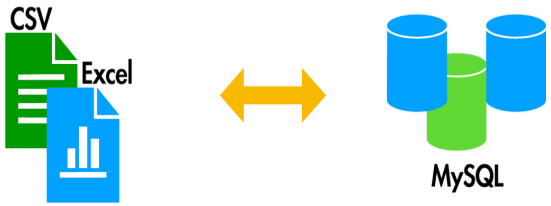
❑ SUPERSTORE : EXCEL AND SQL BASED PROJECT

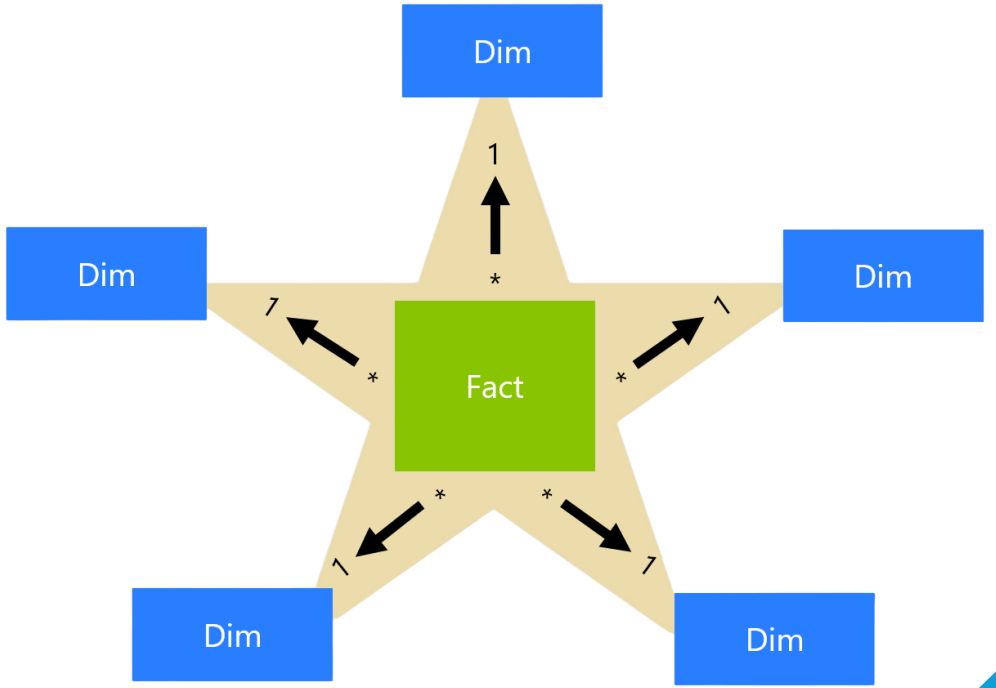
❑ BRIDGING EXCEL AND SQL : A DATA MODELING AND ANALYSIS ENDEAVOR



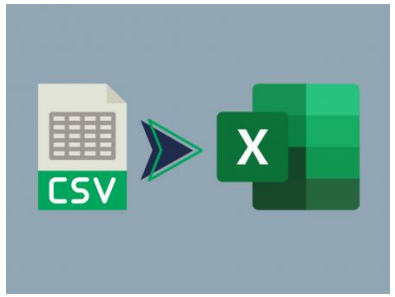
❖ SQL Star Schema Task

1. Take the Superstore.csv Dataset and Convert that into Star Schema Designing using Excel
2. Customer Table: Customer Name & Customer ID
3. Location Table: City, Country, Region, Postal Code, State, Location ID
4. Manufacturer Table: Manufacturer Name, Manufacturer ID
5. Product Table: Product Name, Product ID
6. Segment Table: Segment Name, Segment ID
7. Fact Table: Order ID, Order Date, Ship Date, Discount, Profit, Quantity, Sales, Customer ID, Location ID, Manufacturer ID, Product ID, Segment ID

1.Take the Superstore.csv Data set and Convert that into 🌟Star Schema Designing using Excel



IMPORING CSV DATA INTO EXCEL



superstore

Search for tools, help, and more (Alt + Q)

