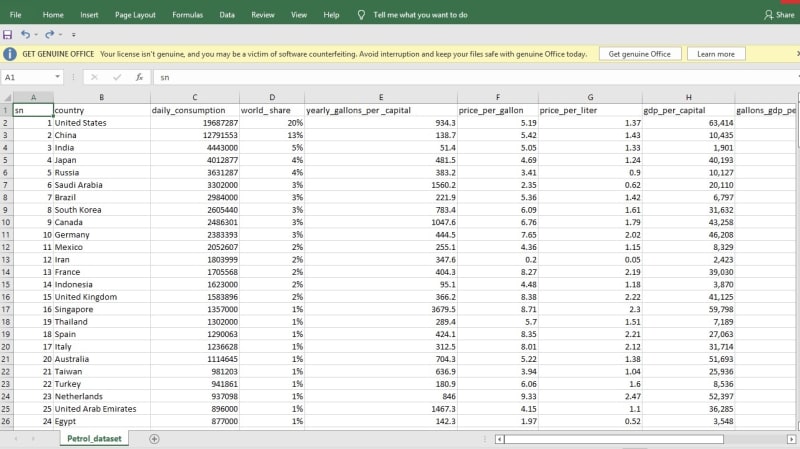
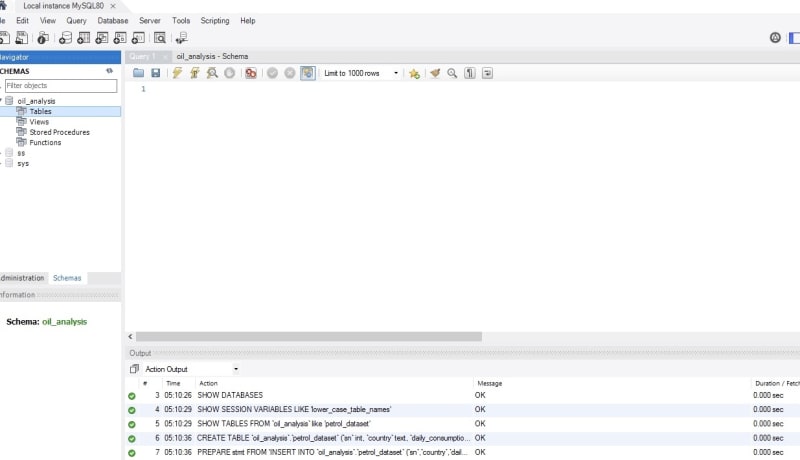
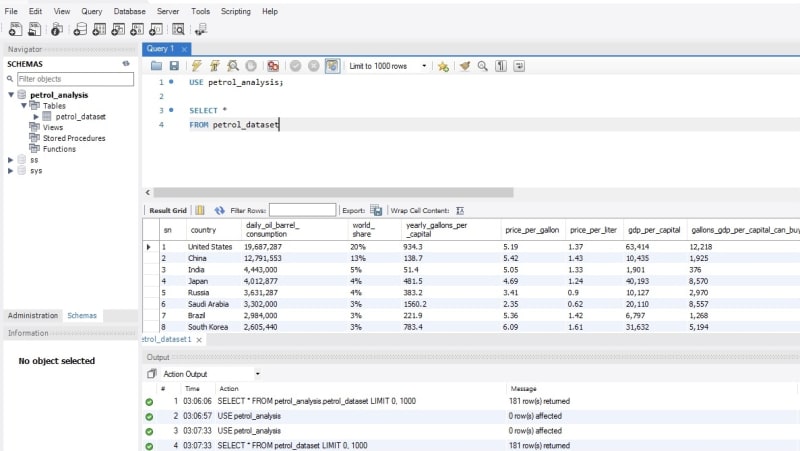
# DATAWAREHOUSING WITH IBM CLOUD DB2 WAREHOUSE

# Exploratory Data Analysis Using SQL.

The dataset was first loaded into Microsoft Excel in order to “clean” the dataset, rename the columns, remove outliers and to check for consistency in the dataset.  
Here is a picture of the dataset in Microsoft Excel:  
[](https://res.cloudinary.com/practicaldev/image/fetch/s--PIpRVqzf--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/j5j4hjwuglk94s39o115.jpeg)

A database was then created in MYSQL, and the dataset was imported into the SQL workbench in order to begin analysis.  
Picture of the dataset after being imported into the SQL workbench:

[](https://res.cloudinary.com/practicaldev/image/fetch/s--k2w1wFPe--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/kts78mudvre3zlt0lzuo.jpeg)

