SQL CODE FOR DATA CLEANING:

SQL CODE FOR DATA CLEANING:

Create database New\_project;

use New\_project;

CREATE TABLE Ecommerce\_data (

customer\_id varchar(100),

customer\_first\_name varchar(100),

ustomer\_last\_name varchar(100),

category\_name varchar(100),

product\_name Varchar(250),

customer\_segment varchar(100),

customer\_city varchar(100),

customer\_state Varchar(100),

customer\_country varchar(100),

customer\_region varchar(100),

delivery\_status varchar(100),

order\_date varchar(100),

order\_id varchar(100),

ship\_date varchar(100),

shipping\_type varchar(100),

days\_for\_shipment\_scheduled INT,

days\_for\_shipment\_real INT,

order\_item\_discount DECIMAL(10, 2),

sales\_per\_order DECIMAL(10, 2),

order\_quantity INT,

profit\_per\_order DECIMAL(10, 2)

);

Drop table Ecommerce\_data;

select \* from Ecommerce\_data;

Create table Lt\_lg\_of\_states (

state varchar(30),

latitude DECIMAL(10, 2),

longitude DECIMAL(10, 2),

`name` varchar(30));

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/ecommerce\_data.csv'

INTO TABLE Ecommerce\_data

CHARACTER SET latin1

FIELDS TERMINATED BY ',' ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

select \* from Lt\_lg\_of\_states;

select \* from Ecommerce\_data;

UPDATE Ecommerce\_data

SET order\_date = CASE

WHEN order\_date LIKE '\_\_-\_\_-\_\_\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(order\_date, "%d-%m-%Y"), "%Y-%m-%d")

WHEN order\_date LIKE '\_\_\_\_-\_\_-\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(order\_date, "%Y-%m-%d"), "%Y-%m-%d")

END; ///-- changed date format for order date.

alter table Ecommerce\_data

add temp\_date varchar(30); ///added temporary column to alter ship date

update Ecommerce\_data

set temp\_date = ship\_date; /// took dates to temp\_date column

update Ecommerce\_data

set temp\_date = replace(temp\_date,"/","-"); /// replaced de-limiter

UPDATE Ecommerce\_data

SET ship\_date = CASE

WHEN temp\_date LIKE '\_-\_-\_\_\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(temp\_date, "%d-%m-%Y"), "%Y-%m-%d")

WHEN temp\_date LIKE '\_\_-\_\_-\_\_\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(temp\_date, "%d-%m-%Y"), "%Y-%m-%d")

WHEN temp\_date LIKE '\_-\_\_-\_\_\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(temp\_date, "%d-%m-%Y"), "%Y-%m-%d")

WHEN temp\_date LIKE '\_\_-\_-\_\_\_\_' THEN DATE\_FORMAT(STR\_TO\_DATE(temp\_date, "%d-%m-%Y"), "%Y-%m-%d")

END; ///as ship\_date was in many formates, changed all into yyyy-mm-dd

select \* from Ecommerce\_data

alter table Ecommerce\_data

drop temp\_date