LUMENDATA

Data Sheet

Snowflake + Streamlit

For Data Analysis and Visualization

5201 GREAT AMERICAN PARKWAY, SUITE 320
SANTA CLARA, CA 95054
Tel: (855) 695-8636
E-mail: info@lumendata.com
Website: www.lumendata.com

We leverage Snowflake's cloud-based infrastructure for robust data querying and Streamlit's user-friendly interface for dynamic visualizations.

Together, we navigate a Gaming Dataset, extracting insights into ratings, votes, genres, and certifications. Witness the seamless integration of analytical capabilities and interactive presentation, unveiling a comprehensive understanding of the gaming landscape.

Data Loading into Snowflake

The data is present in a local machine and is loaded with the help of internal stage. Install SnowSQL and log in through the command prompt. Establish a connection with Snowflake using the following commands.

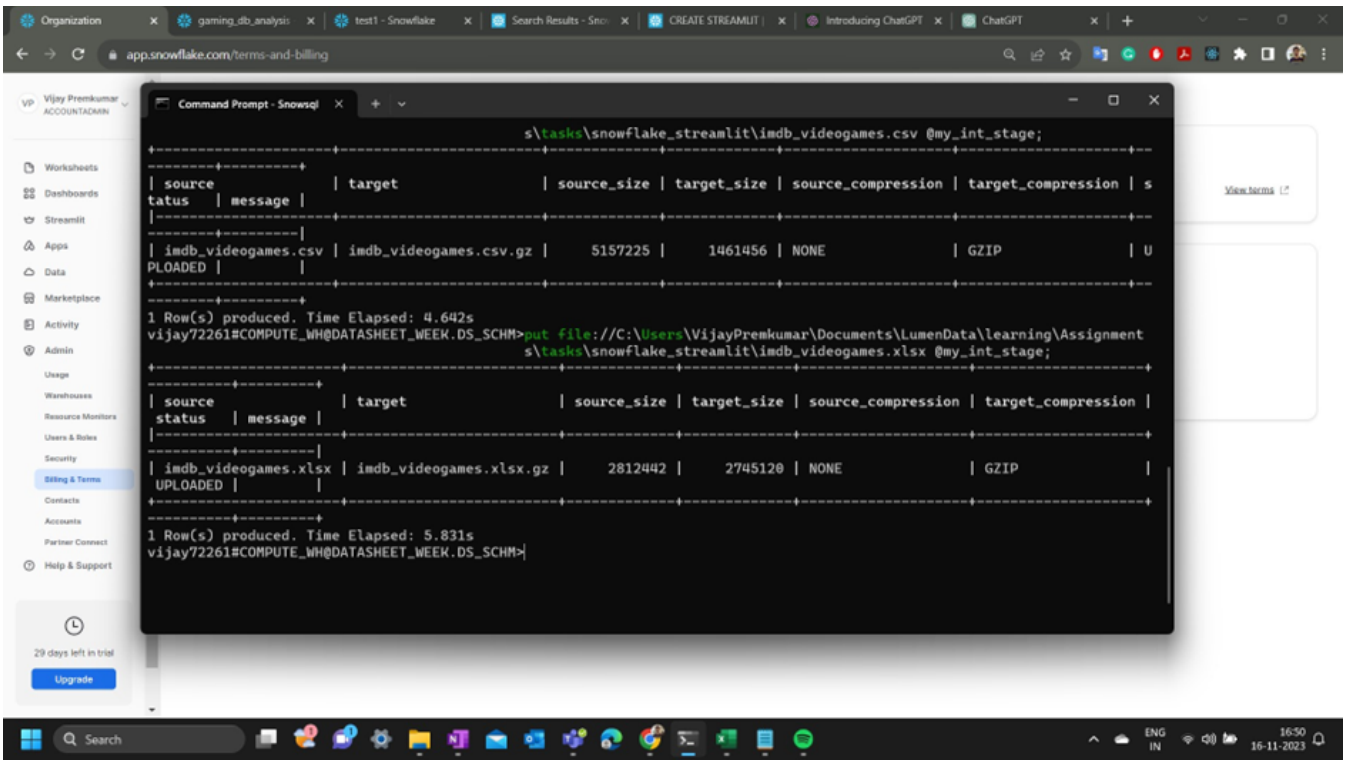
Snowsql -a <account_name> -u <loin_name>

