**SQL Assignment Project**

**Superstores Database**

**Task 1:- Understanding the Data**

1. **Describe the data in hand in your own words.**

* **This database contains Sales details of transaction of a superstore.**
* **The structure has 5 tables**

1. **Cust\_dimen (containing details about customer and their respective locations)**

**Customer\_Name (TEXT):** Name of the customer

**Province (TEXT):** Province of the customer

**Region (TEXT):** Region of the customer

**Customer\_Segment (TEXT**): Segment of the customer

**Cust\_id (TEXT):** Unique Customer ID

1. **Prod\_dimen (containing product category and their subcategories)**

**Product\_Category (TEXT):** Product Category

**Product\_Sub\_Category (TEXT):** Product Sub Category

**Prod\_id (TEXT):** Unique Product ID

1. **Orders\_dimen (with order no, date, and priority)**

**Order\_ID (INT):** Order ID

**Order\_Date (TEXT):** Order Date

**Order\_Priority (TEXT):** Priority of the Order

**Ord\_id (TEXT):** Unique Order ID

1. **Shipping\_dimen (with ship date, order and shipping mode)**

**Order\_ID (INT):** Order ID

**Ship\_Mode (TEXT):** Shipping Mode

**Ship\_Date (TEXT):** Shipping Date

**Ship\_id (TEXT):** Unique Shipment ID

1. **market\_fact (orderwise customerwise marketwise orderquantity, sales value, discount profit and shipping cost details).**

**Ord\_id (TEXT):** Order ID

**Prod\_id (TEXT):** Prod ID

**Ship\_id (TEXT):** Shipment ID

**Cust\_id (TEXT):** Customer ID

**Sales (DOUBLE):** Sales from the Item sold

**Discount (DOUBLE):** Discount on the Item sold

**Order\_Quantity (INT):** Order Quantity of the Item sold

**Profit (DOUBLE):** Profit from the Item sold

**Shipping\_Cost (DOUBLE):** Shipping Cost of the Item sold

**Product\_Base\_Margin (DOUBLE):** Product Base Margin on the Item sold

1. **Identify and list the Primary Keys and Foreign Keys for this dataset provided to**

**you(In case you don’t find either primary or foreign key, then specially mention**

**this in your answer)**

1. **Cust\_dimen**

**Primary Key:** Cust\_id

**Foreign Key:** NA

1. **Prod\_dimen**

**Primary Key:** Prod\_id

**Foreign Key:** NA

1. **Orders\_dimen**

**Primary Key:** Ord\_id

**Foreign Key:** NA

1. **Shipping\_dimen**

**Primary Key:** Ship\_id

**Foreign Key:** NA

1. **Market\_fact**

**Primary Key:** NA

**Foreign Key:** Ord\_id, Prod\_id, Ship\_id, Cust\_id

**Task 2: Basics & Advance Analysis (Queries)**

1. **Write a query to display the Customer\_Name and Customer Segment using alias name “Customer Name", "Customer Segment" from table Cust\_dimen.**

* select customer\_name as 'Customer Name' , customer\_segment as 'Customer Segment' from cust\_dimen;

1. **Write a query to find all the details of the customer from the table cust\_dimen order by desc.**

* select \* from cust\_dimen order by cust\_id desc;

1. **Write a query to get the Order ID, Order date from table orders\_dimen where ‘Order Priority’ is high.**

* select order\_id, order\_date from orders\_dimen where Order\_Priority='High';

1. **Find the total and the average sales (display total\_sales and avg\_sales)**

* select avg(sales) as 'Avg\_sales' , sum(sales) as 'total\_sales' from market\_fact;

1. **Write a query to get the maximum and minimum sales from maket\_fact table.**

* select max(sales), min(sales) from market\_fact;

1. **Display the number of customers in each region in decreasing order of no\_of\_customers. The result should contain columns Region, no\_of\_customers.**

* select region, count(\*) as 'No\_of\_customers' from cust\_dimen group by region order by count(\*) desc;

1. **Find the region having maximum customers (display the region name and max(no\_of\_customers)**

* select region, count(\*) as 'Max(No\_of\_customers' from cust\_dimen group by region order by count(\*) desc limit 1;

