

# PROJECT: Data Warehouse Design for E-commerce Environments

## › AIM:

You will be constructing a data warehouse for a retail e-commerce store in this project. You would also be expected to answer a few particular issues about pricing optimization and inventory allocation in terms of design and implementation. In this project, you'll be attempting to answer the following two questions:

- Were the higher-priced items more prevalent in some markets?
- Should inventory be reallocated or prices adjusted based on location?

## › INTRODUCTION:

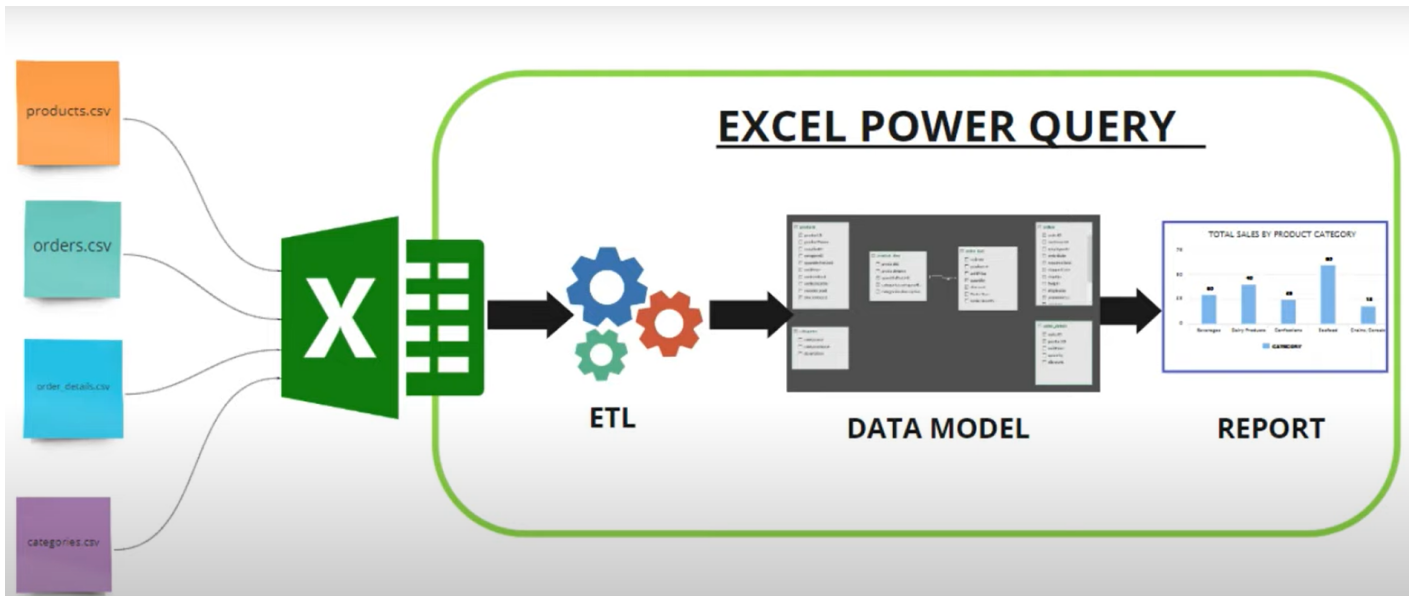
### › Definition:

The primary concept of data warehousing is that the data stored for business analysis can most effectively be accessed by separating it from the data in the operational systems. A data warehouse is a collection of computer-based information that is critical to successful execution of enterprise initiatives. A data warehouse is more than an archive for corporate data and more than a new way of accessing corporate data. A data warehouse is a subject-oriented repository designed with enterprise-wide access in mind. It provides tools to satisfy the information needs of the employees organizational levels- not just for complex data queries, but as general facility for getting quick, accurate and often insightful information. A data warehouse is designed so that its users can recognize the information they want and access that information using simple tools.

One of the principal reasons for developing a data warehouse is to integrate operational data from various sources into a single and consistent architecture that supports analysis and decision-making within the enterprise. Operational systems create, update and delete production data that feed the data warehouse. A data warehouse is analogous to a physical warehouse. Operational systems create data 'parts' that are loaded into the warehouse. Some of those parts are summarised into information 'components' and are stored in the warehouse. Data warehouse users make requests and are delivered information 'products' that are created from the components and parts stored in the warehouse. A data warehouse is typically a blending of technologies, including relational and multidimensional databases, client/ server architecture, extraction / transformation

programs, graphical user interfaces, and more.

