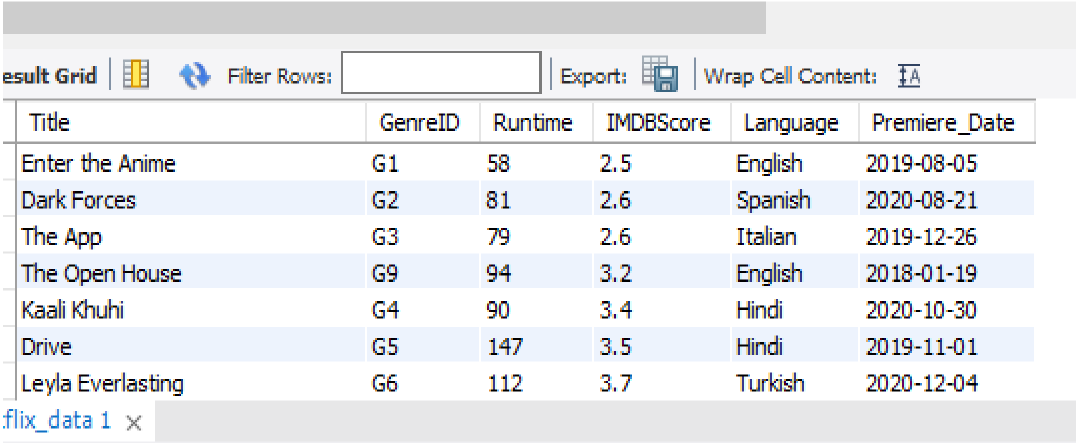
My SQL Project

Netflix Originals Data Exploration and Analysis

- **Creating Database:**
  Created a database named Netflix_originals .

- **Creating Table :**
  Created a table named Netflix_data .

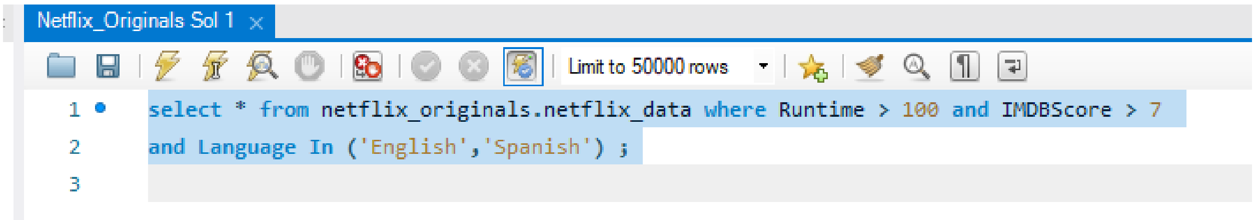
- **Data Importing :**
  Imported the existing CSV file ‘Netflix_Originals - Netflix_Originals’ into the table netflix_data.



Title	GenreID	Runtime	IMDBScore	Language	Premiere_Date
Enter the Anime	G1	58	2.5	English	2019-08-05
Dark Forces	G2	81	2.6	Spanish	2020-08-21
The App	G3	79	2.6	Italian	2019-12-26
The Open House	G9	94	3.2	English	2018-01-19
Kaali Khuhi	G4	90	3.4	Hindi	2020-10-30
Drive	G5	147	3.5	Hindi	2019-11-01
Leyla Everlasting	G6	112	3.7	Turkish	2020-12-04

Tasks

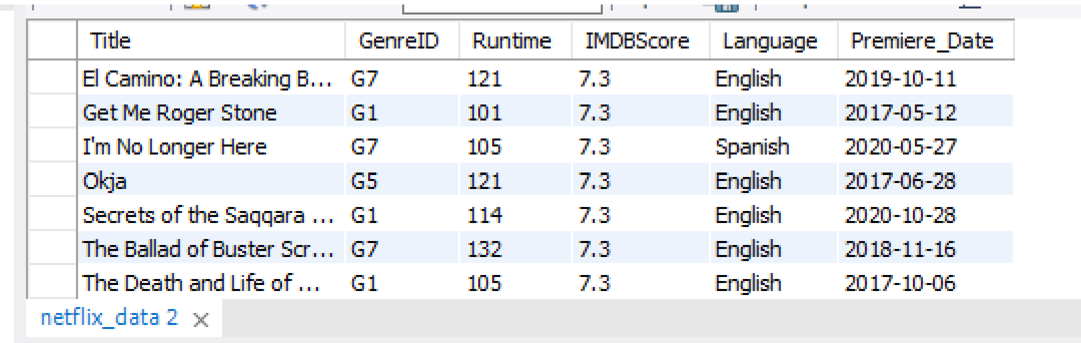
Task 1 : Retrieve all Netflix Originals with an IMDb score greater than 7, runtime greater than 100 minutes, and the language is either English or Spanish.



```sql
select * from netflix_originals.netflix_data where Runtime > 100 and IMDBScore > 7
and Language In ('English','Spanish') ;
```

