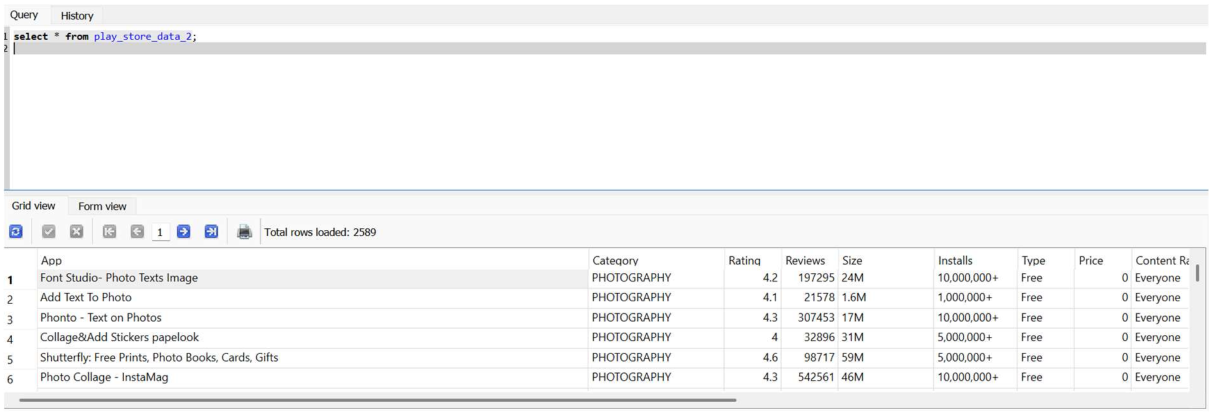
SQL for Data Analysis

Objective: Use SQL queries to extract and analyze data from a database.

1. Select query is use to extract the data from database.

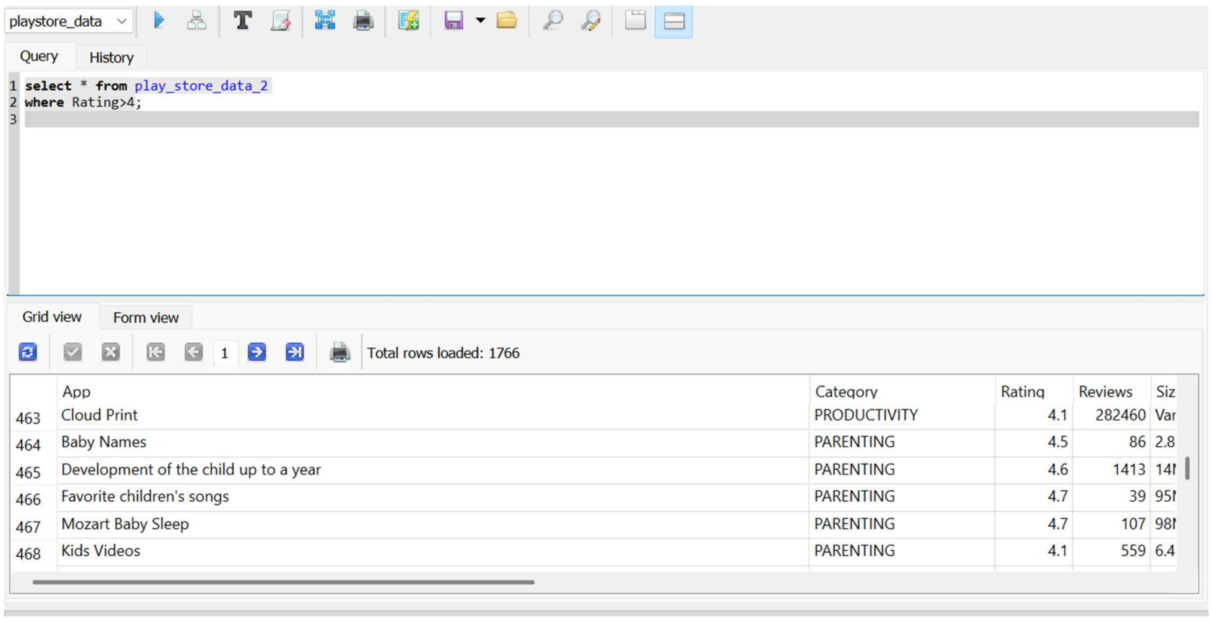


```
select * from play_store_data_2;
```

Total rows loaded: 2589

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ra
1	Font Studio- Photo Texts Image	PHOTOGRAPHY	4.2	197295	24M	10,000,000+	Free	0	Everyone
2	Add Text To Photo	PHOTOGRAPHY	4.1	21578	1.6M	1,000,000+	Free	0	Everyone
3	Phonto - Text on Photos	PHOTOGRAPHY	4.3	307453	17M	10,000,000+	Free	0	Everyone
4	Collage&Add Stickers papelook	PHOTOGRAPHY	4	32896	31M	5,000,000+	Free	0	Everyone
5	Shutterfly: Free Prints, Photo Books, Cards, Gifts	PHOTOGRAPHY	4.6	98717	59M	5,000,000+	Free	0	Everyone
6	Photo Collage - InstaMag	PHOTOGRAPHY	4.3	542561	46M	10,000,000+	Free	0	Everyone

2. Select query and where is use exract the data from database.
   Where is used to specify conditions.



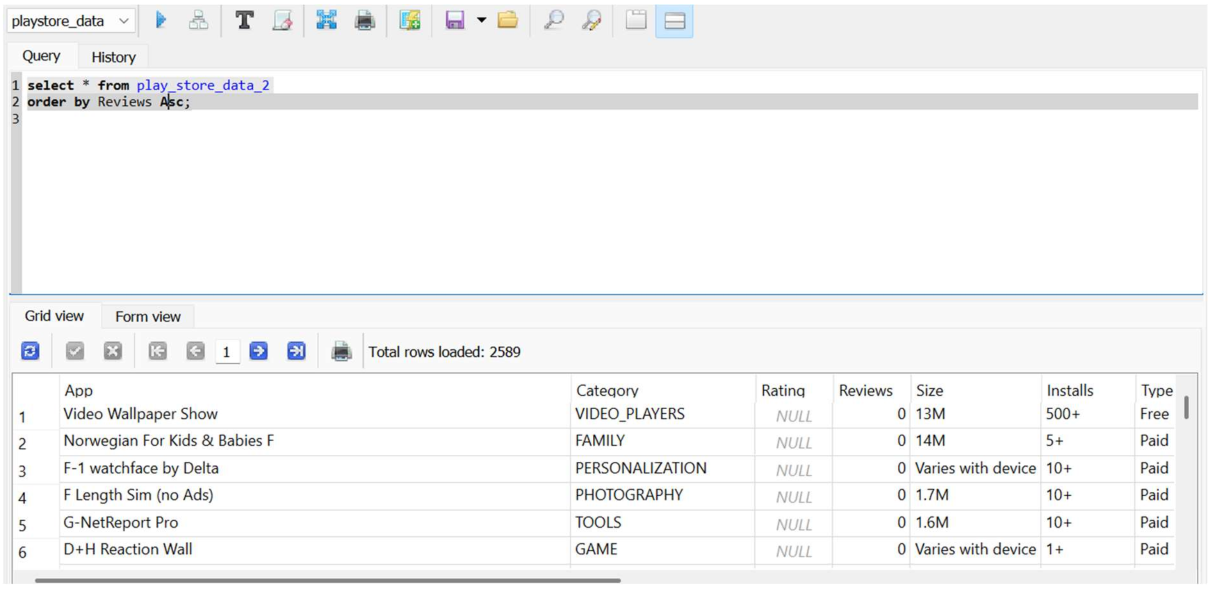
```
select * from play_store_data_2
where Rating>4;
```

Total rows loaded: 1766

	App	Category	Rating	Reviews	Siz
463	Cloud Print	PRODUCTIVITY	4.1	282460	Var
464	Baby Names	PARENTING	4.5	86	2.8
465	Development of the child up to a year	PARENTING	4.6	1413	14
466	Favorite children's songs	PARENTING	4.7	39	95
467	Mozart Baby Sleep	PARENTING	4.7	107	98
468	Kids Videos	PARENTING	4.1	559	6.4