Use database dbName;

Use schema schemaName

Create or replace stage stageName;

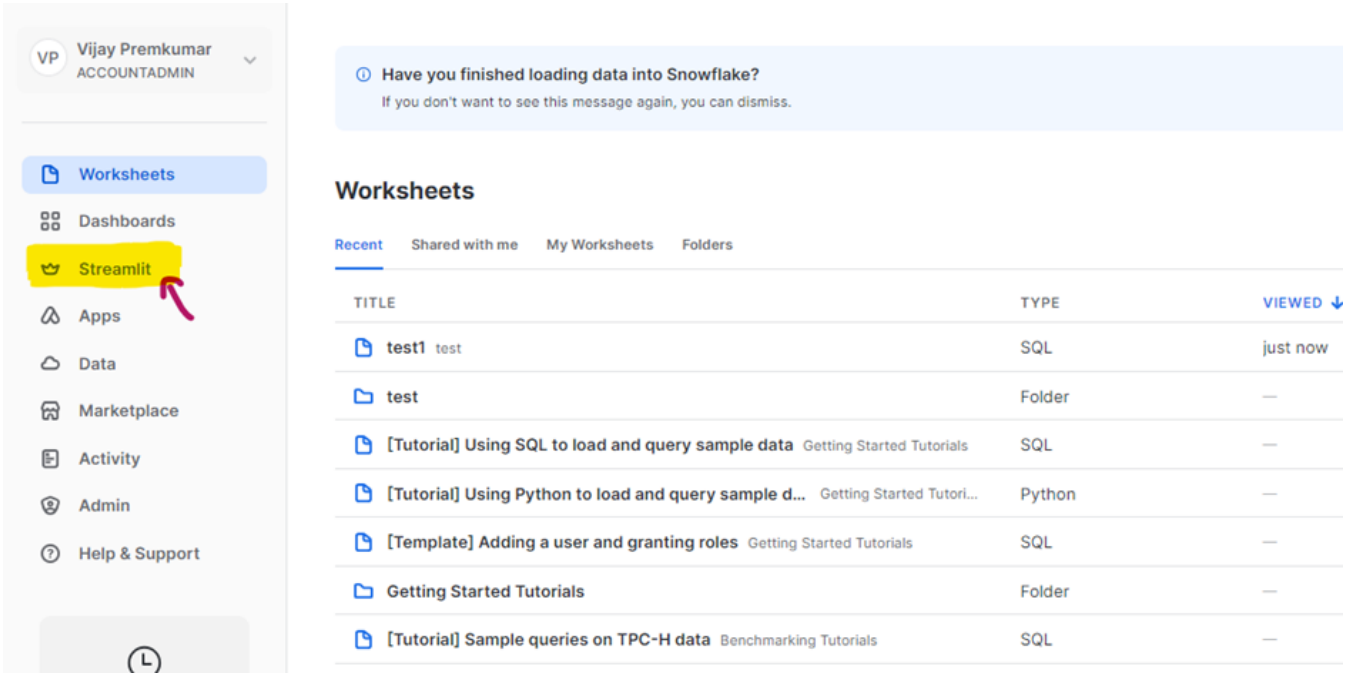
put file://C:\Users\VijayPremkumar\Documents\LumenData\snowflake\filename.csv



