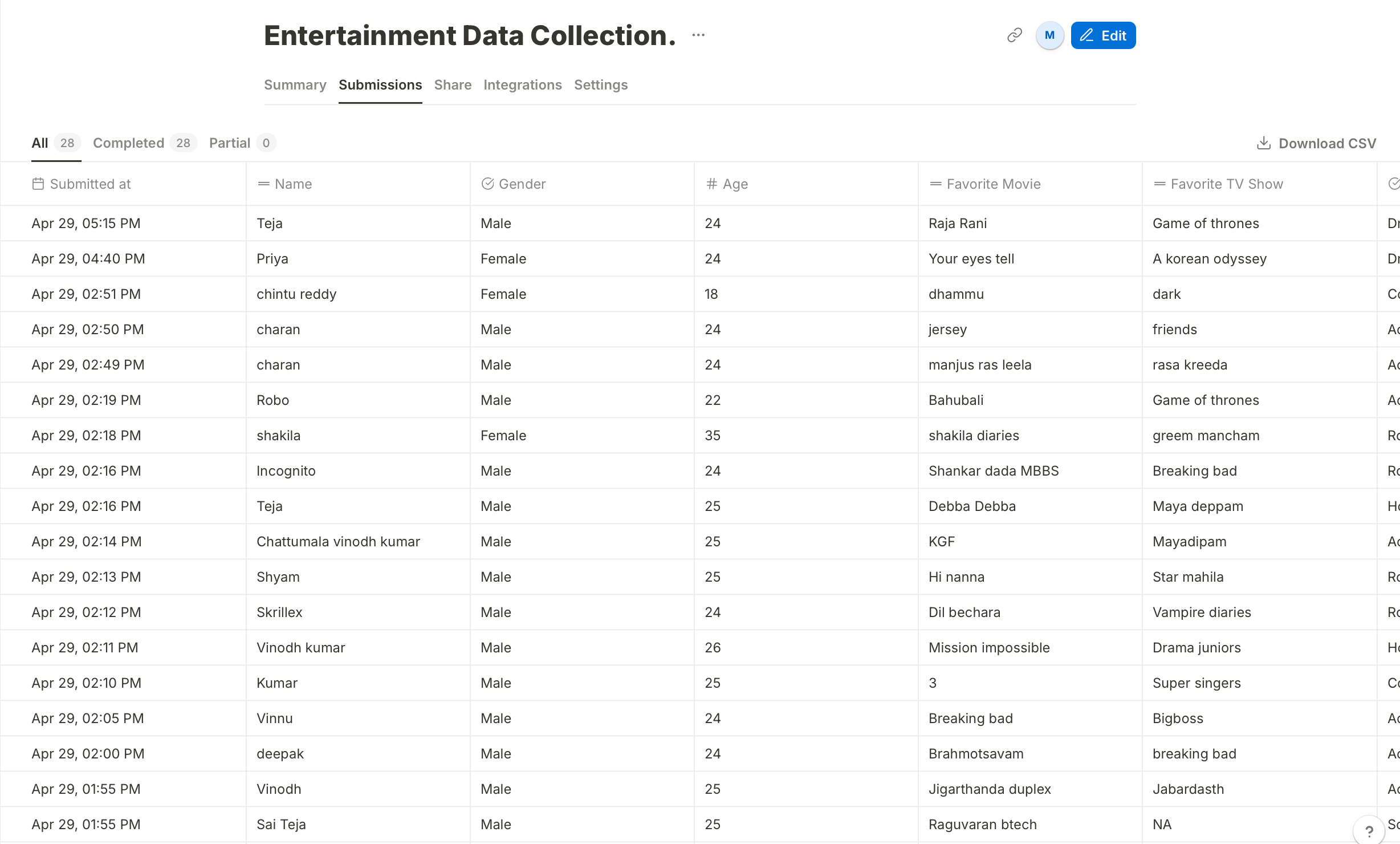
Final SQL Assignment - From Form to SQL: Designing and Analyzing a Complete Data Pipeline