Buy Microsoft 365

AM

File Home Insert Share Page Layout Formulas Data Review View Help Draw

Comments Editing Share

Aptos Narrow ... 11 B General

AC24

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Profit	Quantity	Sales
1	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	Ship Mode	State	Sub-Category	Discount			
2	Furniture	Henderson	United States	Claire Gute	Bush	11-12-2020	CA-2020-105816	42420	Bush Somerset Collection Bookcase	South	Consumer	11-11-2020	Second Class	Kentucky	Bookcases	0	41.9136	2	261.96
3	Furniture	Henderson	United States	Claire Gute	Hon	11-12-2020	CA-2020-105816	42420	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	South	Consumer	11-11-2020	Second Class	Kentucky	Chairs	0	219.582	3	731.94
4	Office Supplies	Los Angeles	United States	Darrin Van Huff	Universal	18-04-2019	CA-2019-115742	90036	Self-Adhesive Address Labels for Typewriters by Universal	West	Corporate	16-06-2020	Second Class	California	Labels	0	6.8714	2	14.62
5	Furniture	Fort Lauderdale	United States	Sean O'Donnell	Bretford	18-04-2019	CA-2019-115742	33311	Bretford CR4500 Series Slim Rectangular Table	South	Consumer	18-10-2019	Standard Class	Florida	Tables	0.45	-383.031	5	957.5775
6	Office Supplies	Fort Lauderdale	United States	Sean O'Donnell	Eldon	18-04-2019	CA-2019-115742	33311	Eldon Fold 'N Roll Cart System	South	Consumer	18-10-2019	Standard Class	Florida	Storage	0.2	2.5164	2	22.368
7	Furniture	Los Angeles	United States	Brosina Hoffman	Eldon	18-04-2019	CA-2019-115742	90032	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	West	Consumer	14-06-2018	Standard Class	California	Furnishings	0	14.1694	7	48.86
8	Office Supplies	Los Angeles	United States	Brosina Hoffman	Newell	20-06-2020	CA-2020-169194	90032	Newell 322	West	Consumer	14-06-2018	Standard Class	California	Art	0	1.9656	4	7.28
9	Technology	Los Angeles	United States	Brosina Hoffman	Mitel	20-06-2020	CA-2020-169194	90032	Mitel 5320 IP Phone VoIP phone	West	Consumer	14-06-2018	Standard Class	California	Phones	0.2	90.7152	6	907.152
10	Office Supplies	Los Angeles	United States	Brosina Hoffman	DXL	20-10-2018	CA-2018-146703	90032	DXL Angle-View Binders with Locking Rings by Samsill	West	Consumer	14-06-2018	Standard Class	California	Binders	0.2	5.7825	3	18.504
11	Office Supplies	Los Angeles	United States	Brosina Hoffman	Belkin	11-03-2020	CA-2020-118255	90032	Belkin F5C206VTEL 6 Outlet Surge	West	Consumer	14-06-2018	Standard Class	California	Appliances	0	34.47	5	114.9
12	Furniture	Los Angeles	United States	Brosina Hoffman	Chromcraft	11-03-2020	CA-2020-118255	90032	Chromcraft Rectangular Conference Tables	West	Consumer	14-06-2018	Standard Class	California	Tables	0.2	85.3092	9	1706.184
13	Technology	Los Angeles	United States	Brosina Hoffman	Other	19-09-2021	CA-2021-139619	90032	Konftel 250 Conference phone - Charcoal black	West	Consumer	14-06-2018	Standard Class	California	Phones	0.2	68.3568	4	911.424
14	Office Supplies	Concord	United States	Andrew Allen	Xerox	17-07-2020	CA-2020-101343	28027	Xerox 1967	South	Consumer	20-04-2021	Standard Class	North Carolina	Paper	0.2	5.4432	3	15.552
15	Office Supplies	Seattle	United States	Irene Maddox	Fellowes	10-09-2021	CA-2021-120999	98103	Fellowes PB200 Plastic Comb Binding Machine	West	Consumer	10-12-2020	Standard Class	Washington	Binders	0.2	132.5922	3	407.976
16	Office Supplies	Fort Worth	United States	Harold Pawlan	Holmes	27-12-2019	CA-2019-117415	76106	Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter	Central	Home Office	26-11-2019	Standard Class	Texas	Appliances	0.8	-123.858	5	68.81
17	Office Supplies	Fort Worth	United States	Harold Pawlan	Storex	27-12-2019	CA-2019-117415	76106	Storex DuraTech Recycled Plastic Frosted Binders	Central	Home Office	26-11-2019	Standard Class	Texas	Binders	0.8	-3.816	3	2.544
18	Office Supplies	Madison	United States	Pete Kriz	Other	27-12-2019	CA-2019-117415	53711	Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D	Central	Consumer	18-11-2018	Standard Class	Wisconsin	Storage	0	13.3176	6	665.88
19	Office Supplies	West Jordan	United States	Alejandro Grove	Fellowes	27-12-2019	CA-2019-117415	84084	Fellowes Super Stor/Drawer	West	Consumer	15-05-2018	Second Class	Utah	Storage	0	9.99	2	55.5
20	Office Supplies	San Francisco	United States	Zuschuss Donatelli	Newell	08-12-2020	CA-2020-117590	94109	Newell 341	West	Consumer	01-09-2018	Second Class	California	Art	0	2.4824	2	8.56
21	Technology	San Francisco	United States	Zuschuss Donatelli	Cisco	08-12-2020	CA-2020-117590	94109	Cisco SPA 501G IP Phone	West	Consumer	01-09-2018	Second Class	California	Phones	0.2	16.011	3	213.48
22	Office Supplies	San Francisco	United States	Zuschuss Donatelli	Wilson Jones	19-10-2021	CA-2021-107727	94109	Wilson Jones Hanging View Binder, White, 1"	West	Consumer	01-09-2018	Second Class	California	Binders	0.2	7.384	4	22.72
23	Office Supplies	Fremont	United States	Ken Black	Newell	17-09-2019	US-2019-150630	68025	Newell 318	Central	Corporate	13-12-2020	Standard Class	Nebraska	Art	0	5.0596	7	19.46
24	Office Supplies	Fremont	United States	Ken Black	Acco	17-09-2019	US-2019-150630	68025	Acco Six-Outlet Power Strip, 4' Cord Length	Central	Corporate	13-12-2020	Standard Class	Nebraska	Appliances	0	15.6884	7	60.34
25	Furniture	Philadelphia	United States	Sandra Flanagan	Global	17-09-2019	US-2019-150630	19140	Global Deluxe Stacking Chair, Gray	East	Consumer	18-07-2021	Second Class	Pennsylvania	Chairs	0.3	-1.0196	2	71.372
26	Furniture	Orem	United States	Emily Burns	Bretford	17-09-2019	US-2019-150630	84057	Bretford CR4500 Series Slim Rectangular Table	West	Consumer	30-09-2019	Standard Class	Utah	Tables	0	240.2649	3	1044.63
27	Office Supplies	Los Angeles	United States	Eric Hoffmann	Wilson Jones	17-09-2019	US-2019-150630	90049	Wilson Jones Active Use Binders	West	Consumer	20-01-2020	Second Class	California	Binders	0.2	4.2224	2	11.648
28	Technology	Los Angeles	United States	Eric Hoffmann	Other	17-09-2019	US-2019-150630	90049	Imation 8GB Mini TravelDrive USB 2.0 Flash Drive	West	Consumer	20-01-2020	Second Class	California	Accessories	0	11.7741	3	90.57
29	Furniture	Philadelphia	United States	Tracy Blumstein	Riverside	17-09-2019	US-2019-150630	19140	Riverside Palais Royal Lawyers Bookcase, Royale Cherry Finish	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Bookcases	0.5	-1665.0522	7	3083.43
30	Office Supplies	Philadelphia	United States	Tracy Blumstein	Avery	16-01-2020	CA-2020-121755	19140	Avery Recycled Flexi-View Covers for Binding Systems	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Binders	0.7	-7.0532	2	9.618
31	Furniture	Philadelphia	United States	Tracy Blumstein	Howard Miller	16-01-2020	CA-2020-121755	19140	Howard Miller 13-3/4" Diameter Brushed Chrome Round Wall Clock	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Furnishings	0.2	15.525	3	124.2
32	Office Supplies	Philadelphia	United States	Tracy Blumstein	Poly	25-09-2019	CA-2019-106320	19140	Poly String Tie Envelopes	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Envelopes	0.2	1.1016	2	3.264
33	Office Supplies	Philadelphia	United States	Tracy Blumstein	Boston	16-07-2021	US-2021-156909	19140	BOSTON Model 1800 Electric Pencil Sharpeners, Putty/Woodgrain	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Art	0.2	9.7092	6	86.304
34	Office Supplies	Philadelphia	United States	Tracy Blumstein	Acco	09-12-2020	CA-2020-137330	19140	Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11", Executive Red	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Binders	0.7	-5.715	6	6.858
35	Office Supplies	Philadelphia	United States	Tracy Blumstein	Other	09-12-2020	CA-2020-137330	19140	Lumber Crayons	East	Consumer	21-09-2019	Standard Class	Pennsylvania	Art	0.2	3.546	2	15.76
36	Office Supplies	Houston	United States	Matt Abelman	Easy-staple pape	27-08-2018	CA-2018-143336	77095	Easy-staple paper	Central	Home Office	23-10-2021	Second Class	Texas	Paper	0.2	9.9468	3	29.472