## DATA PIPELINE MODEL:



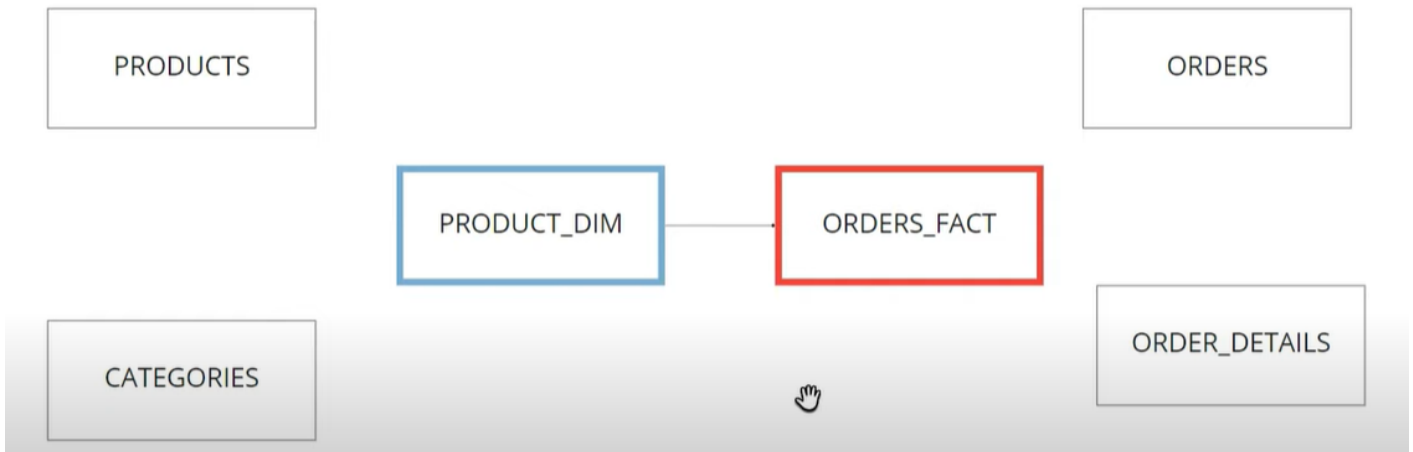
## LOGICAL DATA MODEL:

### LOGICAL DATA MODEL



## REPORTING DATA MODEL:

# REPORTING DATA MODEL



## › CONTENTS:

In this project, we'll be creating the Data Warehouse and Data Pipeline using the Excel Power Query. We can also use tools like AWS Cloud, Microsoft SQL Server Management Studio, Power BI or Ploomberg to create the Pipeline.

## › SOFTWARE REQUIREMENTS:

1. Microsoft Excel  
or
2. Power BI  
or
3. AWS Cloud  
or
4. Microsoft SQL Server Management Studio

## › Steps:

1. Create and Extract DataBase Tables with Dimensions for the E-commerce sales in the Microsoft excel or Power BI/ Microsoft SQL Server Management Studio/ AWS Cloud

### a. EXCEL

productID	productNa	supplierID	categoryID	quantityPe	unitPrice	unitsInStoc	unitsOnOr	reorderLev	discontinued
1	Chai	1	1	10 boxes x	18	39	0	10	0
2	Chang	1	1	24 - 12 oz	19	17	40	25	0
3	Aniseed Sy	1	2	12 - 550 m	10	13	70	25	0
4	Chef Anton	2	2	48 - 6 oz ja	22	53	0	0	0
5	Chef Anton	2	2	36 boxes	21.35	0	0	0	1
6	Grandma's	3	2	12 - 8 oz ja	25	120	0	25	0
7	Uncle Bob	3	7	12 - 1 lb pk	30	15	0	10	0
8	Northwoo	3	2	12 - 12 oz	40	6	0	0	0
9	Mishi Kobe	4	6	18 - 500 g	97	29	0	0	1
10	Ikura	4	8	12 - 200 m	31	31	0	0	0
11	Queso Cab	5	4	1 kg pkg.	21	22	30	30	0
12	Queso Ma	5	4	10 - 500 g	38	86	0	0	0
13	Konbu	6	8	2 kg box	6	24	0	5	0
14	Tofu	6	7	40 - 100 g	23.25	35	0	0	0
15	Genen Sho	6	2	24 - 250 m	15.5	39	0	5	0
16	Pavlova	7	3	32 - 500 g	17.45	29	0	10	0
17	Alice Mutt	7	6	20 - 1 kg ti	39	0	0	0	1
18	Carnarvon	7	8	16 kg pkg.	62.5	42	0	0	0
19	Teatime Cl	8	3	10 boxes x	9.2	25	0	5	0
20	Sir Rodney	8	3	30 gift box	81	40	0	0	0
21	Sir Rodney	8	3	24 pkgs. x	10	3	40	5	0
22	Gustaf's Kr	9	5	24 - 500 g	21	104	0	25	0
23	TunnbrÄŕc	9	5	12 - 250 g	9	61	0	25	0
24	GuaranÄŕj	10	1	12 - 355 m	4.5	20	0	0	1
25	NuNuCa N	11	3	20 - 450 g	14	76	0	30	0
26	GumbÄŕr c	11	3	100 - 250 g	31.23	15	0	0	0

