data analysis adventure.

I defined and built five critical tables in my Oracle SQL Developer environment: "game\_sales," "game\_reviews," "top\_critic\_score," "top\_critic\_scores\_more\_than\_four\_games," and "top\_user\_scores\_more\_than\_four\_games." These tables were carefully designed to capture various characteristics of the dataset. "game\_sales" stores information on game sales, whereas "game\_reviews" stores user reviews and ratings. "top\_critic\_score" highlights games with high critical scores, while "top\_critic\_scores\_more\_than\_four\_games" and "top\_user\_scores\_more\_than\_four\_games" tables highlight games with high critical and user scores over many titles, respectively.

**Table Name – game\_sales**

A screenshot of a computer

Description automatically generated

**Table Name – reviews**

A screenshot of a computer

Description automatically generated

**Table Name – top\_critic\_years**

A screenshot of a computer

Description automatically generated

**Table Name – top\_critic\_years\_four\_games**

A screenshot of a computer

Description automatically generated

**Table Name – top\_user\_years\_four\_games**

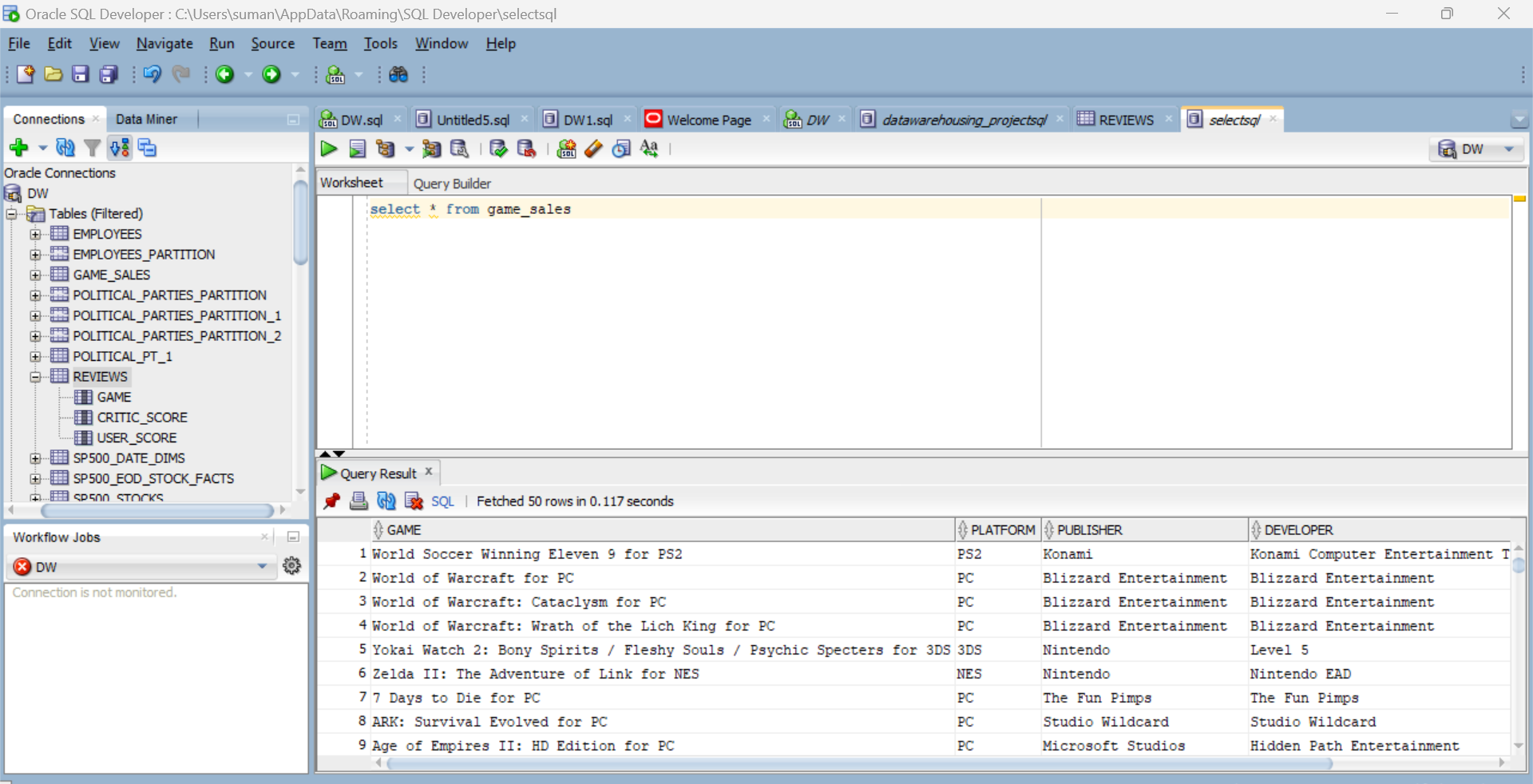
A screenshot of a computer

Description automatically generated

**Importing the Data:**

Importing data into these tables brought the dataset to life within my database. Each table was populated with pertinent data from the Golden Age of Video Games dataset, allowing for seamless data analysis and insights. By generating and filling these tables with data, I created the framework for executing complex SQL queries, data cubes, and exploratory data analysis to gather vital information about this critical period in video game history. My database architecture serves not only as an organized archive for this one-of-a-kind dataset, but also as a powerful tool for uncovering the tales and trends of the Golden Age of Video Games.

**Snapshot of game\_sales data:**



**Snapshot of reviews data:**

A screenshot of a computer

Description automatically generated

**Snapshot of top\_critic\_years data:**

A screenshot of a computer

Description automatically generated

**Snapshot of top\_critic\_years\_four\_games data:**

A screenshot of a computer

Description automatically generated

**Snapshot of top\_user\_years\_four\_games data:**

A screenshot of a computer

Description automatically generated