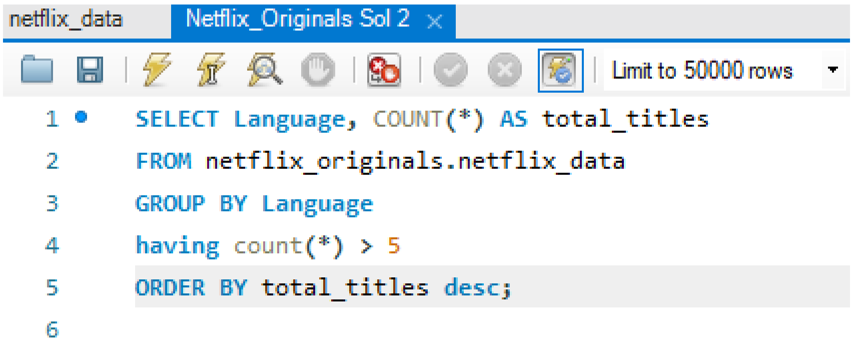
Query used

Output:



Title	GenreID	Runtime	IMDBScore	Language	Premiere_Date
El Camino: A Breaking B...	G7	121	7.3	English	2019-10-11
Get Me Roger Stone	G1	101	7.3	English	2017-05-12
I'm No Longer Here	G7	105	7.3	Spanish	2020-05-27
Okja	G5	121	7.3	English	2017-06-28
Secrets of the Saqqara ...	G1	114	7.3	English	2020-10-28
The Ballad of Buster Scr...	G7	132	7.3	English	2018-11-16
The Death and Life of ...	G1	105	7.3	English	2017-10-06

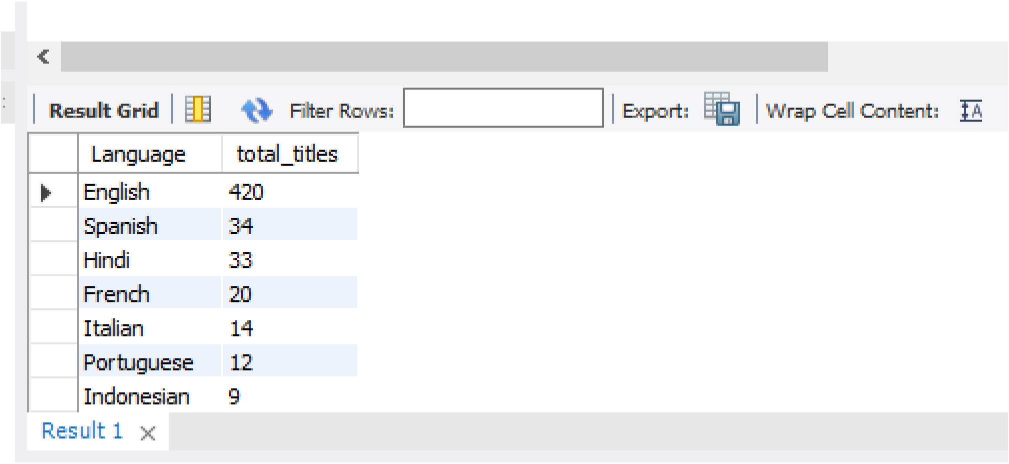
Task 2 : Find the total number of Netflix Originals in each language, but only show those languages that have more than 5 titles.



```sql
SELECT Language, COUNT(*) AS total_titles
FROM netflix_originals.netflix_data
GROUP BY Language
having count(*) > 5
ORDER BY total_titles desc;
```