3. The ORDER BY **clause** in SQL is used to **sort the results of a query** based on one or more columns.

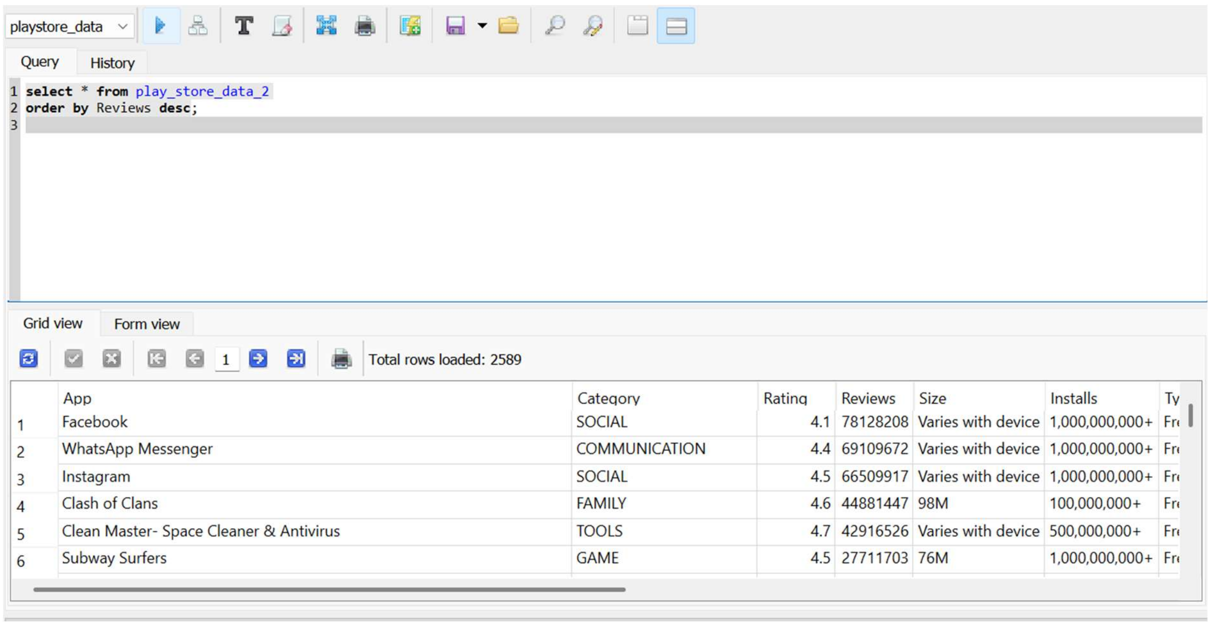
- By default, it sorts in **ascending order (ASC)**.
- You can explicitly specify **descending order (DESC)**.
- You can sort by **multiple columns** (first by one, then by another).



```
select * from play_store_data_2
order by Reviews Asc;
```

Total rows loaded: 2589

	App	Category	Rating	Reviews	Size	Installs	Type
1	Video Wallpaper Show	VIDEO_PLAYERS	NULL	0	13M	500+	Free
2	Norwegian For Kids & Babies F	FAMILY	NULL	0	14M	5+	Paid
3	F-1 watchface by Delta	PERSONALIZATION	NULL	0	Varies with device	10+	Paid
4	F Length Sim (no Ads)	PHOTOGRAPHY	NULL	0	1.7M	10+	Paid
5	G-NetReport Pro	TOOLS	NULL	0	1.6M	10+	Paid
6	D+H Reaction Wall	GAME	NULL	0	Varies with device	1+	Paid



```
select * from play_store_data_2
order by Reviews desc;
```

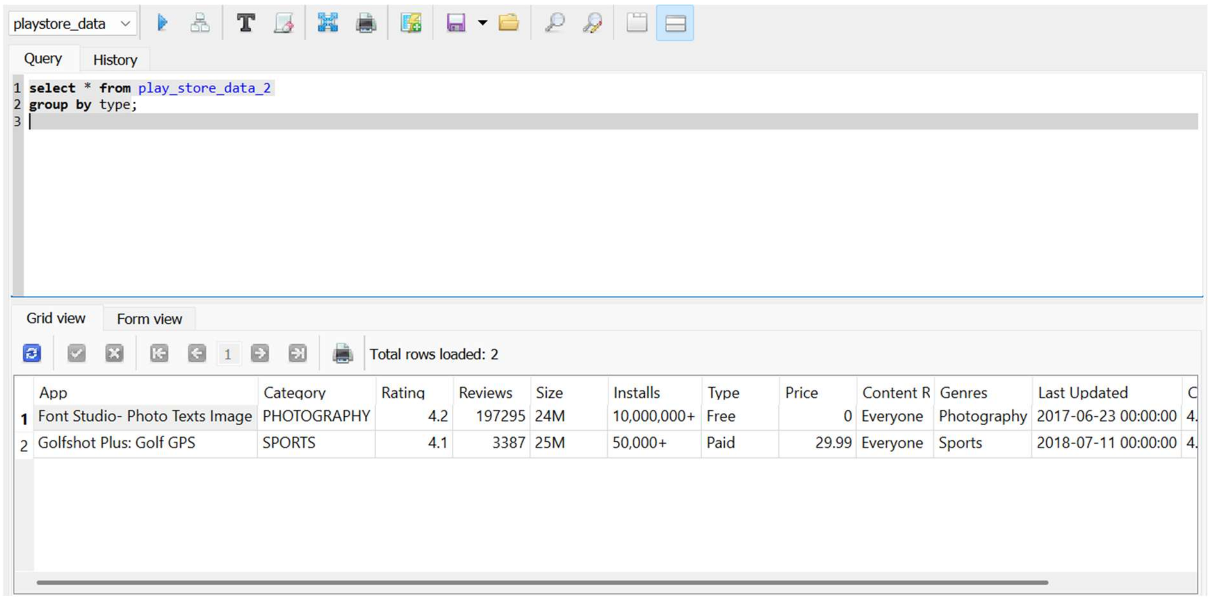
Total rows loaded: 2589

	App	Category	Rating	Reviews	Size	Installs	Ty
1	Facebook	SOCIAL	4.1	78128208	Varies with device	1,000,000,000+	Fr
2	WhatsApp Messenger	COMMUNICATION	4.4	69109672	Varies with device	1,000,000,000+	Fr
3	Instagram	SOCIAL	4.5	66509917	Varies with device	1,000,000,000+	Fr
4	Clash of Clans	FAMILY	4.6	44881447	98M	100,000,000+	Fr
5	Clean Master- Space Cleaner & Antivirus	TOOLS	4.7	42916526	Varies with device	500,000,000+	Fr
6	Subway Surfers	GAME	4.5	27711703	76M	1,000,000,000+	Fr

4. The GROUP BY **clause** is used to **arrange identical data into groups**.

   It’s almost always paired with **aggregate functions** (SUM, AVG, COUNT, MIN, MAX).

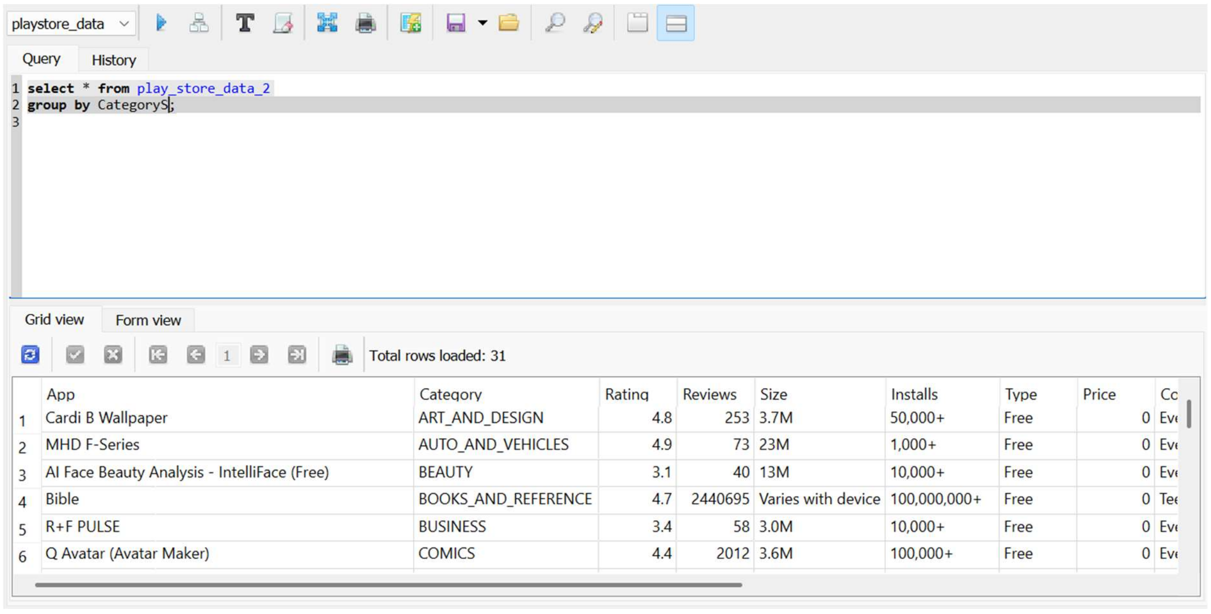
   Instead of returning every row, it summarizes data **per group**.



```
select * from play_store_data_2
group by type;
```

Total rows loaded: 2

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content R	Genres	Last Updated	C
1	Font Studio- Photo Texts Image	PHOTOGRAPHY	4.2	197295	24M	10,000,000+	Free	0	Everyone	Photography	2017-06-23 00:00:00	4.
2	Golfshot Plus: Golf GPS	SPORTS	4.1	3387	25M	50,000+	Paid	29.99	Everyone	Sports	2018-07-11 00:00:00	4.



```
select * from play_store_data_2
group by CategoryS;
```

