Name : Mehnaz Shafeek

Reg No: 22BHI10211

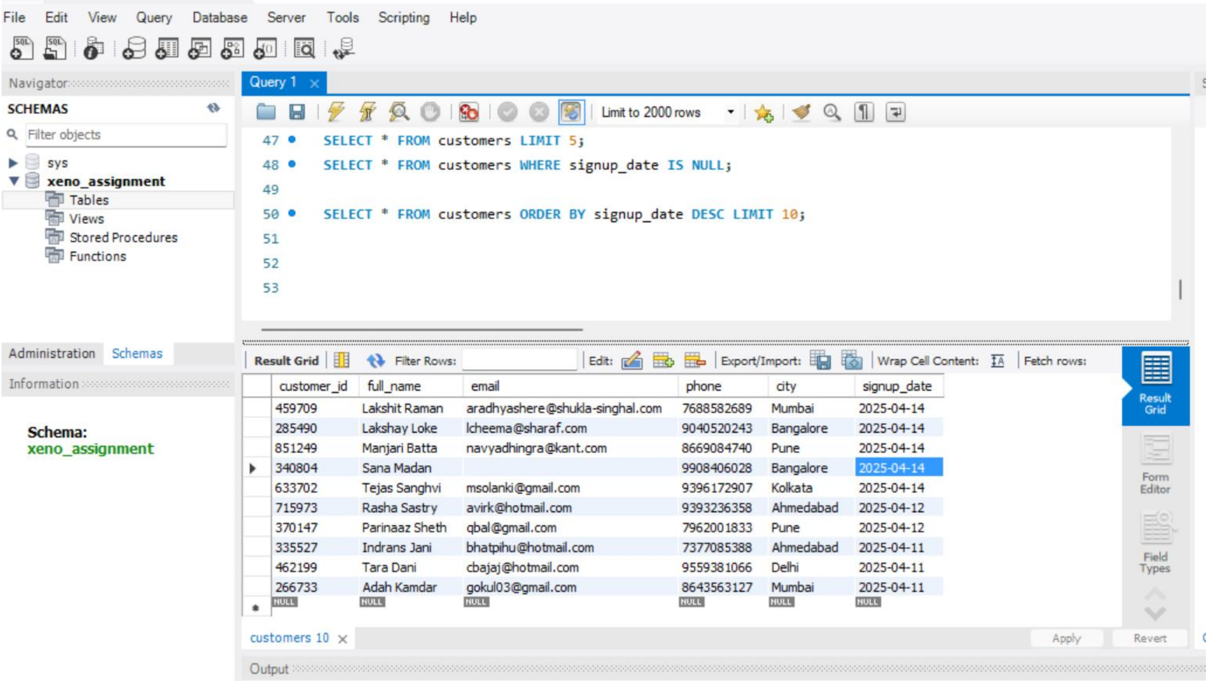
Implementation Internship Assignment

1. **SQL & Data Familiarity**

1. What steps would you take to review this data before importing it into a system? Please explain the process in not more than 3 lines.

Ans. I would validate the structure of the dataset, encoding (UTF-8), and consistent delimiters and headers before importing. After which, I will validate the data types, formats (particularly for dates), and conformity to the target schema. I would identify, then resolve, missing, duplicate, or invalid records to maintain referential integrity. Finally, I would do a small test import to confirm that the mapping was successful and compatible with the system.