## Queries & Connections

Queries

Connections

4 queries

categories

8 rows loaded.

orders

830 rows loaded.

order\_details

2,155 rows loaded.

products

77 rows loaded.

## b. MICROSOFT SQL SERVER MANAGEMENT STUDIO

Createdatabase Retail\_DW  
Go

Use Retail\_DW

Go

Create table DimProduct

```
(  
ProductKey int primary key identity,  
ProductAltKey varchar(10)not null,  
ProductName varchar(100),  
ProductActualCost money,  
ProductSalesCost money
```

)

Go

Create table DimOrder

```
(  
StoreID int primary key identity,  
StoreAltID varchar(10)not null,  
StoreName varchar(100),  
StoreLocation varchar(100),  
City varchar(100),  
State varchar(100),  
Country varchar(100)
```

)

Go

## c. AWS CLOUD

```
$ python create_tables.py
```

2. Transform the data base tables, merge the dimensions with the fact tables of the data base

## a. EXCEL

The screenshot displays the Power Query Editor window titled 'product\_dim - Power Query Editor'. The main area shows a data table with the following columns: categoryID, categoryName, products.productID, and products.productName. The table contains 17 rows of data. The ribbon at the top includes tabs for File, Home, Transform, Add Column, and View. The right sidebar shows the 'Query Settings' pane with the 'Name' set to 'product\_dim' and 'Applied Steps' listed as 'Source', 'Expanded products', and 'Removed Columns'.

categoryID	categoryName	products.productID	products.productName
1	Beverages	1	Chai
2	Beverages	2	Chang
3	Condiments	3	Aniseed Syrup
4	Condiments	4	Chef Anton's Cajun Seasoning
5	Condiments	5	Chef Anton's Gumbo Mix
6	Condiments	6	Grandma's Boysenberry Spread
7	Condiments	8	Northwoods Cranberry Sauce
8	Meat/Poultry	9	Mishi Kobe Niku
9	Produce	7	Uncle Bob's Organic Dried Pears
10	Seafood	10	Ikura
11	Dairy Products	11	Queso Cabrales
12	Dairy Products	12	Queso Manchego La Pastora
13	Seafood	13	Konbu
14	Produce	14	Tofu
15	Condiments	15	Genen Shouyu
16	Confections	16	Pavlova
17	Meat/Poultry	17	Alice Mutton

File Home Transform Add Column View

Column From Custom Invoke Custom Examples Column Function

General

From Text From Number From Date & Time

Queries [6]

- categories
- orders
- order\_details
- products
- product\_dim
- order\_dim

fx = Table.TransformColumns(#"Extracted Year", {"order month", each Date.MonthName(\_), type text})

orderID	productID
1	10248
2	10248
3	10248
4	10249
5	10249
6	10250
7	10250
8	10250
9	10251
10	10251
11	10251
12	10252
13	10252
14	10252
15	10253
16	10253
17	10253
18	10254
19	10254
20	10254
21	10255
22	10255
23	10255

### Custom Column

Add a column that is computed from the other columns.

New column name  
Sales Amount

Custom column formula  
= [unitPrice]\*[quantity]

Available columns  
orderID  
productID  
unitPrice  
quantity  
order year  
order month

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

fx = Table.AddColumn(#"Extracted Month Name", "Sales Amount", each [unitPrice]\*[quantity])

orderID	productID	unitPrice	quantity	order year	order month	Sales Amount
1	10248	11	14	1996	July	168
2	10248	42	9.8	1996	July	98
3	10248	72	34.8	1996	July	174
4	10249	14	18.6	1996	July	167.4
5	10249	51	42.4	1996	July	1696
6	10250	41	7.7	1996	July	77
7	10250	51	42.4	1996	July	1484
8	10250	65	16.8	1996	July	252
9	10251	22	16.8	1996	July	100.8
10	10251	57	15.6	1996	July	234
11	10251	65	16.8	1996	July	336
12	10252	20	64.8	1996	July	2592
13	10252	33	2	1996	July	50
14	10252	60	27.2	1996	July	1088
15	10253	31	10	1996	July	200
16	10253	39	14.4	1996	July	604.8
17	10253	49	16	1996	July	640
18	10254	24	3.6	1996	July	54
19	10254	55	19.2	1996	July	403.2