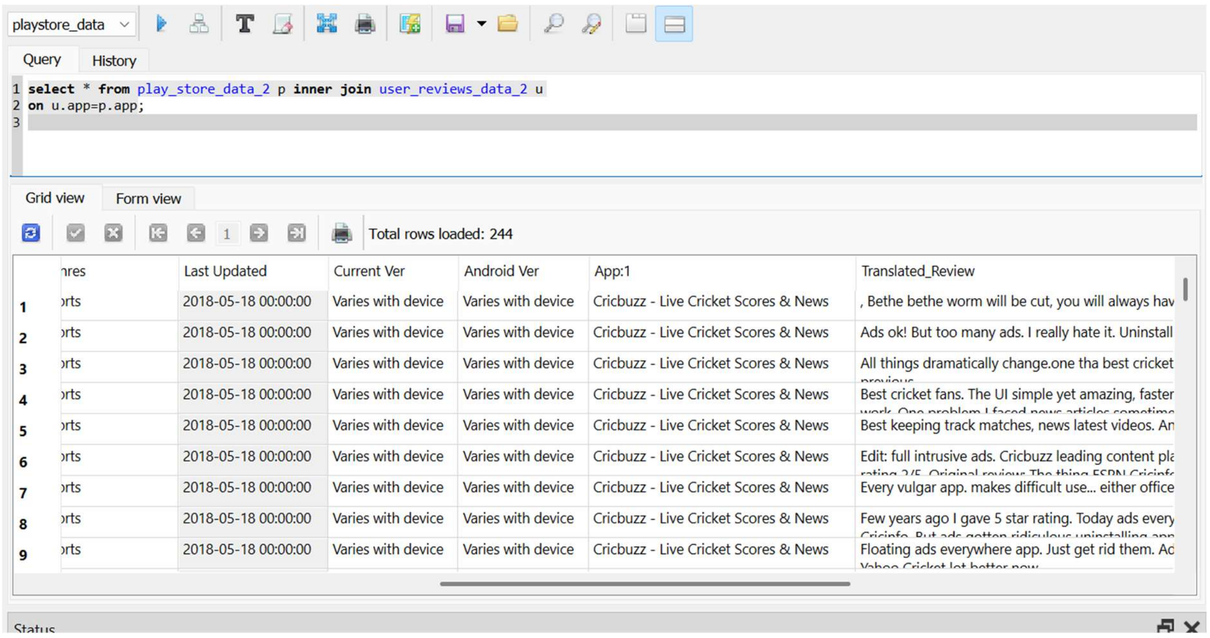
Total rows loaded: 31

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Cc
1	Cardi B Wallpaper	ART_AND_DESIGN	4.8	253	3.7M	50,000+	Free	0	Eve
2	MHD F-Series	AUTO_AND_VEHICLES	4.9	73	23M	1,000+	Free	0	Eve
3	AI Face Beauty Analysis - IntelliFace (Free)	BEAUTY	3.1	40	13M	10,000+	Free	0	Eve
4	Bible	BOOKS_AND_REFERENCE	4.7	2440695	Varies with device	100,000,000+	Free	0	Tee
5	R+F PULSE	BUSINESS	3.4	58	3.0M	10,000+	Free	0	Eve
6	Q Avatar (Avatar Maker)	COMICS	4.4	2012	3.6M	100,000+	Free	0	Eve

5. An **INNER JOIN** returns rows that have **matching values in both tables**.

If there's no match, the row is excluded.

It's the most commonly used join for analysis.



```
select * from play_store_data_2 p inner join user_reviews_data_2 u
on u.app=p.app;
```

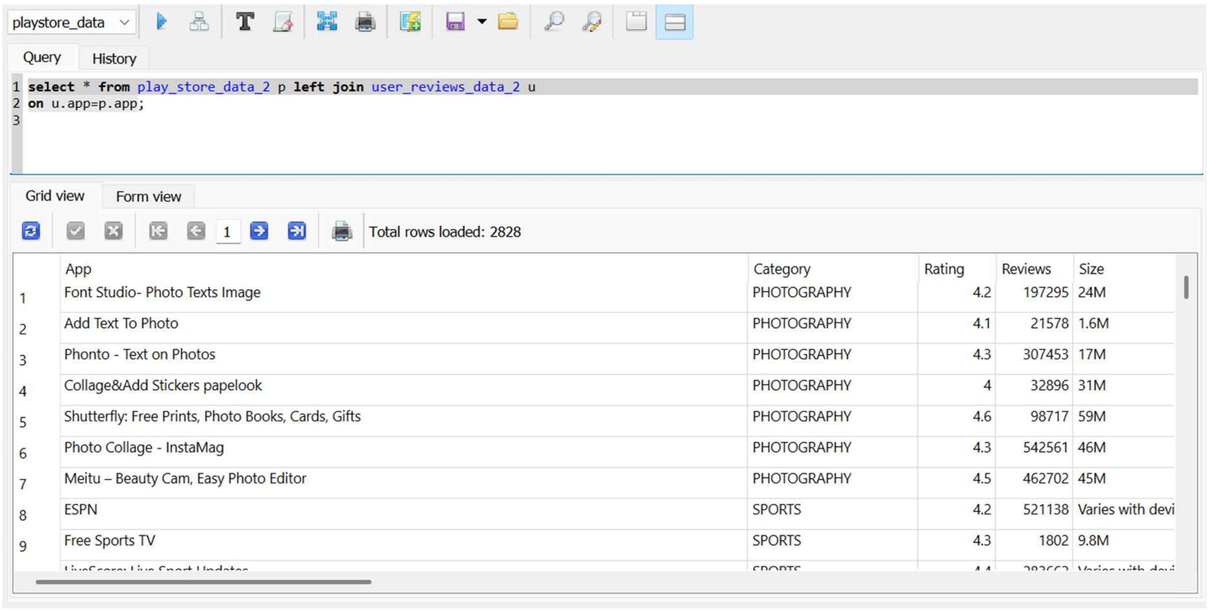
Total rows loaded: 244

	ıres	Last Updated	Current Ver	Android Ver	App:1	Translated_Review
1	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	, Bethe bethe worm will be cut, you will always hav
2	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Ads ok! But too many ads. I really hate it. Uninstall
3	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	All things dramatically change.one tha best cricket
4	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Best cricket fans. The UI simple yet amazing, faster
5	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Best keeping track matches, news latest videos. An
6	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Edit: full intrusive ads. Cricbuzz leading content pla
7	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Every vulgar app. makes difficult use... either office
8	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Few years ago I gave 5 star rating. Today ads every
9	ɔrts	2018-05-18 00:00:00	Varies with device	Varies with device	Cricbuzz - Live Cricket Scores & News	Floating ads everywhere app. Just get rid them. Ad

6. A **LEFT JOIN** returns **all rows from the left table** and the **matching rows from the right table**.

If there's no match in the right table, the result will still include the left table row, but with **NULLs** for the right table's columns.

It's useful when you want to see everything from one table, even if related data doesn't exist in the other.



```
select * from play_store_data_2 p left join user_reviews_data_2 u
on u.app=p.app;
```

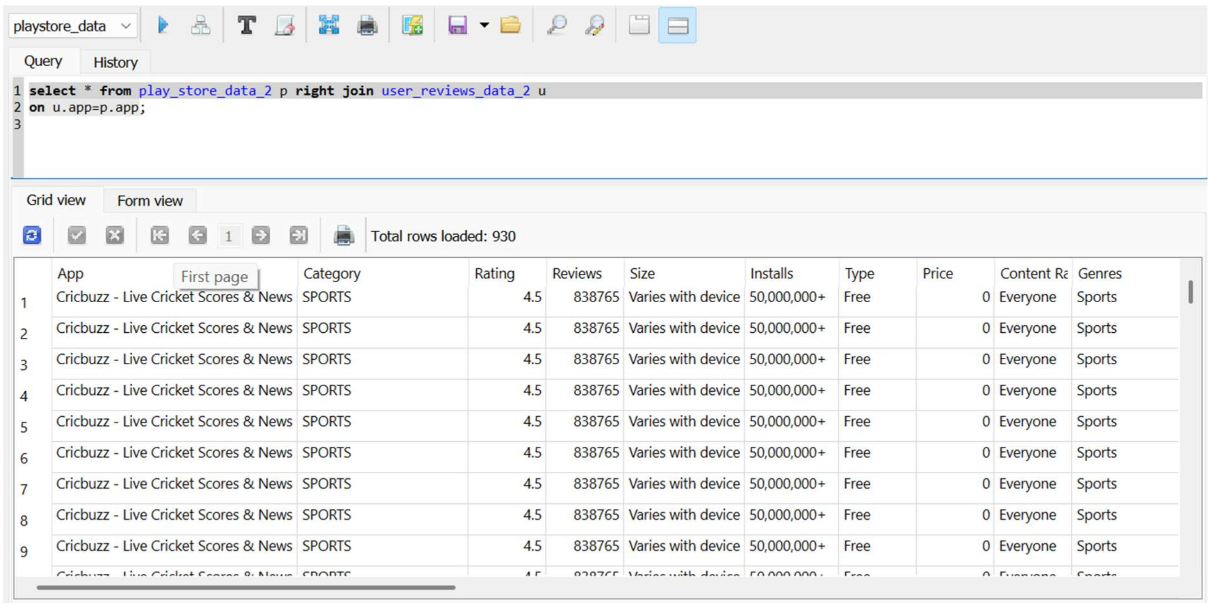
Total rows loaded: 2828

	App	Category	Rating	Reviews	Size
1	Font Studio- Photo Texts Image	PHOTOGRAPHY	4.2	197295	24M
2	Add Text To Photo	PHOTOGRAPHY	4.1	21578	1.6M
3	Phonto - Text on Photos	PHOTOGRAPHY	4.3	307453	17M
4	Collage&Add Stickers papelook	PHOTOGRAPHY	4	32896	31M
5	Shutterfly: Free Prints, Photo Books, Cards, Gifts	PHOTOGRAPHY	4.6	98717	59M
6	Photo Collage - InstaMag	PHOTOGRAPHY	4.3	542561	46M
7	Meitu – Beauty Cam, Easy Photo Editor	PHOTOGRAPHY	4.5	462702	45M
8	ESPN	SPORTS	4.2	521138	Varies with devi
9	Free Sports TV	SPORTS	4.3	1802	9.8M

7. A **RIGHT JOIN** returns **all rows from the right table** and the **matching rows from the left table**.

If there’s no match in the left table, the result will still include the right table row, but with **NULLs** for the left table’s columns.

