SHREYA N RAJ(AF0333789)

SPRINT 2

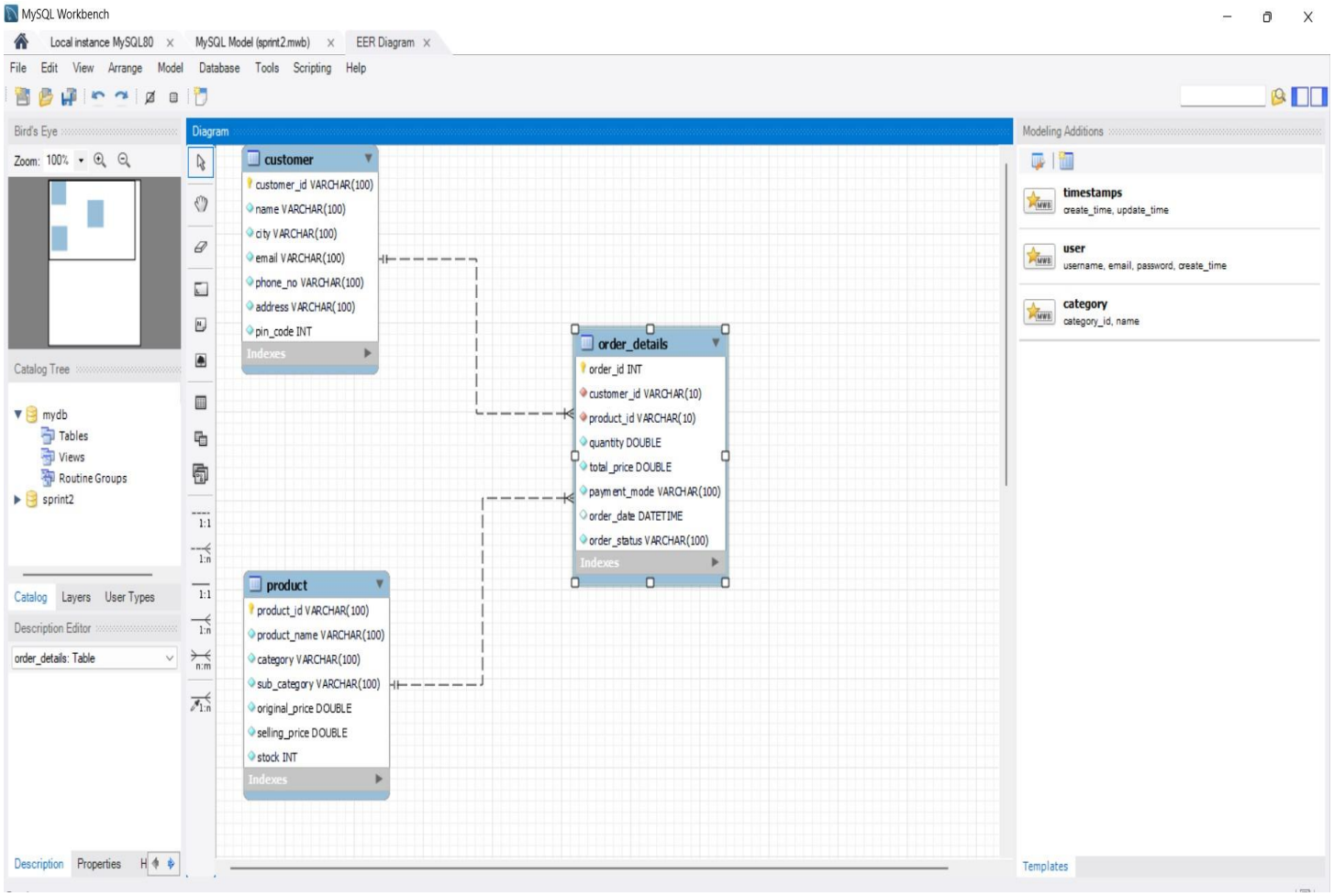
SUPERSTORE DATA ANALYSIS

Raw Dataset

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Na	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases
2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United State	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs
3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van H	Corporate	United State	Los Angeles	California	90036	West	OFF-LA-1000	Office Suppli	Labels
4	US-2015-108966	11-10-2015	18-10-2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	FUR-TA-1000	Furniture	Tables
5	US-2015-108966	11-10-2015	18-10-2015	Standard Cla	SO-20335	Sean O'Donn	Consumer	United State	Fort Lauderd	Florida	33311	South	OFF-ST-1000	Office Suppli	Storage
6	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings
7	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	OFF-AR-1000	Office Suppli	Art
8	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones
9	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	OFF-BI-1000	Office Suppli	Binders
10	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	OFF-AP-1000	Office Suppli	Appliances
11	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables
12	CA-2014-115812	09-06-2014	14-06-2014	Standard Cla	BH-11710	Brosina Hoffi	Consumer	United State	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones
13	CA-2017-114412	15-04-2017	20-04-2017	Standard Cla	AA-10480	Andrew Aller	Consumer	United State	Concord	North Caroli	28027	South	OFF-PA-1000	Office Suppli	Paper
14	CA-2016-161389	05-12-2016	10-12-2016	Standard Cla	IM-15070	Irene Maddo	Consumer	United State	Seattle	Washington	98103	West	OFF-BI-1000	Office Suppli	Binders
15	US-2015-118983	22-11-2015	26-11-2015	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Suppli	Appliances
16	US-2015-118983	22-11-2015	26-11-2015	Standard Cla	HP-14815	Harold Pawla	Home Office	United State	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Suppli	Binders
17	CA-2014-105893	11-11-2014	18-11-2014	Standard Cla	PK-19075	Pete Kriz	Consumer	United State	Madison	Wisconsin	53711	Central	OFF-ST-1000	Office Suppli	Storage
18	CA-2014-167164	13-05-2014	15-05-2014	Second Class	AG-10270	Alejandro Gr	Consumer	United State	West Jordan	Utah	84084	West	OFF-ST-1000	Office Suppli	Storage
19	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-AR-1000	Office Suppli	Art
20	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	TEC-PH-1000	Technology	Phones
21	CA-2014-143336	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United State	San Francisc	California	94109	West	OFF-BI-1000	Office Suppli	Binders
22	CA-2016-137330	09-12-2016	13-12-2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AR-1000	Office Suppli	Art
23	CA-2016-137330	09-12-2016	13-12-2016	Standard Cla	KB-16585	Ken Black	Corporate	United State	Fremont	Nebraska	68025	Central	OFF-AP-1000	Office Suppli	Appliances
24	US-2017-156909	16-07-2017	18-07-2017	Second Class	SF-20065	Sandra Flana	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-CH-1000	Furniture	Chairs
25	CA-2015-106320	25-09-2015	30-09-2015	Standard Cla	EB-13870	Emily Burns	Consumer	United State	Orem	Utah	84057	West	FUR-TA-1000	Furniture	Tables
26	CA-2016-121755	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffman	Consumer	United State	Los Angeles	California	90049	West	OFF-BI-1000	Office Suppli	Binders
27	CA-2016-121755	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffman	Consumer	United State	Los Angeles	California	90049	West	TEC-AC-1000	Technology	Accessories
28	US-2015-150630	17-09-2015	21-09-2015	Standard Cla	TB-21520	Tracy Blumst	Consumer	United State	Philadelphia	Pennsylvania	19140	East	FUR-BO-1000	Furniture	Bookcases

