



Non Graded Challenge : Data Wrangling in SQL

FTDS Phase 0 - Week 2

Contents



Objective	_____	03
Task	_____	04
Data	_____	05
Output	_____	06
Submission	_____	07

Non Graded Challenge

Objective

This Non Graded Challenge was created to evaluate the concept of Data Wrangling in SQL with the aim of:

1. Able to convert data types when doing Data Query Language (DQL).
2. Able to grouping and sorting data.
3. Able to perform subqueries.

Non Graded Challenge

Task

You are a Data Scientist who works for the Ministry of Law and Human Rights in a department that deals with patent archives. Your supervisor wants to see the cumulative number of patents related to cancer on an annual basis in the ministry database with condition :

- the patent was registered in 2012 to 2023
- the patent contains the keyword "*data mining*"

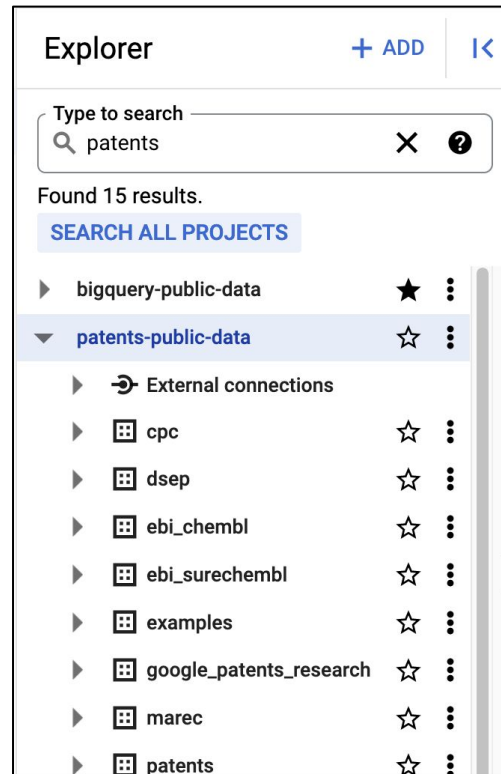
Display the cumulative number of patents sorted by largest year and smallest month.

Non Graded Challenge

Data

To answer this question, you will use dataset from Google BigQuery.

1. Log in to Google Cloud Console and select Google BigQuery.
2. In the Google BigQuery search bar, please type “*patents*”. You will see the result as shown beside.
3. Under *patents-public-data*, you will use dataset with name *uspto_oce_cancer* and table *publications*.



Non Graded Challenge

Output

Below is the expected output from your query. The query will output 12 entries of data.

Row	year	month	total_patent
1	2014	4	1
2	2014	12	1
3	2013	3	1
4	2013	8	2
5	2013	11	1
6	2012	1	3
7	2012	2	1
8	2012	4	1
9	2012	6	3
10	2012	10	3
11	2012	11	1
12	2012	12	1

Non Graded Challenge

Submission

1. Create a file to store your answers. This file can have extension `.sql` or `.txt`.
2. Save this assignment with filename : `h8dsft_ngc_data_wrangling_in_sql.sql`.
3. Push your answer into your own GitHub repository.



Thank You