It’s basically the opposite of a **LEFT JOIN**.



```sql
select * from play_store_data_2 p right join user_reviews_data_2 u
on u.app=p.app;
```

Total rows loaded: 930

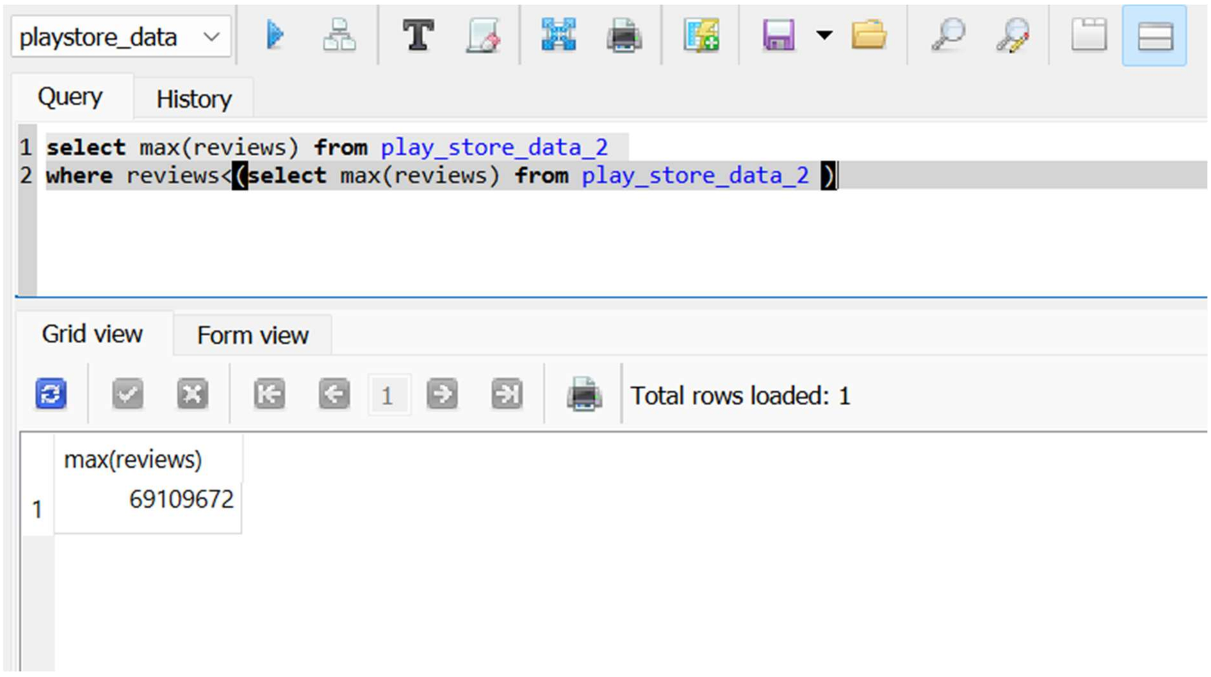
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ra	Genres
1	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
2	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
3	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
4	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
5	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
6	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
7	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
8	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports
9	Cricbuzz - Live Cricket Scores & News	SPORTS	4.5	838765	Varies with device	50,000,000+	Free	0	Everyone	Sports

8. A **subquery** is a query **inside another query**.

It's enclosed in parentheses ().

It can return a **single value**, a **list of values**, or even a **table**.

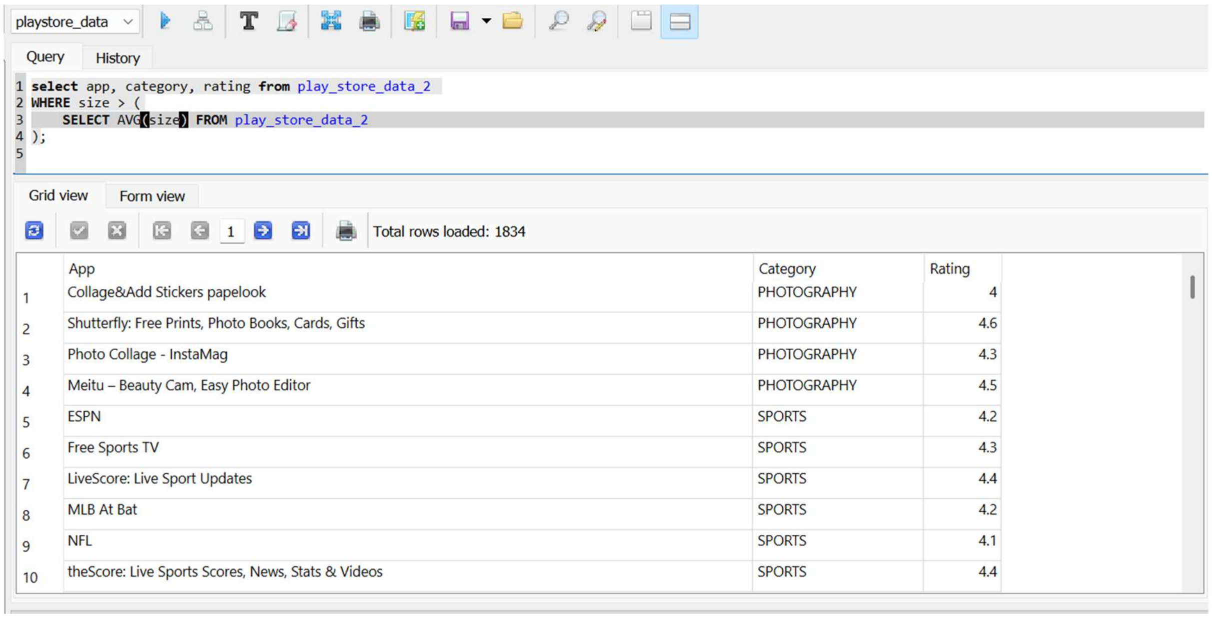
Subqueries are often used in WHERE, FROM, or SELECT clauses.



```
select max(reviews) from play_store_data_2
where reviews<(select max(reviews) from play_store_data_2 )
```

Total rows loaded: 1

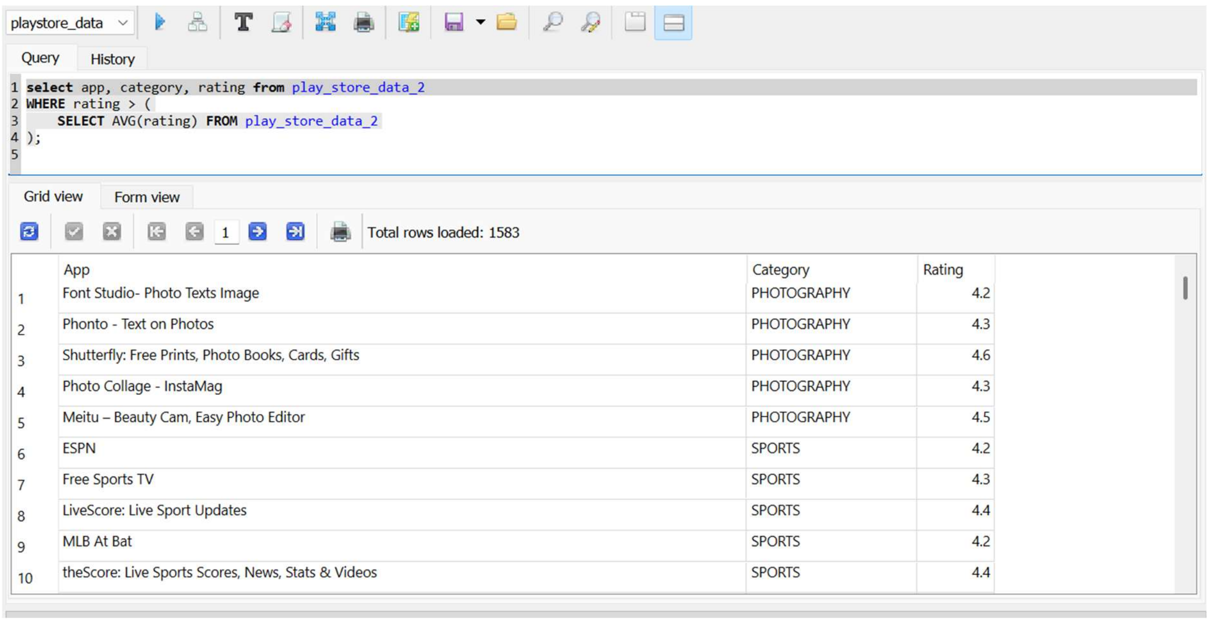
	max(reviews)
1	69109672



```
select app, category, rating from play_store_data_2
WHERE size > (
    SELECT AVG(size) FROM play_store_data_2
);
```