Orders | Returns | People

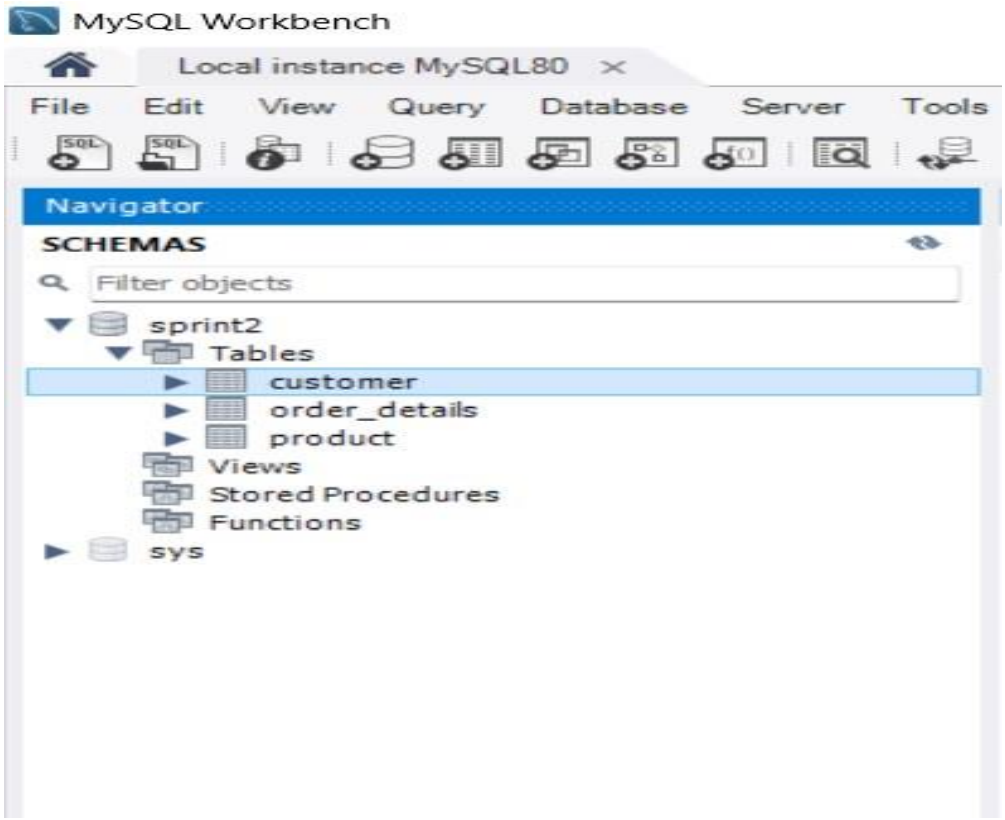
EEE Diagram

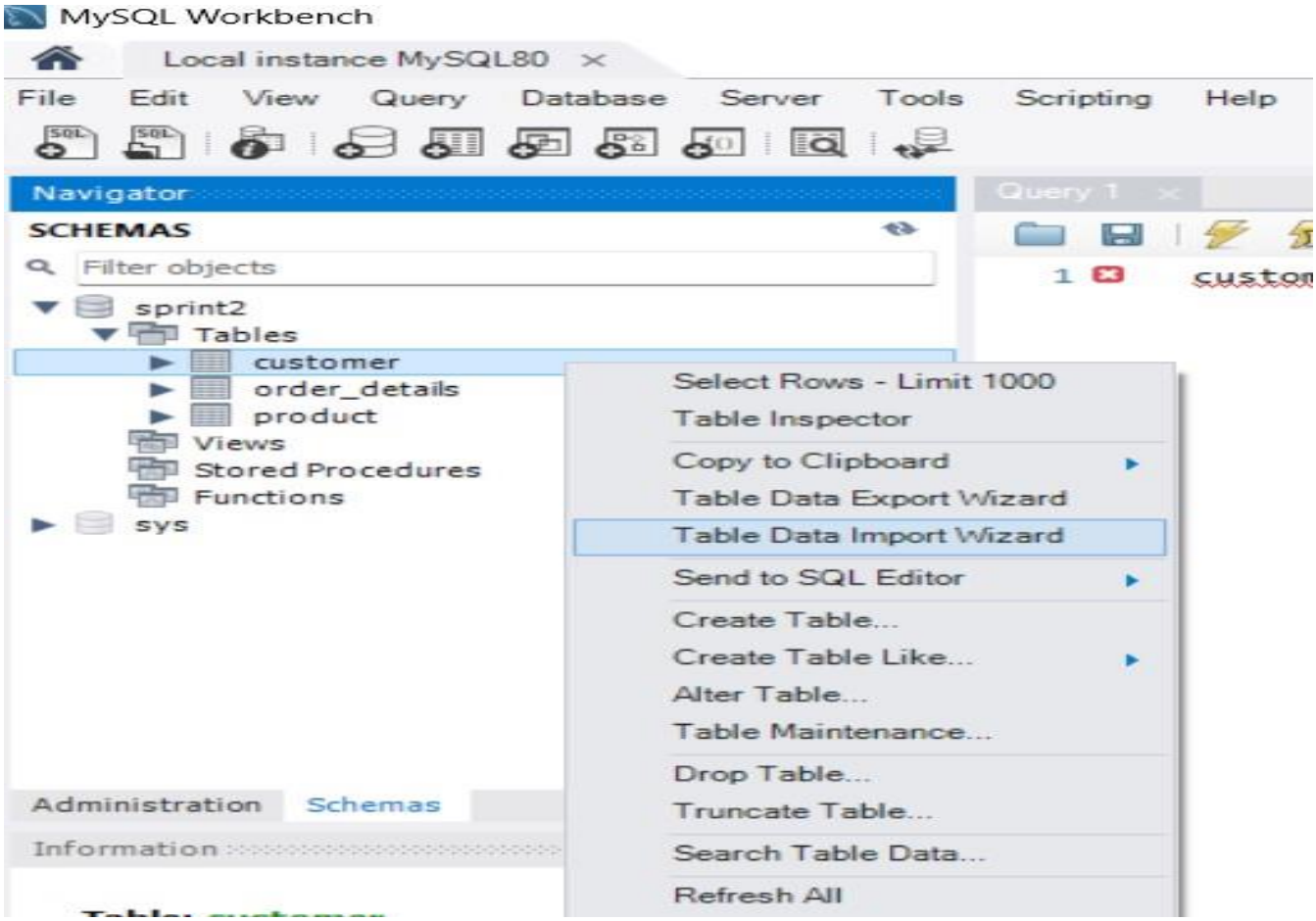


SHREYA N RAJ(AF0333789)