# Data Collection Form

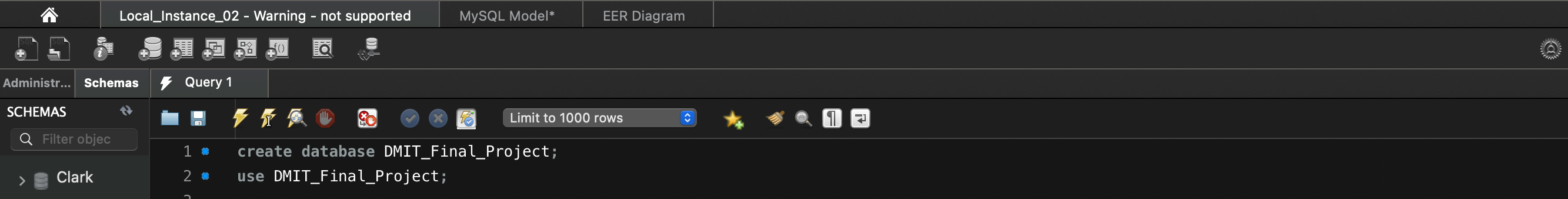
<https://tally.so/r/w2VzBD>

# Submissions from the formScreenshot 2025-04-29 at 20.44.54.pngScreenshot 2025-04-29 at 20.46.05.png



# CSV File used for the projectScreenshot 2025-04-29 at 20.47.39.png

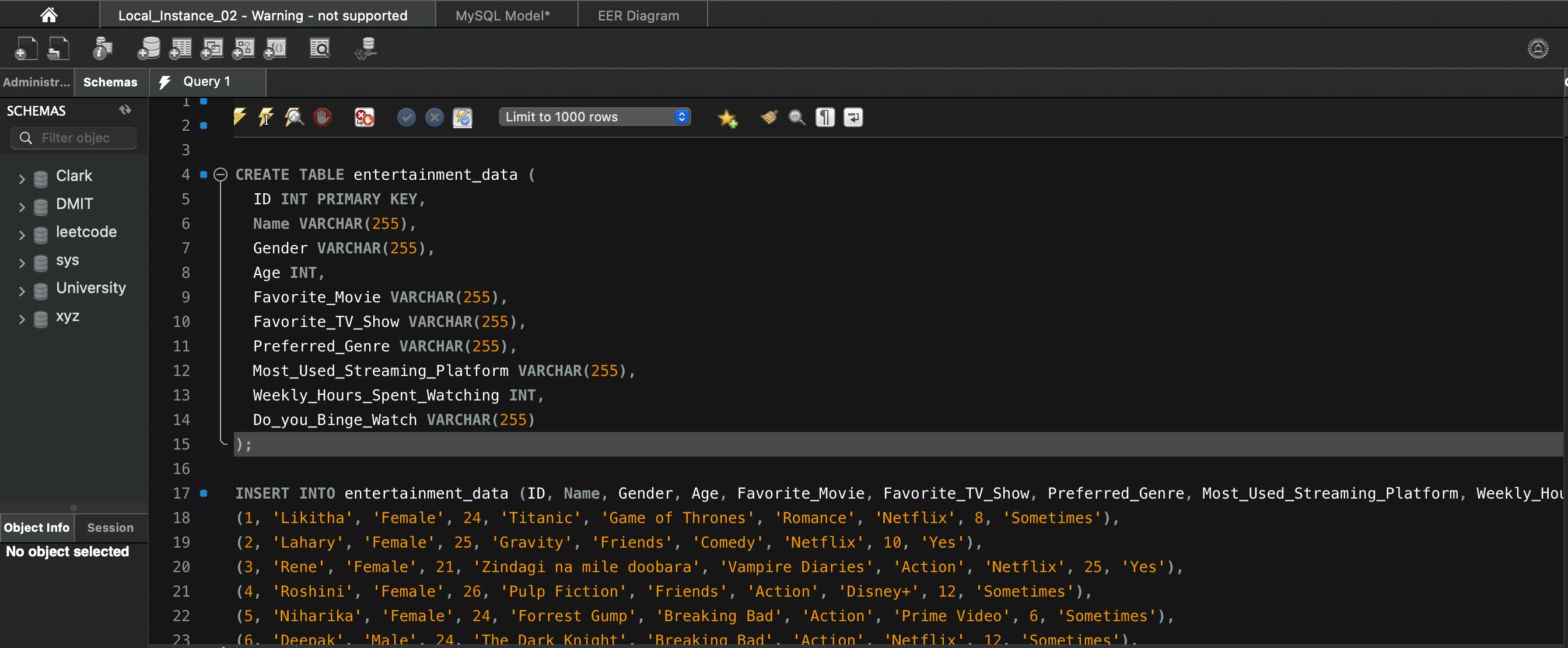
# SQL

I used MySQL Workbench to create a database for this project and used it to create the table and insert the data

Query :

*create database DMIT\_Final\_Project;*

use DMIT\_Final\_Project;

Created table to insert the data according to the CSV file.

Query :

CREATE TABLE entertainment\_data (

ID INT PRIMARY KEY,

Name VARCHAR(255),

Gender VARCHAR(255),

Age INT,

Favorite\_Movie VARCHAR(255),

Favorite\_TV\_Show VARCHAR(255),

Preferred\_Genre VARCHAR(255),

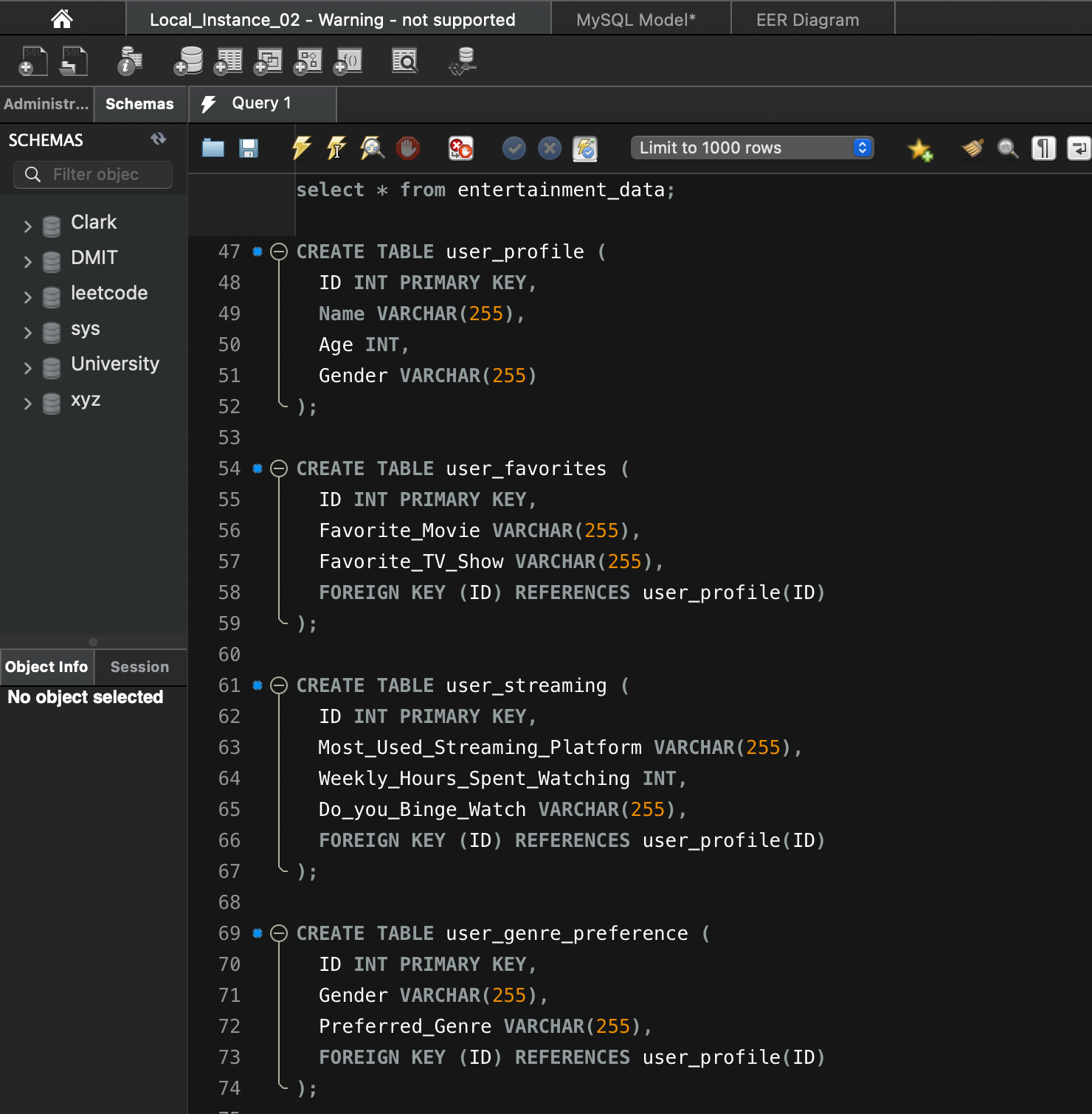
Most\_Used\_Streaming\_Platform VARCHAR(255),