2. Creating Customer Table : Customer Name & Unique Customer ID

❑ Creating Unique Customer ID :

Utilized customer names to generate unique customer IDs. Ensured data integrity and identification by assigning each customer a distinct identifier.

❑ Removing Duplicates :

Employed Excel's "Remove Duplicates" function to streamline data. Eliminated redundant entries and optimized dataset for analysis. Detected and eliminated 637 duplicate entries, resulting in 363 distinct Customer values.

	A	B
1	**Customer Name**	**Customer ID**
2	Claire Gute	C1
3	Darrin Van Huff	C2
4	Sean O'Donnell	C3
5	Brosina Hoffman	C4
6	Andrew Allen	C5
7	Irene Maddox	C6
8	Harold Pawlan	C7
9	Pete Kriz	C8
10	Alejandro Grove	C9
11	Zuschuss Donatelli	C10
12	Ken Black	C11
13	Sandra Flanagan	C12
14	Emily Burns	C13
15	Eric Hoffmann	C14
16	Tracy Blumstein	C15
17	Matt Abelman	C16

3. Creating Location Table : City, Region, Postal Code, State, Location ID

❑ Combining Location Attributes :

Utilized a combination of city, region, postal code, and state data. Formulated a comprehensive representation of each location by integrating multiple attributes.

❑ Generating Unique Location IDs:

Applied the combined location attributes to create unique location identifiers. Ensured each location is distinctly identified by the generated unique IDs. Detected and eliminated 778 duplicate entries, resulting in 222 distinct location values using Excel's "Remove Duplicates" function .