SPRINT 2

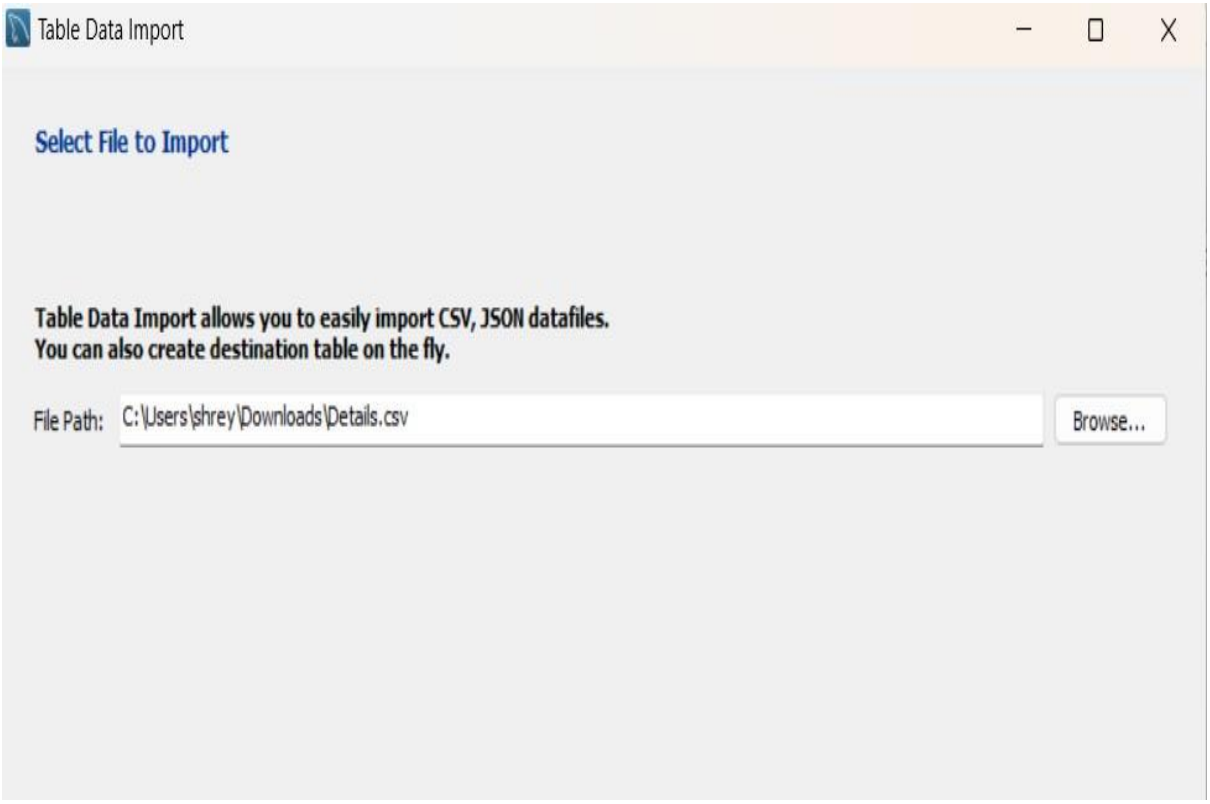
MYSQL WORKBENCH

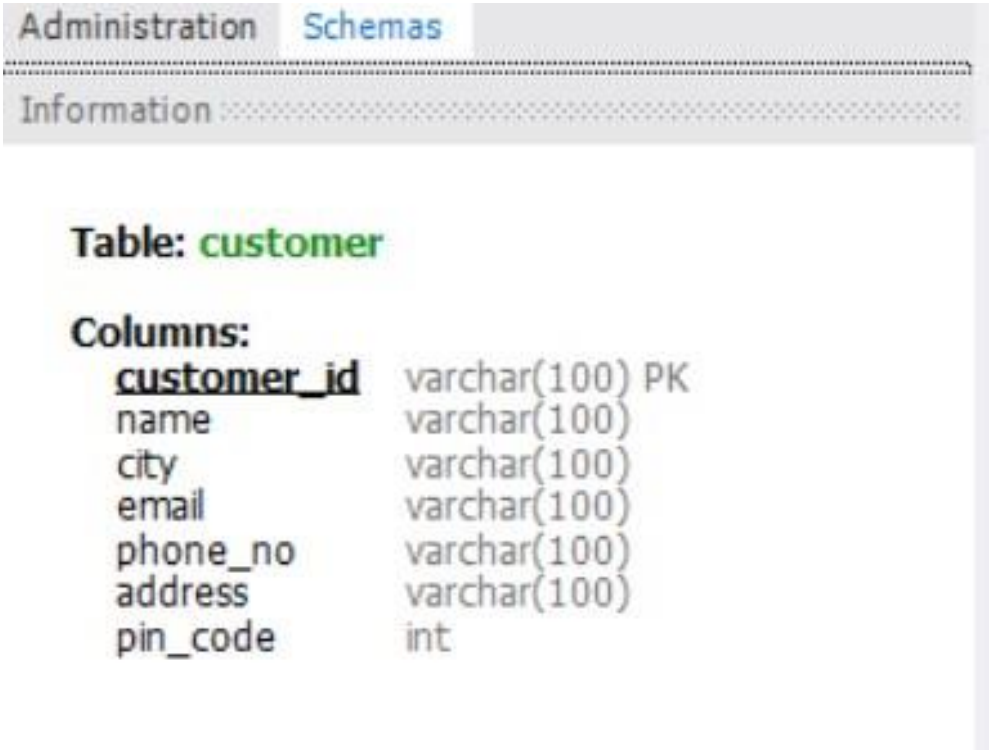




SHREYA N RAJ(AF0333789)

