*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. As a Beev Data Engineer, you are asked by the Sales team to aggregate their numbers into a robust environment where they can easily scale and enrich the existing data, they provide you with two csv files. 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 or DBeaver.*

**Steps to Implement the Database:**

**1**. First we download docker:***https://www.docker.com/products/docker-desktop/***

**2**. In the terminal CMD:

We navigate to the folder where **docker-compose.yml** file is located:

***cd path\desktop\amarildo\docker-compose-file***

**3**.To start the database (and download the files needed):

***docker compose up –d***

**4**.Test: ***docker compose version***

**5**. TEST:database container is running:***docker ps***

### **Connect to the Database Using DBeaver**

DBeaver (For Developers/Data Analysts: more powerful and flexible).

**1**.Downolad DBeaver:***https://dbeaver.io/download/***

**2.New Database Connection** → **PostgreSQL**.

Details(docker-compose.yml file):

* 1. **Host**: localhost
  2. **Port**: 5432
  3. **Database**: test\_db
  4. **Username**: admin
  5. **Password**: admin

------------------------------------------------------------------------------------------

*In the CMD:To run the python script we have created*

*1.Download Python:* https://www.python.org/downloads/

py -m pip install pandas

py -m pip install psycopg2

-------------------------------

py insert\_data.py

*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 and should be flexible enough to easily understand, scale & enrich the data.*