Query Settings

PROPERTIES

Name  
order\_dim

APPLIED STEPS

- Source
- Expanded orders
- Duplicated Column
- Removed Columns
- Renamed Columns
- Extracted Year
- Extracted Month Name
- Added Custom

fx = Table.TransformColumnTypes(#"Sorted Rows",{"Sales Amount", Currency.Type})

product_dim.categoryName	Sales Amount
Beverages	15,810.00
Beverages	15,810.00
Beverages	10,540.00
Beverages	10,540.00
Beverages	10,540.00
Beverages	10,329.20
Meat/Poultry	9,903.20
Beverages	8,432.00
Beverages	7,905.00
Beverages	7,905.00
Beverages	7,905.00
Meat/Poultry	7,427.40

Query Settings

PROPERTIES

Name  
Total Sales

APPLIED STEPS

- Source
- Expanded product\_dim
- Removed Columns
- Reordered Columns
- Sorted Rows
- Changed Type

Total_Sales	Remove Duplicates	Insert Slicer	Export Refresh	Open in Browser	Total Row	Last Column	Table Styles									
Resize Table	Convert to Range	Tools	External Table Data	Unlink	Banded Rows	Banded Columns	Table Style Options									
product_dim.categoryName																
A1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	product_dim.categoryName	Sales Amount														
2	Beverages	15810														
3	Beverages	15810														
4	Beverages	10540														
5	Beverages	10540														
6	Beverages	10540														
7	Beverages	10329.2														
8	Beverages	8432														
9	Beverages	7905														
10	Beverages	7905														
11	Beverages	7905														
12	Beverages	6587.5														
13	Beverages	6324														
14	Beverages	4216														
15	Beverages	4216														
16	Beverages	3952.5														
17	Beverages	3952.5														
18	Beverages	3952.5														
19	Beverages	2760														
20	Beverages	2635														
21	Beverages	2340														
22	Beverages	2108														
23	Beverages	2108														
24	Beverages	1900														
25	Beverages	1840														
26	Beverages	1656														