SPRINT 2

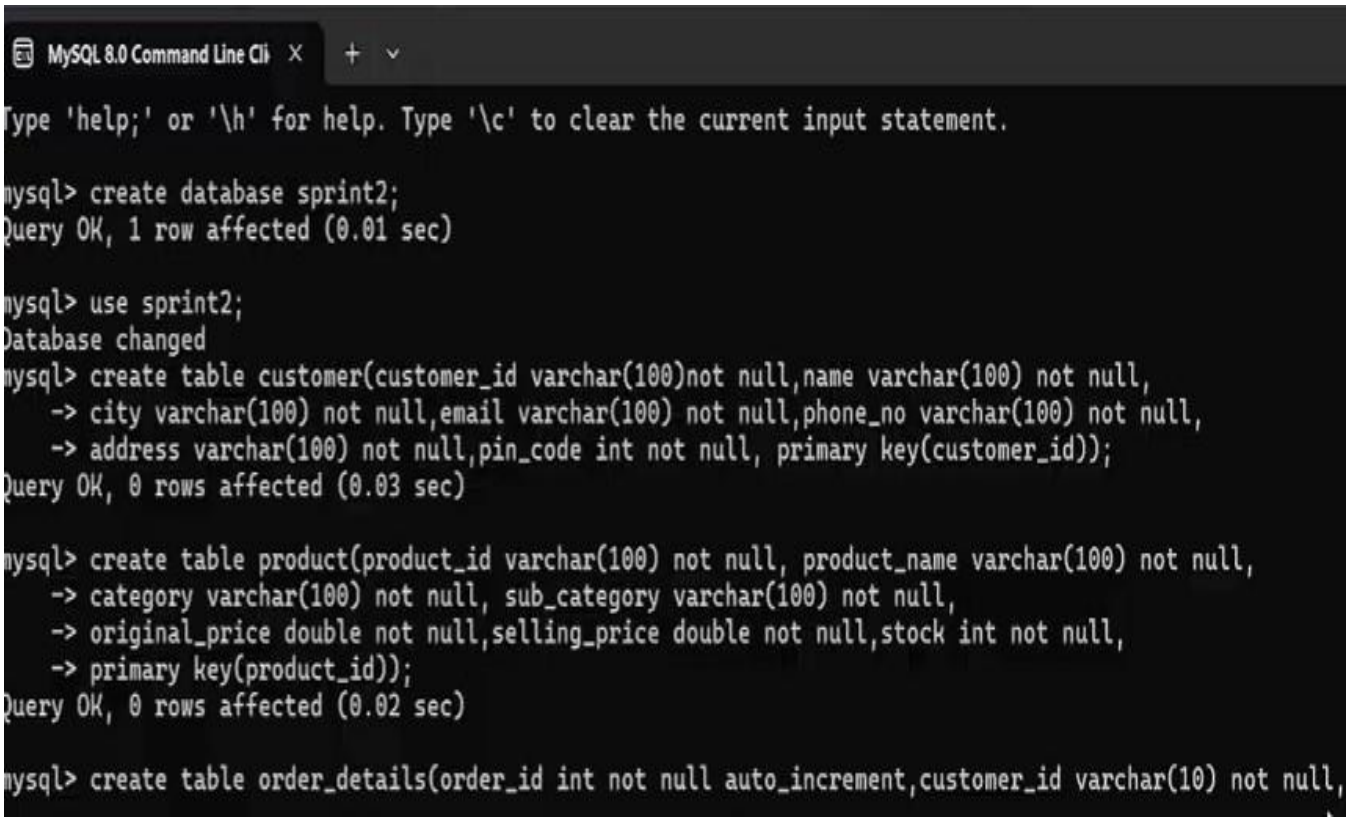




