Data Engineering Use Case

The automotive industry has a very large model base that keeps growing, and we at Beev want to try and get a better understanding of the market shares in a few key countries.

We have extracted a bunch of sample data on which we would like to run a few tests before we generalize the solution.

In the following exercise you are asked to:

1. Create a database that can handle the sample data. You are provided with a simple docker file that allows you to run a virtual database on your pc. You need to install docker desktop first, then run the docker-compose commands “up -d” and “down” to start and shut down the database. You can try to connect to the database with the pgAdmin tool.

2. Once the database is up and running we need a python script that will be able to read the provided csv files and insert the data into the database. The data model is up to you to decide.

3. Once the data is in the database, we want to run a few sql queries to check some basic data quality:(write the answers in a .sql file)
    a. We want to find the total number of cars by model by country.
    b. We want to know which country has the most of each model
    c. We want to know if any model is sold in the USA but not in France.
    d. We want to know how much the average car costs in every country by engine type.
    e. We want to know the average ratings of electric cars vs thermal cars

4. Bonus: Write a script that can read data from the database and return a graph that would show the amount of electric vs thermal cars sold per year.

You are free to take all the technical decisions you deem necessary to fulfill this exercise, all you need to do is be able to justify your decision later on.

Hope you have fun doing this !

For any question plz contact us by mail on Welcome To The Jungle.