Weekly\_Hours\_Spent\_Watching INT,

Do\_you\_Binge\_Watch VARCHAR(255)

);

Normalized into four tables

1. user\_profile
2. user\_favorites
3. user\_streaming
4. user\_genre\_preference

CREATE TABLE user\_profile (

ID INT PRIMARY KEY,

Name VARCHAR(255),

Age INT,

Gender VARCHAR(255)

);

CREATE TABLE user\_favorites (

ID INT PRIMARY KEY,

Favorite\_Movie VARCHAR(255),

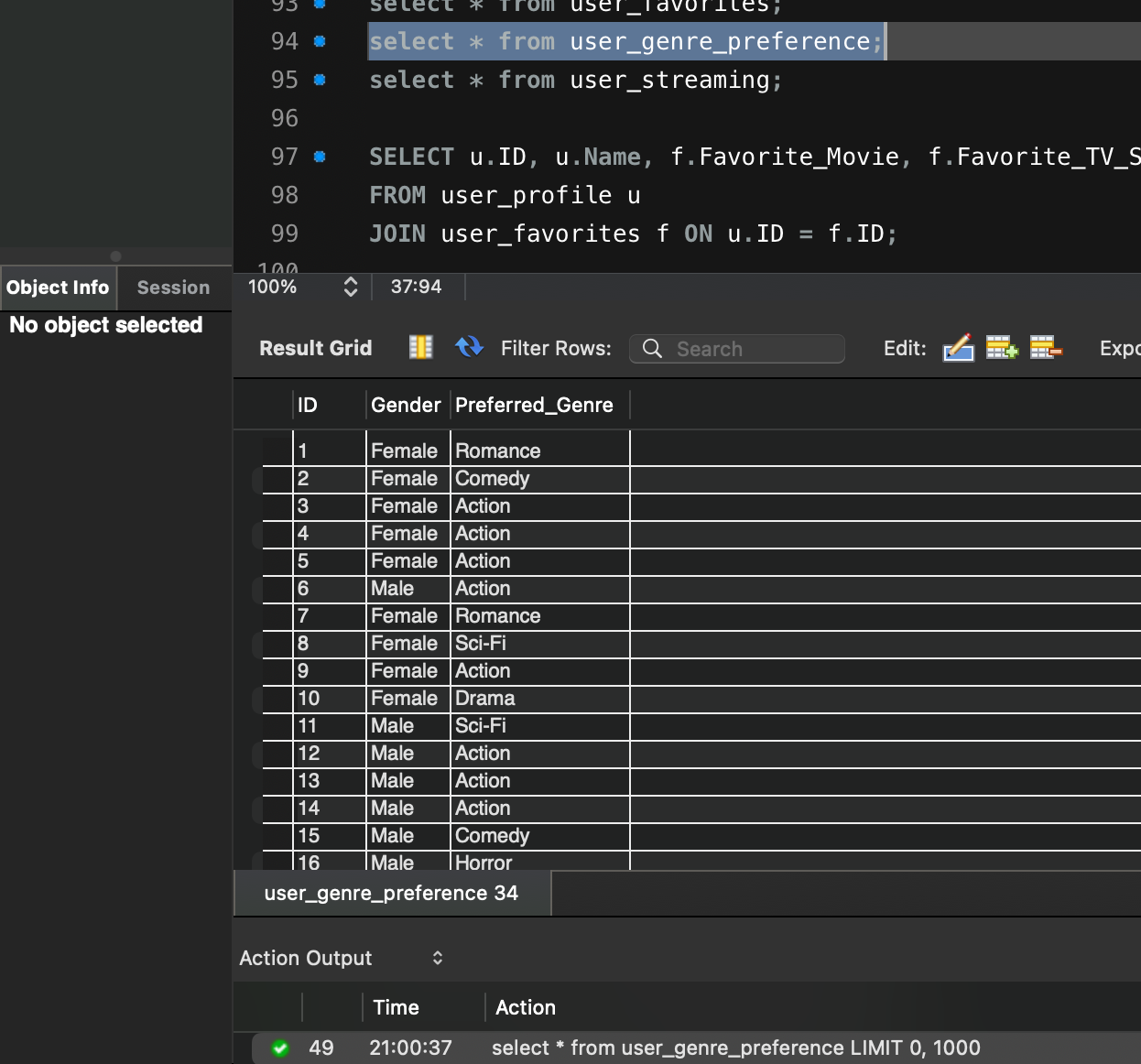
Favorite\_TV\_Show VARCHAR(255),

FOREIGN KEY (ID) REFERENCES user\_profile(ID)

);

CREATE TABLE user\_streaming (

ID INT PRIMARY KEY,

Most\_Used\_Streaming\_Platform VARCHAR(255),

Weekly\_Hours\_Spent\_Watching INT,

Do\_you\_Binge\_Watch VARCHAR(255),

FOREIGN KEY (ID) REFERENCES user\_profile(ID)