Total rows loaded: 1834

	App	Category	Rating
1	Collage&Add Stickers papelook	PHOTOGRAPHY	4
2	Shutterfly: Free Prints, Photo Books, Cards, Gifts	PHOTOGRAPHY	4.6
3	Photo Collage - InstaMag	PHOTOGRAPHY	4.3
4	Meitu – Beauty Cam, Easy Photo Editor	PHOTOGRAPHY	4.5
5	ESPN	SPORTS	4.2
6	Free Sports TV	SPORTS	4.3
7	LiveScore: Live Sport Updates	SPORTS	4.4
8	MLB At Bat	SPORTS	4.2
9	NFL	SPORTS	4.1
10	theScore: Live Sports Scores, News, Stats & Videos	SPORTS	4.4


playstore_data

Query History

```
select app, category, rating from play_store_data_2
WHERE rating > (
    SELECT AVG(rating) FROM play_store_data_2
);
```

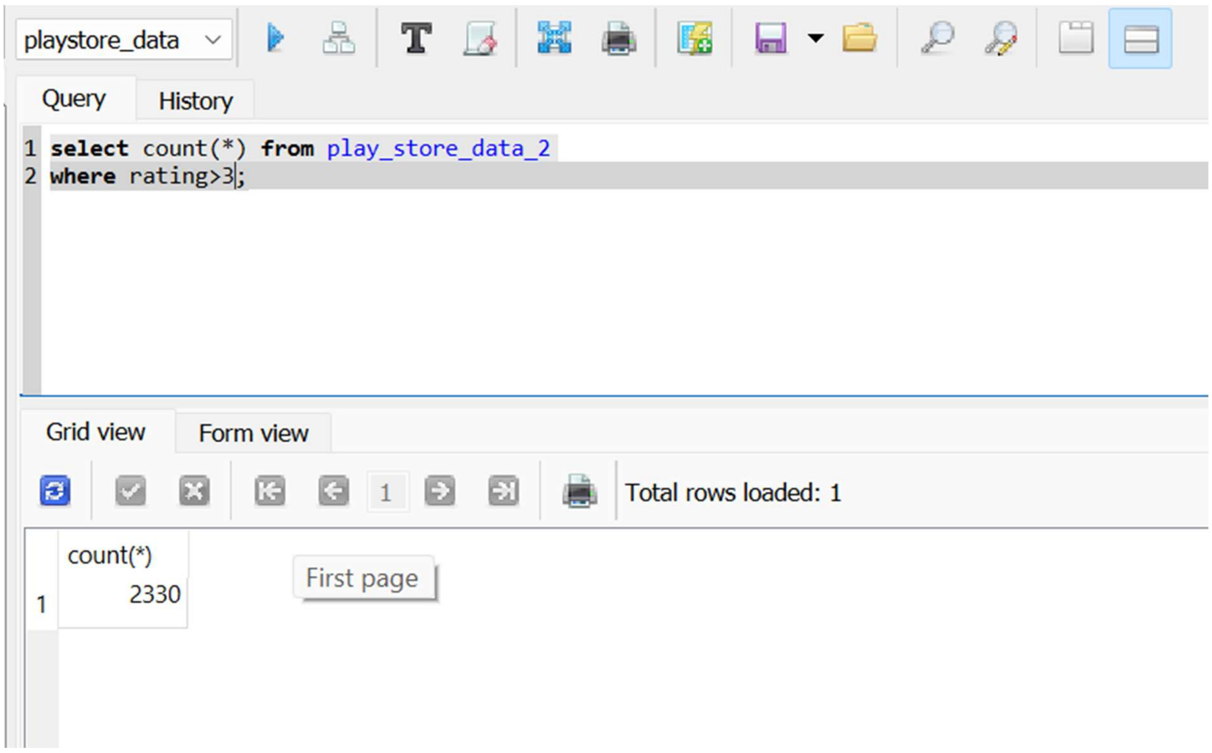
Grid view Form view

Total rows loaded: 1583

	App	Category	Rating
1	Font Studio- Photo Texts Image	PHOTOGRAPHY	4.2
2	Phonto - Text on Photos	PHOTOGRAPHY	4.3
3	Shutterfly: Free Prints, Photo Books, Cards, Gifts	PHOTOGRAPHY	4.6
4	Photo Collage - InstaMag	PHOTOGRAPHY	4.3
5	Meitu – Beauty Cam, Easy Photo Editor	PHOTOGRAPHY	4.5
6	ESPN	SPORTS	4.2
7	Free Sports TV	SPORTS	4.3
8	LiveScore: Live Sport Updates	SPORTS	4.4
9	MLB At Bat	SPORTS	4.2
10	theScore: Live Sports Scores, News, Stats & Videos	SPORTS	4.4

9. COUNT() returns the number of rows that match a condition.

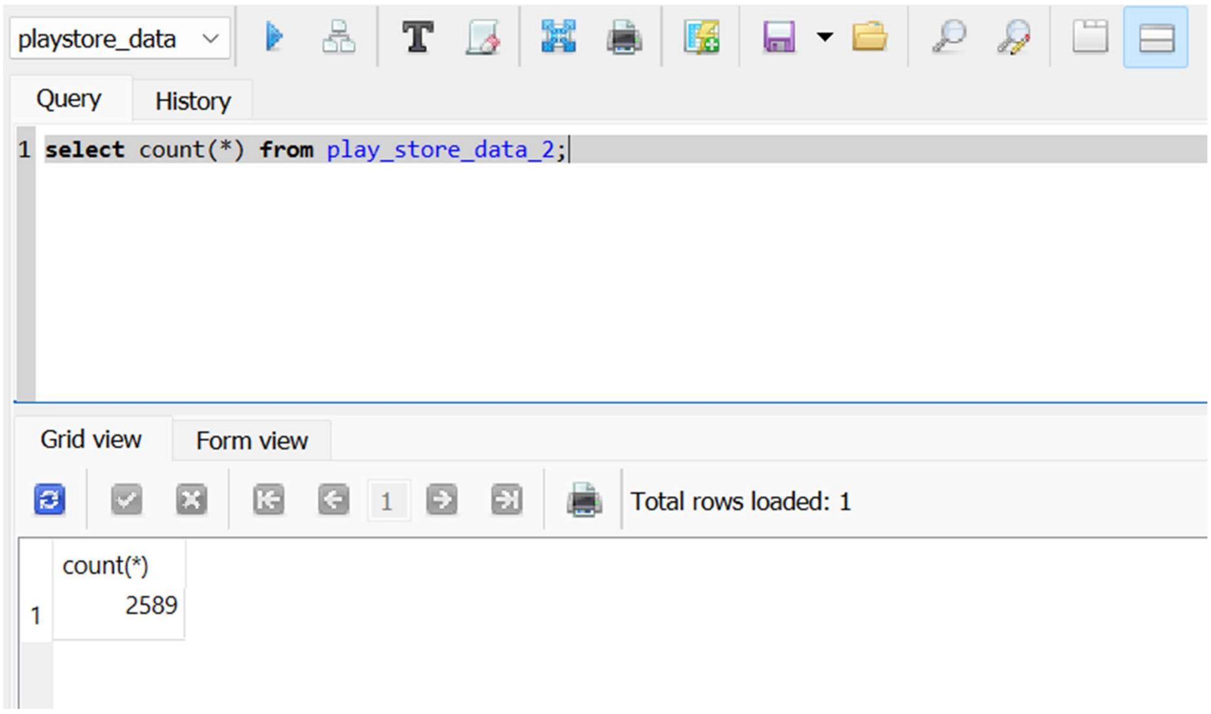
   It can count **all rows** (COUNT(*)) or **non-NULL values** in a specific column(COUNT(column_name)).



```
select count(*) from play_store_data_2
where rating>3;
```

Total rows loaded: 1

```
  count(*)
1     2330
```

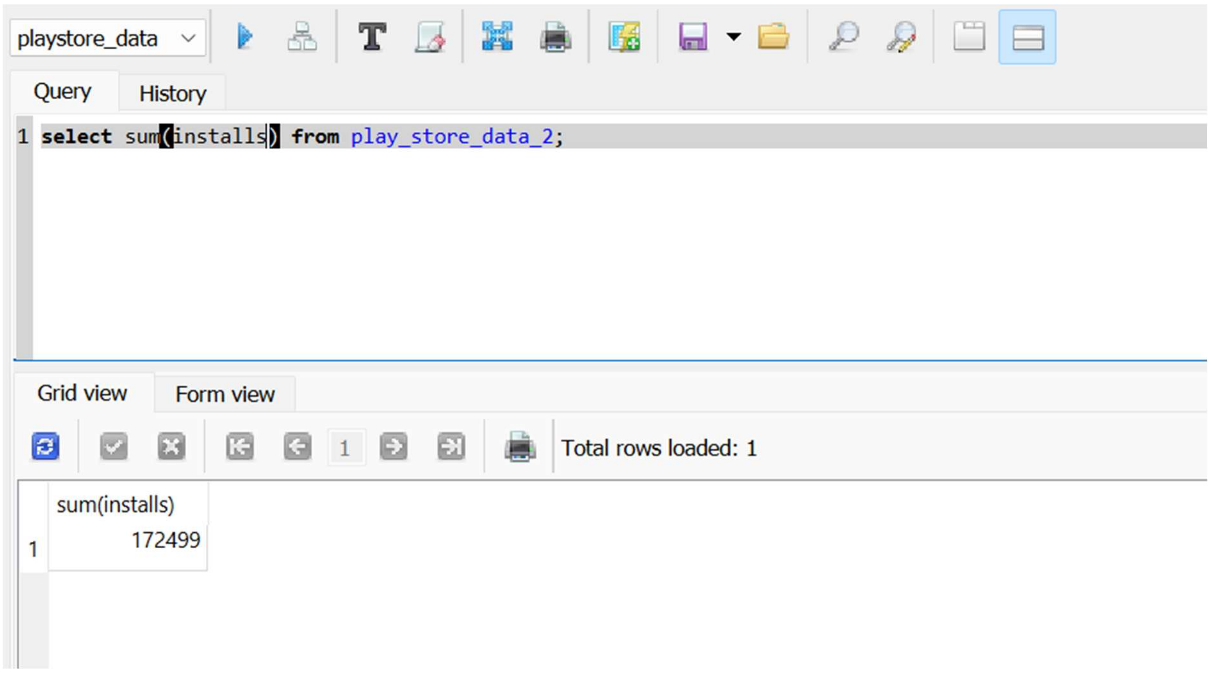


```
select count(*) from play_store_data_2;
```