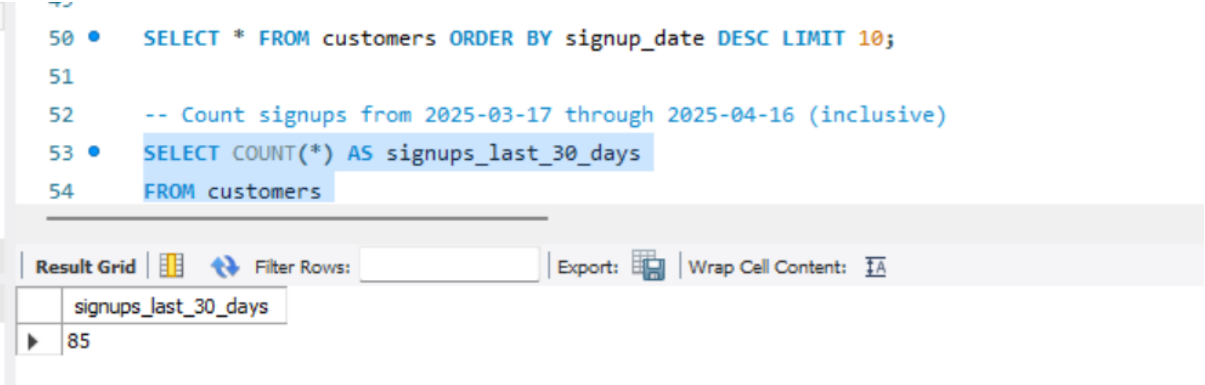
2.1 Write a query to display all customers from the city 'Delhi'.



```sql
SELECT * FROM customers LIMIT 5;
SELECT * FROM customers WHERE signup_date IS NULL;

SELECT * FROM customers ORDER BY signup_date DESC LIMIT 10;
```

customer_id	full_name	email	phone	city	signup_date
459709	Lakshit Raman	aradhyashere@shukla-singhal.com	7688582689	Mumbai	2025-04-14
285490	Lakshay Loke	lcheema@sharaf.com	9040520243	Bangalore	2025-04-14
851249	Manjari Batta	navyadhingra@kant.com	8669084740	Pune	2025-04-14
340804	Sana Madan		9908406028	Bangalore	2025-04-14
633702	Tejas Sanghvi	msolanki@gmail.com	9396172907	Kolkata	2025-04-14
715973	Rasha Sastry	avirk@hotmail.com	9393236358	Ahmedabad	2025-04-12
370147	Parinaaz Sheth	qbal@gmail.com	7962001833	Pune	2025-04-12
335527	Indrans Jani	bhatpihu@hotmail.com	7377085388	Ahmedabad	2025-04-11
462199	Tara Dani	cbajaj@hotmail.com	9559381066	Delhi	2025-04-11
266733	Adah Kamdar	gokul03@gmail.com	8643563127	Mumbai	2025-04-11
NULL	NULL	NULL	NULL	NULL	NULL

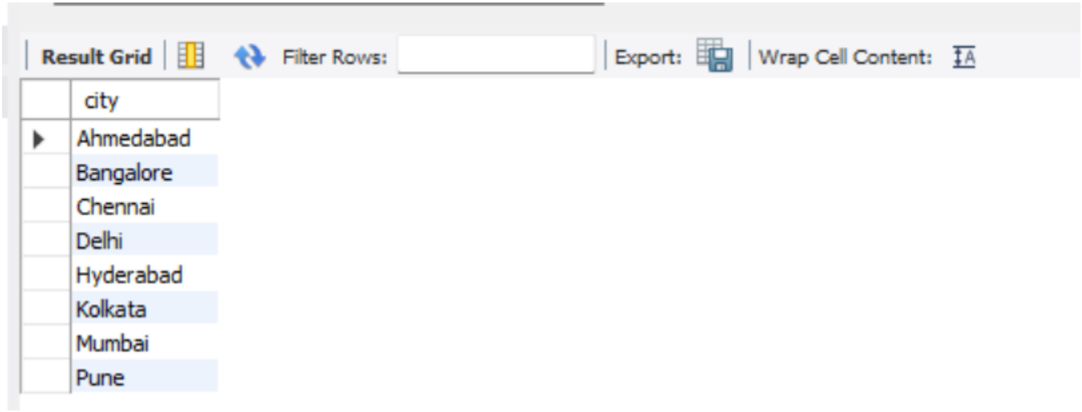
2.2 Count the number of signups in the last 30 days. Assume today to be 16th April 2025



```sql
SELECT * FROM customers ORDER BY signup_date DESC LIMIT 10;

-- Count signups from 2025-03-17 through 2025-04-16 (inclusive)
SELECT COUNT(*) AS signups_last_30_days
FROM customers
```

