

Instagram User Analytics

Project Description

For this project, I analyzed Instagram user data using SQL and MySQL Workbench in order to respond to inquiries from the management team. My observations will assist the team's product manager and other members in making defensible choices on the Instagram app's future.

This project aims to leverage my SQL expertise to get insightful conclusions from the data. My research could have an impact on how one of the most widely used social media platforms in the world develops in the future.

Approach

The approach towards this project is to use SQL queries to analyze the database provided.

Tech-Stack Used

MySQL Workbench 8.0 CE is the program that I used to do this project. It is employed for writing SQL queries and building databases. Additionally, I make Word file using Microsoft Word and store SQL queries and snapshots of output in the same.

Insights

Below are the insights that I have brought into light for the management team for their clear understanding about the analytics for the instagram users data.

- Top 5 Oldest Users of Instagram
- Users who never posted photos on Instagram
- Most liked photo on Instagram
- Top 5 most commonly used has-tags on Instagram
- Total number of users on Instagram
- Total number of photos on Instagram
- Average number of photos per user
- Bots and fake accounts on Instagram

SQL Tasks :

A) Marketing Analysis:

1. **Loyal User Reward:** The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
Your Task: Identify the five oldest users on Instagram from the provided database

CODE:

```
SELECT * FROM users
```

```
ORDER BY created_at
```

```
LIMIT 5;
```

OUTPUT:

The screenshot shows a SQL IDE interface. The top toolbar includes icons for file operations, a 'Limit to 1000 rows' dropdown, and search/execution icons. The SQL editor contains the following code:

```
359 -- Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the  
360 -- 1 Task: Identify the five oldest users on Instagram from the provided database.  
361 • SELECT * FROM users  
362 ORDER BY created_at  
363 LIMIT 5;  
364  
365  
366
```

Below the editor is the 'Result Grid' showing the output of the query. It has columns for 'id', 'username', and 'created_at'.

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26

At the bottom, the 'Output' pane shows the 'Action Output' for the query execution:

#	Time	Action	Message	Duration / Fetch
15	16:05:38	INSERT INTO tags(tag_name) VALUES ('sunset'), ('photography'), ('sunset'), ('lands...	21 row(s) affected Records: 21 Duplicates: 0 Warnings: 0	0.000 sec
16	16:05:45	INSERT INTO photo_tags(photo_id, tag_id) VALUES (1, 18), (1, 17), (1, 21), (1, 13)...	501 row(s) affected Records: 501 Duplicates: 0 Warnings: 0	0.015 sec
17	16:10:39	SELECT * FROM users ORDER BY created_at LIMIT 5	5 row(s) returned	0.000 sec / 0.000 sec

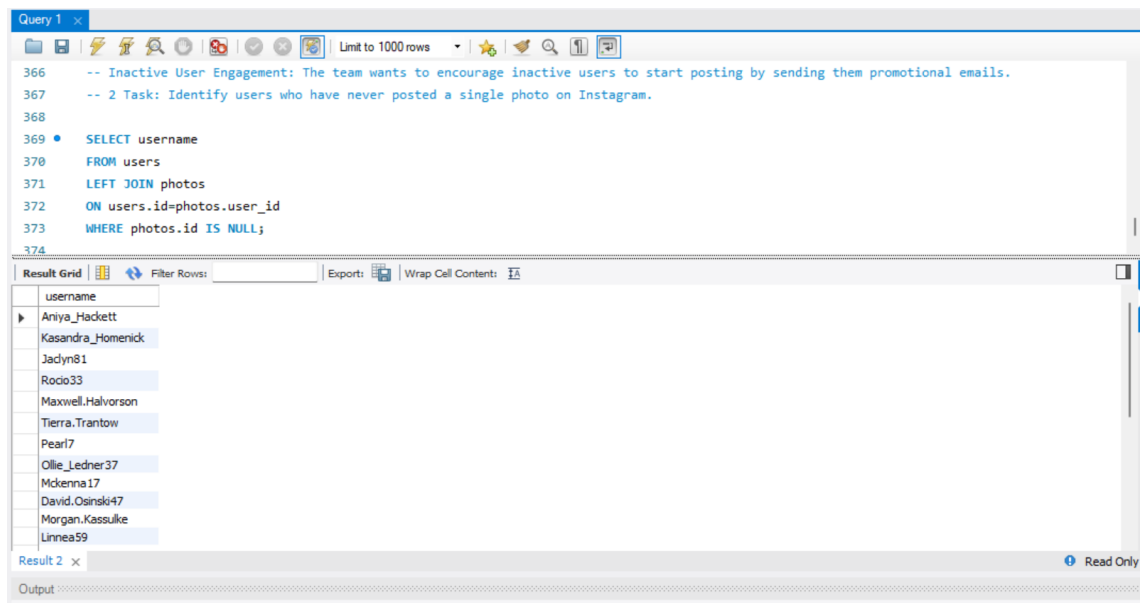
2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