);

CREATE TABLE user\_genre\_preference (

ID INT PRIMARY KEY,

Gender VARCHAR(255),

Preferred\_Genre VARCHAR(255),

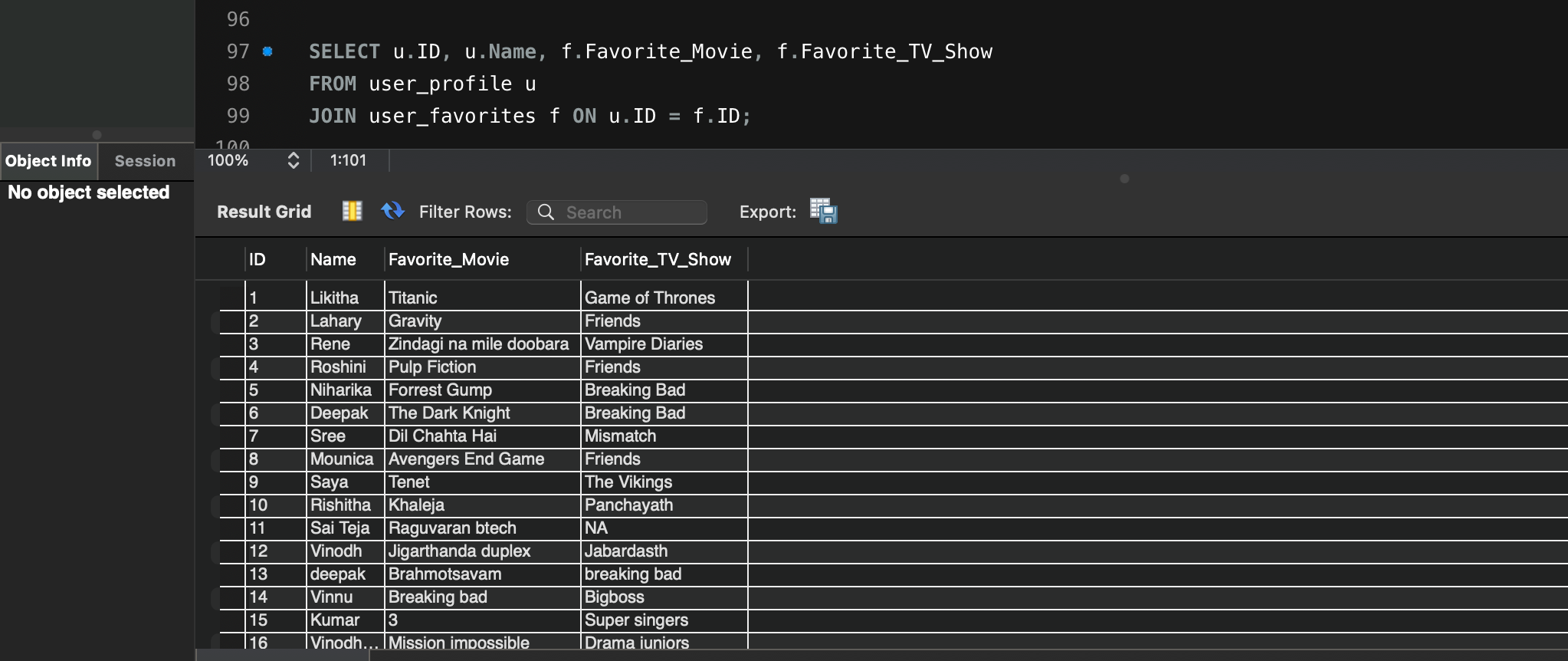
FOREIGN KEY (ID) REFERENCES user\_profile(ID)

);