1. **Find all the customers from Atlantic region who have ever purchased ‘TABLES’ and the number of tables purchased (display the customer name, no\_of\_tables purchased)**

* select cust\_dimen.Customer\_Name, count(market\_fact.Order\_Quantity) as no\_of\_tables\_purchased from market\_fact inner join cust\_dimen on market\_fact.Cust\_id=cust\_dimen.Cust\_id inner join prod\_dimen on prod\_dimen.Prod\_id=market\_fact.Prod\_id where cust\_dimen.region='Atlantic' and prod\_dimen.product\_sub\_category='TABLES' group by Customer\_Name order by count(market\_fact.Order\_Quantity) desc;

1. **Find all the customers from Ontario province who own Small Business. (display the customer name, no of small business owners)**

* select customer\_name, count(Customer\_Segment) as 'No Of Small Business Owners' from cust\_dimen where customer\_segment='SMALL BUSINESS' and province='Ontario' group by Customer\_Name;

1. **Find the number and id of products sold in decreasing order of products sold (display product id, no\_of\_products sold)**

* select prod\_id, count(Order\_Quantity) from market\_fact group by Prod\_id order by count(Order\_Quantity) desc;

1. **Display product Id and product sub category whose produt category belongs to Furniture and Technlogy. The result should contain columns product id, product sub category**.

* select prod\_id, product\_sub\_category from prod\_dimen where Product\_Category in ('FURNITURE' , 'TECHNOLOGY');

1. **Display the product categories in descending order of profits (display the product category wise profits i.e. product\_category, profits)?**

* select prod\_dimen.Product\_Category , market\_fact.Profit from prod\_dimen inner join market\_fact on prod\_dimen.Prod\_id=market\_fact.Prod\_id order by profit desc;

1. **Display the product category, product sub-category and the profit within each subcategory in three columns.**

* select prod\_dimen.Product\_Category, prod\_dimen.Product\_Sub\_Category, sum(market\_fact.Profit) from prod\_dimen inner join market\_fact on prod\_dimen.Prod\_id=market\_fact.Prod\_id group by Product\_Sub\_Category order by sum(market\_fact.Profit) desc;

1. **Display the order date, order quantity and the sales for the order.**

* select orders\_dimen.Order\_Date, market\_fact.Order\_Quantity, market\_fact.Sales from orders\_dimen inner join market\_fact on orders\_dimen.Ord\_id=market\_fact.Ord\_id;

1. **Display the names of the customers whose name contains the**

**i) Second letter as ‘R’**

**ii) Fourth letter as ‘D’**

1. Second letter as ‘R’

* select customer\_name from cust\_dimen where Customer\_Name like '\_R%';

1. Fourth letter as ‘D’

* select customer\_name from cust\_dimen where Customer\_Name like '\_\_\_D%';

1. **Write a SQL query to to make a list with Cust\_Id, Sales, Customer Name and their region where sales are between 1000 and 5000.**

* select market\_fact.Cust\_id, market\_fact.Sales, cust\_dimen.Customer\_Name, cust\_dimen.Region from market\_fact inner join cust\_dimen on market\_fact.Cust\_id=cust\_dimen.Cust\_id where sales between '1000' and '5000';

1. **Write a SQL query to find the 3rd highest sales.**

* select sales from market\_fact order by sales desc limit 2,1;

1. **Where is the least profitable product subcategory shipped the most? For the least profitable product sub-category, display the region-wise no\_of\_shipments and the profit made in each region in decreasing order of profits (i.e. region, no\_of\_shipments, profit\_in\_each\_region)**

**→ Note: You can hardcode the name of the least profitable product subcategory**

* select cust\_dimen.Region as "Region", count(market\_fact.ship\_id) as "no\_of\_shipment",round(sum(market\_fact.profit),2) as "profit in each region" from market\_fact join cust\_dimen on market\_fact.cust\_id=cust\_dimen.cust\_id join prod\_dimen on market\_fact.prod\_id=prod\_dimen.prod\_id where product\_sub\_category=(select prod\_dimen.product\_sub\_category from market\_fact join prod\_dimen on market\_fact.prod\_id=prod\_dimen.prod\_id group by product\_sub\_category order by sum(market\_fact.profit) limit 1) group by cust\_dimen.Region order by sum(market\_fact.profit);