Your Task: Identify users who have never posted a single photo on Instagram.

CODE:

```
SELECT username  
  
FROM users  
  
LEFT JOIN photos  
  
ON users.id=photos.user_id  
  
WHERE photos.id IS NULL;
```

OUTPUT:



The screenshot shows a SQL query editor with a query to find inactive users. The query is as follows:

```
366 -- Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.  
367 -- 2 Task: Identify users who have never posted a single photo on Instagram.  
368  
369 • SELECT username  
370 FROM users  
371 LEFT JOIN photos  
372 ON users.id=photos.user_id  
373 WHERE photos.id IS NULL;  
374
```

The results are displayed in a table with the following usernames:

username
Aniya_Hackett
Kassandra_Homenick
Jacyn81
Rodo33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59

The interface also includes a "Result Grid" tab, a "Filter Rows" button, an "Export" button, and a "Wrap Cell Content" button. The status bar at the bottom indicates "Read Only" and "Output".

3. Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

Your Task: Determine the winner of the contest and provide their details to the team.

CODE:

```
SELECT username, photos.id, photos.image_url, COUNT(likes.user_id) AS
total

FROM photos

INNER JOIN likes

ON likes.photo_id=photos.id

INNER JOIN users

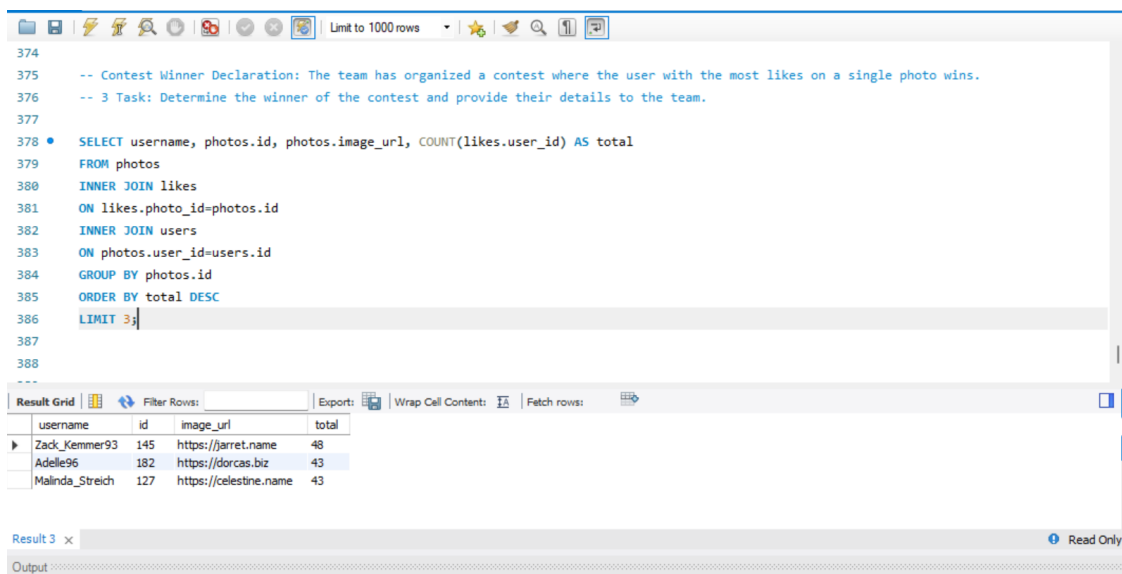
ON photos.user_id=users.id

GROUP BY photos.id

ORDER BY total DESC

LIMIT 3;
```