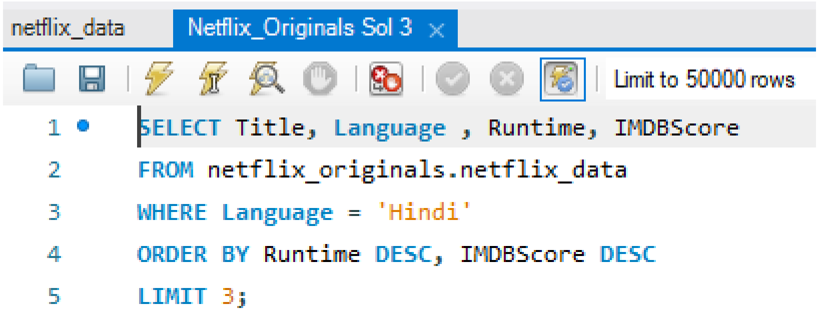
Query used

Output:



Language	total_titles
English	420
Spanish	34
Hindi	33
French	20
Italian	14
Portuguese	12
Indonesian	9

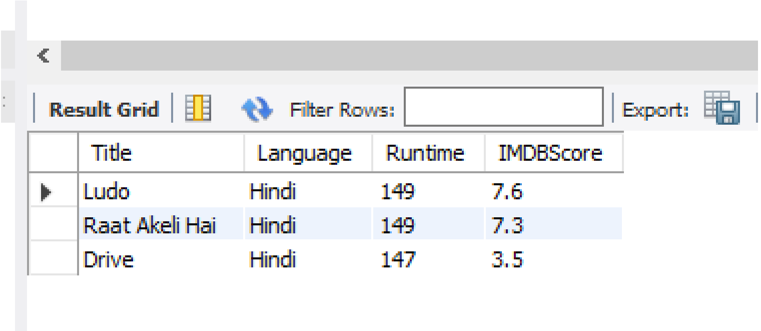
Task 3 :. Get the top 3 longest-running movies in Hindi language sorted by IMDb score in descending order.



```sql
SELECT Title, Language , Runtime, IMDBScore
FROM netflix_originals.netflix_data
WHERE Language = 'Hindi'
ORDER BY Runtime DESC, IMDBScore DESC
LIMIT 3;
```

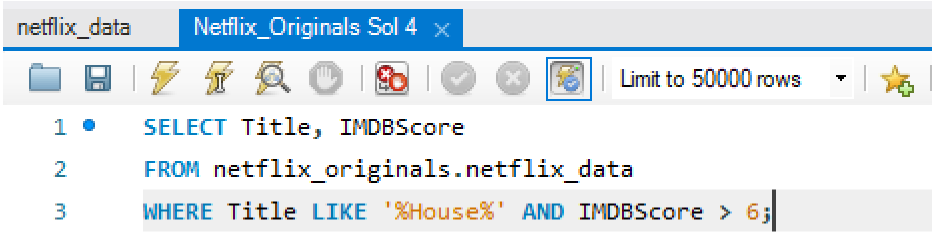
Query used

Output:



Title	Language	Runtime	IMDBScore
Ludo	Hindi	149	7.6
Raat Akeli Hai	Hindi	149	7.3
Drive	Hindi	147	3.5

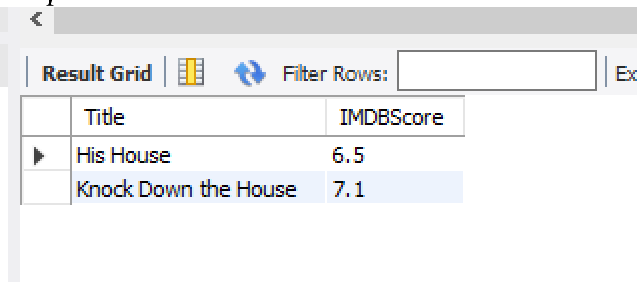
Task 4 : Retrieve all titles that contain the word "House" in their name and have an IMDb score greater than 6.



```sql
SELECT Title, IMDBScore
FROM netflix_originals.netflix_data
WHERE Title LIKE '%House%' AND IMDBScore > 6;
```