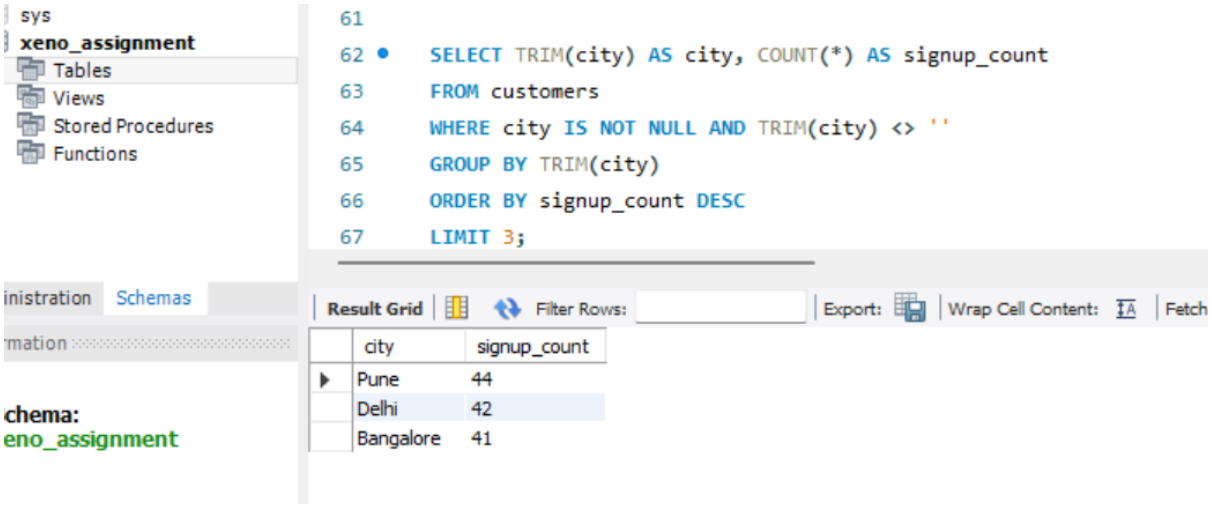
signups_last_30_days
85

2.3 List unique cities where customers are based



city
Ahmedabad
Bangalore
Chennai
Delhi
Hyderabad
Kolkata
Mumbai
Pune

2.4 List the top 3 cities by number of signups.



```
SELECT TRIM(city) AS city, COUNT(*) AS signup_count
FROM customers
WHERE city IS NOT NULL AND TRIM(city) <> ''
GROUP BY TRIM(city)
ORDER BY signup_count DESC
LIMIT 3;
```

city	signup_count
Pune	44
Delhi	42
Bangalore	41

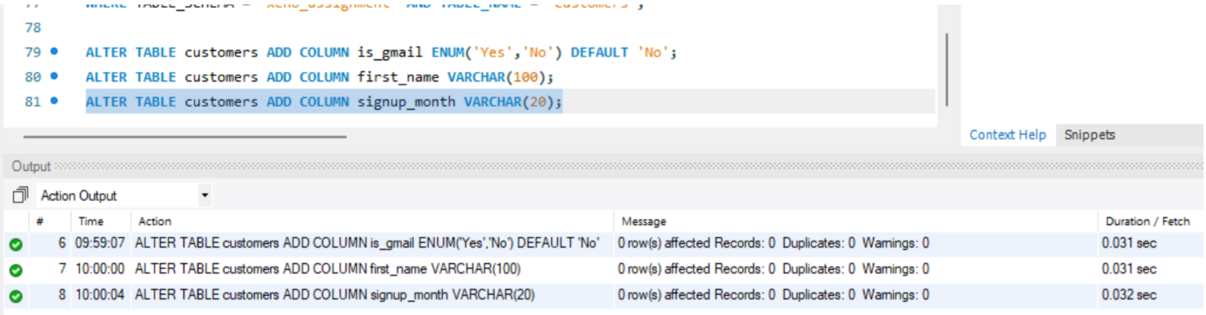
2.5 Assume there’s another table orders (customer_id, order_id, amount). How would you find customers who have never placed an order?