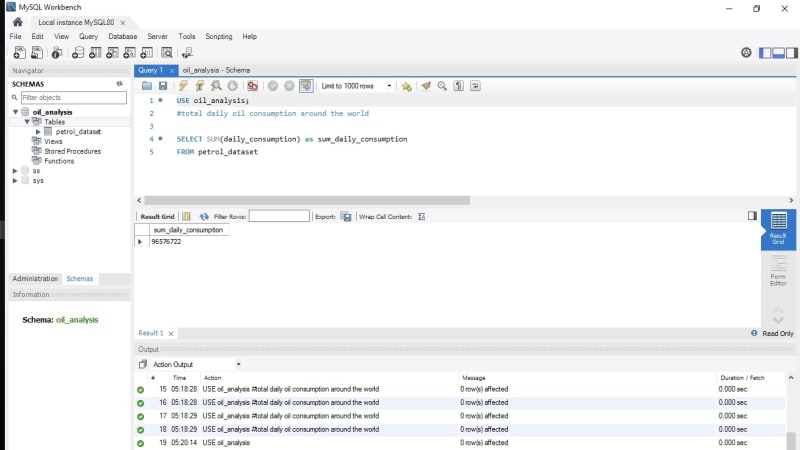
[](https://res.cloudinary.com/practicaldev/image/fetch/s--vhi6D_FU--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/ext18ng3zifdv3bhya9x.jpeg)

**ANALYSIS**

Analysis was done to get answers to some very important questions and to get an understanding of the dataset.

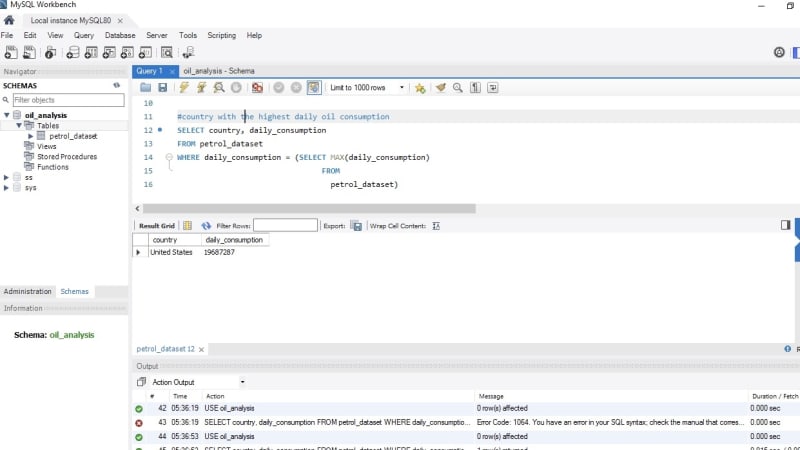
**PETROL/ GAS CONSUMPTION AROUND THE WORLD.**

1)To start with, I wanted to know the total sum of barrel of petrol/gas consumed by the entire world on a daily basis.

[](https://res.cloudinary.com/practicaldev/image/fetch/s--xBBJ6F1b--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/7fqun73mslfjvfh89u0d.jpeg)

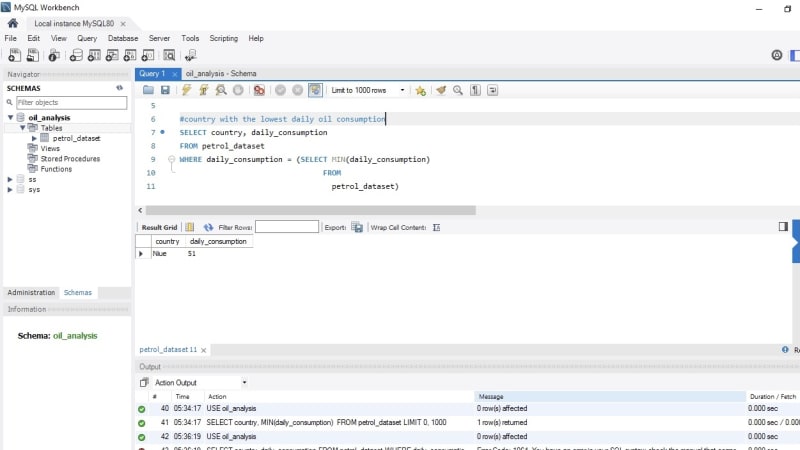
The query above shows the world consumes 96,576,722 barrels of petrol/ gas on a daily basis, as of June 2022.

2.Country with the highest petrol/gas consumption daily:

[](https://res.cloudinary.com/practicaldev/image/fetch/s--GatBmBYL--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/t393ojzhse66hyev92dy.jpeg)

The United States of America consumes the most petrol/gas daily.  
19,687,287 barrels per day.