	A	B	C	D	E
1	**City**	**Region**	**Postal Code**	**State**	**Location ID**
2	Henderson	South	42420	Kentucky	L1
3	Los Angeles	West	90036	California	L2
4	Fort Lauderdale	South	33311	Florida	L3
5	Los Angeles	West	90032	California	L4
6	Concord	South	28027	North Carolina	L5
7	Seattle	West	98103	Washington	L6
8	Fort Worth	Central	76106	Texas	L7
9	Madison	Central	53711	Wisconsin	L8
10	West Jordan	West	84084	Utah	L9
11	San Francisco	West	94109	California	L10
12	Fremont	Central	68025	Nebraska	L11
13	Philadelphia	East	19140	Pennsylvania	L12
14	Orem	West	84057	Utah	L13
15	Los Angeles	West	90049	California	L14
16	Houston	Central	77095	Texas	L15
17	Richardson	Central	75080	Texas	L16
18	Houston	Central	77041	Texas	L17
19	Naperville	Central	60540	Illinois	L18
20	Melbourne	South	32935	Florida	L19
21	Eagan	Central	55122	Minnesota	L20
22	Westland	Central	48185	Michigan	L21
23	Dover	East	19901	Delaware	L22
24	New Albany	Central	47150	Indiana	L23

4. Creating Manufacturer Table: Manufacturer Name, Manufacturer ID

❑ Creating Unique Manufacturer ID :

Utilized Manufacturer names to generate unique Manufacturer IDs. Ensured data integrity and identification by assigning each Manufacturer a distinct identifier.

❑ Removing Duplicates :

Employed Excel's "Remove Duplicates" function to streamline data. Eliminated redundant entries and optimized dataset for analysis. Detected and eliminated 850 duplicate entries, resulting in 150 distinct Manufacturer values.

	A	B
1	**Manufacturer Name**	**Manufacturer ID**
2	Bush	M1
3	Hon	M2
4	Universal	M3
5	Bretford	M4
6	Eldon	M5
7	Newell	M6
8	Mitel	M7
9	DXL	M8
10	Belkin	M9
11	Chromcraft	M10
12	Other	M11
13	Xerox	M12
14	Fellowes	M13
15	Holmes	M14
16	Storex	M15
17	Cisco	M16
18	Wilson Jones	M17

5. Creating Product Table: Product Name, Product ID

❑ Creating Unique Product ID :

Utilized Product names to generate unique Product IDs. Ensured data integrity and identification by assigning each Product a distinct identifier.

❑ Removing Duplicates :

Employed Excel's "Remove Duplicates" function to streamline data. Eliminated redundant entries and optimized dataset for analysis.

Detected and eliminated duplicate 236 entries, resulting in 764 distinct Product values.

	A	B
1	**Product Name**	**Product ID**
2	#10- 4 1/8" x 9 1/2" Recycled Envelopes	P1
3	#10- 4 1/8" x 9 1/2" Security-Tint Envelopes	P2
4	#10 Gummed Flap White Envelopes, 100/Box	P3
5	#10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes	P4
6	1.7 Cubic Foot Compact "Cube" Office Refrigerators	P5
7	12-1/2 Diameter Round Wall Clock	P6
8	2300 Heavy-Duty Transfer File Systems by Perma	P7
9	3M Hangers With Command Adhesive	P8
10	3M Office Air Cleaner	P9
11	3-ring staple pack	P10
12	4009 Highlighters by Sanford	P11
13	50 Colored Long Pencils	P12
14	6" Cubicle Wall Clock- Black	P13
15	9-3/4 Diameter Round Wall Clock	P14
16	Aastra 57i VoIP phone	P15
17	Acco 3-Hole Punch	P16
18	Acco 7-Outlet Masterpiece Power Center_Wihtout Fax/Phone Line Protection	P17
19	Acco D-Ring Binder w/DublLock	P18
20	Acco Hanging Data Binders	P19

6. Creating Segment Table: Segment Name, Segment ID

❑ Creating Unique Segment ID :

Utilized Segment names to generate unique Segment IDs. Ensured data integrity and identification by assigning each Segment a distinct identifier.

❑ Removing Duplicates :

Employed Excel's "Remove Duplicates" function to streamline data. Eliminated redundant entries and optimized dataset for analysis.

Detected and eliminated duplicate 997 entries, resulting in 3 distinct Segment values.

	A	B
1	**Segment**	**Segment ID**
2	Consumer	S1
3	Corporate	S2
4	Home Office	S3
5		

7. Fact Table: Order ID, Order Date, Ship Date, Discount, Profit, Quantity, Sales, Customer ID, Location ID, Manufacturer ID, Product ID, Segment ID

❑ **Fact Table Creation:**

Utilized various attributes to construct a comprehensive fact table. Integrated key data points including order ID, order date, ship date, discount, profit, quantity, and sales.