This SQL query makes it possible to identify those customers who have never placed an order.It uses the NOT EXISTS condition to compare the customers table against the orders table. Results will include only those customers who do not have any matching records in the orders table.

```
SELECT 1

FROM orders o

WHERE o.customer_id = c.customer_id
```

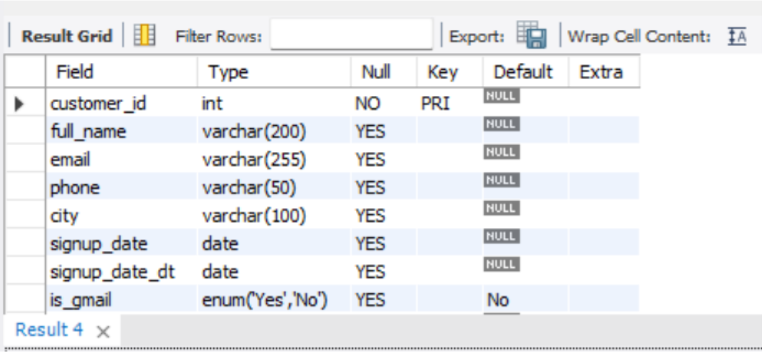
2.Data Transformation and Enrichment

1. Add a new column to show if the email domain is ‘gmail.com’ or not. Fill it with ‘Yes’ or ‘No.



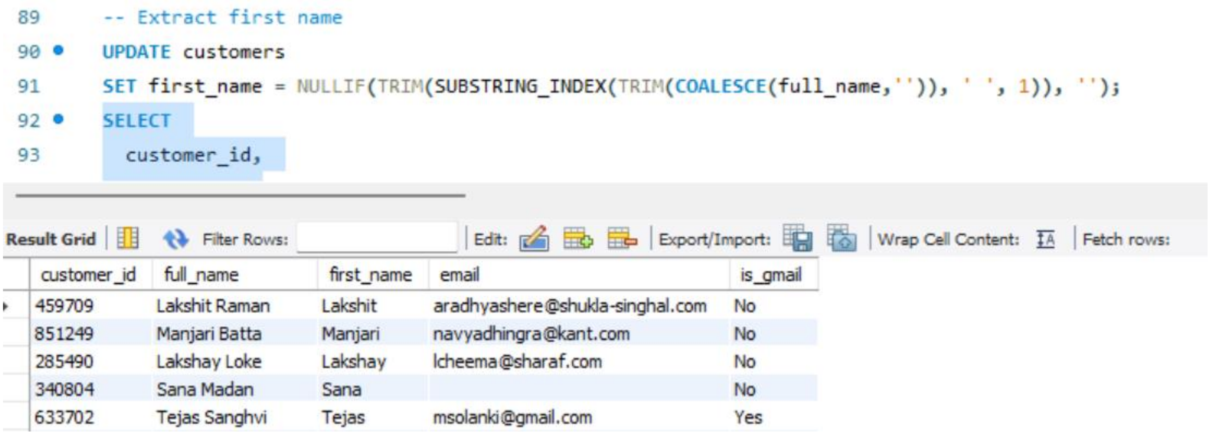
```sql
    WHERE TABLE_SCHEMA = 'xeno_assignment' AND TABLE_NAME = 'customers';

ALTER TABLE customers ADD COLUMN is_gmail ENUM('Yes','No') DEFAULT 'No';
ALTER TABLE customers ADD COLUMN first_name VARCHAR(100);
ALTER TABLE customers ADD COLUMN signup_month VARCHAR(20);
```