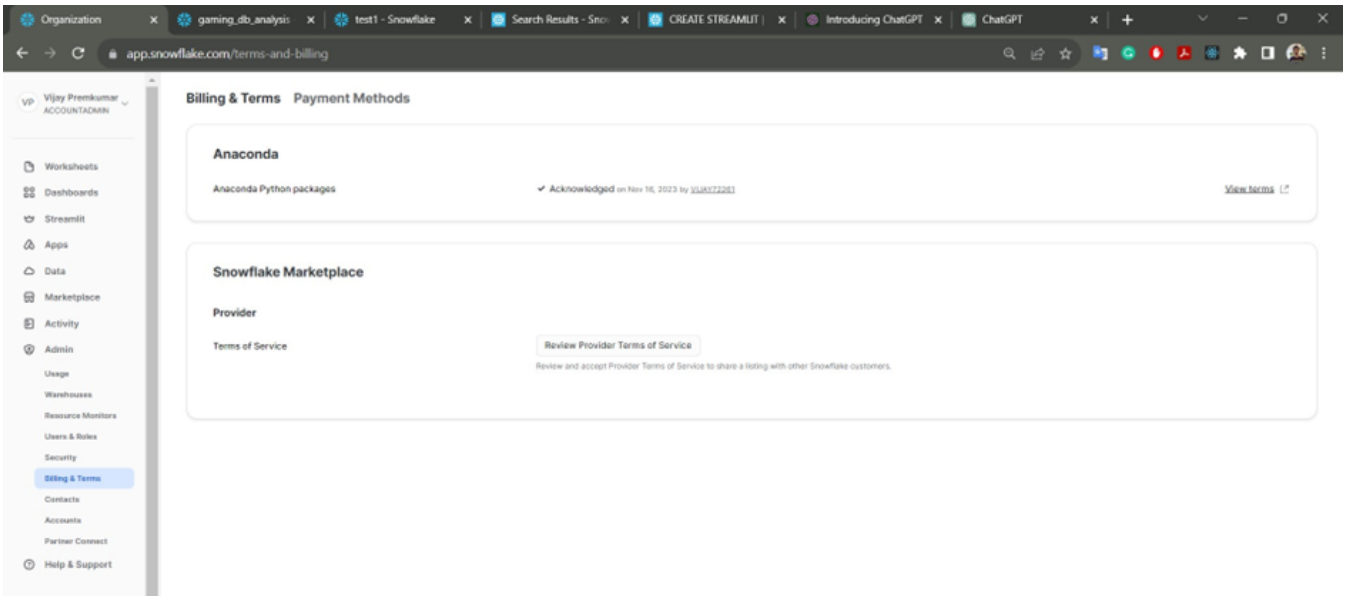
Data Analysis and Visualization:

Create a database, schema, and tables before loading the data in the form of **CSV**. Use copy into the command for loading the data into the table in a structured format, ensuring it's ready for analysis and visualization.

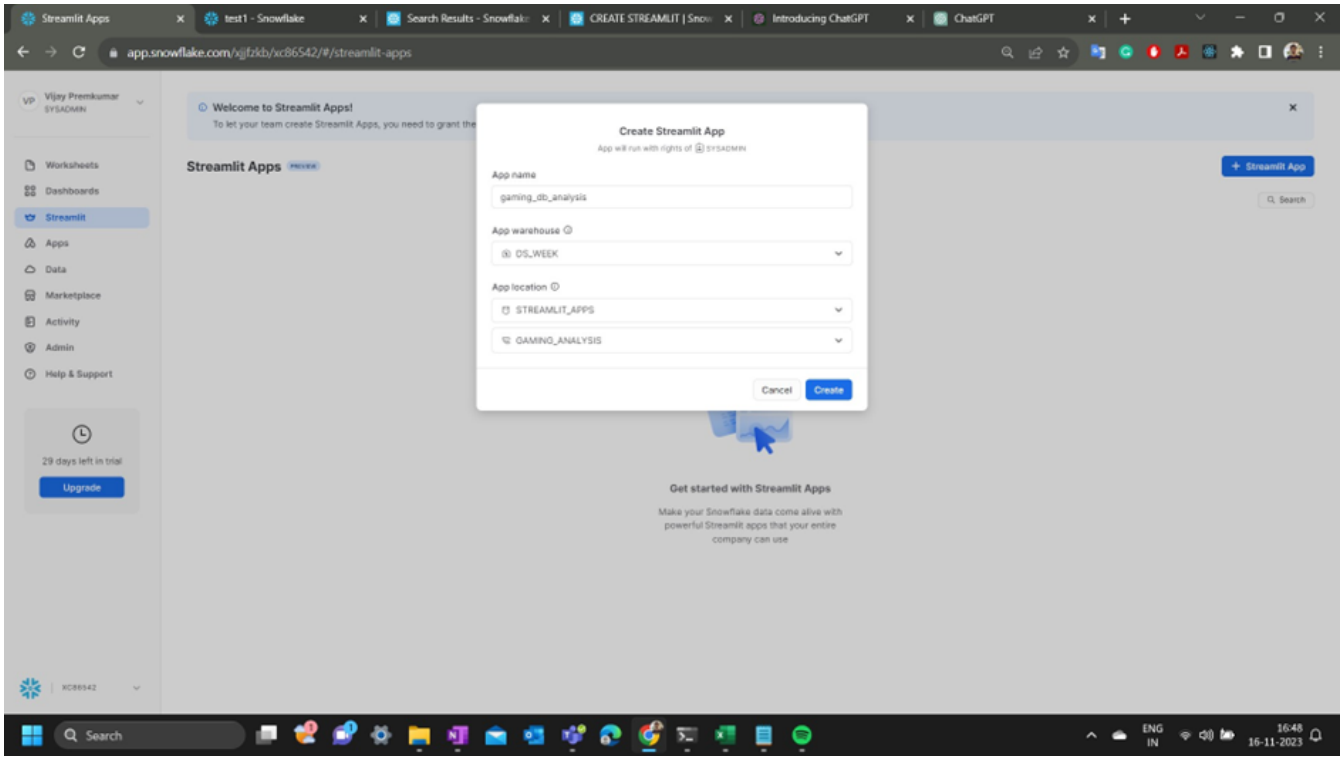
Next comes the creation of the Streamlit app.