# Join Queries

1. SELECT u.ID, u.Name, f.Favorite\_Movie, f.Favorite\_TV\_Show

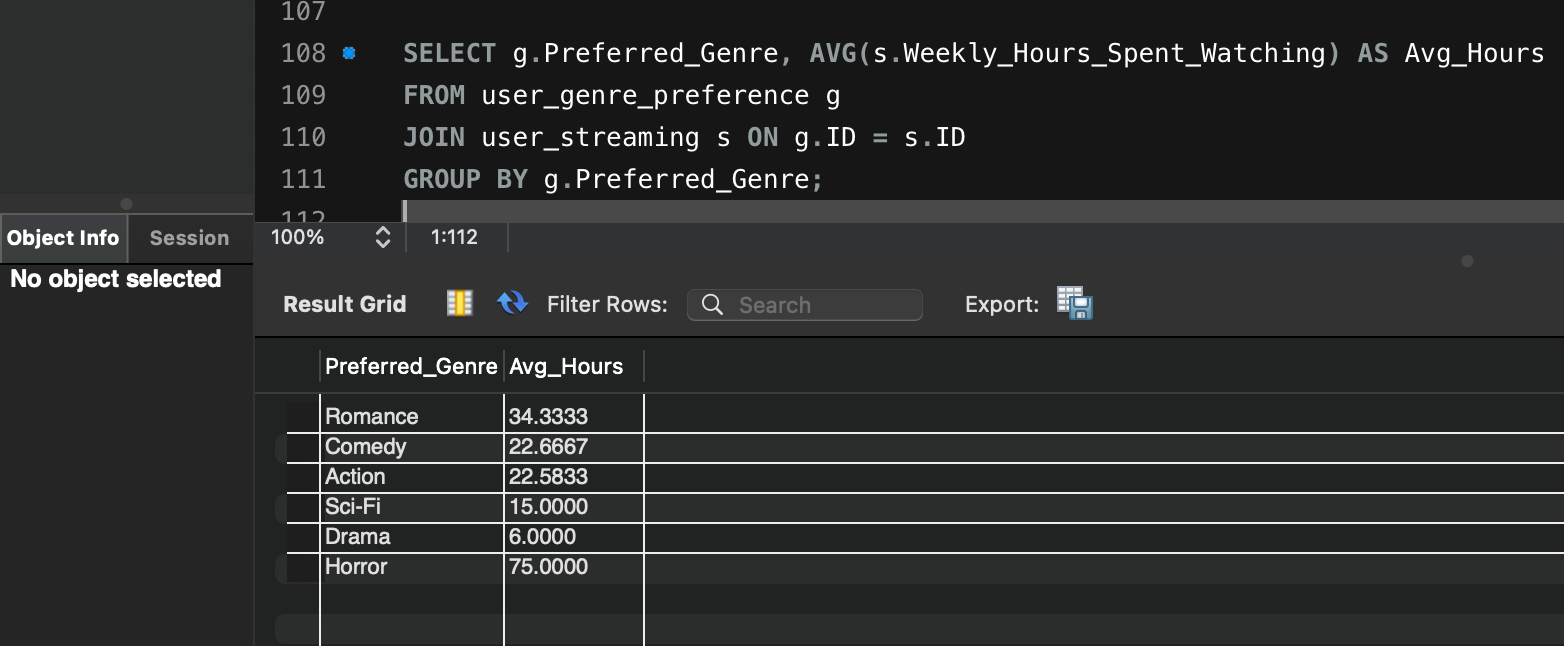
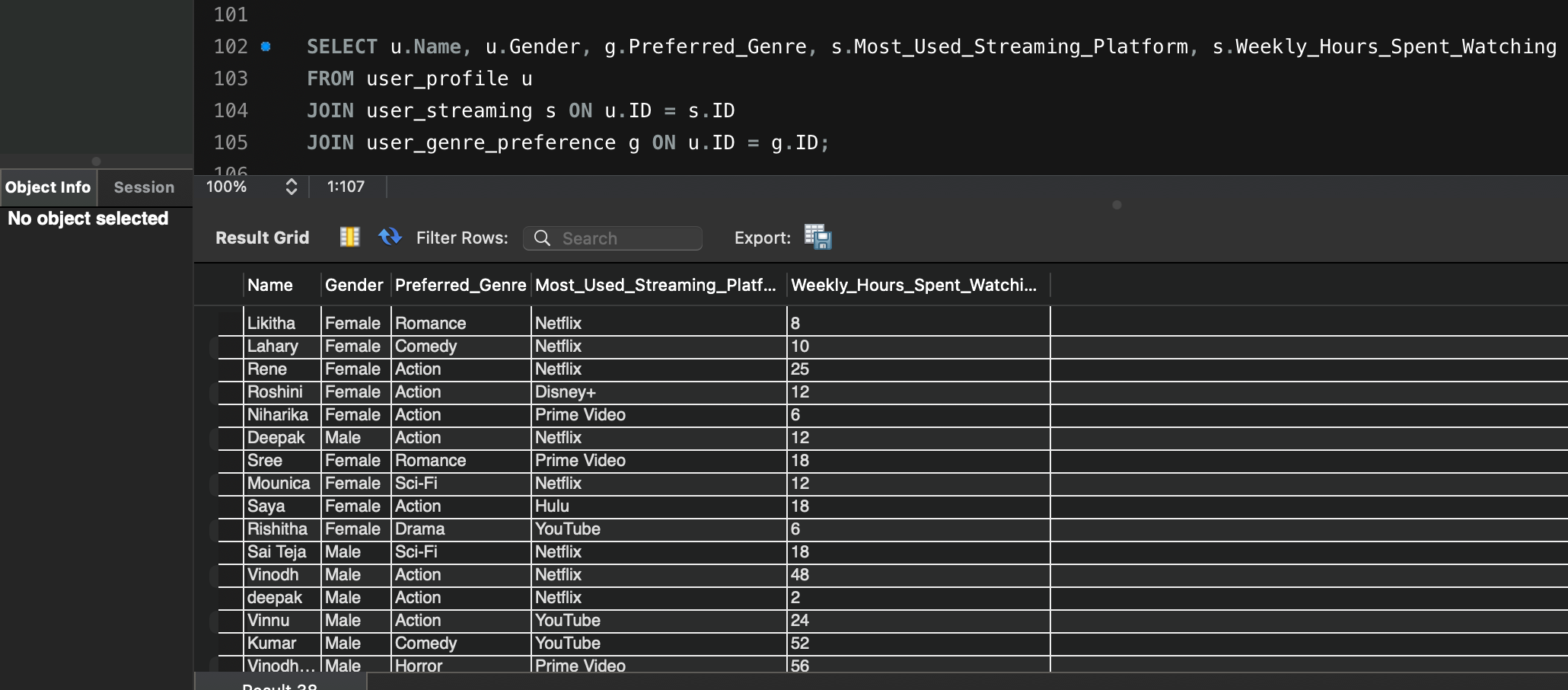
FROM user\_profile u

JOIN user\_favorites f ON u.ID = f.ID;

SELECT u.Name, u.Gender, g.Preferred\_Genre, s.Most\_Used\_Streaming\_Platform, s.Weekly\_Hours\_Spent\_Watching

FROM user\_profile u

JOIN user\_streaming s ON u.ID = s.ID

JOIN user\_genre\_preference g ON u.ID = g.ID;

# Group By Query

SELECT g.Preferred\_Genre, AVG(s.Weekly\_Hours\_Spent\_Watching) AS Avg\_Hours

FROM user\_genre\_preference g

JOIN user\_streaming s ON g.ID = s.ID

GROUP BY g.Preferred\_Genre;