	#	Time	Action	Message	Duration / Fetch
✓	6	09:59:07	ALTER TABLE customers ADD COLUMN is_gmail ENUM('Yes','No') DEFAULT 'No'	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.031 sec
✓	7	10:00:00	ALTER TABLE customers ADD COLUMN first_name VARCHAR(100)	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.031 sec
✓	8	10:00:04	ALTER TABLE customers ADD COLUMN signup_month VARCHAR(20)	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.032 sec



Field	Type	Null	Key	Default	Extra
customer_id	int	NO	PRI	NULL	
full_name	varchar(200)	YES		NULL	
email	varchar(255)	YES		NULL	
phone	varchar(50)	YES		NULL	
city	varchar(100)	YES		NULL	
signup_date	date	YES		NULL	
signup_date_dt	date	YES		NULL	
is_gmail	enum('Yes','No')	YES		No	

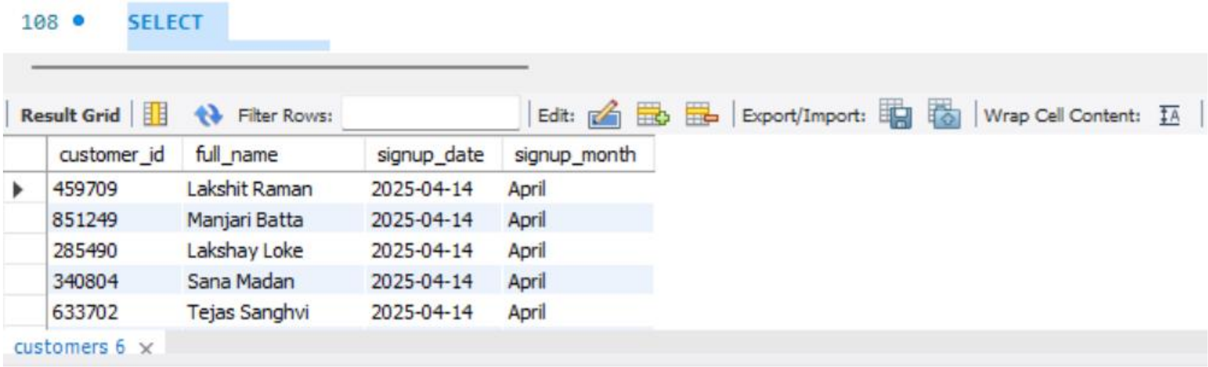
2. Extract the first name from the name column and store it in ‘first_name’



```sql
-- Extract first name
UPDATE customers
SET first_name = NULLIF(TRIM(SUBSTRING_INDEX(TRIM(COALESCE(full_name,'')), ' ', 1)), '');
SELECT
  customer_id,
```

customer_id	full_name	first_name	email	is_gmail
459709	Lakshit Raman	Lakshit	aradhyashere@shukla-singhal.com	No
851249	Manjari Batta	Manjari	navyadhingra@kant.com	No
285490	Lakshay Loke	Lakshay	lcheema@sharaf.com	No
340804	Sana Madan	Sana		No
633702	Tejas Sanghvi	Tejas	msolanki@gmail.com	Yes

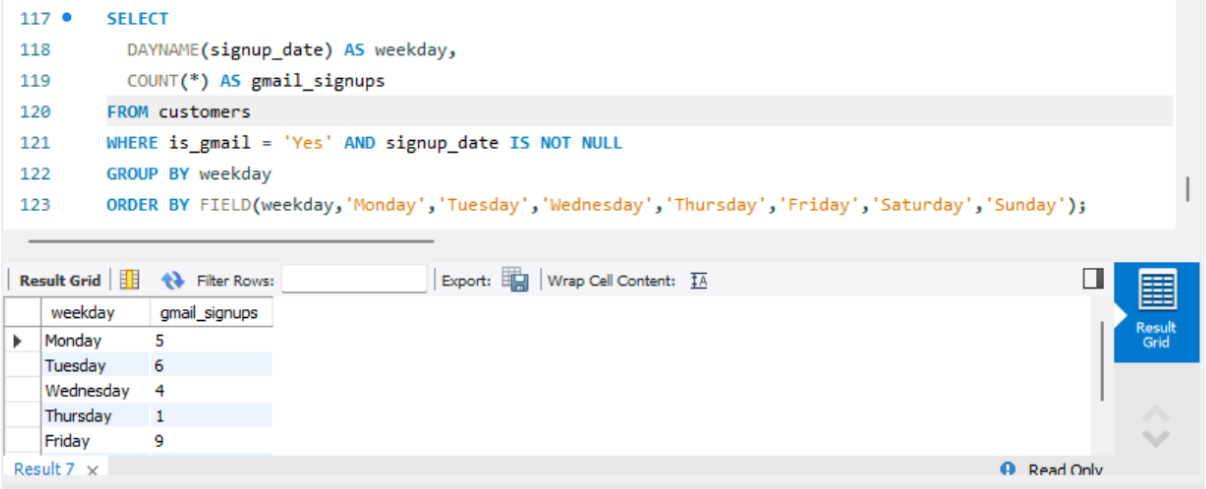
3. Add a column ‘signup_month’ to capture the month name the customer signed up.



```sql
SELECT
```