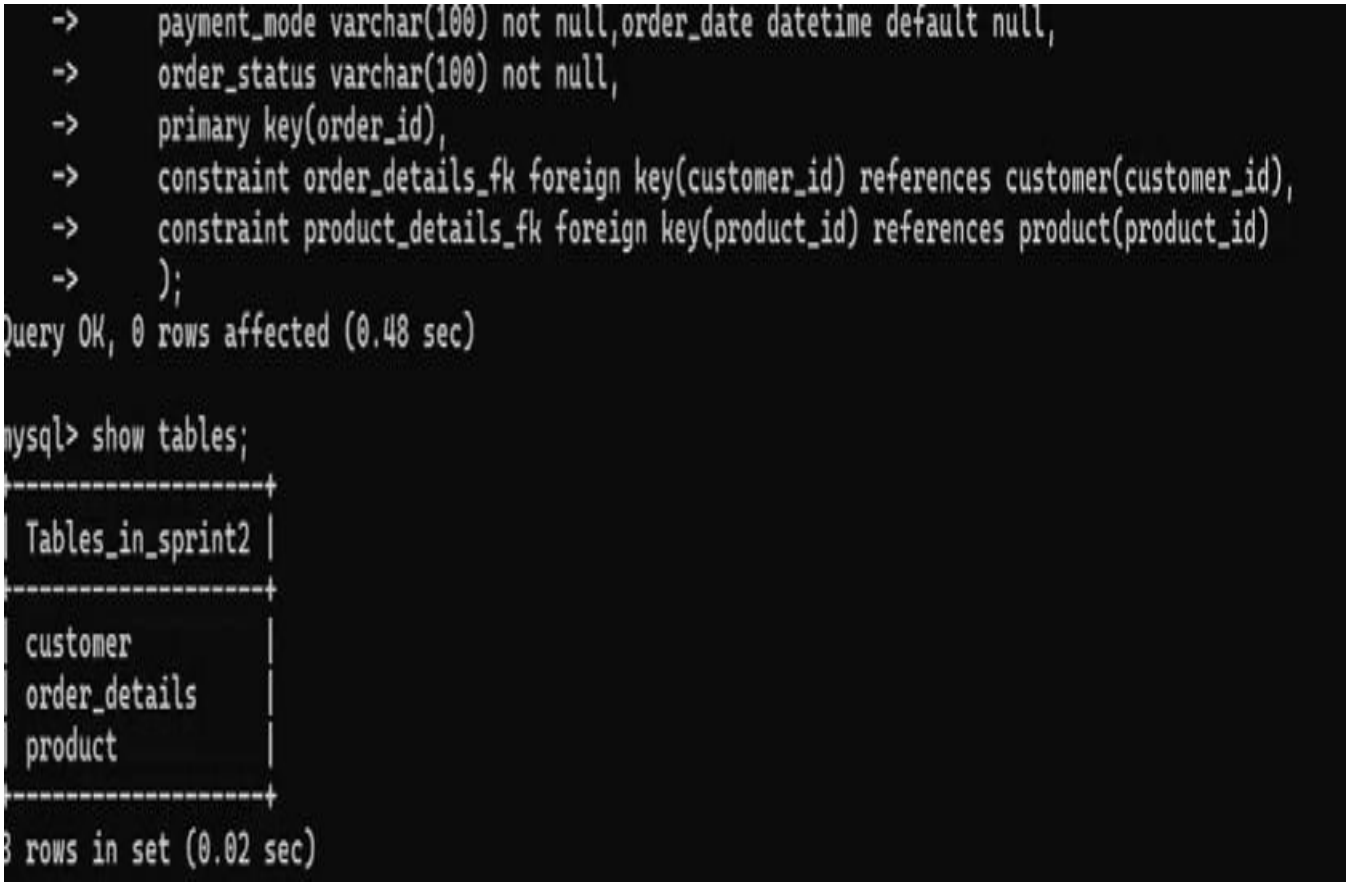
SHREYA N RAJ(AF0333789)