# Case Query

SELECT u.Name, s.Weekly\_Hours\_Spent\_Watching,

CASE

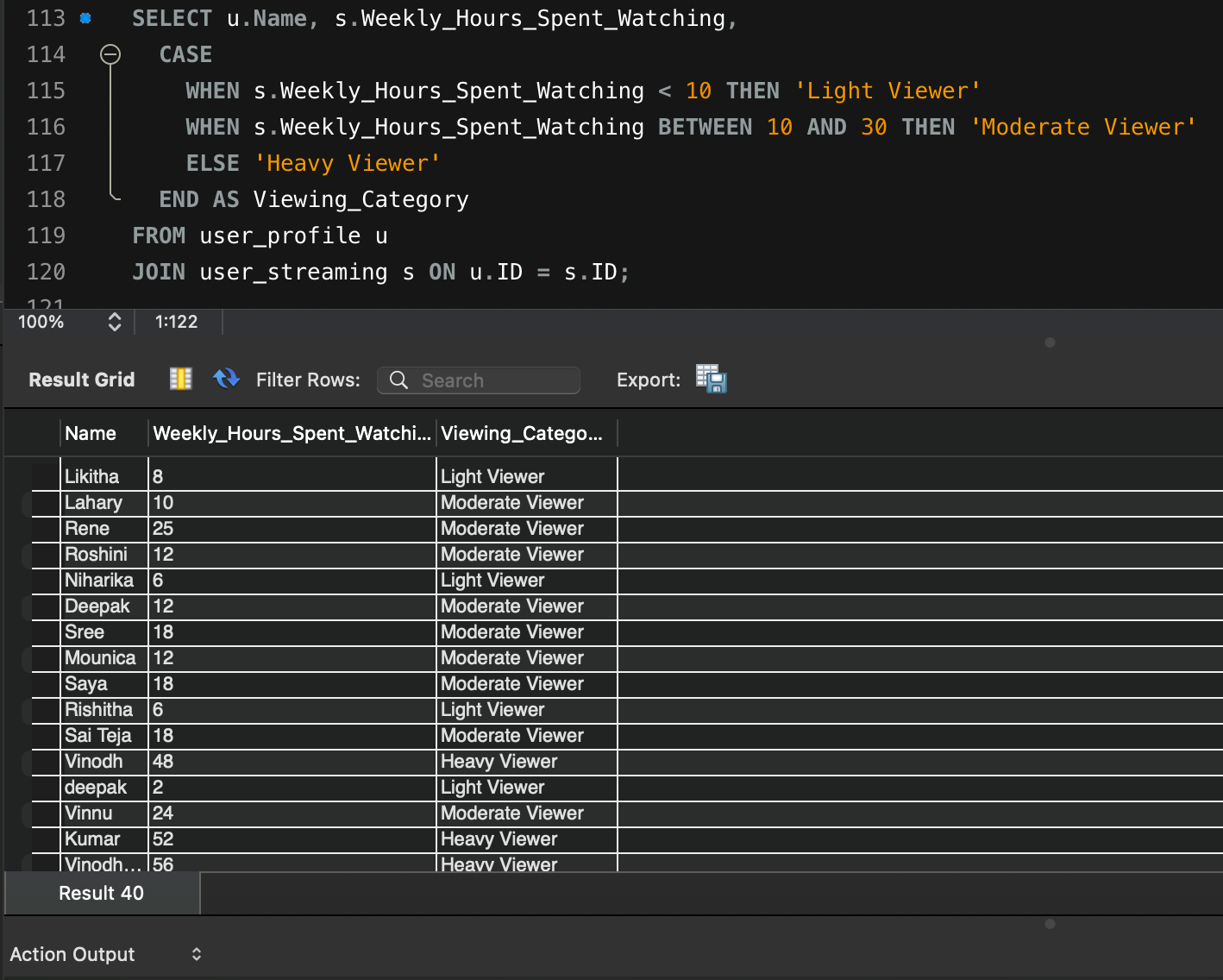
WHEN s.Weekly\_Hours\_Spent\_Watching < 10 THEN 'Light Viewer'

WHEN s.Weekly\_Hours\_Spent\_Watching BETWEEN 10 AND 30 THEN 'Moderate Viewer'

ELSE 'Heavy Viewer'

END AS Viewing\_Category

FROM user\_profile u

JOIN user\_streaming s ON u.ID = s.ID;

# View Query

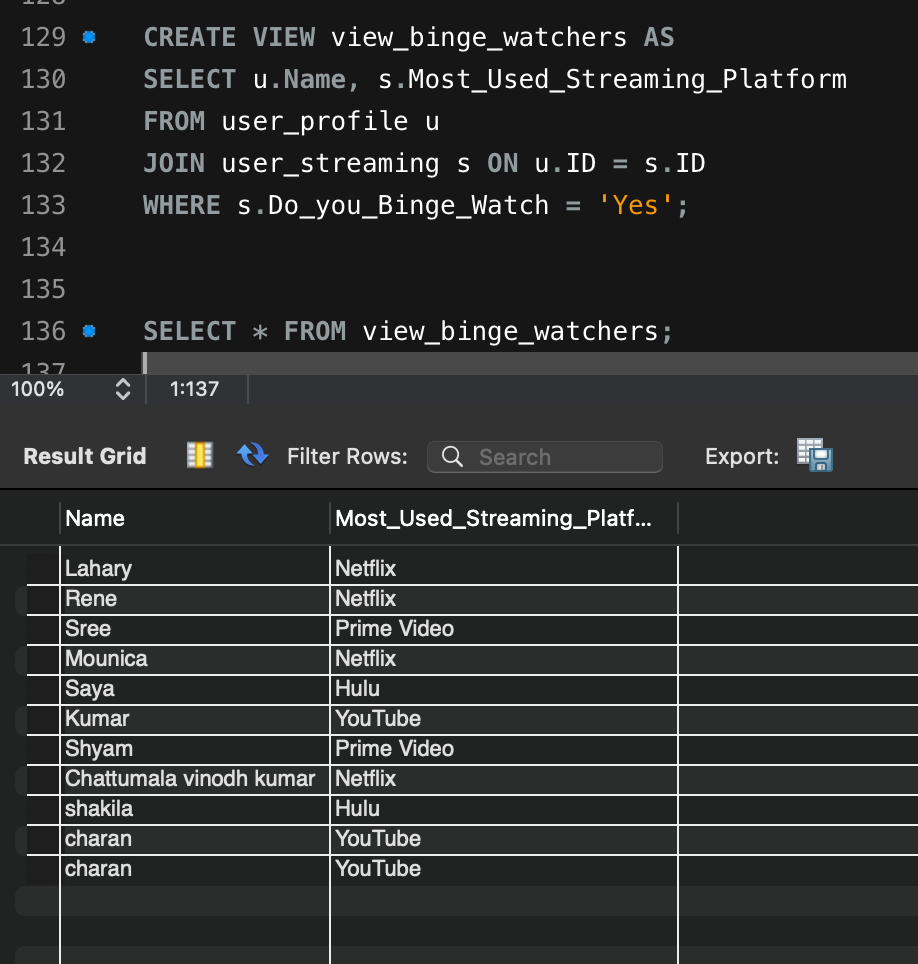
CREATE VIEW view\_binge\_watchers AS

SELECT u.Name, s.Most\_Used\_Streaming\_Platform

FROM user\_profile u

JOIN user\_streaming s ON u.ID = s.ID

WHERE s.Do\_you\_Binge\_Watch = ‘Yes';