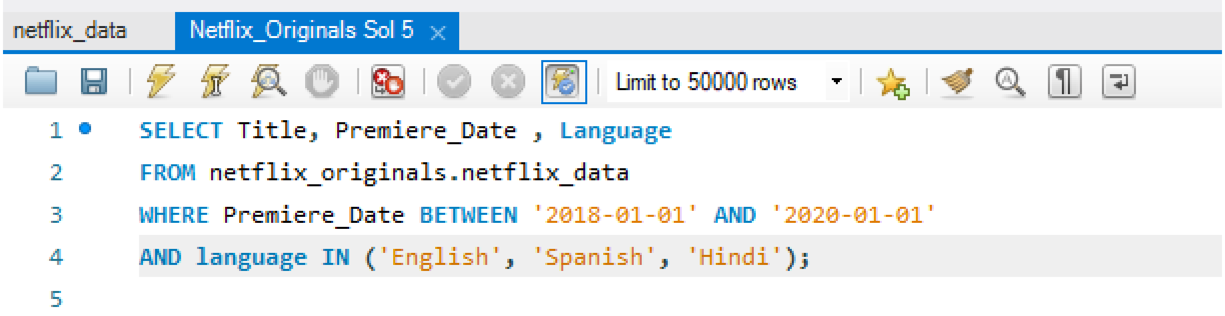
Query used

Output:



Title	IMDBScore
His House	6.5
Knock Down the House	7.1

Task 5 : Find all Netflix Originals released between the years 2018 and 2020 that are ineither English, Spanish, or Hindi.

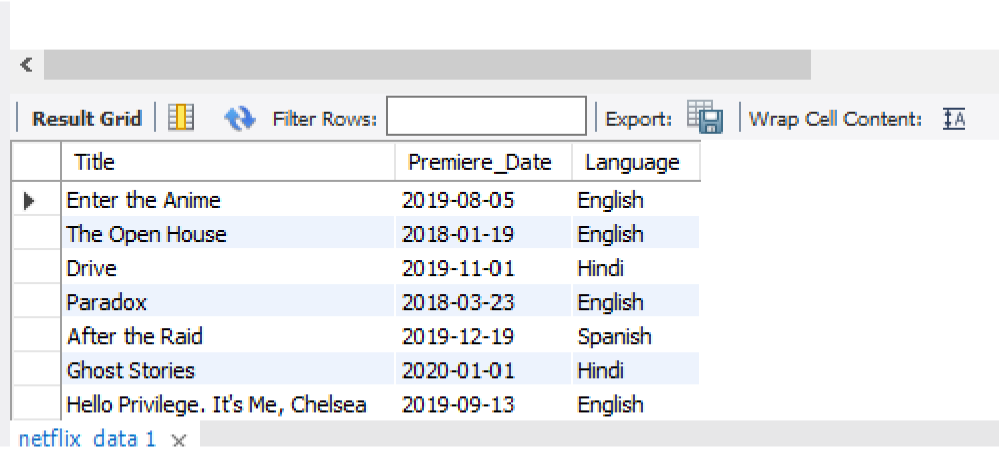


```sql
SELECT Title, Premiere_Date , Language
FROM netflix_originals.netflix_data
WHERE Premiere_Date BETWEEN '2018-01-01' AND '2020-01-01'
AND language IN ('English', 'Spanish', 'Hindi');

```

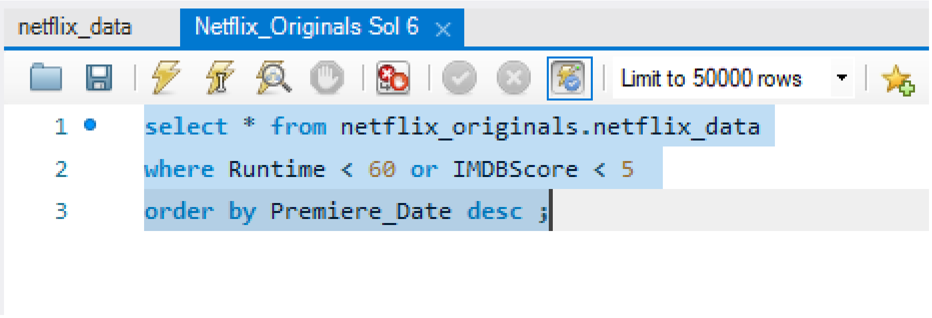
Query used

Output:



Title	Premiere_Date	Language
Enter the Anime	2019-08-05	English
The Open House	2018-01-19	English
Drive	2019-11-01	Hindi
Paradox	2018-03-23	English
After the Raid	2019-12-19	Spanish
Ghost Stories	2020-01-01	Hindi
Hello Privilege. It's Me, Chelsea	2019-09-13	English