❑ **Incorporated Additional Tables:**

Augmented the fact table by incorporating identifiers from related tables. Integrated customer ID from the customer table, location ID from the location table, manufacturer ID from the manufacturer table, product ID from the product table, and segment ID from the segment table.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Order ID	Order Date	Ship Date	Discoun	Profit	Quantity	Sales	Product ID	Customer ID	Manufacturer ID	Segment ID	Location ID
2	CA-2020-105816	11-12-2020	11-11-2020	0%	41.9136	2	$ 261.96	P168	C1	M1	S1	L1
3	CA-2020-105816	11-12-2020	11-11-2020	0%	219.582	3	$ 731.94	P360	C1	M2	S1	L1
4	CA-2019-115742	18-04-2019	16-06-2020	0%	6.8714	2	$ 14.62	P600	C2	M3	S2	L2
5	CA-2019-115742	18-04-2019	18-10-2019	45%	-383.031	5	$ 957.58	P159	C3	M4	S1	L3
6	CA-2019-115742	18-04-2019	18-10-2019	20%	2.5164	2	$ 22.37	P248	C3	M5	S1	L3
7	CA-2019-115742	18-04-2019	14-06-2018	0%	14.1694	7	$ 48.86	P247	C4	M5	S1	L2
8	CA-2020-169194	20-06-2020	14-06-2018	0%	1.9656	4	$ 7.28	P483	C4	M6	S1	L2
9	CA-2020-169194	20-06-2020	14-06-2018	20%	90.7152	6	$ 907.15	P469	C4	M7	S1	L2
10	CA-2018-146703	20-10-2018	14-06-2018	20%	5.7825	3	$ 18.50	P234	C4	M8	S1	L2
11	CA-2020-118255	11-03-2020	14-06-2018	0%	34.47	5	$ 114.90	P126	C4	M9	S1	L2
12	CA-2020-118255	11-03-2020	14-06-2018	20%	85.3092	9	$1,706.18	P184	C4	M10	S1	L2
13	CA-2021-139619	19-09-2021	14-06-2018	20%	68.3568	4	$ 911.42	P412	C4	M11	S1	L2
14	CA-2020-101343	17-07-2020	20-04-2021	20%	5.4432	3	$ 15.55	P732	C5	M12	S1	L5
15	CA-2021-120999	10-09-2021	10-12-2020	20%	132.5922	3	$ 407.98	P279	C6	M13	S1	L6
16	CA-2019-117415	27-12-2019	26-11-2019	80%	-123.858	5	$ 68.81	P352	C7	M14	S3	L7
17	CA-2019-117415	27-12-2019	26-11-2019	80%	-3.816	3	$ 2.54	P638	C7	M15	S3	L7
18	CA-2019-117415	27-12-2019	18-11-2018	0%	13.3176	6	$ 665.88	P641	C8	M11	S1	L8
19	CA-2019-117415	27-12-2019	15-05-2018	0%	9.99	2	$ 55.50	P287	C9	M13	S1	L9
20	CA-2020-117590	08-12-2020	01-09-2018	0%	2.4824	2	$ 8.56	P492	C10	M6	S1	L10
21	CA-2020-117590	08-12-2020	01-09-2018	20%	16.011	3	$ 213.48	P189	C10	M16	S1	L10
22	CA-2021-107727	19-10-2021	01-09-2018	20%	7.384	4	$ 22.72	P688	C10	M17	S1	L10
23	US-2019-150630	17-09-2019	13-12-2020	0%	5.0596	7	$ 19.46	P481	C11	M6	S2	L11

Utilization Of "VLOOKUP" 🔎 Function In Fact Table Creation

➢ **Objective :** Demonstrating the strategic use of VLOOKUP function in amalgamating various data elements for constructing the fact table.

➢ **Methodology :** Leveraged Excel for data manipulation and coordination. Employed the VLOOKUP function to match distinct IDs (Customer, Location, Manufacturer, Segment, Product) with corresponding entries in their respective tables.

➢ **Importance :** Ensured meticulous alignment and synchronization of data from disparate sources. Facilitated the seamless integration of essential data components into the fact table, enhancing its comprehensiveness and coherence.

➢ **Example :**

Sample Formulae : Customer ID: =VLOOKUP(A2, 'Customer Table'!A:B, 2, FALSE) ,Location ID: =VLOOKUP(CONCATENATE(A2,B2,C2,D2) ,'Location Table'!A2:E222, 2, FALSE) ,Manufacturer ID: =VLOOKUP(C2, 'Manufacturer Table'!A:D, 2, FALSE),Segment ID: =VLOOKUP(D2, 'Segment Table'!A:E, 2, FALSE),Product ID: =VLOOKUP(A2, 'Product Table'!A2:E764, 5, FALSE)

➢ **Conclusion :** The strategic utilization of the VLOOKUP function served as a cornerstone in expediting the creation process of the fact table by seamlessly linking essential data attributes from diverse tables.

SQL Data Modelling Task

1. Load all these Tables into MySQL.
2. Perform Below Data Retrieval Tasks.

✓a. Top 5 Customers based on Sales.

✓b. Bottom 10 Customers based on Profits.

✓c. Running Sum of Sales of Corporate Segment from Past Three-Month Range.

✓d. Moving Average of December Month from West Region.

✓e. Retrieve the Data of Customers whose sales are greater than 500 and who are from California.

✓f. Top 5 Products based on Profits with City of the Customers.

✓ g. Retrieve the SUM of sales of each Product and Arrange the Sales in DESC Order.

✓h. Retrieve the SUM of Profits of Each Customers and Arrange the Profits in DESC Order.

✓i.categorize customers based on their total sales value into 'High Value', 'Medium Value', and 'Low Value' segments, and then calculate the number of customers and total sales for each segment.