SPRINT 2

SQL QUESRIES



```
MySQL 8.0 Command Line Cli

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create database sprint2;
Query OK, 1 row affected (0.01 sec)

mysql> use sprint2;
Database changed
mysql> create table customer(customer_id varchar(100)not null,name varchar(100) not null,
    -> city varchar(100) not null,email varchar(100) not null,phone_no varchar(100) not null,
    -> address varchar(100) not null,pin_code int not null, primary key(customer_id));
Query OK, 0 rows affected (0.03 sec)

mysql> create table product(product_id varchar(100) not null, product_name varchar(100) not null,
    -> category varchar(100) not null, sub_category varchar(100) not null,
    -> original_price double not null,selling_price double not null,stock int not null,
    -> primary key(product_id));
Query OK, 0 rows affected (0.02 sec)

mysql> create table order_details(order_id int not null auto_increment,customer_id varchar(10) not null,
```



```
    ->     payment_mode varchar(100) not null,order_date datetime default null,
    ->     order_status varchar(100) not null,
    ->     primary key(order_id),
    ->     constraint order_details_fk foreign key(customer_id) references customer(customer_id),
    ->     constraint product_details_fk foreign key(product_id) references product(product_id)
    ->     );
Query OK, 0 rows affected (0.48 sec)

mysql> show tables;
+------------------+
| Tables_in_sprint2 |
+------------------+
| customer         |
| order_details    |
| product          |
+------------------+
3 rows in set (0.02 sec)
```

SHREYA N RAJ(AF0333789)

SPRINT 2

```
mysql> create table product(product_id varchar(100) not null, product_name varchar(100) not null,
    -> category varchar(100) not null, sub_category varchar(100) not null,
    -> original_price double not null,selling_price double not null,stock int not null,
    -> primary key(product_id));
Query OK, 0 rows affected (0.02 sec)

mysql> create table order_details(order_id int not null auto_increment,customer_id varchar(10) not null,
    -> product_id varchar(100) not null,quantity double not null,total_price not null,
    -> payment_mode varchar(100) not null,order_date datetime default null,
    -> order_status varchar(100) not null,
    -> primary key(order_id),
    -> constraint 'order_details_fk' foreign key(customer_id) references customer(customer_id),
    -> constraint 'product_details_fk' foreign key(product_id) references product(product_id)
    -> );
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL
t syntax to use near 'not null,
payment_mode varchar(100) not null,order_date datetime default null,
o' at line 2
```

```
mysql> DESCRIBE customer;
+-------------+--------------+------+-----+---------+-------+
| Field       | Type         | Null | Key | Default | Extra |
+-------------+--------------+------+-----+---------+-------+
| customer_id | varchar(100) | NO   | PRI | NULL    |       |
| name        | varchar(100) | NO   |     | NULL    |       |
| city        | varchar(100) | NO   |     | NULL    |       |
| email       | varchar(100) | NO   |     | NULL    |       |
| phone_no    | varchar(100) | NO   |     | NULL    |       |
| address     | varchar(100) | NO   |     | NULL    |       |
| pin_code    | int          | NO   |     | NULL    |       |
+-------------+--------------+------+-----+---------+-------+
7 rows in set (0.02 sec)

mysql> DESCRIBE product;
+----------------+--------------+------+-----+---------+-------+
| Field          | Type         | Null | Key | Default | Extra |
+----------------+--------------+------+-----+---------+-------+
| product_id     | varchar(100) | NO   | PRI | NULL    |       |
| product_name   | varchar(100) | NO   |     | NULL    |       |
| category       | varchar(100) | NO   |     | NULL    |       |
| sub_category   | varchar(100) | NO   |     | NULL    |       |
| original_price | double       | NO   |     | NULL    |       |
| selling_price  | double       | NO   |     | NULL    |       |
| stock          | int          | NO   |     | NULL    |       |
+----------------+--------------+------+-----+---------+-------+
7 rows in set (0.00 sec)
```

SHREYA N RAJ(AF0333789)

SPRINT 2

Jupyter Notebook Data Analysis

jupyter sprint2 Last Checkpoint: Last Thursday at 7:13 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [1]:
```
pip install mysql-connector-python
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting mysql-connector-python
  Downloading mysql_connector_python-8.4.0-cp39-cp39-win_amd64.whl (14.4 MB)
Installing collected packages: mysql-connector-python
Successfully installed mysql-connector-python-8.4.0
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
import mysql.connector
```

In [3]:
```
con = mysql.connector.connect(
user='root',
password='Shreya@452',
host='localhost',
database='sprint2'
)
```

In [4]:
```
cursor = con.cursor()
```

In [5]:
```
cursor.execute('SELECT * FROM customer')
```

In [6]:
```
import pandas as pd
```

In [7]:
```
customer_data = pd.DataFrame(cursor.fetchall(), columns=[desc[0] for desc in
cursor.description])
```

In [8]:
```
customer_data
```

Out[8]:

customer_id	name	city	email	phone_no	address	pin_code

In [9]:
```
cursor.execute('SELECT * FROM product')
product_data = pd.DataFrame(cursor.fetchall(), columns=[desc[0] for desc in
cursor.description])
product_data
```

Out[9]:

product_id	product_name	category	sub_category	original_price	selling_price	stock

In [10]:
```
cursor.execute('SELECT * FROM order_details')
order_data = pd.DataFrame(cursor.fetchall(), columns=[desc[0] for desc in
cursor.description])
order_data
```

Out[10]:

order_id	customer_id	product_id	quantity	total_price	payment_mode	order_date	order_status