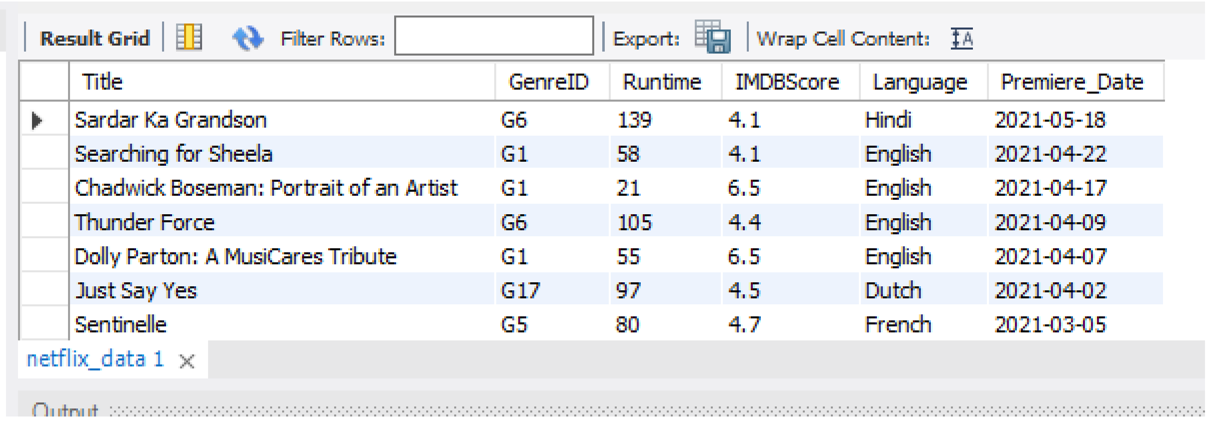
Task 6 : Find all movies that either have a runtime less than 60 minutes or an IMDb score less than 5, sorted by Premiere Date.



```sql
select * from netflix_originals.netflix_data
where Runtime < 60 or IMDBScore < 5
order by Premiere_Date desc ;
```

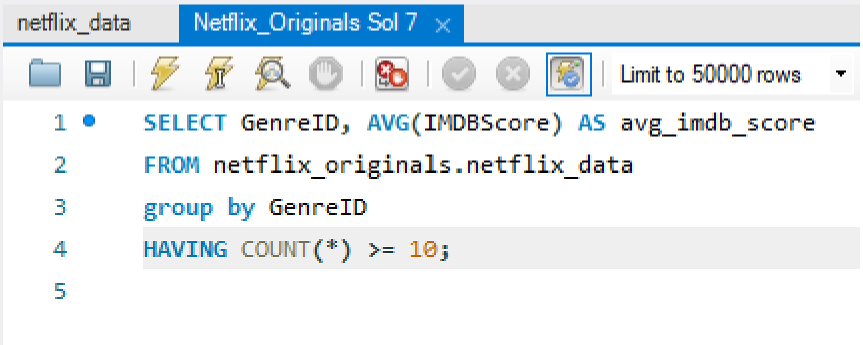
Query used

Output:



Title	GenreID	Runtime	IMDBScore	Language	Premiere_Date
Sardar Ka Grandson	G6	139	4.1	Hindi	2021-05-18
Searching for Sheela	G1	58	4.1	English	2021-04-22
Chadwick Boseman: Portrait of an Artist	G1	21	6.5	English	2021-04-17
Thunder Force	G6	105	4.4	English	2021-04-09
Dolly Parton: A MusiCares Tribute	G1	55	6.5	English	2021-04-07
Just Say Yes	G17	97	4.5	Dutch	2021-04-02
Sentinelle	G5	80	4.7	French	2021-03-05