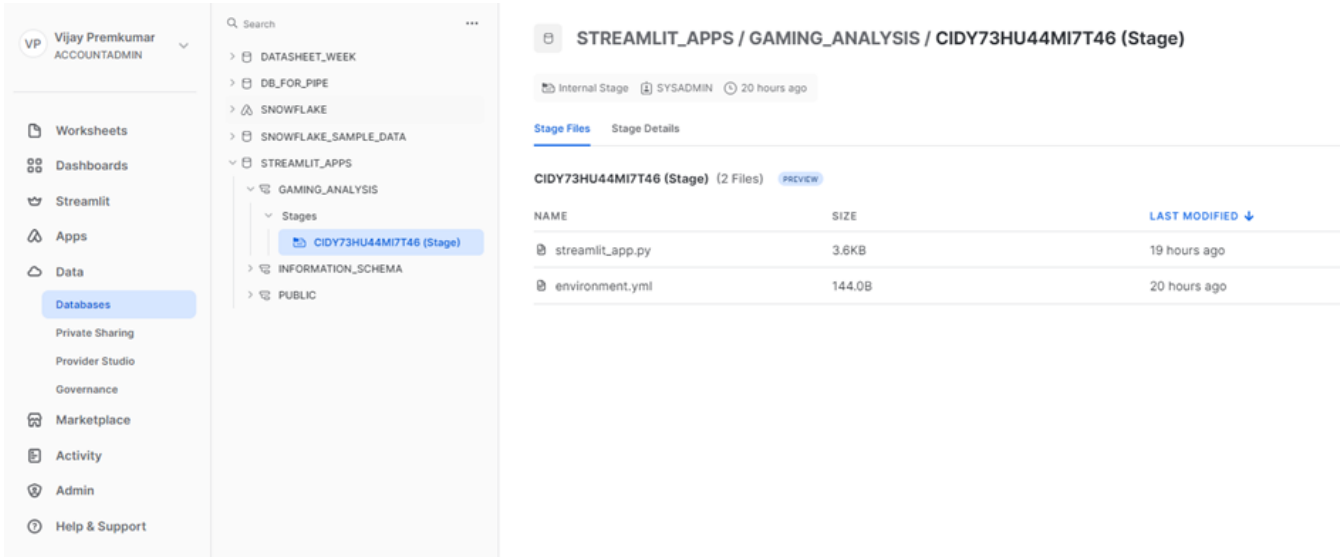
Use the existing database or create a new one and grant the required privilege to the role that you are currently using. Go to **Admin-> Billing** and acknowledge Anaconda Python packages. You will be able to create the Streamlit app.