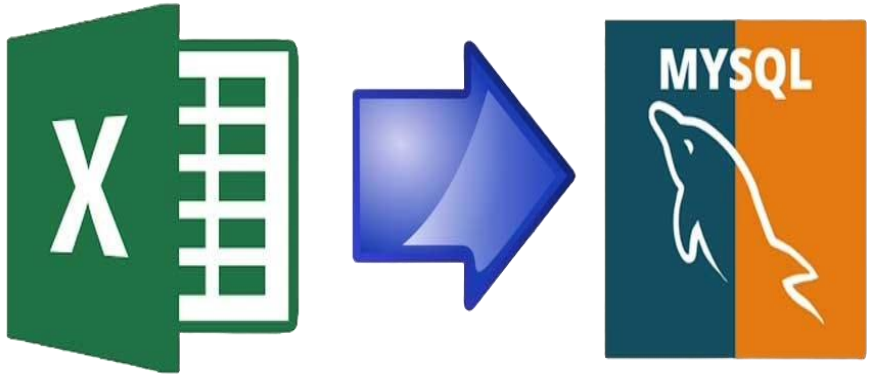
1. Load all these Tables into MySQL

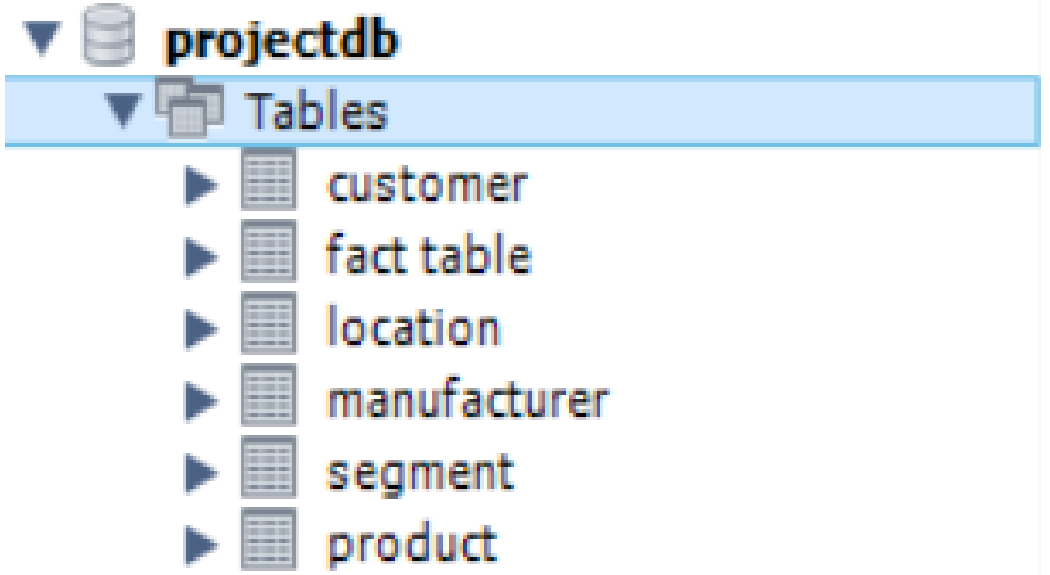
❖ Data Preparation :

Transformed and Cleaned the data.Ensured that unique id's in the dimension tables matches with the fact table id 's .

❖ Importing:

- ❖ **Prepare CSV File :** Save Excel file as CSV.
- ❖ **Connect to SQL Database:** Open SQL database management tool.
- ❖ **Open Import Tool:** Access data import feature.
- ❖ **Select CSV File:** Choose CSV file from local system.
- ❖ **Specify Target Table:** Select target table in SQL database.
- ❖ **Map Columns:** Map CSV columns to database table columns.
- ❖ **Configure Settings:** Specify import settings.
- ❖ **Initiate Import:** Start import process.
- ❖ **Monitor Progress:** Ensure successful completion.
- ❖ **Review Data:** Verify imported data in SQL table.





Establishing Data Integrity and Relationships

Introduction:

Importance of data integrity and relationships in databases.

1. Defining Primary Keys : Assigning unique identifiers to key tables (customer, location, product, manufacturer, segment).

2. Establishing Foreign Keys: Creating references between tables using foreign keys in the fact table.Ensuring data consistency with foreign key constraints.

3. Data Type Optimization : Optimizing data types for efficient storage and processing.Aligning with storage requirements and performance considerations.

4. Ensuring Consistency : Utilizing primary and foreign keys for data consistency.Applying referential integrity constraints and additional validations.

5.Conclusion : Recap of steps taken for robust data integrity and relationships.Importance of these steps for reliable and accurate database management.

a. Top 5 Customers based on Sales

➢ QUERY :

```
SELECT c.`Customer ID`, c.`Customer Name`,SUM(f.`Sales`) as TotalSales from `fact table` f join customer c on c.`Customer ID` = f.`Customer ID` GROUP BY c.`Customer ID`,c.`Customer Name` ORDER BY TotalSales DESC LIMIT 5;
```

➢ SQL QUERY BREAKDOWN : TOP 5 CUSTOMERS BY SALES

- **SELECT :** Chooses the columns to display in the query result.
- **SUM(f.Sales) :** Calculates the total sales for each customer.
- **FROM :** Specifies the source table for data retrieval.
- **JOIN :** Combines data from two tables based on a shared column.
- **ON :** Defines the join condition.
- **GROUP BY :** Groups rows based on specified columns.
- **ORDER BY :** Sorts the result set based on specified columns.
- **LIMIT :** Restricts the number of rows returned by the query.