SHREYA N RAJ(AF0333789)

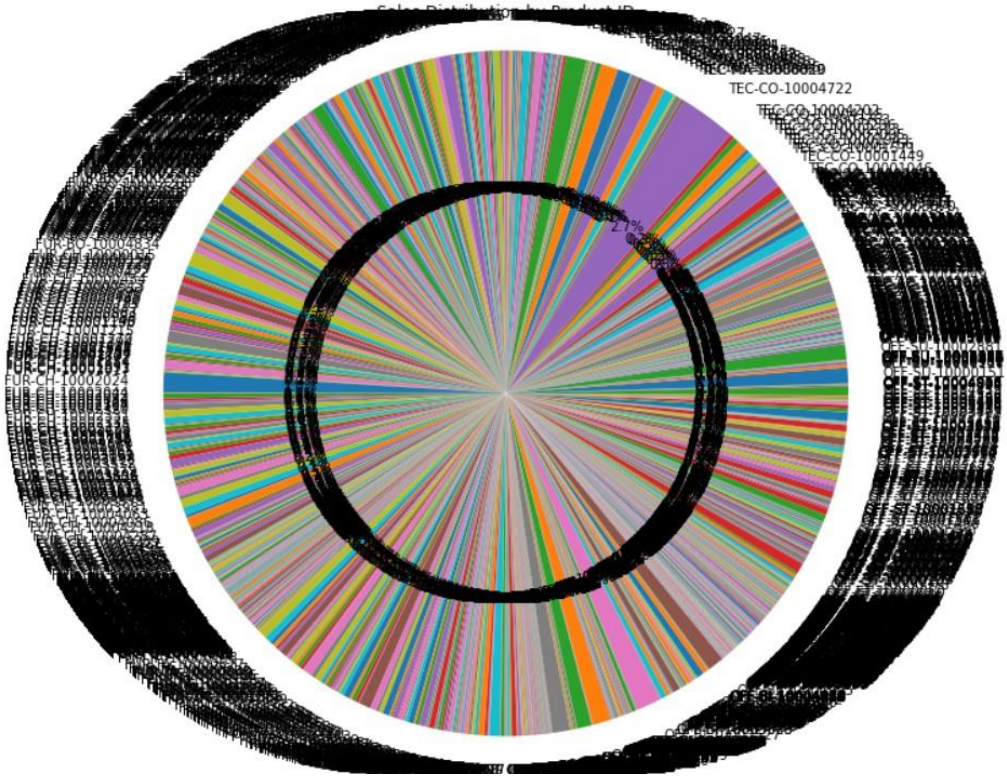
SPRINT 2

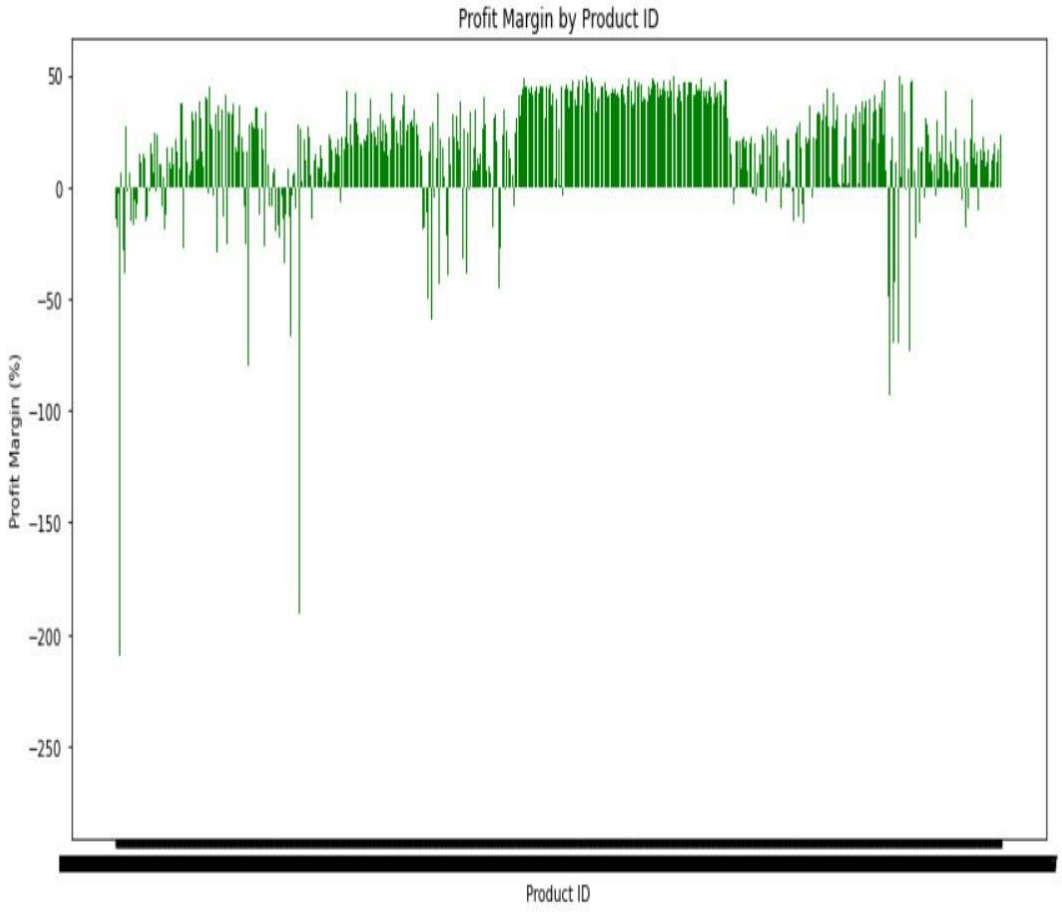
In [3]:
```
# Importing necessary libraries
import pandas as pd
import matplotlib.pyplot as plt

# Loading the data
file_path = (r'C:\Users\shrey\Downloads\PYTHON ANUDIP\Details.xlsx')
data = pd.read_excel(file_path)

# Displaying the first few rows of the data
print("First few rows of the dataset:")
print(data.head())

# Summary Statistics
print("\nSummary Statistics:")
print(data.describe())

# Data Manipulation: Aggregating total sales and profit by Product ID
product_summary = data.groupby('Product ID').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()

# Bar Graph - Total Sales and Profit by Product ID
plt.figure(figsize=(14, 7))
plt.bar(product_summary['Product ID'], product_summary['Sales'], color='blue', label='Sales')
plt.bar(product_summary['Product ID'], product_summary['Profit'], color='orange', label='Profit', alpha=0.7)
plt.xlabel('Product ID')
plt.ylabel('Amount')
plt.title('Total Sales and Profit by Product ID')
plt.legend()
plt.show()

# Pie Chart - Sales Distribution by Product ID
plt.figure(figsize=(10, 10))
plt.pie(product_summary['Sales'], labels=product_summary['Product ID'], autopct='%1.1f%%', startangle=140)
plt.title('Sales Distribution by Product ID')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```