Hit the create button and start naming your Streamlit app. Select the warehouse for your app and fill in all the required details. Click '**CREATE**.'

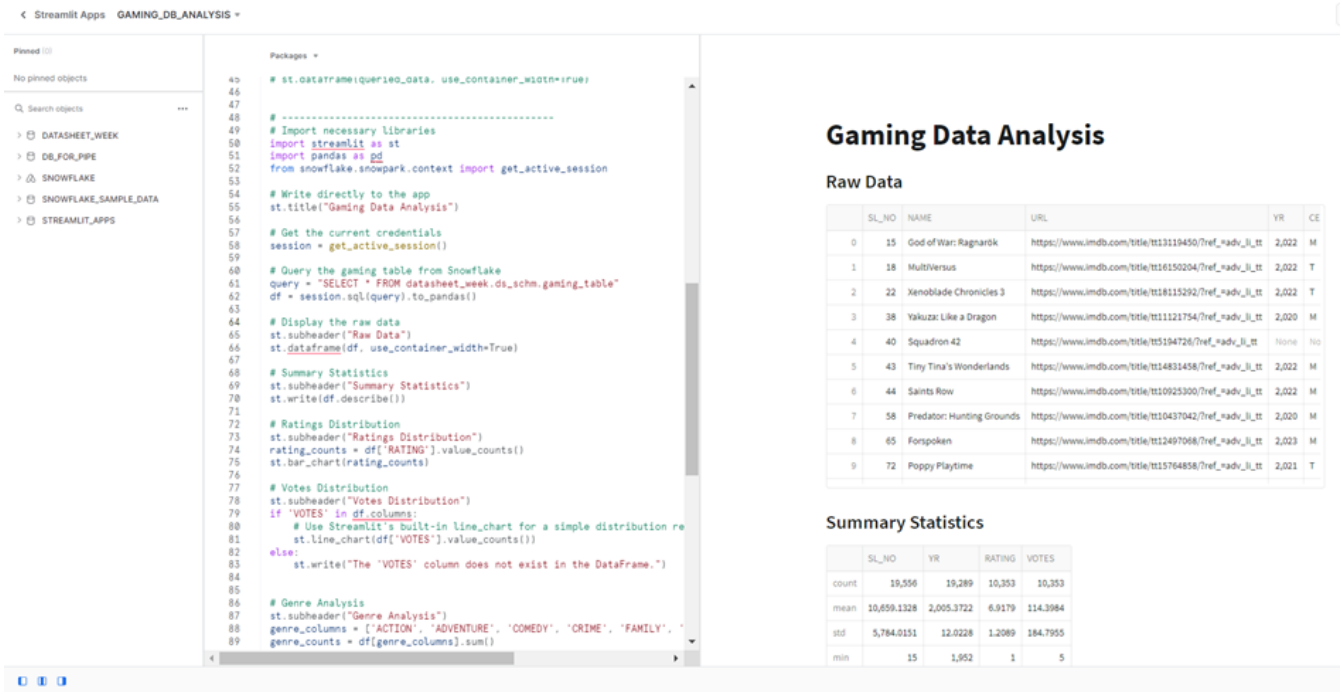


Once it is created, you can access it anytime through the same tab (Streamlit). The analysis and visualization will be done with the help of Python. We can see the **streamlit_app.py** file and an **environment.yml** file inside the database and schema.



When you open the Streamlit app for the first time, it will show a sample code and output on the right side of your file.

You can start coding/ editing the file for the current data of gaming dataset which we just loaded into the table. We start with summary statistics, ratings distribution, votes distribution, genre analysis, and certificate analysis. We need to store the result of the query in a data frame and then, we can use the data frame for various kinds of plots or a graph.