Total rows loaded: 1

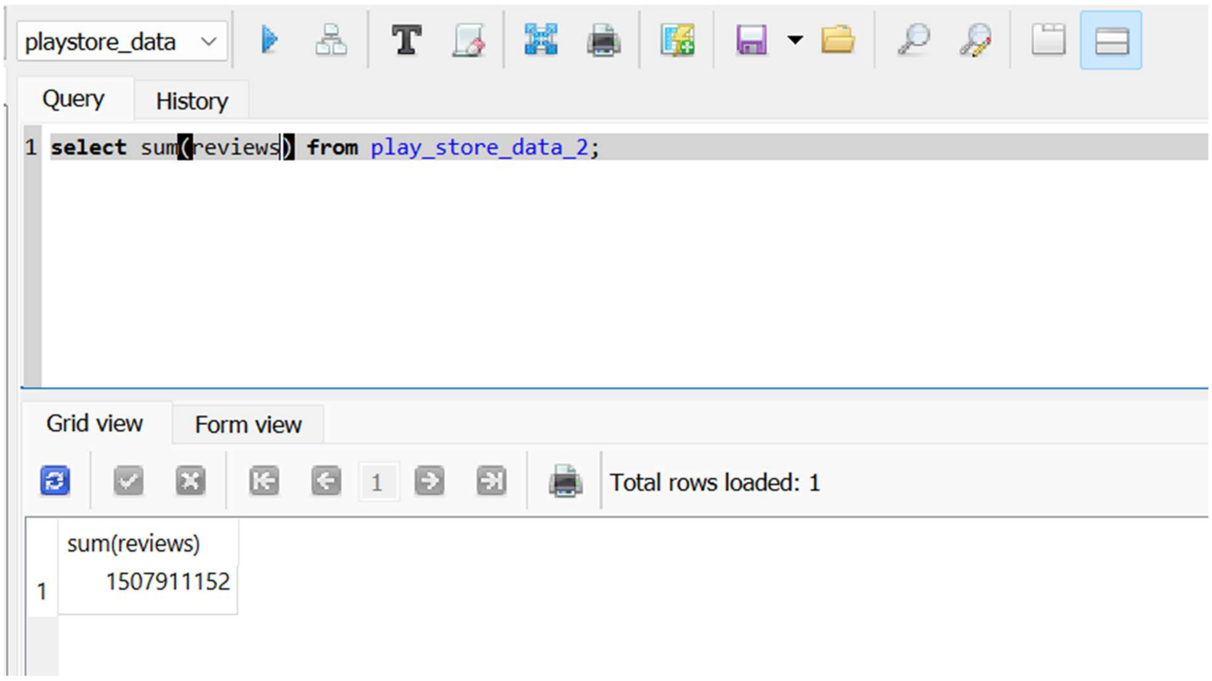
```
  count(*)
1     2589
```

10. Sum() adds up the values in a numeric column.
Often used with `GROUP BY` to calculate totals per category, customer, or product.
Works only on numeric data types (e.g., integers, decimals).



```
select sum(installs) from play_store_data_2;
```

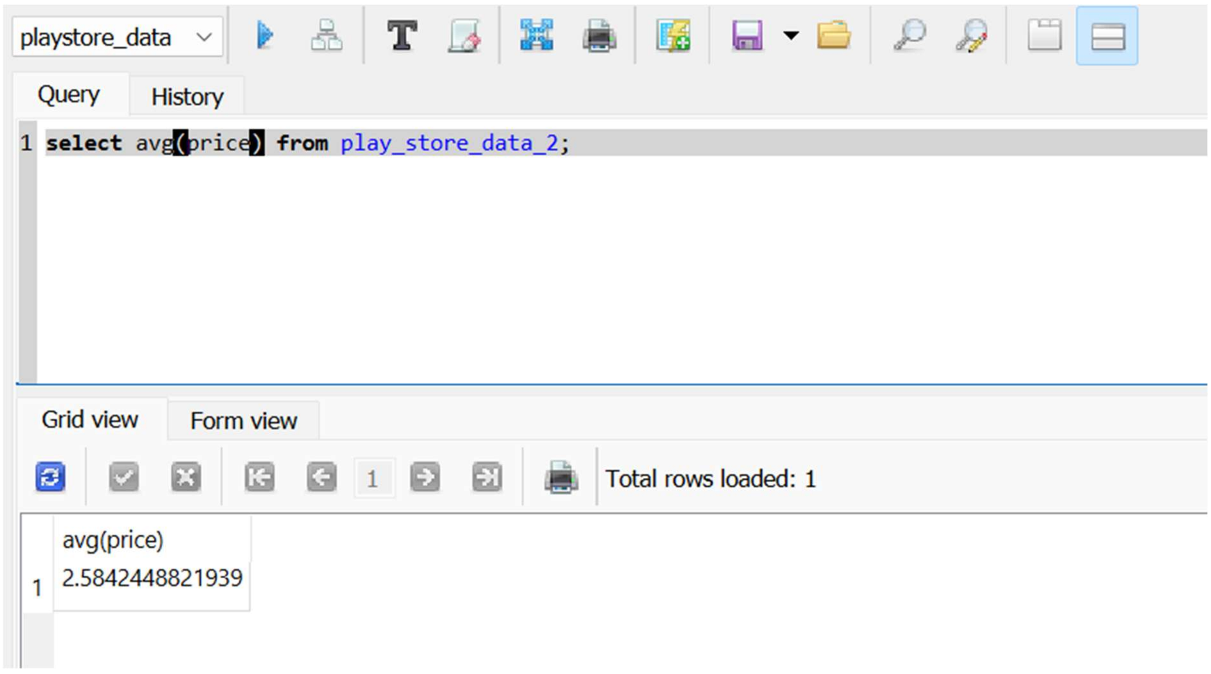
```
  sum(installs)
1        172499
```



```
select sum(reviews) from play_store_data_2;
```

```
  sum(reviews)
1   1507911152
```

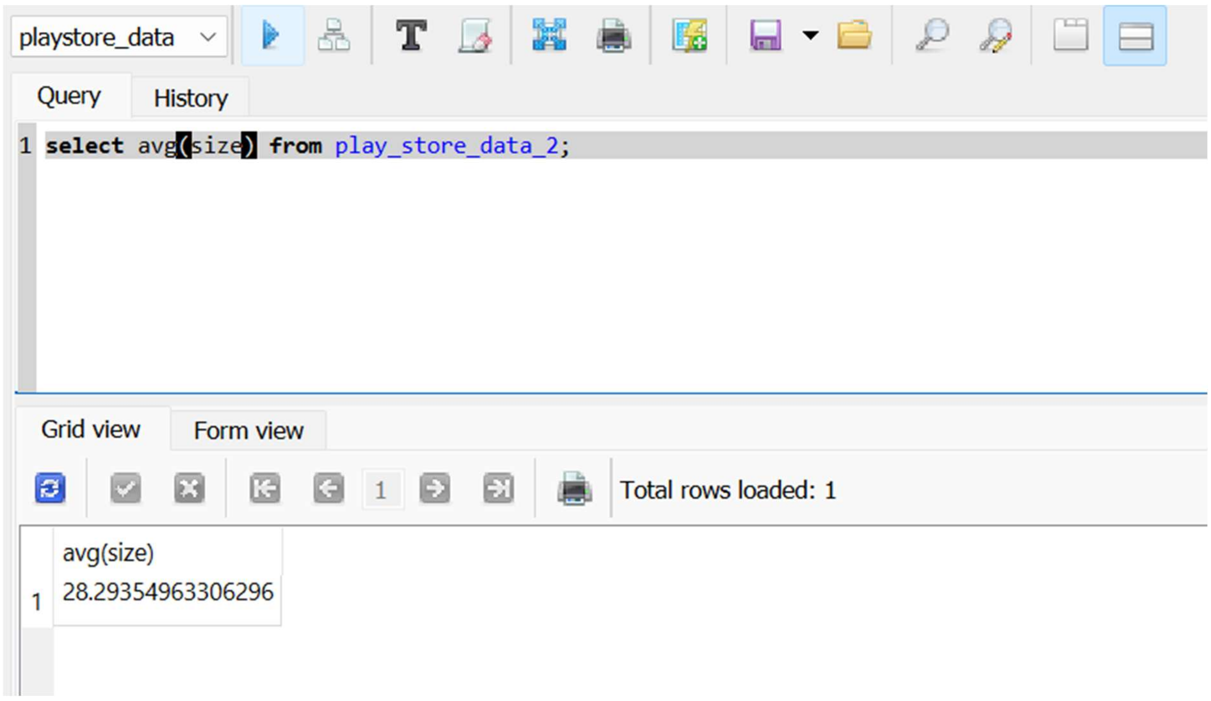
11. `AVG()` calculates the **average value** of a numeric column.
Often used with `GROUP BY` to find averages per category, customer, or product.
Ignores `NULL` values automatically.



```
select avg(price) from play_store_data_2;
```