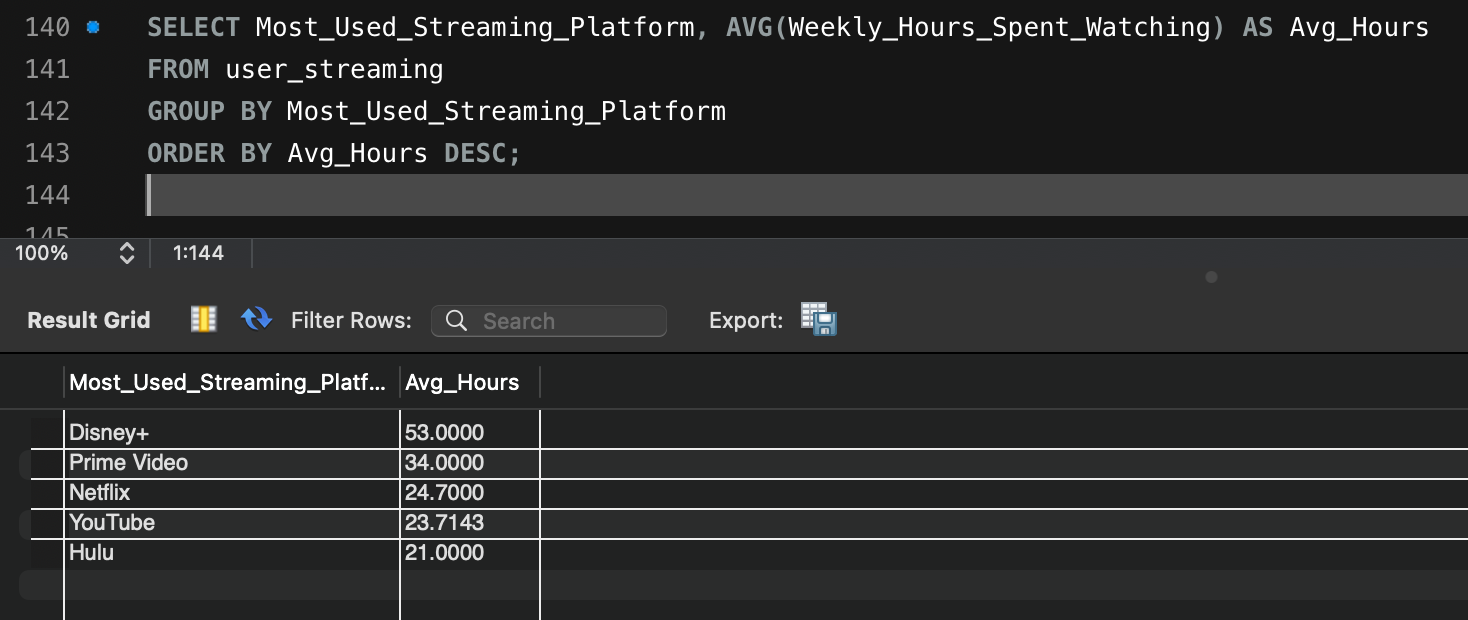
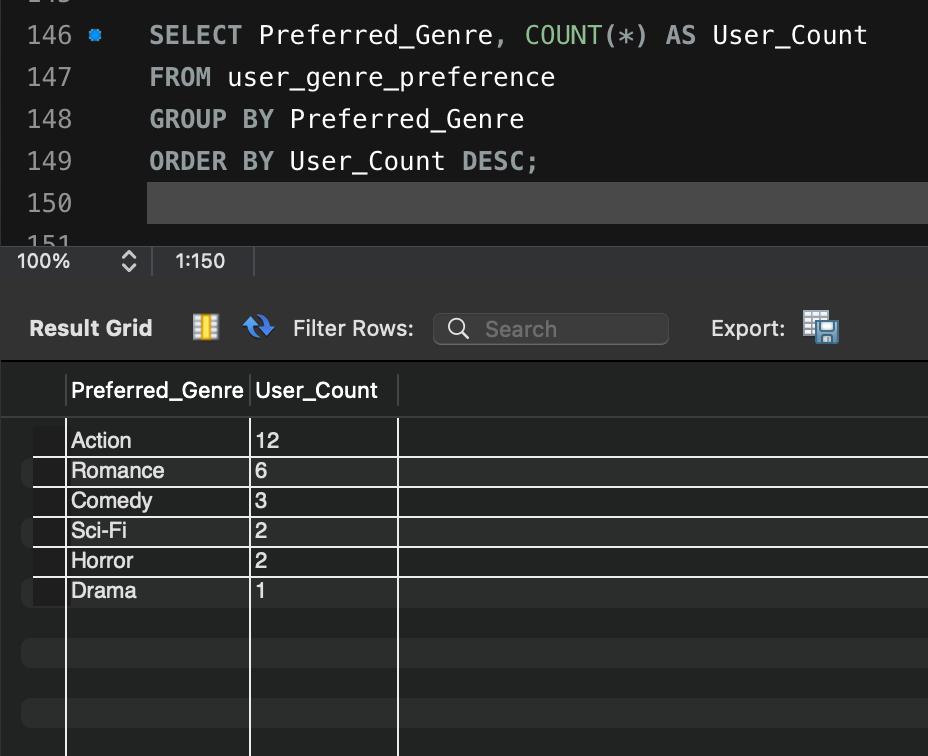
SELECT \* FROM view\_binge\_watchers;