Summary Statistics:

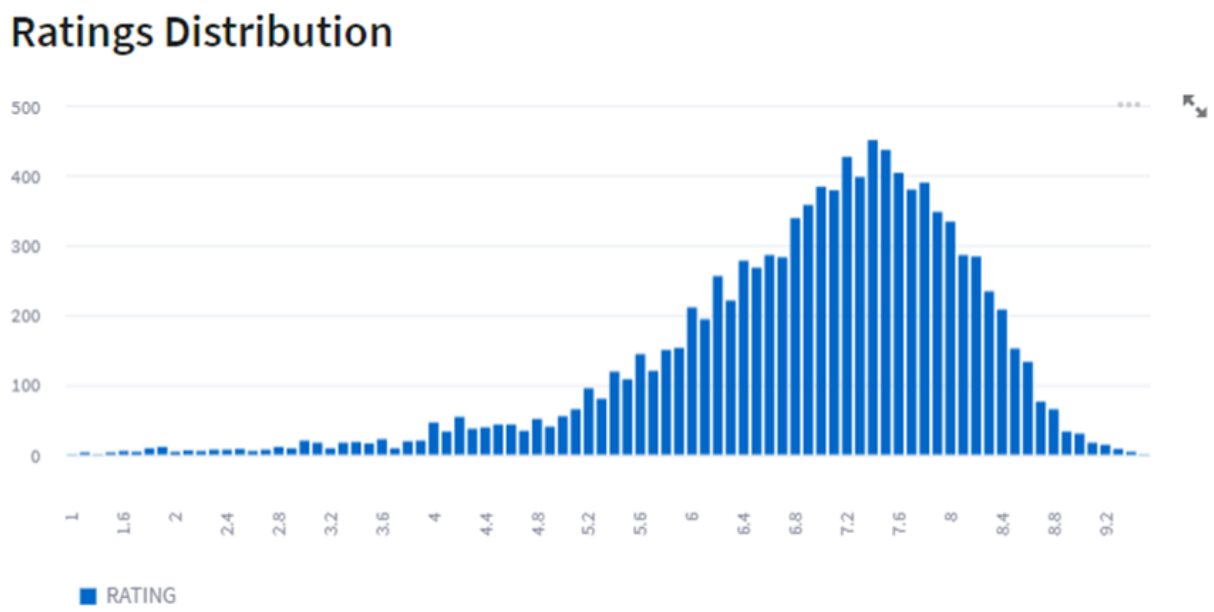
The sum of all votes and ratings is displayed in the graph, along with the mean and standard deviation.

Summary Statistics

	SL_NO	YR	RATING	VOTES
count	19,556	19,289	10,353	10,353
mean	10,659.1328	2,005.3722	6.9179	114.3984
std	5,784.0151	12.0228	1.2089	184.7955
min	15	1,952	1	5
25%	5,675.75	1,996	6.3	15
50%	10,683.5	2,007	7.1	37
75%	15,615.25	2,016	7.8	115
max	20,802	2,027	9.5	999

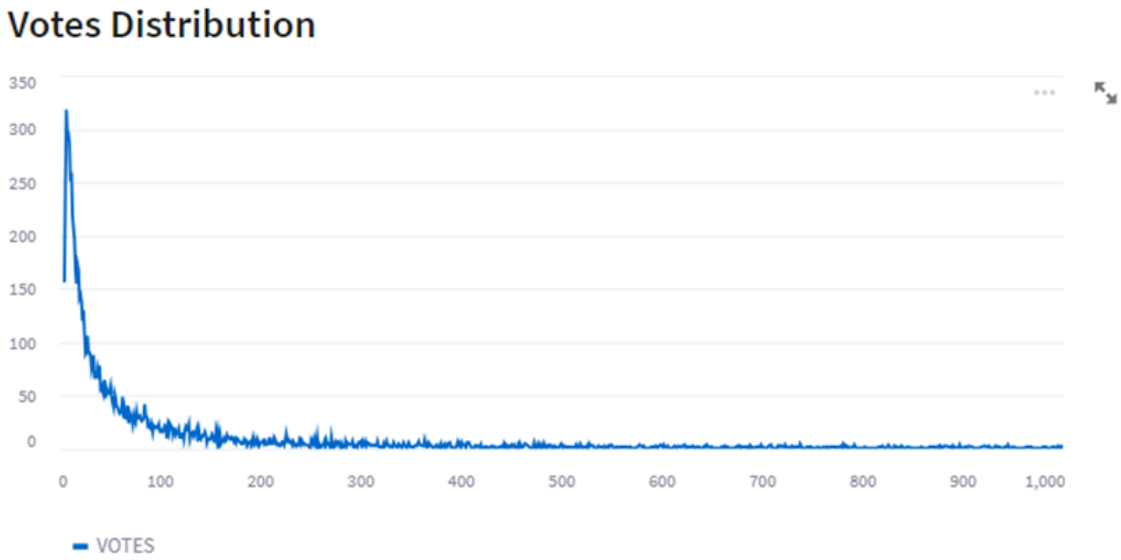
Rating Distribution:

The count and sum of the total ratings are displayed in a bar chart by reusing the same data frame which holds the result of the query. For example, if the dataset has a variety of ratings (e.g., 1 to 10), the chart will display the number of games that fall into each rating category.



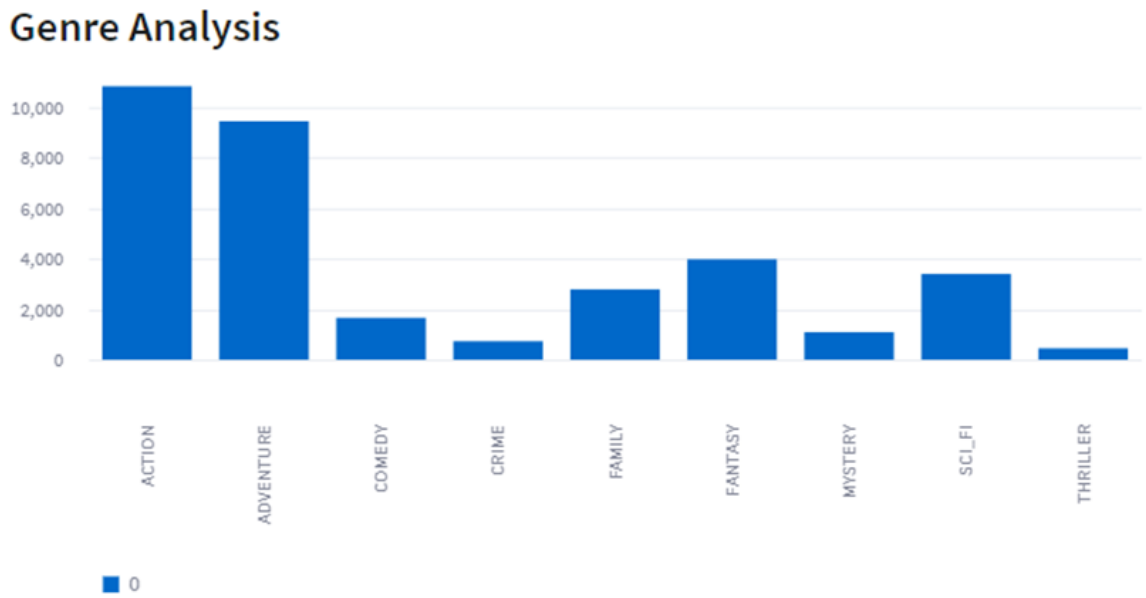
Votes Distribution:

This visualization illustrates the distribution of votes or reviews for the video games in the dataset. It could be represented as a histogram or line chart, showcasing the frequency of different vote counts. The goal is to understand how many games received a specific number of votes.



