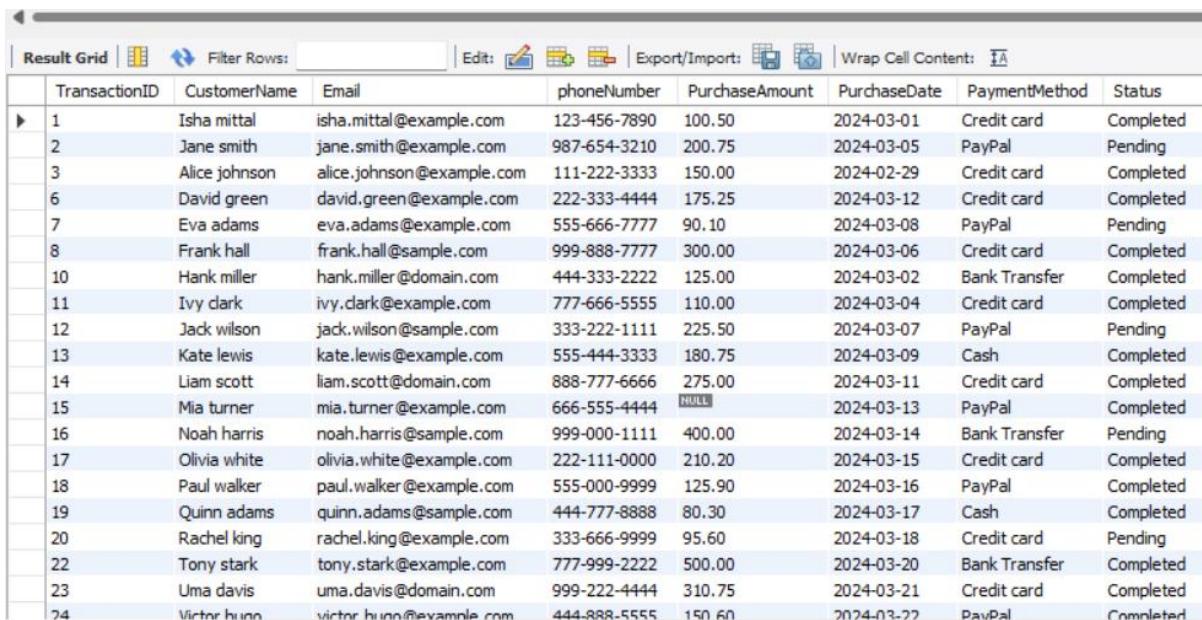# MySQL Data Cleaning Project

Overview

I have completed my **Data Cleaning & Preprocessing Project using MySQL**.
This project focuses on cleaning and preprocessing customer transaction data imported from Excel into MySQL. The goal was to fix missing data, remove duplicates, correct formatting issues, and prepare the dataset for analysis.