OUTPUT:



The screenshot shows a SQL IDE interface. The top toolbar includes icons for file operations, a search icon, and a 'Limit to 1000 rows' dropdown. The main editor area contains a SQL query with line numbers 374 to 388. The query is a SELECT statement that joins the photos, likes, and users tables to find the top 3 users by the number of likes on their photos. The bottom toolbar includes icons for result grid, filter rows, export, wrap cell contents, and fetch rows. Below the toolbar is a 'Result Grid' table with 4 columns: username, id, image_url, and total. The table contains 3 rows of data. At the bottom, there is a 'Result 3' tab and an 'Output' section.

```
374
375 -- Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.
376 -- 3 Task: Determine the winner of the contest and provide their details to the team.
377
378 • SELECT username, photos.id, photos.image_url, COUNT(likes.user_id) AS total
379 FROM photos
380 INNER JOIN likes
381 ON likes.photo_id=photos.id
382 INNER JOIN users
383 ON photos.user_id=users.id
384 GROUP BY photos.id
385 ORDER BY total DESC
386 LIMIT 3;
387
388
```

username	id	image_url	total
Zack_Kemmer93	145	https://jarret.name	48
Adelle96	182	https://dorcus.biz	43
Malinda_Streich	127	https://celestine.name	43

Result 3 x Read Only

Output :

4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

CODE:

```
select * from tags;

with top_tags as

(select tag_id from photo_tags

group by tag_id

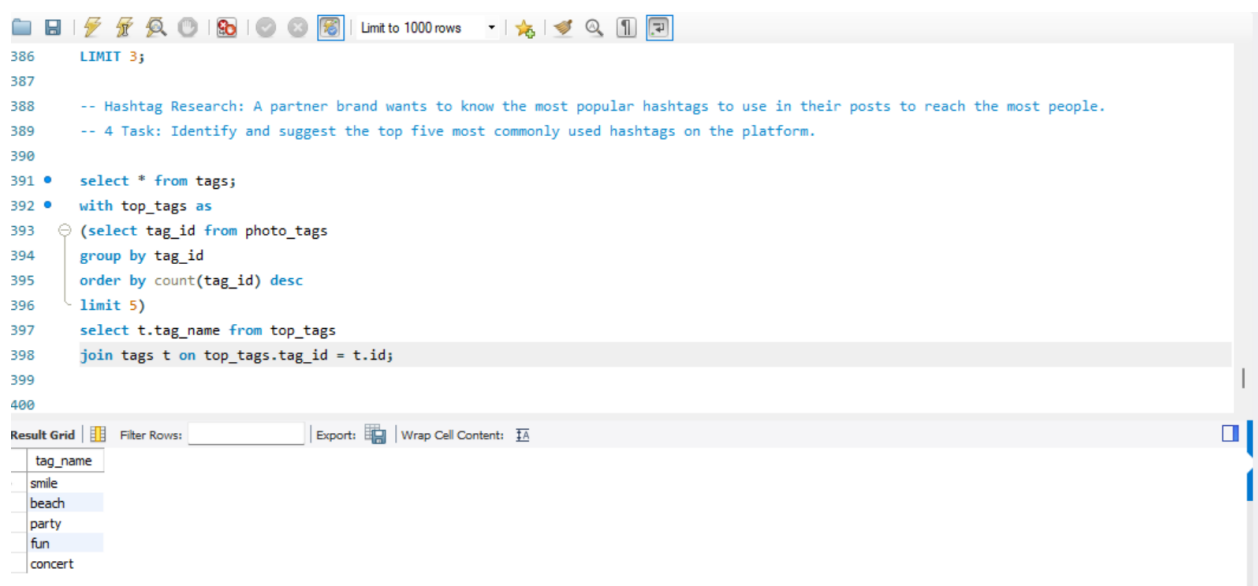
order by count(tag_id) desc

limit 5)

select t.tag_name from top_tags

join tags t on top_tags.tag_id = t.id;
```

