

# TV Shows Ratings Survey

AUTHOR

Pascal Hermann Kouogang Tafo

## Introduction

This assignment was designed to help us learn how to use R to analyze a data set store in a SQL database. In my work, i will focus on creating a table that contains the rating of the 6 recent popular TV-Shows, store it in a database and load it in R as dataFrame.

## APPROACH

- In order to complete my assignment, i will use an LLM to help me write a code that will create the ratings table of **five people rating the most recent popular TV-Shows in a scale of 1 to 5** in a SQL database. To generate a more concise code, i will specify in my prompt the characteristics of selection of the shows which will include their relevance, popularity, and availability across major streaming platforms to ensure that the 5 people have already watched at least some, thereby increasing the probability of getting significant ratings. I will also make sure to have to add some missing value since people are selected randomly therefore they might not have watched every show.
- Once, the ratings table is stored in my SQL database, i will try to load the data from the database into my R environment and that will require me to install some R packages such as **"DBI", "RPostgres"** which are required when connecting to my SQL Database through my R environment.
- Finally, i will open the loaded data into a data frame and use **tidyverse functions** to clean, and validate before any potential analysis.

## Step1: INSTALL the useful R-packages

---

I installed **"DBI", "RPostgres"** directly in the console in order to render the file easily.

## Step 2: Load the useful R-packages

---

```
library(DBI)
```

Warning: package 'DBI' was built under R version 4.5.2

```
library(RPostgres)
```

Warning: package 'RPostgres' was built under R version 4.5.2

## Step 3: Connect to my Database

```
con <- dbConnect(  
  RPostgres::Postgres(),  
  dbname = "Pascal2026",  
  host = "localhost",    # or your server IP  
  port = 5432,           # default for PostgreSQL  
  user = "postgres",  
  password = Sys.getenv("DB_PASSWORD")  
)
```

## Table creation in PostgreSQL

```
CREATE TABLE tv_show_ratings (  
  rating_id SERIAL PRIMARY KEY,  
  user_id INT NOT NULL,  
  show_id INT NOT NULL,  
  show_name VARCHAR(100),  
  rating INT CHECK (rating BETWEEN 1 AND 5)  
);  
  
INSERT INTO tv_show_ratings (user_id, show_id, show_name, rating) VALUES  
(1, 1, 'Stranger Things', 5),  
(1, 2, 'Breaking Bad', 4),  
(1, 3, 'The Crown', 5),  
(1, 4, 'The Witcher', NULL),  
(1, 5, 'Money Heist', 4),  
(1, 6, 'Squid Game', NULL);  
  
INSERT INTO tv_show_ratings (user_id, show_id, show_name, rating) VALUES  
(2, 1, 'Stranger Things', 4),  
(2, 2, 'Breaking Bad', 5),  
(2, 3, 'The Crown', NULL),  
(2, 4, 'The Witcher', 3),  
(2, 5, 'Money Heist', 4),  
(2, 6, 'Squid Game', 5);  
  
INSERT INTO tv_show_ratings (user_id, show_id, show_name, rating) VALUES  
(3, 1, 'Stranger Things', 5),  
(3, 2, 'Breaking Bad', NULL),  
(3, 3, 'The Crown', 4),  
(3, 4, 'The Witcher', 4),  
(3, 5, 'Money Heist', NULL),  
(3, 6, 'Squid Game', 5);  
  
INSERT INTO tv_show_ratings (user_id, show_id, show_name, rating) VALUES  
(4, 1, 'Stranger Things', NULL),  
(4, 2, 'Breaking Bad', 5),
```

```
(4, 3, 'The Crown', 4),
(4, 4, 'The Witcher', NULL),
(4, 5, 'Money Heist', 3),
(4, 6, 'Squid Game', 4);
```

```
INSERT INTO tv_show_ratings (user_id, show_id, show_name, rating) VALUES
(5, 1, 'Stranger Things', 4),
(5, 2, 'Breaking Bad', 5),
(5, 3, 'The Crown', NULL),
(5, 4, 'The Witcher', 3),
(5, 5, 'Money Heist', 4),
(5, 6, 'Squid Game', NULL);
```

## Step 4: Load Table into R Data Frame

```
TV_Shows_Ratings_df <- dbReadTable(con, "tv_show_ratings")
```

```
TV_Shows_Ratings_df
```

	rating_id	user_id	show_id	show_name	rating
1	1	1	1	Stranger Things	5
2	2	1	2	Breaking Bad	4
3	3	1	3	The Crown	5
4	4	1	4	The Witcher	NA
5	5	1	5	Money Heist	4
6	6	1	6	Squid Game	NA
7	7	2	1	Stranger Things	4
8	8	2	2	Breaking Bad	5
9	9	2	3	The Crown	NA
10	10	2	4	The Witcher	3
11	11	2	5	Money Heist	4
12	12	2	6	Squid Game	5
13	13	3	1	Stranger Things	5
14	14	3	2	Breaking Bad	NA
15	15	3	3	The Crown	4
16	16	3	4	The Witcher	4
17	17	3	5	Money Heist	NA
18	18	3	6	Squid Game	5
19	19	4	1	Stranger Things	NA
20	20	4	2	Breaking Bad	5
21	21	4	3	The Crown	4
22	22	4	4	The Witcher	NA
23	23	4	5	Money Heist	3
24	24	4	6	Squid Game	4
25	25	5	1	Stranger Things	4
26	26	5	2	Breaking Bad	5
27	27	5	3	The Crown	NA
28	28	5	4	The Witcher	3

29	29	5	5	Money Heist	4
30	30	5	6	Squid Game	NA

## Group and Order the data frame by User\_id

```
TV_Shows_Ratings_df <- "
SELECT
  user_id,
  MAX(CASE WHEN show_name = 'Stranger Things' THEN rating END) AS Stranger_Things,
  MAX(CASE WHEN show_name = 'Breaking Bad' THEN rating END) AS Breaking_Bad,
  MAX(CASE WHEN show_name = 'The Crown' THEN rating END) AS The_Crown,
  MAX(CASE WHEN show_name = 'The Witcher' THEN rating END) AS The_Witcher,
  MAX(CASE WHEN show_name = 'Money Heist' THEN rating END) AS Money_Heist,
  MAX(CASE WHEN show_name = 'Squid Game' THEN rating END) AS Squid_Game
FROM tv_show_ratings
GROUP BY user_id
ORDER BY user_id;
"

TV_Shows_Ratings_df <- dbGetQuery(con, TV_Shows_Ratings_df)

TV_Shows_Ratings_df
```

```
  user_id stranger_things breaking_bad the_crown the_witcher money_heist
1       1              5           4           5          NA           4
2       2              4           5          NA           3           4
3       3              5          NA           4           4          NA
4       4             NA           5           4          NA           3
5       5              4           5          NA           3           4
  squid_game
1          NA
2           5
3           5
4           4
5          NA
```