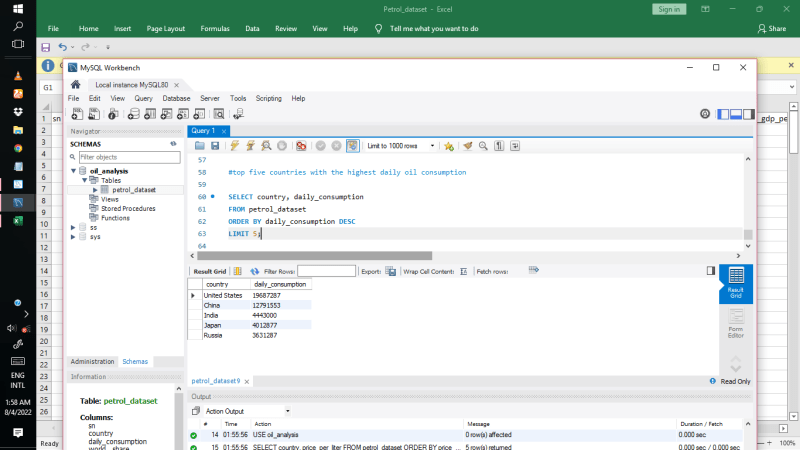
3.Country with the least petrol/ gas consumption daily:

[](https://res.cloudinary.com/practicaldev/image/fetch/s--ozs6Rz5---/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/ju7eav6v8gcv5bt2n4ny.jpeg)

Niue consumes the least petrol/ gas in the world. 51 barrels per day.  
Until while carrying out this analysis, I didn’t know of the country

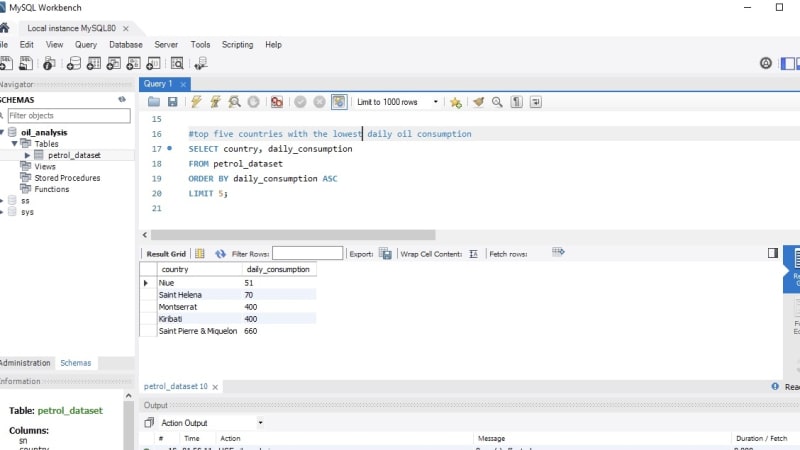
called Niue, had to google it up and realized it’s a country with an average of 1,620 people, located in the Oceania continent.

4.Top five countries with the highest daily oil consumption:

[](https://res.cloudinary.com/practicaldev/image/fetch/s--FC9pkiKR--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/jv5jb4vkhwoihv1tgi83.png)

USA, China, India, Japan and Russia consumes the most petrol/gas daily .

5.Bottom five countries with the lowest oil consumption around the world:

[](https://res.cloudinary.com/practicaldev/image/fetch/s--inNnGklC--/c_limit%2Cf_auto%2Cfl_progressive%2Cq_auto%2Cw_800/https:/dev-to-uploads.s3.amazonaws.com/uploads/articles/q7t5tp1l60ogswqioyk2.jpeg)

### Advantages of ETL process in data warehousing:

1. **Improved data quality:** ETL process ensures that the data in the data warehouse is accurate, complete, and up-to-date.
2. **Better data integration:**ETL process helps to integrate data from multiple sources and systems, making it more accessible and usable.
3. **Increased data security:** ETL process can help to improve data security by controlling access to the data warehouse and ensuring that only authorized users can access the data.
4. **Improved scalability: E**TL process can help to improve scalability by providing a way to manage and analyse large amounts of data.
5. **Increased automation:**ETL tools and technologies can automate and simplify the ETL process, reducing the time and effort required to load and update data in the warehouse.

### Disadvantages of ETL process in data warehousing:

1. **High cost:** ETL process can be expensive to implement and maintain, especially for organizations with limited resources.
2. C**omplexity:**ETL process can be complex and difficult to implement, especially for organizations that lack the necessary expertise or resources.
3. **Limited flexibility:**ETL process can be limited in terms of flexibility, as it may not be able to handle unstructured data or real-time data streams.
4. **Limited scalability**: ETL process can be limited in terms of scalability, as it may not be able to handle very large amounts of data.
5. **Data privacy concerns**: ETL process can raise concerns about data privacy, as large amounts of data are collected, stored, and analysed