OUTPUT

Customer ID	Customer Name	TotalSales
C91	Cassandra Brandow	2658.48
C33	Joel Eaton	2371.08
C4	Brosina Hoffman	2009.11
C38	Christopher Schild	1966.23
C127	Kelly Lampkin	1761.70

b. Bottom 10 Customers based on Profits

➢ QUERY :

```
SELECT c.`Customer ID`, c.`Customer Name`,SUM(f.`Profit`) AS TotalProfit FROM customer c
LEFT JOIN `fact table` f ON c.`Customer ID` = f.`Customer ID` GROUP BY c.`Customer
ID`,c.`Customer Name` ORDER BY TotalProfit ASC LIMIT 10;
```

➢ SQL QUERY BREAKDOWN : BOTTOM 10 CUSTOMERS BY PROFIT

- **SELECT :** Chooses the columns to display in the query result.
- **SUM(f.Profit) :** Calculates the total profit for each customer.
- **FROM :** Specifies the source table for data retrieval.
- **JOIN :** Combines data from two tables based on a shared column.
- **ON :** Defines the join condition.
- **GROUP BY :** Groups rows based on specified columns.
- **ORDER BY :** Sorts the result set based on specified columns.
- **LIMIT :** Restricts the number of rows returned by the query.

OUTPUT

Customer ID	Customer Name	TotalProfit
C266	Grant Thornton	-3825.339478492737
C79	Becky Martin	-1878.7891430854797
C15	Tracy Blumstein	-1515.2652465105057
C101	Victoria Wilson	-1196.0596222877502
C185	Tanja Norvell	-857.1918148994446
C325	Paul Knutson	-807.7769813537598
C116	Sean Braxton	-761.9928131103516
C177	Ross Baird	-558.3462073206902
C108	Ken Lonsdale	-460.6215977668762
C38	Christopher Schild	-410.52539920806885

C.Running Sum of Sales of Corporate Segment from Past Three-Month Range

➢QUERY :

```
select f.`Order Date`,s.`Segment` as SegmentName ,SUM(f.`Sales`) over (order by f.`Order Date` range between interval 3 month preceding and current row) as RunningSum from `fact table` f join segment s on f.`Segment ID`= s.`Segment ID` where s.`Segment` = 'Corporate' order by f.`Order Date`;
```

➢SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

SUM(f.Sales) : Calculate the running sum of sales.

OVER : The OVER clause defines a window function that specifies how to partition and order the rows for the calculation.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

WHERE : Filter rows based on conditions.

ORDER BY : Arrange rows in a specific order.

OUTPUT LINK

d. Moving Average of December Month from West Region

➢QUERY :

```
select f.`Order Date`,l.`Region` as Region ,AVG(f.`Sales`) over (order by f.`Order Date` range between interval 1 month preceding and interval 1 month following) as MovingAverage from `fact table` f join location l on f.`Location ID` = l.`Location ID` where l.`Region` = 'West' and month(f.`Order Date`) = 12 order by f.`Order Date`;
```

➢SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

AVG(f.Sales) : Calculate the running average of sales.

OVER : The OVER clause defines a window function that specifies how to partition and order the rows for the calculation.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

WHERE : Filter rows based on conditions.

ORDER BY : Arrange rows in a specific order.

OUTPUT LINK :

e. Retrieve the Data of Customers whose sales are greater than 500 andwho are from California

➢QUERY :

```
select c.`Customer ID`,c.`Customer Name`,l.`State` as CustomerState,f.`Sales` from `fact table` f join customer c on f.`Customer ID`= c.`Customer ID` join location l on f.`Location ID`=l.`Location ID` where f.`Sales` >500 and l.`State` = 'California'; select f.`Order ID`,f.`Sales`,l.`State` from `fact table` f join location l on f.`Location ID` = l.`Location ID` where l.`State` = 'California';
```

➢SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

WHERE : Filter rows based on conditions.

OUTPUT

Customer ID	Customer Name	CustomerState	Sales
C121	Kristen Hastings	California	725.84
C142	Alyssa Crouse	California	698.35
C188	Jeremy Pistek	California	703.97
C126	Nick Zandusky	California	544.01
C4	Brosina Hoffman	California	907.15
C4	Brosina Hoffman	California	911.42
C141	Doug Bickford	California	959.98
C193	Michelle Tran	California	514.17
C67	Helen Andreada	California	671.93
C187	Maya Herman	California	567.12

f. Top 5 Products based on Profits with City of the Customers

➢ QUERY :

```
select p.`Product Name` ,sum(f.`Profit`) as TotalProfit , l.`City` as CustomerCity from `fact table` f join product p on f.`Product ID`= p.`Product ID` join Customer c on f.`Customer ID`= c.`Customer ID` join location l on f.`Location ID`=l.`Location ID` group by p.`Product Name` ,l.`City` order by TotalProfit desc limit 5;
```

➢ SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

SUM(f.Sales) : Calculate the running sum of profit.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

WHERE : Filter rows based on conditions.

ORDER BY : Arrange rows in a specific order.

LIMIT : Restricts the number of rows returned by the query.