```
# Additional: Bar Graph of Profit Margin by Product ID
product_summary['Profit Margin'] = (product_summary['Profit'] / product_summary['Sales']) * 100
plt.figure(figsize=(14, 7))
plt.bar(product_summary['Product ID'], product_summary['Profit Margin'], color='green')
plt.xlabel('Product ID')
plt.ylabel('Profit Margin (%)')
plt.title('Profit Margin by Product ID')
plt.show()
```

```
First few rows of the dataset:
         Order ID       Product ID     Sales  Quantity  Discount    Profit
0  CA-2016-152156  FUR-BO-10001798  261.9600         2      0.00   41.9136
1  CA-2016-152156  FUR-CH-10000454  731.9400         3      0.00  219.5820
2  CA-2016-138688  OFF-LA-10000240   14.6200         2      0.00    6.8714
3  US-2015-108966  FUR-TA-10000577  957.5775         5      0.45 -383.0310
4  US-2015-108966  OFF-ST-10000760   22.3680         2      0.20    2.5164

Summary Statistics:
              Sales     Quantity     Discount       Profit
count   9994.000000  9994.000000  9994.000000  9994.000000
mean     229.858001     3.789574     0.156203    28.656896
std      623.245101     2.225110     0.206452   234.260108
min        0.444000     1.000000     0.000000 -6599.978000
25%       17.280000     2.000000     0.000000     1.728750
50%       54.490000     3.000000     0.200000     8.666500
75%      209.940000     5.000000     0.200000    29.364000
max    22638.480000    14.000000     0.800000  8399.976000
```

```
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```





In [6]:
```
# Importing necessary libraries
import pandas as pd
import matplotlib.pyplot as plt

# Loading the data
file_path = (r'C:\Users\shrey\Downloads\PYTHON ANUDIP\Orders.xlsx')
data = pd.read_excel(file_path)

# Displaying the first few rows of the data
print("First few rows of the dataset:")
print(data.head())

# Summary Statistics
print("\nSummary Statistics:")
print(data.describe())
```

```
First few rows of the dataset:
   Row ID        Order ID Order Date  Ship Date       Ship Mode Customer ID  \
0       1  CA-2016-152156 2016-11-08 2016-11-11    Second Class    CG-12520
1       2  CA-2016-152156 2016-11-08 2016-11-11    Second Class    CG-12520
2       3  CA-2016-138688 2016-06-12 2016-06-16    Second Class    DV-13045
3       4  US-2015-108966 2015-10-11 2015-10-18  Standard Class    SO-20335
4       5  US-2015-108966 2015-10-11 2015-10-18  Standard Class    SO-20335

     Customer Name    Segment        Country             City       State  \
0      Claire Gute   Consumer  United States        Henderson    Kentucky
1      Claire Gute   Consumer  United States        Henderson    Kentucky
2  Darrin Van Huff  Corporate  United States      Los Angeles  California
3   Sean O'Donnell   Consumer  United States  Fort Lauderdale     Florida
4   Sean O'Donnell   Consumer  United States  Fort Lauderdale     Florida

   Postal Code Region
0        42420  South
1        42420  South
2        90036   West
3        33311  South
4        33311  South

Summary Statistics:
            Row ID   Postal Code
count  9994.000000   9994.000000
mean   4997.500000  55190.379428
std    2885.163629  32063.693350
min       1.000000   1040.000000
25%    2499.250000  23223.000000
50%    4997.500000  56430.500000
75%    7495.750000  90008.000000
max    9994.000000  99301.000000
```

SHREYA N RAJ(AF0333789)

SPRINT 2

In [8]:
```
# Additional: Bar Graph of Profit Margin by Product ID
product_summary['Profit Margin'] = (product_summary['Profit'] / product_summary['Sales']) * 100
plt.figure(figsize=(14, 7))
plt.bar(product_summary['Product ID'], product_summary['Profit Margin'], color='green')
plt.xlabel('Product ID')
plt.ylabel('Profit Margin (%)')
plt.title('Profit Margin by Product ID')
plt.show()
```