Total rows loaded: 1

```
   avg(price)
1  2.5842448821939
```



```
select avg(size) from play_store_data_2;
```

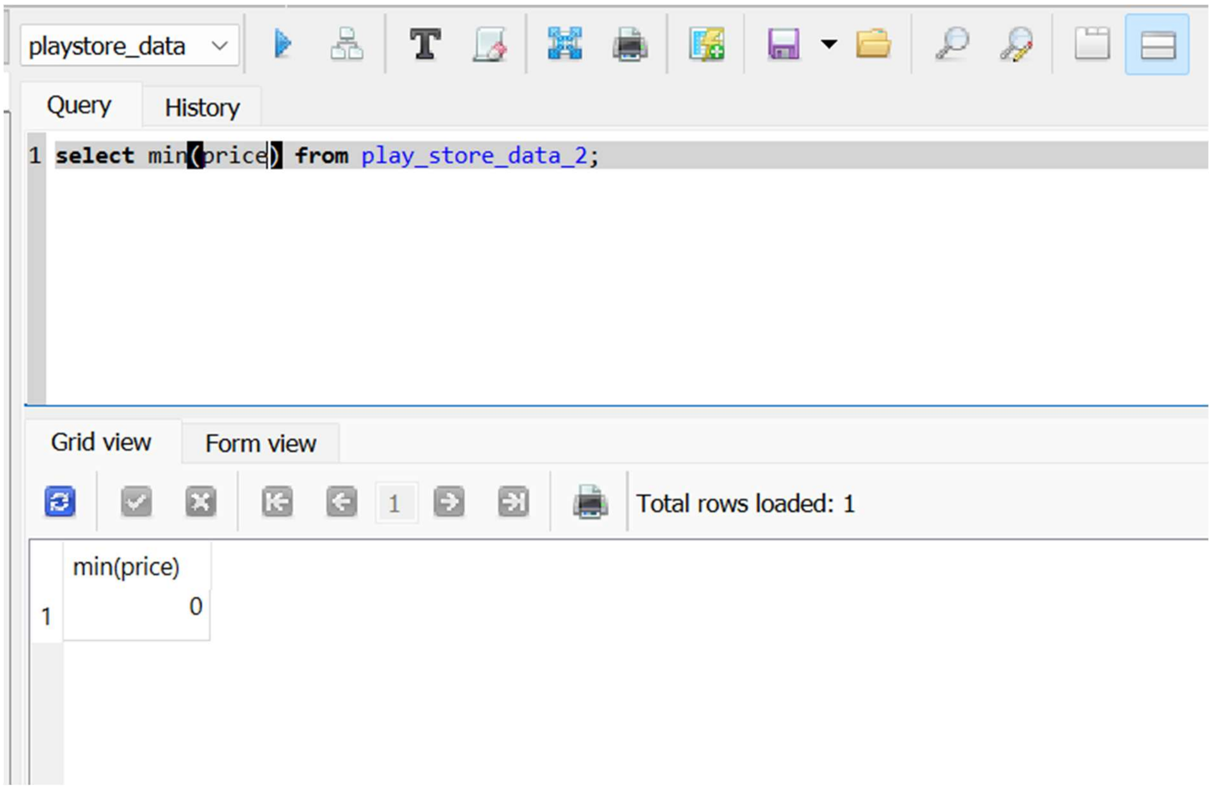
Total rows loaded: 1

```
   avg(size)
1  28.29354963306296
```

12. MIN() returns the **smallest value** in a column.

Often used to find the **earliest date, lowest price, smallest size, or minimum rating**.

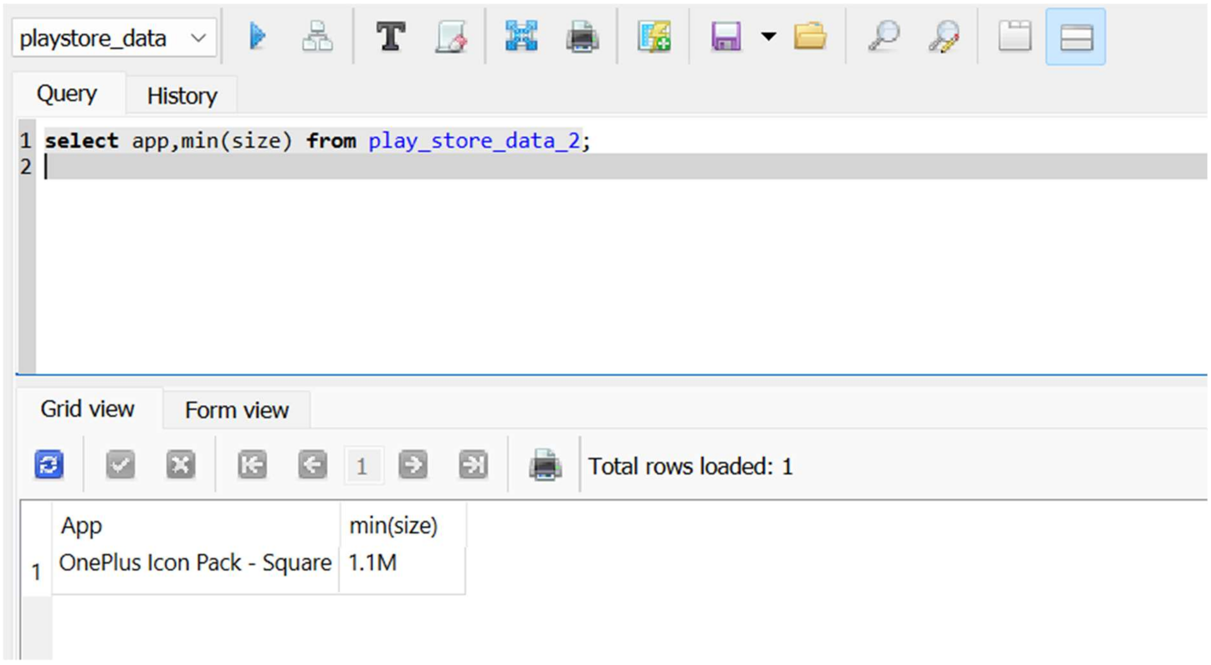
Ignores NULL values automatically.



```
select min(price) from play_store_data_2;
```

Total rows loaded: 1

```
  min(price)
1          0
```



```
select app,min(size) from play_store_data_2;
```

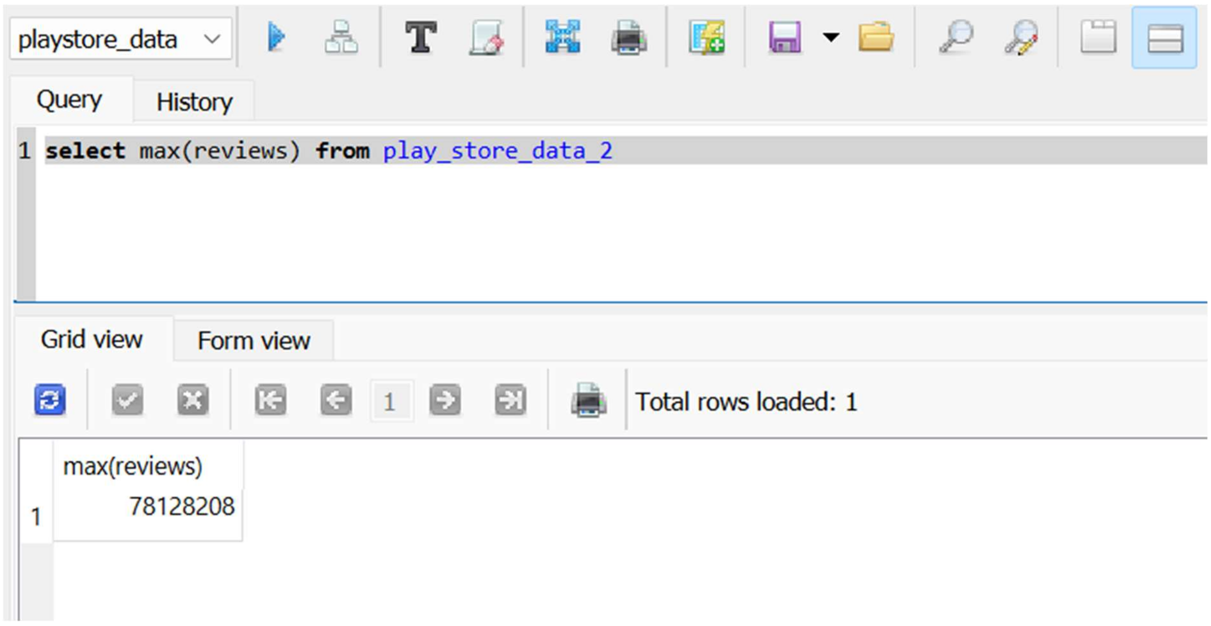
Total rows loaded: 1

```
  App                        min(size)
1 OnePlus Icon Pack - Square 1.1M
```

13. MAX() returns the **largest value** in a column.

Often used to find the **highest rating, most installs, latest date, or maximum price**.

