# Assignment for the Lead Data Engineer position

## Task 1

You are working for a business that owns two DVD rental stores. Both stores use software to manage the DVD rentals, but the software has no reporting capabilities.
The owner wants to better understand various aspects of his business, and has asked you a few questions.
In order to answer his questions, you will design and execute a few queries on the database of the DVD rental software that is used by the stores (there is one database for both stores).
You can view the database schema at the following link:
http://www.postgresqltutorial.com/postgresql-sample-database/
This page also contains instructions to download the full database and restore it in a PostgreSQL instance that you will set up.

Please provide the SQL queries that you will use to answer the owner's questions, as well as the actual answers.

Tip: You may receive a few errors on alter table statements when restoring the database, which can be ignored.
Tip: You may find it useful to use the WITH clause.

1. Which customer has made the most rentals at store 2?
2. What percentage of movies were out of stock at each store on 29/7/2005 at midnight?
3. Is the employee that performs the rental of a DVD usually the one who also takes the Payment?
4. How many rentals do we do per month?
5. What percentage of our customers are active at any given month? We define active as performing at least one rental during that month.
6. Are there some films that are particularly popular and are rented all the time, or do people tend to spread their choice evenly among the available films?
7. Which film category is the most popular among our customers?
8. Are there any other insights that you can gather from the data that would be helpful to the owner of this business?

### Deliverables

For each question, please provide the following:

- Answer
- SQL query
- For question 8, also explain the way in which the insights that you provide will be useful to the business

# Task 2

The objective of this assignment is to implement an application that will retrieve and store movie information.

The application will use the API from The Movie DB (https://developers.themoviedb.org/3/getting-started). Use the following API key: bbb0e77b94b09193e6f32d5fac7a3b9c or create a new one on the site.

The application must retrieve a list of the movies currently in theaters in a configurable list of countries (initially United States, United Kingdom, Canada, Greece).

All the above information will be stored in a relational database. The database will be updated each time the process runs.
The process must update the database with new data once per day.

You should aim to design a well-formed database schema. It should be possible to answer at a minimum the following questions by querying the database (you do not need to provide the queries or the answers):

- Which movies were playing on a particular date in a particular country? [Assume that the question will not be asked for dates before the process started to be regularly executed]
- Which movies are directed by a particular director?
- How many movies of a particular genre are now playing in a particular country?
- Which review authors write the most reviews per country?
- In which country are most movies produced?
- What was the popularity of a particular movie on a particular date?

Lastly, the final requirement is that on the first day of each month, a table is appended with the count of different movies that were playing during the previous month per country (i.e. the table columns are yearmonth|country|numberofmovies).
Implement a stored procedure or a software process that updates this table.

You are free to include any additional features / optimizations that you may find relevant or could showcase your skills but please bear in mind that you should cover the core requirements first before attempting any improvements. Also note that although it may be tempting to use ready-made libraries or gems for querying TMDB we would prefer you to make direct API requests.

## Deliverables

The final deliverable should contain:

- Source code
- DDL for Database schema
- A simple Readme.txt that will describe the way to build and use the application

## Notes

You will need to provide a working version of the application. You are free to implement your project in any technology stack you prefer between Python/Perl/PHP/Ruby/Any scripting language/Java, and also to use any tools that you see fit.

You will be assessed on the following:
a) Fitness for purpose (i.e. your application does what its specifications require)
b) Simplicity (the simpler / smaller the solution, the better)
c) Robustness (how much “production-ready” your solution is)
d) Code quality