**BI/Database Developer**

**Challenge**

**Description**

You work in a movie theatre chain and you are in charge of the company’s data pipeline and storage. You are provided with the company’s transactional data from 2016 and you are given the task to store it in a data warehouse(DWH) towards making it ready for analysis. The available data concerns the chain’s ticket sales, the projected movies, as well as the two cinemas and screens. Your sources of data include:

* loyalty\_program\_users.csv – Contains a list of all clients who are part of the company’s loyalty program.
* loyalty\_sales.xlsx – Includes all the sales that took place by clients in the loyalty program and are performed exclusively by the online reservation platform.
* all\_tickets\_sold.xlsx – All the ticket sales from people that are not subscribed to the loyalty program and take place in person by a cashier.
* tickets\_list.csv – The price catalogue depending on the movie format.
* [TheMovieDB](https://www.themoviedb.org/)- A movie website with an open [API](https://developers.themoviedb.org/3/getting-started) where you can freely access movie-related data.

You are asked to access all available data sources, transform and clean the data, model it as you think best, and then store it appropriately in a DWH to be subsequently analysed.

The main KPIs that the company management is interested in is sales (€) and number of clients. The data should be modelled and stored in the DWH in a way that makes answering the following questions as easy as possible:

* What were the total sales per week, per month, per movie, per projection format, per cinema and screen?
* What were the most used promotions? What was the age group and sex of the people that used them the most?
* What were the most popular movie genres?
* Did the movies’ runtime, budget, or online vote average have any effect on the sales and the number of clients?
* How did the sales evolve in the days after each movie’s release date?

The movies’ genres, release date, runtime, budget, and vote average can be accessed via the movies API. The rest of the data is available in the enclosed file exports. Please keep in mind that the sales data was automatically generated and will not provide any real-life insights, if analysed.

The requested deliverables are:

1. All the scripts implemented in order to gather, transform, model, and load the data.
2. The ER diagram of the developed data warehouse schema.
3. The database export. You may use PostgreSQL, SQLite, or a similar database to play the role of the DWH.