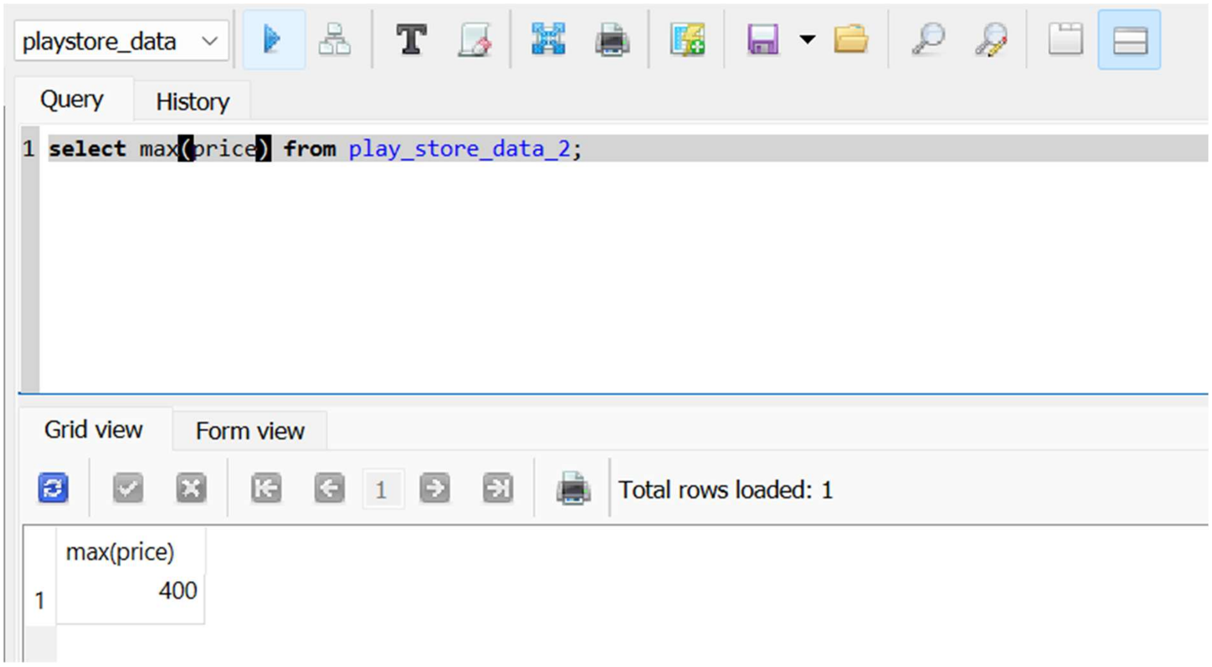
Ignores NULL values automatically.



```
select max(reviews) from play_store_data_2
```

Total rows loaded: 1

```
  max(reviews)
1     78128208
```



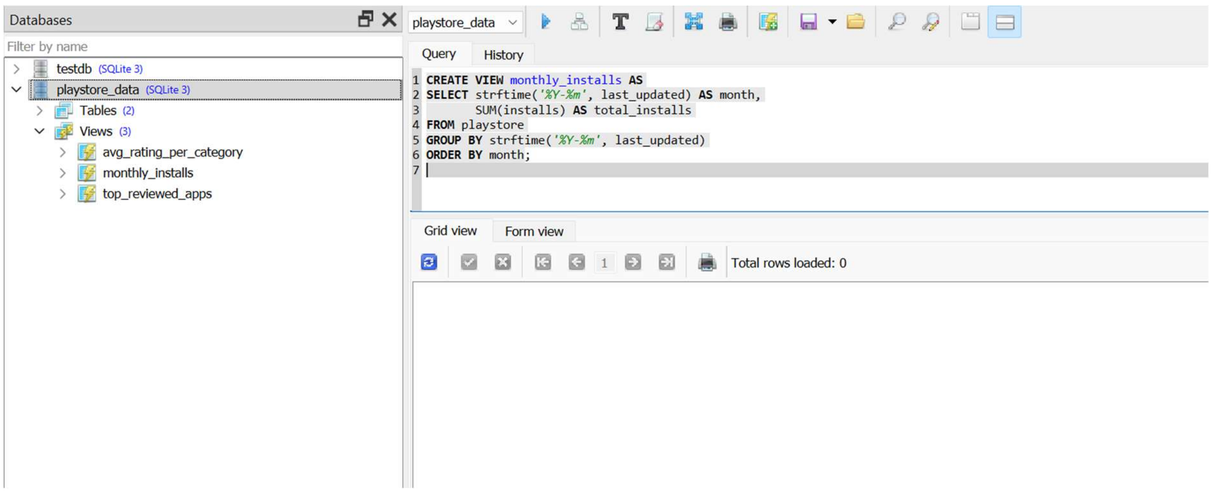
```
select max(price) from play_store_data_2;
```

Total rows loaded: 1

```
  max(price)
1        400
```

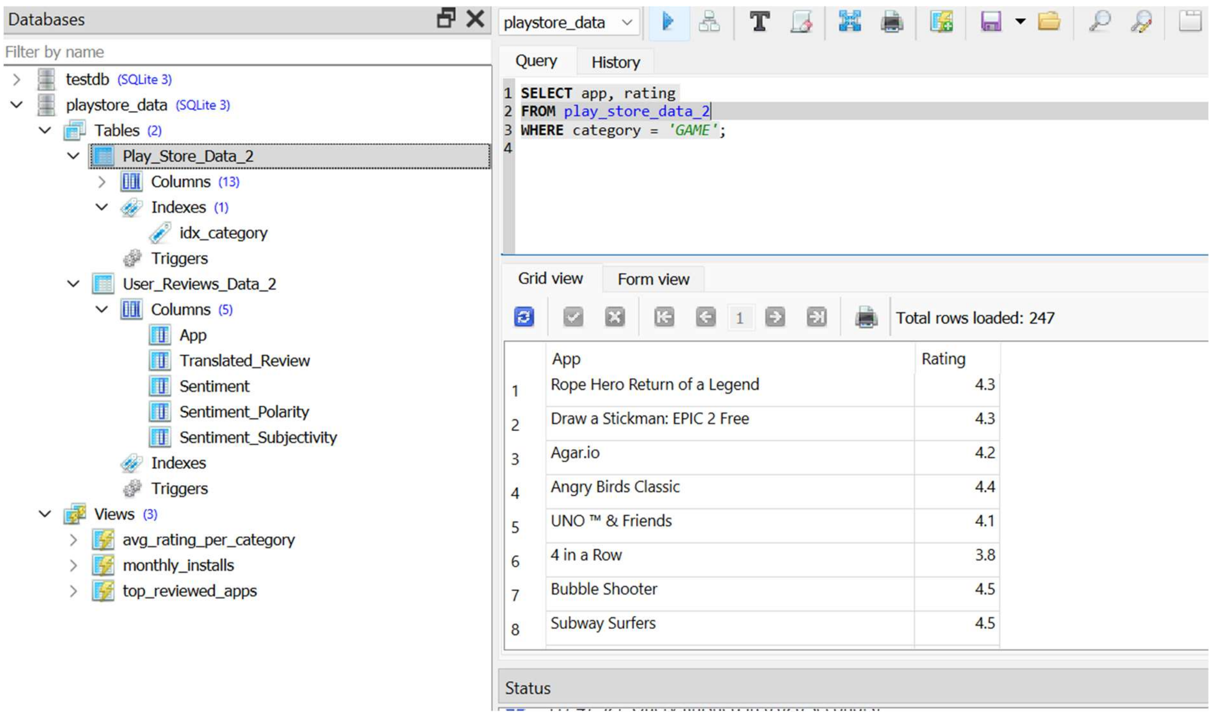
14. **SQL views for analysis**. Views are like reusable "virtual tables" that store complex queries so you can query them directly for insights.

They're especially useful in analytics because they simplify repeated queries and make dashboards easier to build.



```
CREATE VIEW monthly_installs AS
SELECT strftime('%Y-%m', last_updated) AS month,
       SUM(installs) AS total_installs
FROM playstore
GROUP BY strftime('%Y-%m', last_updated)
ORDER BY month;
```

Total rows loaded: 0



```
SELECT app, rating
FROM play_store_data_2
WHERE category = 'GAME';
```

Total rows loaded: 247

	App	Rating
1	Rope Hero Return of a Legend	4.3
2	Draw a Stickman: EPIC 2 Free	4.3
3	Agar.io	4.2
4	Angry Birds Classic	4.4
5	UNO ™ & Friends	4.1
6	4 in a Row	3.8
7	Bubble Shooter	4.5
8	Subway Surfers	4.5