OUTPUT:



The screenshot shows a SQL query editor interface. The top toolbar includes icons for file operations, a 'Limit to 1000 rows' dropdown, and other utility icons. The query text is as follows:

```
386 LIMIT 3;
387
388 -- Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
389 -- 4 Task: Identify and suggest the top five most commonly used hashtags on the platform.
390
391 • select * from tags;
392 • with top_tags as
393   (select tag_id from photo_tags
394    group by tag_id
395    order by count(tag_id) desc
396    limit 5)
397   select t.tag_name from top_tags
398   join tags t on top_tags.tag_id = t.id;
399
400
```

Below the query editor, the 'Result Grid' is visible. It has a 'Filter Rows' input field, an 'Export' button, and a 'Wrap Cell Contents' checkbox. The results are displayed in a table with the following data:

tag_name
smile
beach
party
fun
concert

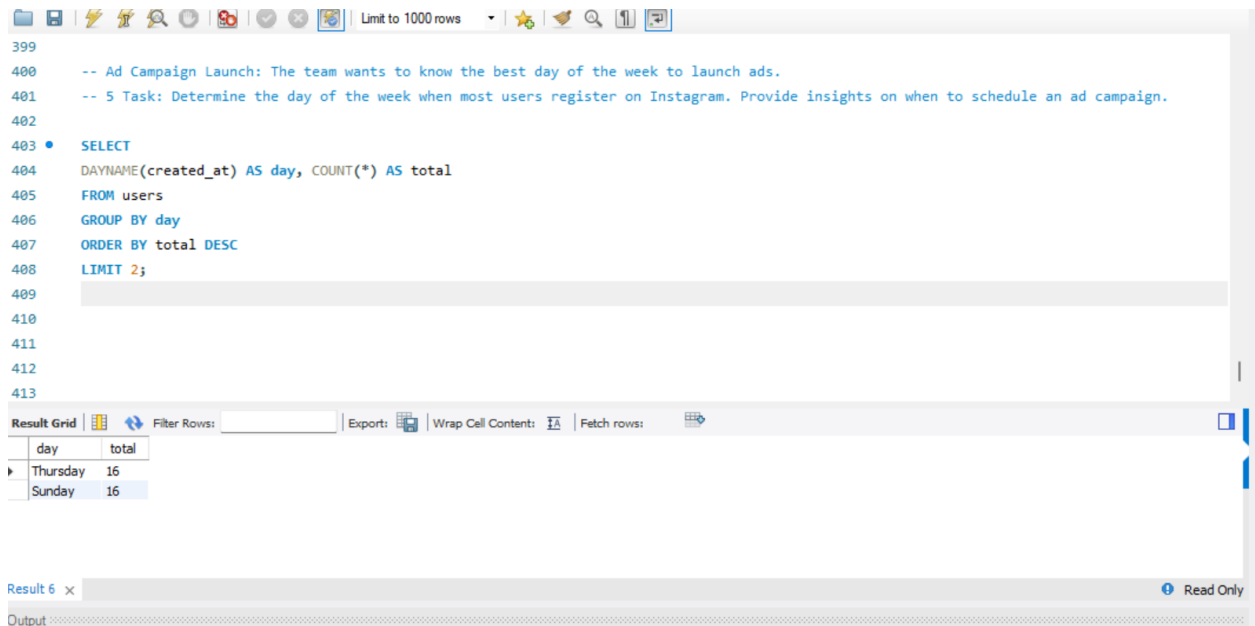
5. Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