customer_id	full_name	signup_date	signup_month
459709	Lakshit Raman	2025-04-14	April
851249	Manjari Batta	2025-04-14	April
285490	Lakshay Loke	2025-04-14	April
340804	Sana Madan	2025-04-14	April
633702	Tejas Sanghvi	2025-04-14	April

4. Create a report which shows the no. of GMAIL customers who have signed up for each day of the week.

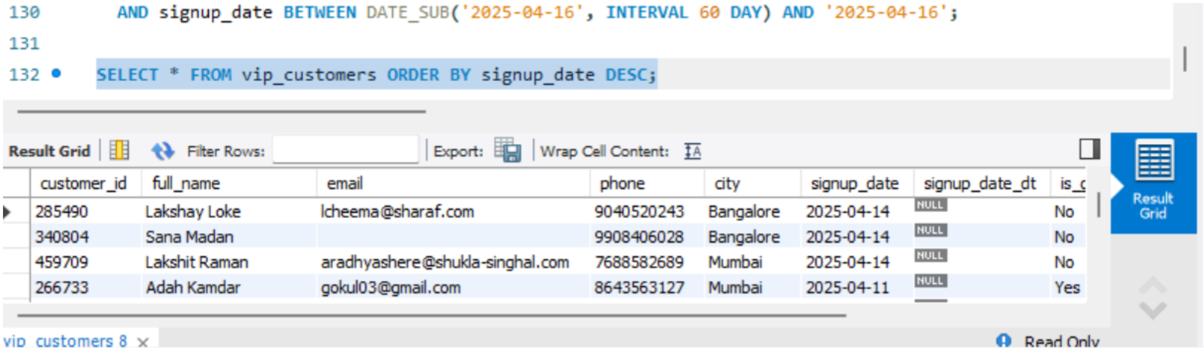


```
SELECT
  DAYNAME(signup_date) AS weekday,
  COUNT(*) AS gmail_signups
FROM customers
WHERE is_gmail = 'Yes' AND signup_date IS NOT NULL
GROUP BY weekday
ORDER BY FIELD(weekday,'Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday');
```

weekday	gmail_signups
Monday	5
Tuesday	6
Wednesday	4
Thursday	1
Friday	9

5. Create a new table ‘vip_customers’ with customers from Delhi, Mumbai and Bangalore

who have signed up in the last 60 days from 16th April 2025.

