

Project Title: Movie Review Analysis**Marks: 30**

IMBD is an online database of movie-related information. IMBD users rate the movies and provide reviews. They rate the movies on a scale of 1 to 5; 1 being the worst and 5 being the best. The dataset also has additional information, such as the release year of the movie. You have to analyze the data collected and answer the following questions:

You need to find:

- 1) The total number of movies
- 2) The maximum rating of movies
- 3) The number of movies that have maximum rating
- 4) The movies with ratings 1 and 2
- 5) The list of years and number of movies released each year
- 6) The number of movies that have a runtime of two hours

Dataset: Project_Dataset_movie

Please use the data and visualize the relevant information which do you think need a Business Analyst.

Project Title: Customer Analysis**Marks: 50**

You have to find:

- 1) The date ranges from employee's table where date range means employee start date and end date
- 2) Employees start date by month (You can use MONTH and group by function)

Hint: select MONTH(employees.startdate), year(employees.startdate), count(1)
from employees
group by MONTH(employees.startdate), year(employees.startdate)
order by 2, 1

- 3) Please find the active employees on a given date = '2008-07-01' (You can use operators/start and end date/NULL function)

- 4) Check maximum and min dates:

For this step you need to check date table and employee's table. You have to check how many employees were active in the given date in date table.

Select min (dates.datekey) , max (dates.datekey)
From dates

Here you can see the oldest and newest dates from the dates table. You have to check which columns are important for the analysis from step 1 to end steps of your task and you can delete the columns accordingly. If you want to work with pgAdmin it may help you. You can also use <https://data.world/schedule-a-demo/> if you are getting problem with pgAdmin platform. But you have to check how you can add Tableau in that case. Otherwise making Tables and columns on pgAdmin and then importing the data will be easier for you.

5) Check the datekey those that have an enddate. That means there is a datekey in dates table, start date and end date in the employee's table. If you see the end date is available that means the employee is not active.

Hint:

```
select dates.datekey, employees.startdate, employees.enddate, count(1)
from employees cross join dates
where dates.datekey BETWEEN employees.startdate and employees.enddate
group by dates.datekey, employees.startdate, employees.enddate
```

We have to use cross Join because the name of the table is changed, and you can use this function to check the result together.

6) Add check statement for those without end-dates

```
select dates.datekey, employees.startdate, employees.enddate, count (1)
from employees cross join dates
where dates.datekey >= employees.startdate
and (dates.datekey <= employees.enddate
    OR
    employees.enddate is null)
group by dates.datekey, employees.startdate, employees.enddate
order by dates.datekey
```

The above steps will help to get the overview of the data where customers were still active at the company or not. You can use any platform to analyse the data which can support cross Join function. Data world can be an option where you can work easily. For the first project you can use Databricks too or may any other platform to give the answer.

After completion of all steps please check the output and according to the output use Tableau for visualization. We already know how to connect Tableau with pgAdmin. In case you are using any other platform try to extract the table from the output code. Hints of the codes are given too. If you are unable to extract it, then according to the information you can separately use the Tableau without connecting with DB. In that case you have to import complete data sets and you need required columns to visualize it.