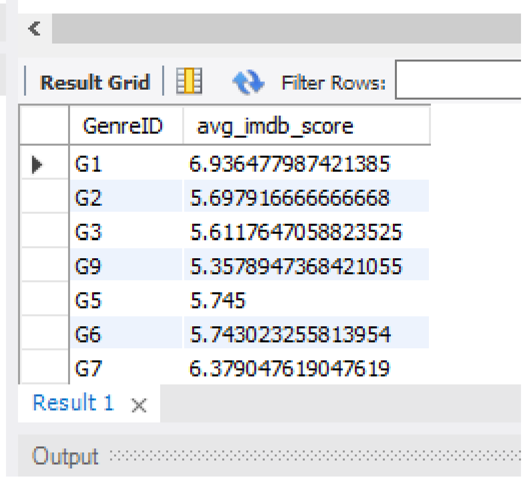
Task 7 : Get the average IMDb score for each genre where the genre has at least 10 movies.



```sql
SELECT GenreID, AVG(IMDBScore) AS avg_imdb_score
FROM netflix_originals.netflix_data
group by GenreID
HAVING COUNT(*) >= 10;
```

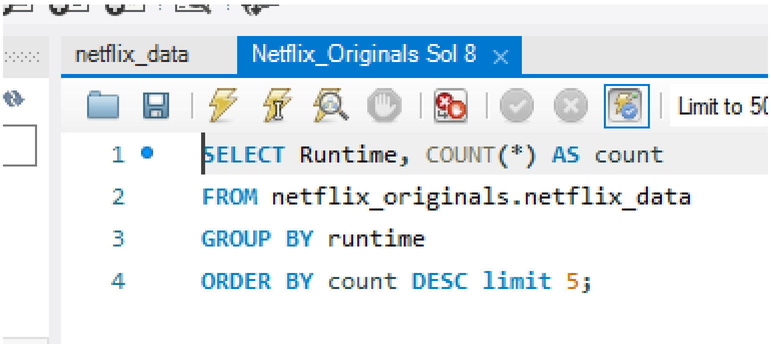
Query used

Output:



GenreID	avg_imdb_score
G1	6.936477987421385
G2	5.697916666666668
G3	5.6117647058823525
G9	5.3578947368421055
G5	5.745
G6	5.743023255813954
G7	6.379047619047619

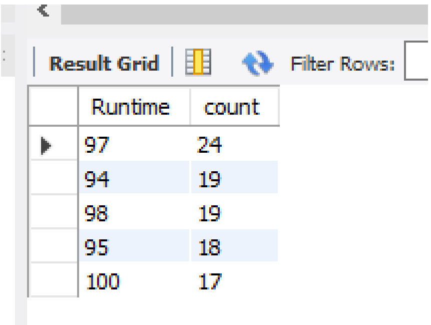
Task 8 : Retrieve the top 5 most common runtimes for Netflix Originals.



```sql
SELECT Runtime, COUNT(*) AS count
FROM netflix_originals.netflix_data
GROUP BY runtime
ORDER BY count DESC limit 5;
```

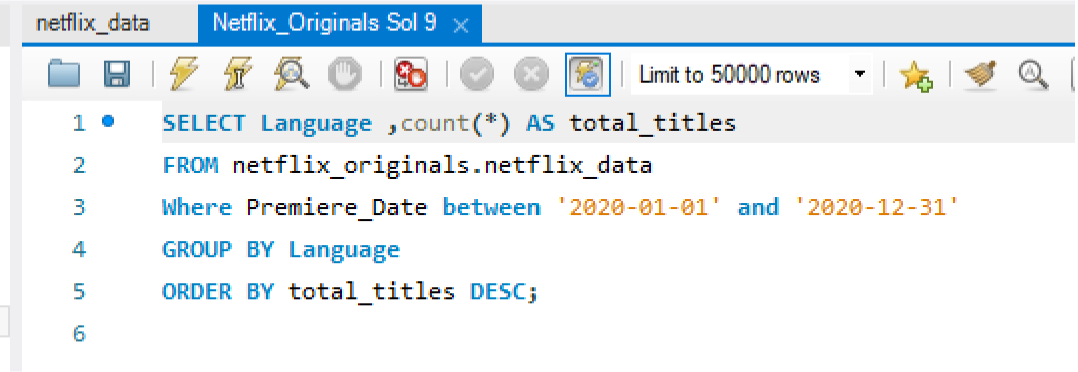
Query used

Output:



Runtime	count
97	24
94	19
98	19
95	18
100	17

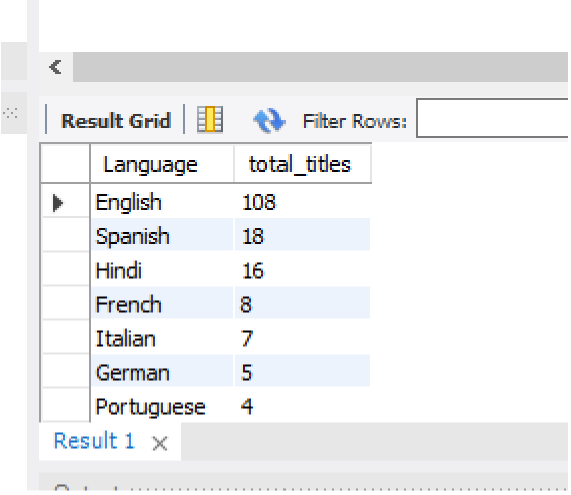
Task 9 : List all Netflix Originals that were released in 2020, grouped by language, and show the total count of titles for each language.



```
SELECT Language ,count(*) AS total_titles
FROM netflix_originals.netflix_data
Where Premiere_Date between '2020-01-01' and '2020-12-31'
GROUP BY Language
ORDER BY total_titles DESC;
```

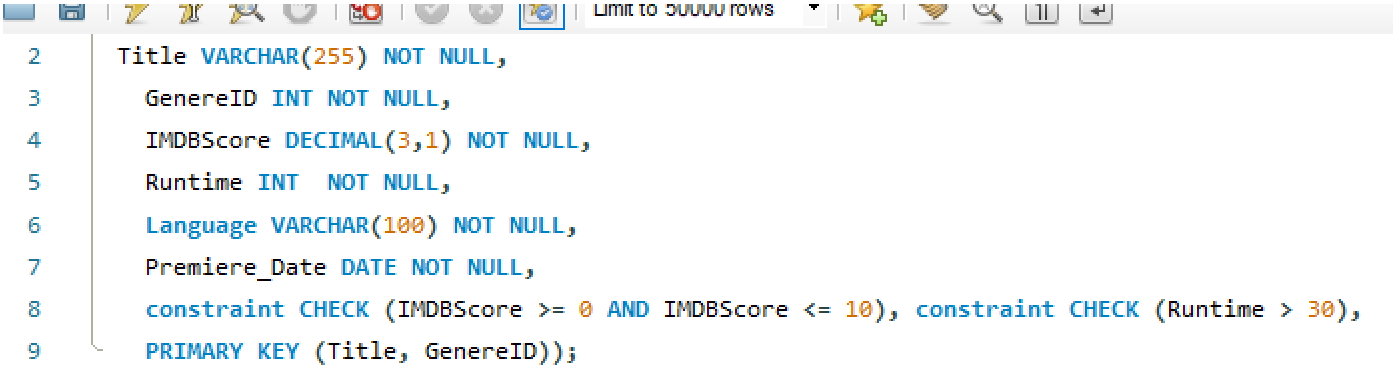
Query used

Output:



Language	total_titles
English	108
Spanish	18
Hindi	16
French	8
Italian	7
German	5
Portuguese	4

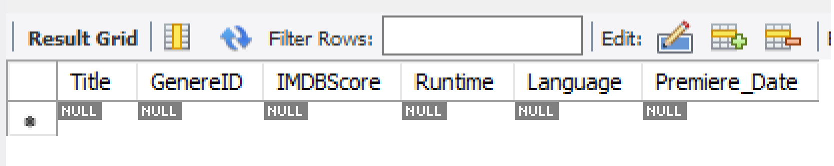
Task 10 : Create a new table that enforces a constraint on the IMDb score to be between 0 and 10 and the runtime to be greater than 30 minutes.



```
Title VARCHAR(255) NOT NULL,
  GenereID INT NOT NULL,
  IMDBScore DECIMAL(3,1) NOT NULL,
  Runtime INT  NOT NULL,
  Language VARCHAR(100) NOT NULL,
  Premiere_Date DATE NOT NULL,
  constraint CHECK (IMDBScore >= 0 AND IMDBScore <= 10), constraint CHECK (Runtime > 30),
  PRIMARY KEY (Title, GenereID));
```

Query used

Output:



Title	GenereID	IMDBScore	Runtime	Language	Premiere_Date
NULL	NULL	NULL	NULL	NULL	NULL

These are the analysed tasks and the insights

Thank You