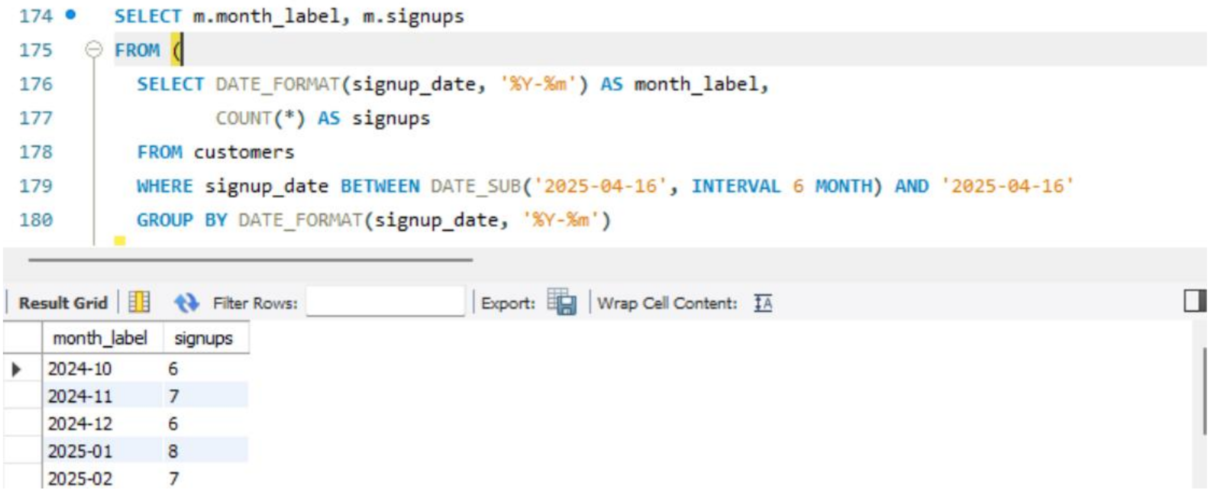


```
  AND signup_date BETWEEN DATE_SUB('2025-04-16', INTERVAL 60 DAY) AND '2025-04-16';

SELECT * FROM vip_customers ORDER BY signup_date DESC;
```

customer_id	full_name	email	phone	city	signup_date	signup_date_dt	is_
285490	Lakshay Loke	lcheema@sharaf.com	9040520243	Bangalore	2025-04-14	NULL	No
340804	Sana Madan		9908406028	Bangalore	2025-04-14	NULL	No
459709	Lakshit Raman	aradhyashere@shukla-singhal.com	7688582689	Mumbai	2025-04-14	NULL	No
266733	Adah Kamdar	gokul03@gmail.com	8643563127	Mumbai	2025-04-11	NULL	Yes

3. Analytics & Reporting

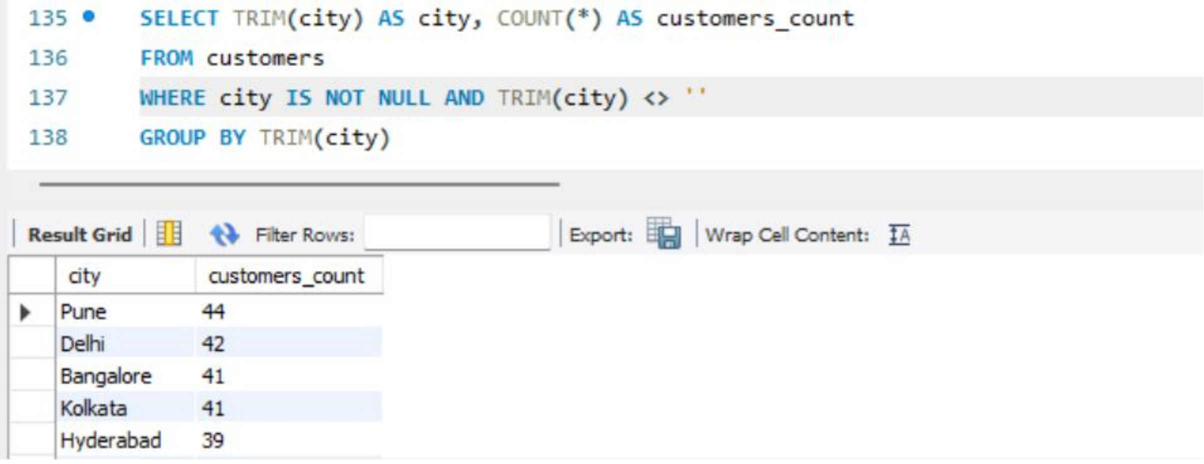
1. Show a monthly signup count for the past 6 months.



```
SELECT m.month_label, m.signups
FROM (
  SELECT DATE_FORMAT(signup_date, '%Y-%m') AS month_label,
         COUNT(*) AS signups
  FROM customers
  WHERE signup_date BETWEEN DATE_SUB('2025-04-16', INTERVAL 6 MONTH) AND '2025-04-16'
  GROUP BY DATE_FORMAT(signup_date, '%Y-%m')
```