TransactionID	CustomerName	Email	phoneNumber	PurchaseAmount	PurchaseDate	PaymentMethod	Status
1	Isha mittal	isha.mittal@example.com	123-456-7890	100.50	2024-03-01	Credit card	Completed
2	Jane smith	jane.smith@example.com	987-654-3210	200.75	2024-03-05	PayPal	Pending
3	Alice johnson	alice.johnson@example.com	111-222-3333	150.00	2024-02-29	Credit card	Completed
6	David green	david.green@example.com	222-333-4444	175.25	2024-03-12	Credit card	Completed
7	Eva adams	eva.adams@example.com	555-666-7777	90.10	2024-03-08	PayPal	Pending
8	Frank hall	frank.hall@sample.com	999-888-7777	300.00	2024-03-06	Credit card	Completed
10	Hank miller	hank.miller@domain.com	444-333-2222	125.00	2024-03-02	Bank Transfer	Completed
11	Ivy clark	ivy.clark@example.com	777-666-5555	110.00	2024-03-04	Credit card	Completed
12	Jack wilson	jack.wilson@sample.com	333-222-1111	225.50	2024-03-07	PayPal	Pending
13	Kate lewis	kate.lewis@example.com	555-444-3333	180.75	2024-03-09	Cash	Completed
14	Liam scott	liam.scott@domain.com	888-777-6666	275.00	2024-03-11	Credit card	Completed
15	Mia turner	mia.turner@example.com	666-555-4444	NULL	2024-03-13	PayPal	Completed
16	Noah harris	noah.harris@sample.com	999-000-1111	400.00	2024-03-14	Bank Transfer	Pending
17	Olivia white	olivia.white@example.com	222-111-0000	210.20	2024-03-15	Credit card	Completed
18	Paul walker	paul.walker@example.com	555-000-9999	125.90	2024-03-16	PayPal	Completed
19	Quinn adams	quinn.adams@sample.com	444-777-8888	80.30	2024-03-17	Cash	Completed
20	Rachel king	rachel.king@example.com	333-666-9999	95.60	2024-03-18	Credit card	Pending
22	Tony stark	tony.stark@example.com	777-999-2222	500.00	2024-03-20	Bank Transfer	Completed
23	Uma davis	uma.davis@domain.com	999-222-4444	310.75	2024-03-21	Credit card	Completed
24	Victor hugo	victor.hugo@example.com	444-888-5555	150.60	2024-03-22	PayPal	Completed

Database & Table Creation

The database and table were created using SQL DDL commands before data cleaning began.

```
1       -- STEP 1 CREATE DATABASE
2  ●    Create database CustomerDB;
3  ●    USE CustomerDB;
4
5       -- Create Table
6  ●    Create table Customer_Transaction
7    ⊖  (TransactionID int primary key,
8       CustomerName varchar(100),
9       Email varchar(100),
10      phoneNumber varchar(20),
11      PurchaseAmount decimal(10,2),
12      PurchaseDate date,
13      PaymentMethod varchar(50),
14      Status varchar (20)
15      );
16
17
```

Tools Used

MySQL

Excel

Steps Performed

1. Handled Missing Emails

Filled blank email fields with a placeholder value.

```sql
-- 1 Fill all missing address with a placeholder
SET SQL_SAFE_UPDATES = 0;


update customer_transaction
SET EMAIL = 'UNKNOWN@EXAMPLE.COM'
where EMAIL = ''
;
```

2. Removed Duplicate Records

Deleted duplicate entries based on phone numbers using SQL grouping logic.

```sql
-- 2 Identify and removed duplicate records based on phone number

delete from customer_transaction
where TransactionID NOT IN
(select * FROM (
SELECT min(TransactionID)
FROM customer_transaction
group by phoneNumber
) AS TEMP_TABLE
);
```

3. Fixed Email Format Issues

Appended '.com' to emails that did not follow the correct format.

```sql
-- 3 Fix email format issues
Update customer_transaction
set email = concat(email, '.com')
where email not like '%.com';
```

4. Standardized Payment Method Names

Converted all card-related payment methods into a single label: 'Credit Card'.

```sql
-- 4 Standardize payment method names

update customer_transaction
set paymentmethod = 'Credit card'
where paymentmethod like '%card%';
```

5. Removed Failed Transactions

Deleted rows where transaction status was marked as 'failed'.

```
-- 5 Remove failed transaction from dataset
Delete from customer_transaction
where status = 'failed';
```

6. Corrected Negative Purchase Amounts

Replaced negative values with NULL to maintain data validity.

```
-- 6 Replace negative purchaseamount with null alter

update customer_transaction
set purchaseamount = null
where purchaseamount < 0;
```

7. Formatted Customer Names

Capitalized the first letter of each customer's name.

```
-- 7 Capitalize the first letter in customername
update customer_transaction
set customername = concat(
upper(left(customername,1)
),lower (substring(customername,2))
)
where customername is not null;
```