

Assignment (II)

Due: March 26, 2023

Assume that you are an employee at a gaming company. Your manager asks you to survey the gaming market. You find a comprehensive dataset from Kaggle¹ about Steam, a video game digital distribution service and a storefront by Valve, and import it to the database server in your company. Now you start to write SQL queries to analyze the data.

- **Environment Setup:** Make sure that you have PostgreSQL installed and read *Tips on using PostgreSQL* carefully. You should execute `ddl.sql` and `data.sql` in PostgreSQL interactive terminal to create tables and import the dataset. Here we only provide you with a small dataset, but your submission will be graded on a larger dataset, which means **you should ensure that your SQL queries can handle possible corner cases correctly.**
- **About the tables:** There are five tables in total and you can check their definitions in `ddl.sql`. **The columns not marked with 'NOT NULL' can have NULL values.**
- **About the submission:** You should write your SQL query in a **separate** file for each question. The file should be named as `i.sql`, where 'i' is the question number. For example, your answer to question 1 should be written in `1.sql`. When you complete all the questions, you should compress all the `.sql` files into a single ZIP named `submission.zip` and upload it to Canvas. **You should also include the given example 1.sql in your submission!**

Let's get it started!

1. **EXAMPLE:** List the names of games lacking English support.

```
SELECT
    name
FROM
    app
WHERE
    NOT english;
```

2. List the names of games that children aged 14 can play and whose ratio of negative ratings is less than 15%.
3. List the names of free games that have the top 5 positive ratings.
HINT: The top three of (1, 2, 3, 3, 3) should be (3, 3, 3).
4. List the websites of games available on Linux or Mac. Each website appears only once and NULL values should be skipped.
5. List the names of all companies along with the corresponding average prices of games they have published.
6. List the names of companies that have published more than 5 games between the year 2005 and 2015 (left inclusive and right exclusive).

HINT: In addition to the basic data types we introduced in the lectures, the SQL standard supports several data types relating to dates and times. The most basic one is the `DATE` type.

¹<https://www.kaggle.com/datasets/nikdavis/steam-store-games>

A value representing the date March 6, 2023 can be written as `DATE '2023-03-06'`. Furthermore, SQL allows comparison operations on the `DATE` type. We can write the following query to list all apps that have been released before March 6, 2023.

```
SELECT * FROM app WHERE release_date < DATE '2023-03-06';
```

Apart from the `DATE` type, SQL also supports date types `TIME` and `TIMESTAMP`. For more details about the date/time data types and their support in PostgreSQL, please refer to <https://www.postgresql.org/docs/current/datatype-datetime.html>.

7. List the names of companies that have never developed games more expensive than 50\$.
8. List the names of games that have won Game of the Year.
HINT: This query is related to table 'description'. The format of the text descriptions is a bit messy. You may use the `LOWER` function to convert them to lowercase characters, e.g.,

```
SELECT * FROM app WHERE LOWER(name) LIKE 'a lowercase name';
```


For more details about string functions support in SQL, please refer to <https://www.postgresql.org/docs/current/functions-string.html>.
9. Find the companies that have the top 3 positive ratings and list their names and corresponding positive ratings. (The positive ratings of a company are the sum of positive ratings of the games they either developed or published.)
HINT: The top three of (1, 2, 3, 3) should be (3, 3, 2).
10. You have spent much time writing SQL queries and analyzing the data, which makes you exhausted. Now you want to play some games listed in the table to have some entertainment. You plan to buy three games and your budget is 100\$. Print the number of your possible choices. (NOTE: you have no interest in free games!)