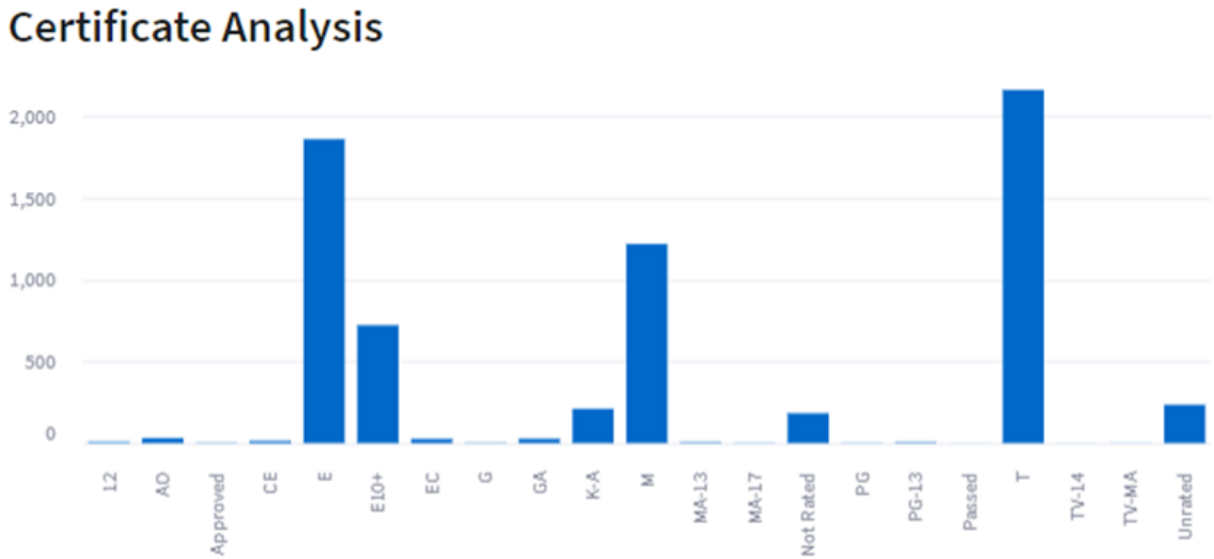
Genre Analysis:

This visualization captures the distribution of different genres within the gaming dataset. It could be presented as a bar chart, showing the count or percentage of games belonging to each genre. For instance, the chart may reveal the popularity of specific genres among the video games in the dataset.



Certificate Analysis:

Just like the genre analysis, this visualization focuses on the distribution of certificates or ratings assigned to the video games. It could be a bar chart displaying the count or percentage of games falling under each certification category. This provides insights into the distribution of content ratings in the dataset.



Queries:

```
create database datasheet_week;
CREATE or replace TABLE datasheet_week.ds_schm.gaming_table (  sl_no int,
name VARCHAR(255),  url VARCHAR(255),  yr INT,certificate VARCHAR(255),rating
DECIMAL(3,1),  votes INT,  plot_summary TEXT,  Action BOOLEAN,  Adventure
BOOLEAN,  Comedy BOOLEAN,  Crime BOOLEAN,  Family BOOLEAN,  Fantasy
BOOLEAN,  Mystery BOOLEAN,  Sci_Fi BOOLEAN,  Thriller BOOLEAN);
COPY INTO datasheet_week.ds_schm.gaming_table FROM
'@"DATASHEET_WEEK"."DS_SCHM"."MY_INT_STAGE"/imdb_videogames.csv.gz'file_
format = (format_name='datasheet_week.ds_schm.my_csv_format')ON_ERROR =
'CONTINUE';
```

Source code : gamingdata_streamlit.py

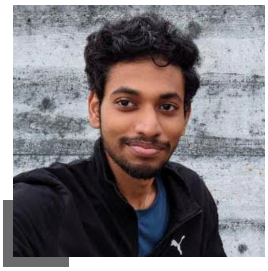
The analysis and visualization of the Gaming Dataset using Snowflake and Streamlit have provided valuable insights into the characteristics of the video games included in the dataset. The combination of Snowflake's powerful querying capabilities and Streamlit's interactive visualizations has facilitated a comprehensive data exploration.

These insights collectively contribute to a more comprehensive understanding of the gaming landscape represented in the dataset.

As technology and data analytics continue to evolve, leveraging tools like Snowflake and Streamlit enhance the efficiency and depth of data exploration, enabling data-driven decision-making in the gaming industry.

Authors

Sai Bharadwaja
Consultant



Vijay Premkumar
Associate Consultant

About LumenData

LumenData is a leading provider of **Enterprise Data Management, Cloud & Analytics** solutions. We help businesses navigate their data visualization and analytics anxieties and enable them to accelerate their innovation journeys.

Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 leading companies. LumenData is **SOC2 certified** and has instituted extensive controls to protect client data, including adherence to GDPR and CCPA regulations.