# Analytical Queries

Query1 : SELECT Most\_Used\_Streaming\_Platform, AVG(Weekly\_Hours\_Spent\_Watching) AS Avg\_Hours

FROM user\_streaming

GROUP BY Most\_Used\_Streaming\_Platform

ORDER BY Avg\_Hours DESC;

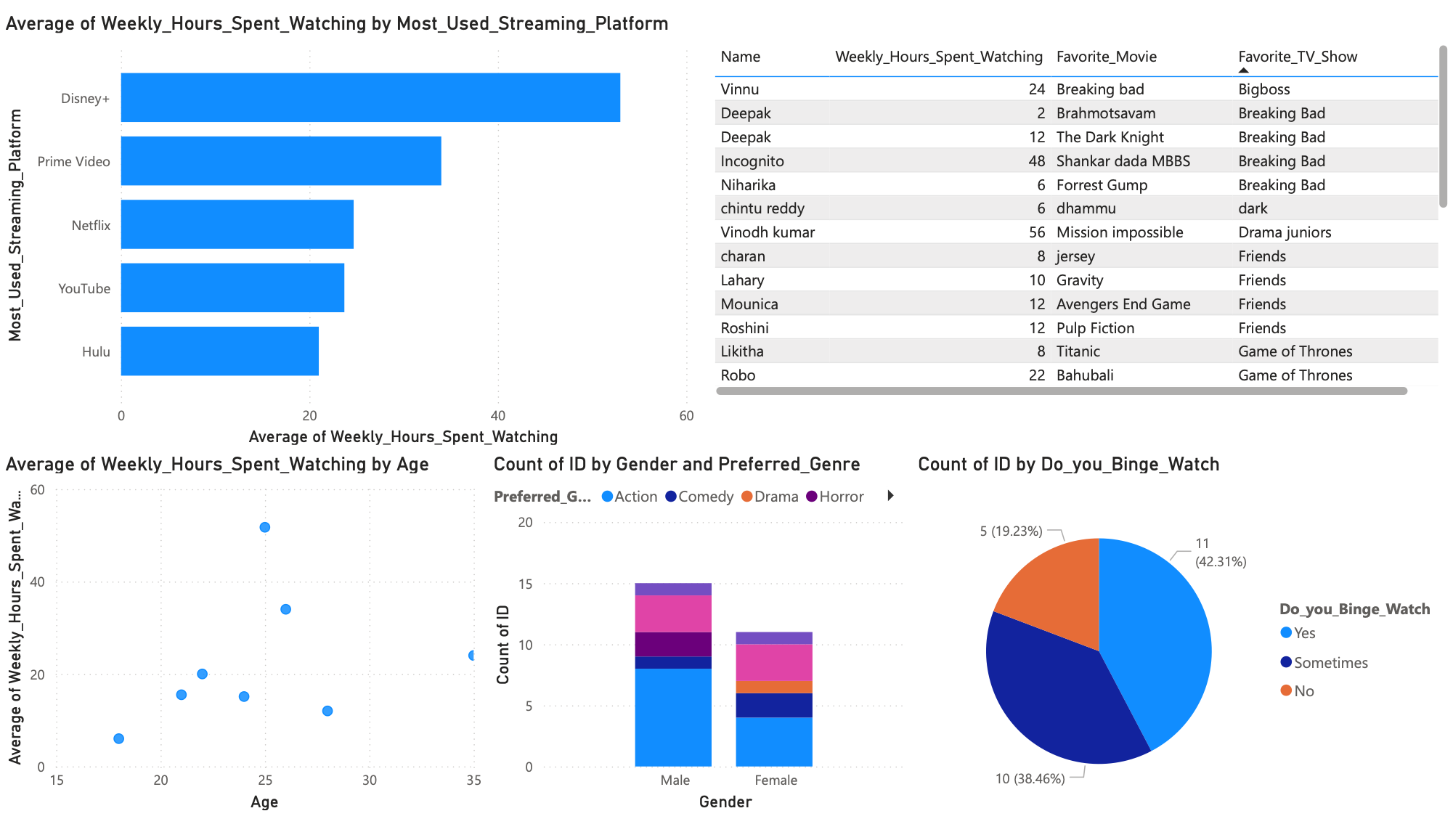
Query2 : SELECT Preferred\_Genre, COUNT(\*) AS User\_Count

FROM user\_genre\_preference

GROUP BY Preferred\_Genre

ORDER BY User\_Count DESC;

# PowerBi Dashboard based on the collected data

<https://app.powerbi.com/links/tBTj8iiJxx?ctid=b5b2263d-68aa-453e-b972-aa1421410f80&pbi_source=linkShare>

# Final Project Summary

1. What was your topic and goal for the data you collected?

My project focused on analyzing entertainment consumption habits which include binge-watching and streaming services consumption trends. It aimed to figure out how age, gender, preferred genres, and weekly hours of viewing correlate with users’ binge-watching habits through structured data analysis and SQL queries.

1. How did you organize the data into a relational schema?

The information was aggregated in a single flat dataset containing various attributes of the user’s activities and preferences. In order to enhance the maintenance of data organization and reduction of redundancy, I split the previously single table into four relational tables:

* 1. user\_profile
  2. user\_favorites
  3. user\_streaming
  4. user\_genre\_preference

1. What were your key SQL takeaways?

Through this project, I learned how to

* Use JOIN queries to extract insights across multiple tables.
* Apply GROUP BY with aggregate functions like AVG and COUNT.
* Write conditional logic using CASE statements.
* Create and query VIEWS to encapsulate repeated logic.
* Structure normalized tables and enforce relationships using foreign keys.

1. What challenges did you face during normalization or querying?

When I wanted to create a dashboard I decided to use PowerBi but the interface was so difficult for me to understand and it took me some time to understand. For scatter chart I was not able to give the data for the VALUES field.

1. What would you improve if you repeated this project?

I would do some PowerBi homework to understand how to create meaningful insights with the data I have. There is so much to learn in it. I feel like I didnt use it enough to create the dashboard.