OUTPUT

Product Name	TotalProfit	CustomerCity
Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	3177.47509765625	Atlanta
Canon imageCLASS MF7460 Monochrome Digital Laser Multifunction Copier	1995.989990234375	New York City
GBC DocuBind P400 Electric Binding System	1415.4295654296875	New York City
Canon Image Class D660 Copier	1379.97705078125	Great Falls
Fellowes PB300 Plastic Comb Binding Machine	1276.487060546875	Virginia Beach

g. Retrieve the SUM of sales of each Product and Arrange the Sales in DESC Order

➢QUERY :

```
select p.`Product Name` ,SUM(f.`Sales`) AS TotalSales FROM `fact table` f join product p on f.`Product ID` = p.`Product ID` group by p.`Product Name` order by TotalSales desc;
```

➢SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

SUM(f.Sales) : Calculate the running sum of sales.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

ORDER BY : Arrange rows in a specific order.

GROUP BY : Groups rows based on specified columns.

OUTPUT LINK

h. Retrieve the SUM of Profits of Each Customers and Arrange the Profits in DESC Order

➢QUERY :

```
select c.`Customer Name` ,sum(f.`Profit`) as TotalProfit FROM `fact table` f join customer c on f.`Customer ID` = c.`Customer ID` group by c.`Customer Name` order by TotalProfit desc ;
```

➢SQL QUERY BREAKDOWN : L

SELECT : Choose columns for the output.

SUM(f.Profit) : Calculate the running sum of profit.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

ORDER BY : Arrange rows in a specific order.

OUTPUT LINK

i.categorize customers based on their total sales value into 'High Value', 'Medium Value', and 'Low Value' segments, and then calculate the number of customers and total sales for each segment

➢QUERY :

```
select case when TotalSales >= 999.43 then 'High Value' when TotalSales >=
500  and TotalSales < 999.43 then 'Medium Value' else 'Low Value' end as
CustomerSegment ,count(distinct `Customer ID`) as
Number_Of_Customers,sum(TotalSales) as TotalSales from (select `Customer
ID`,sum(f.`Sales`) as TotalSales from `fact table` f  group by `Customer ID` ) as
CustomerSales group by CustomerSegment order by TotalSales desc;
```

➢ SQL QUERY BREAKDOWN :

SELECT : Choose columns for the output.

END : Marks the end of the CASE statement block.

FROM : Specify the source tables.

JOIN : Combine data from multiple tables.

GROUP BY : Groups rows based on specified columns.

ORDER BY : Arrange rows in a specific order.

WHEN : Defines conditions to evaluate,determining which result to return based on those conditions.

CASE : Evaluates conditions & returns a value based on the first condition that is true.

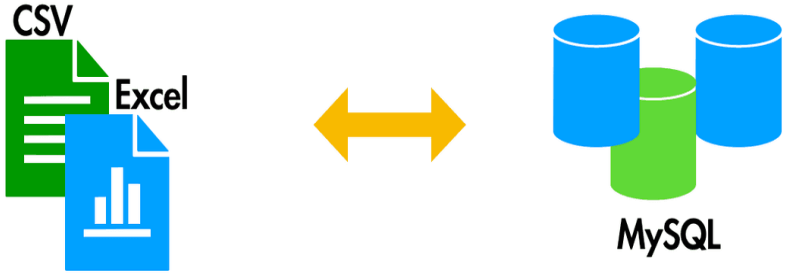
OUTPUT

CustomerSegment	Number_Of_Customers	TotalSales
Medium Value	67	48904.43
Low Value	270	41857.21
High Value	25	36770.88

Navigating the Data Seas: My SQL Journey from Chaos to Clarity

In the dynamic world of retail, I encountered a formidable challenge: the Superstore's vast and chaotic dataset. With sales figures fluctuating and customer insights obscured, I knew that taming this data beast would require skill, determination, and a strategic approach.With SQL as my compass, I set out on a journey to transform chaos into clarity and unlock the store's true potential.First, armed with Excel, I meticulously designed a Star Schema that would serve as the backbone of our data architecture. Each dimension — Customers, Locations, Manufacturers, Products, and Segments — was carefully crafted to provide structure and insight, while the Fact Table anchored our analysis with its wealth of operational data.Having laid the foundation, I ventured into the realm of MySQL, where I loaded the tables and breathed life into my schema. The database became a beacon of organization, ready to be mined for valuable insights.Then came the exhilarating task of data retrieval. With each SQL query, I uncovered hidden treasures within the dataset:I identified the top 5 customers based on sales, shining a spotlight on our most valuable patrons.I delved into profitability, discovering the bottom 10 customers based on profits and areas for improvement.I forecasted future trends by calculating the running sum of sales for the corporate segment and the moving average of December sales in the West region.I segmented our customer base, identifying high-value customers from California with sales exceeding $500.I analyzed product performance, determining the top 5 products based on profits and their associated cities.With each query executed, I gained deeper insights into our operations and paved the way for informed decision-making.

My journey from data chaos to clarity stands as a testament to the power of strategic data analysis and SQL expertise in driving business success. As I continue to navigate the data seas, I am confident that the lessons learned and insights gained will guide us toward a brighter future.



THANK YOU !