CODE:

```
SELECT  
  
DAYNAME(created_at) AS day, COUNT(*) AS total  
  
FROM users  
  
GROUP BY day  
  
ORDER BY total DESC  
  
LIMIT 2;
```

OUTPUT:



The screenshot shows a SQL IDE interface. The top toolbar includes icons for file operations, execution, and settings, along with a 'Limit to 1000 rows' dropdown. The main editor area contains a SQL query with line numbers 399 to 413. The query is as follows:

```
399  
400 -- Ad Campaign Launch: The team wants to know the best day of the week to launch ads.  
401 -- 5 Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.  
402  
403 • SELECT  
404 DAYNAME(created_at) AS day, COUNT(*) AS total  
405 FROM users  
406 GROUP BY day  
407 ORDER BY total DESC  
408 LIMIT 2;  
409  
410  
411  
412  
413
```

Below the editor is the 'Result Grid' section, which displays the query results in a table format. The table has two columns: 'day' and 'total'. The results are:

day	total
Thursday	16
Sunday	16

At the bottom of the interface, there is a 'Result 6' tab and an 'Output' section.

B) Investor Metrics:

1. **User Engagement:** Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

Your Task: Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

CODE:

```
SELECT
```

```
(SELECT COUNT(*) FROM photos) / ( SELECT COUNT(*) FROM users) AS  
avg;
```

```
-- Calculate the average number of posts per user on Instagram post count by  
user
```

```
select user_id, count(*) as posts_count from photos
```

```
group by user_id
```

```
order by posts_count desc;
```

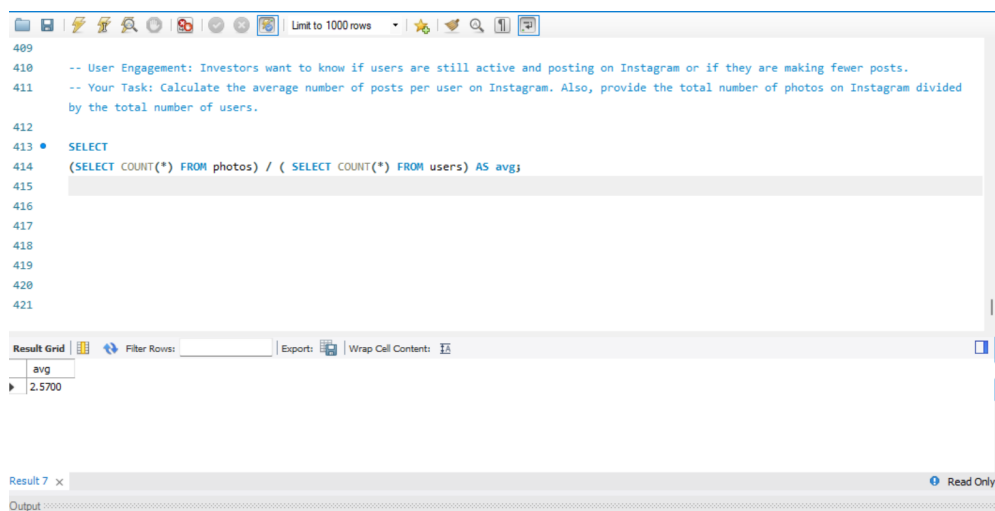
```
-- average_post per user
```

```
SELECT AVG(posts_count) as avg_posts_per_user FROM ( select user_id,  
count(*) as posts_count from photos
```

```
group by user_id
```

```
order by posts_count desc) as user_posts;
```

OUTPUT:



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
409  
410 -- User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.  
411 -- Your Task: Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided  
412 by the total number of users.  
413 • SELECT  
414 (SELECT COUNT(*) FROM photos) / ( SELECT COUNT(*) FROM users) AS avg;  
415  
416  
417  
418  
419  
420  
421
```

The output is displayed in a table below the query:

Result Grid
avg
2.5700

At the bottom, there is a tab labeled "Result 7" and a "Read Only" indicator.

2. **Bots & Fake Accounts:** Investors want to know if the platform is crowded with fake and dummy accounts.

Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

CODE:

```
SELECT users.id,username, COUNT(users.id) As total_likes_by_user

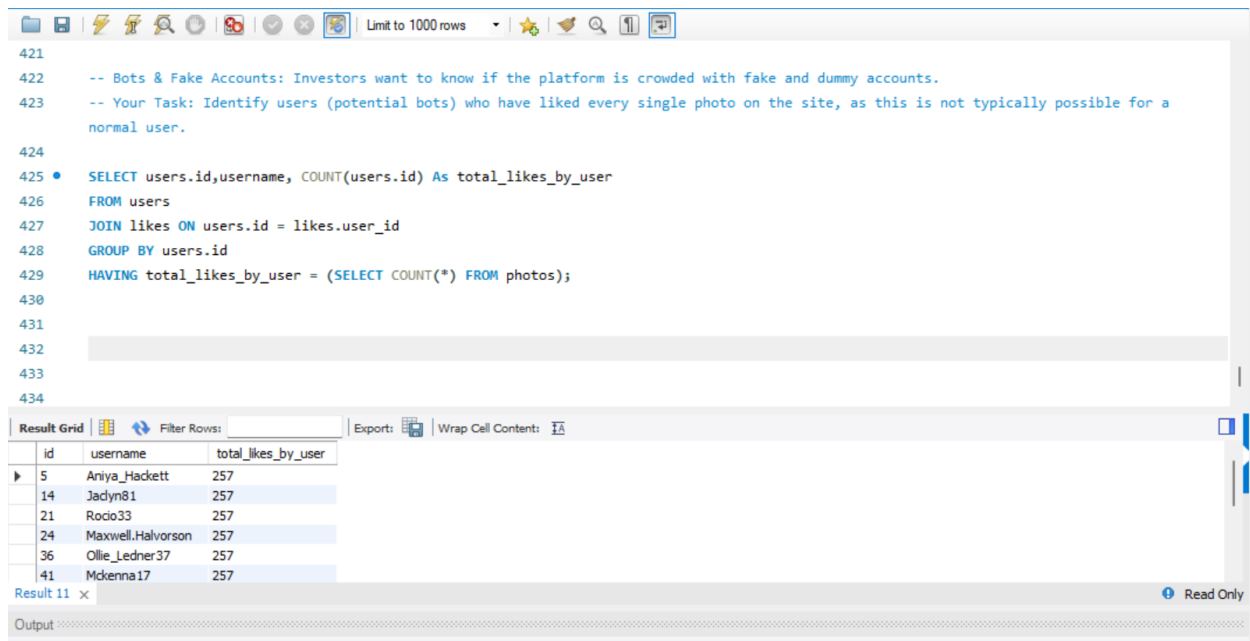
FROM users

JOIN likes ON users.id = likes.user_id

GROUP BY users.id

HAVING total_likes_by_user = (SELECT COUNT(*) FROM photos);
```

OUTPUT:



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
-- Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.
-- Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

SELECT users.id,username, COUNT(users.id) As total_likes_by_user
FROM users
JOIN likes ON users.id = likes.user_id
GROUP BY users.id
HAVING total_likes_by_user = (SELECT COUNT(*) FROM photos);
```

Below the query, a "Result Grid" is displayed with the following data:

	id	username	total_likes_by_user
▶	5	Aniya_Hackett	257
	14	Jadyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	McKenna17	257

At the bottom, it says "Result 11 x" and "Read Only".

Result

I learned from this assignment how data analysts and business professionals work with real-time data to make data-driven decisions. I gather that this project had a pretty modest dataset in terms of rows and columns, but even so, working on a project of this nature was a great experience. It greatly aided in my comprehension of the analysis process and offered guidance for making the best choice.