month_label	signups
2024-10	6
2024-11	7
2024-12	6
2025-01	8
2025-02	7

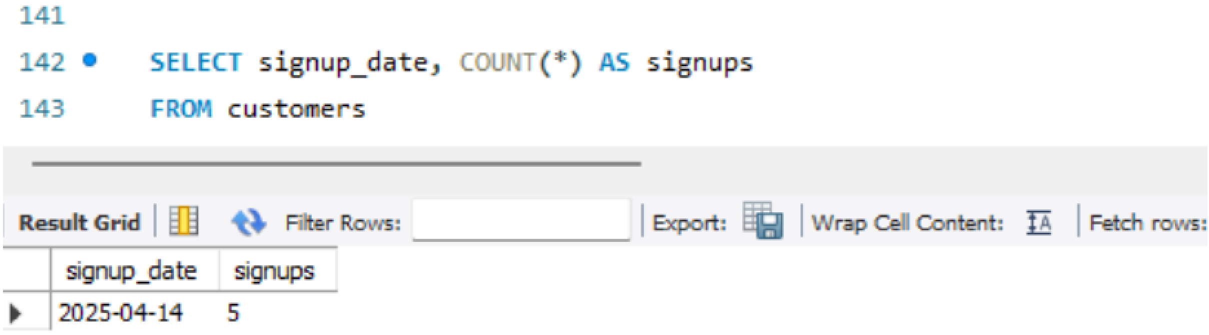
2.Get a list of cities with more than 20 customers.



```sql
SELECT TRIM(city) AS city, COUNT(*) AS customers_count
FROM customers
WHERE city IS NOT NULL AND TRIM(city) <> ''
GROUP BY TRIM(city)
```

city	customers_count
Pune	44
Delhi	42
Bangalore	41
Kolkata	41
Hyderabad	39

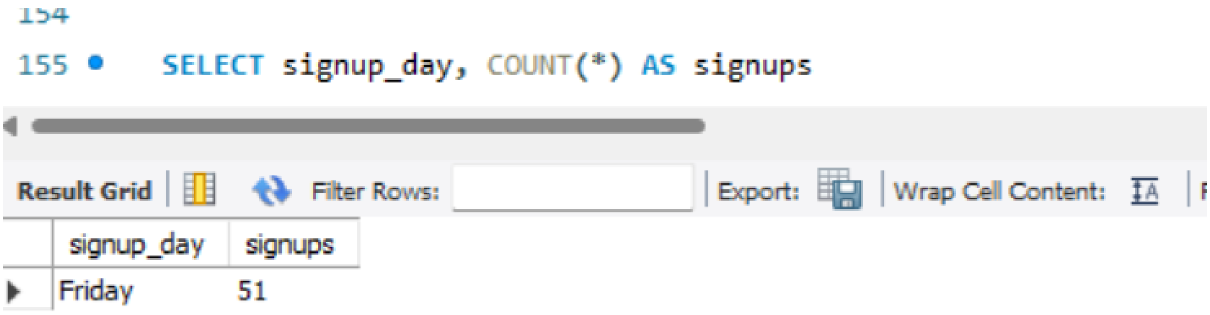
3.Find the date with the highest number of signups.



```sql

SELECT signup_date, COUNT(*) AS signups
FROM customers
```

signup_date	signups
2025-04-14	5

4. Add a new column to show the day of the signup date. Find the day with the highest number of signups.



```sql
SELECT signup_day, COUNT(*) AS signups
```

signup_day	signups
Friday	51