## b. MICROSOFT SQL SERVER MANAGEMENT STUDIO

```

Create Table FactProduct
(
    TransactionID bigint primary key identity,
    SalesInvoiceNumber int not null,
    SalesDateKey int,
    SalesTimeKey int,
    SalesTimeAltKey int,
    StoreID int not null,
    CustomerID int not null,
    ProductID int not null,
    SalesPersonID int not null,
    Quantity float,
    SalesTotalCost money,
    ProductActualCost money,
    Deviation float
)
Go

-- Add relation between fact table foreign keys to Primary keys of Dimensions
ALTER TABLE FactProductSales ADD CONSTRAINT _
FK_StoreID FOREIGN KEY (StoreID)REFERENCES DimStores(StoreID);
ALTER TABLE FactProductSales ADD CONSTRAINT _
FK_CustomerID FOREIGN KEY (CustomerID)REFERENCES Dimcustomer(CustomerID);
ALTER TABLE FactProductSales ADD CONSTRAINT _
FK_ProductKey FOREIGN KEY (ProductID)REFERENCES Dimproduct(ProductKey);
ALTER TABLE FactProductSales ADD CONSTRAINT _
FK_SalesPersonID FOREIGN KEY (SalesPersonID)REFERENCES Dimsalesperson(SalesPersonID);
Go
ALTER TABLE FactProductSales ADD CONSTRAINT _
FK_SalesDateKey FOREIGN KEY (SalesDateKey)REFERENCES DimDate(DateKey);
Go
ALTER TABLE FactProductSales ADD CONSTRAINT _

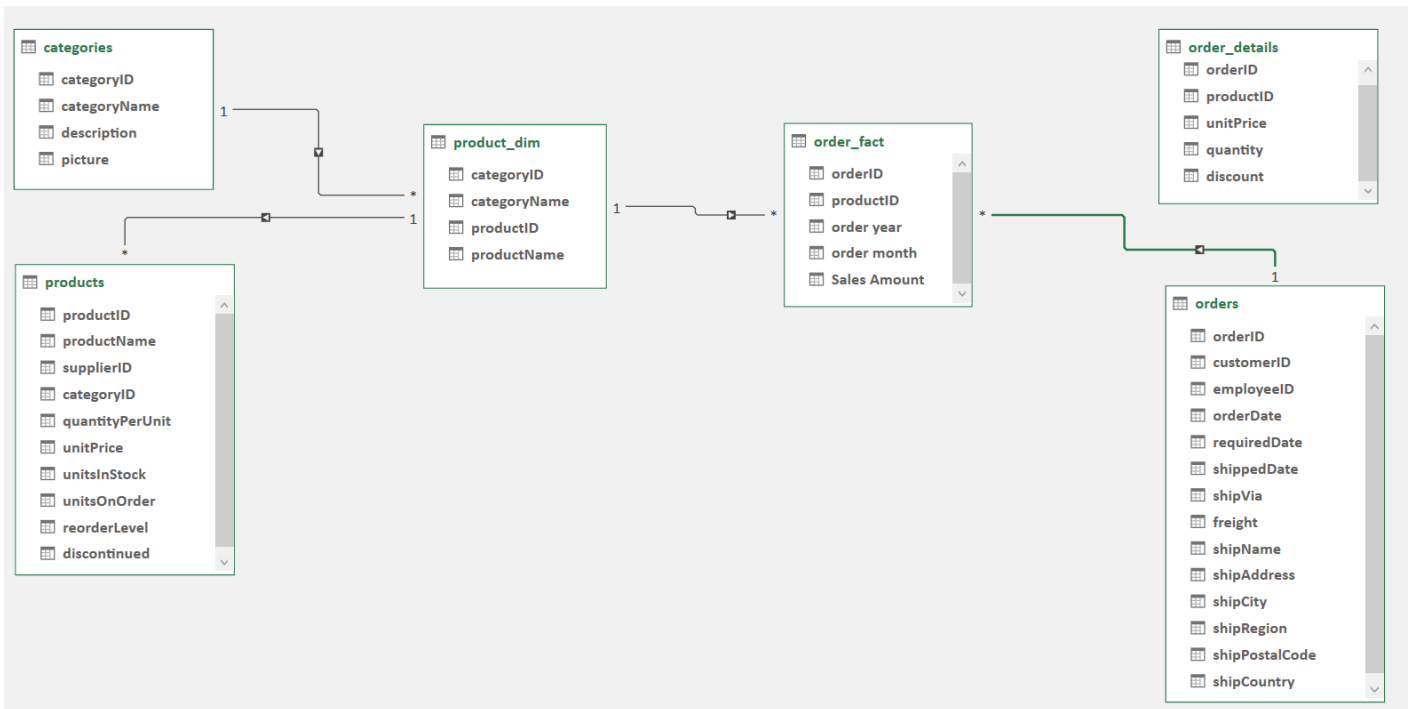
```

```
FK_SalesTimeKey FOREIGN KEY (SalesTimeKey)REFERENCES DimDate(TimeKey);  
Go
```

### c. AWS CLOUD

```
$ python etl.py
```

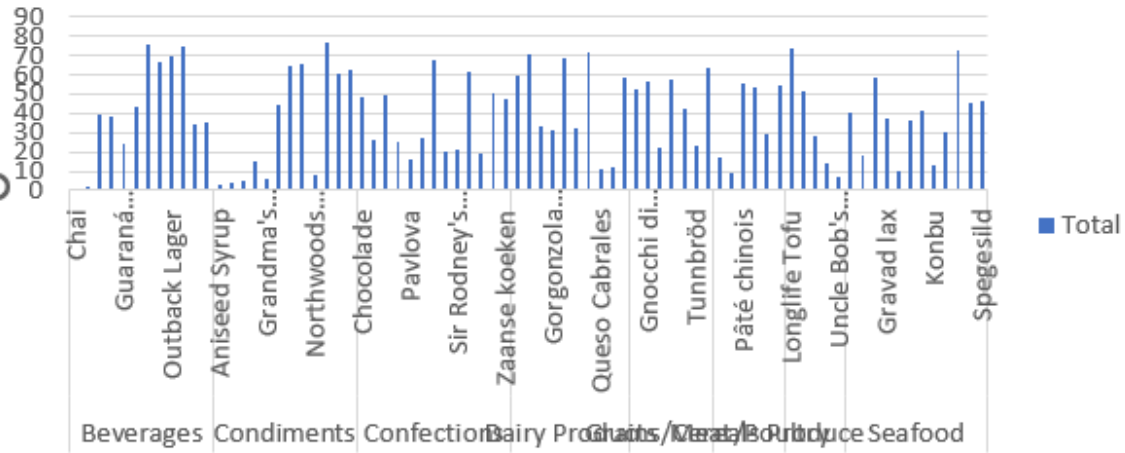
### 3. Visualaize and get the report of the data using Chart or Power BI





Sum of productID

Total



categoryName

productName

+ -

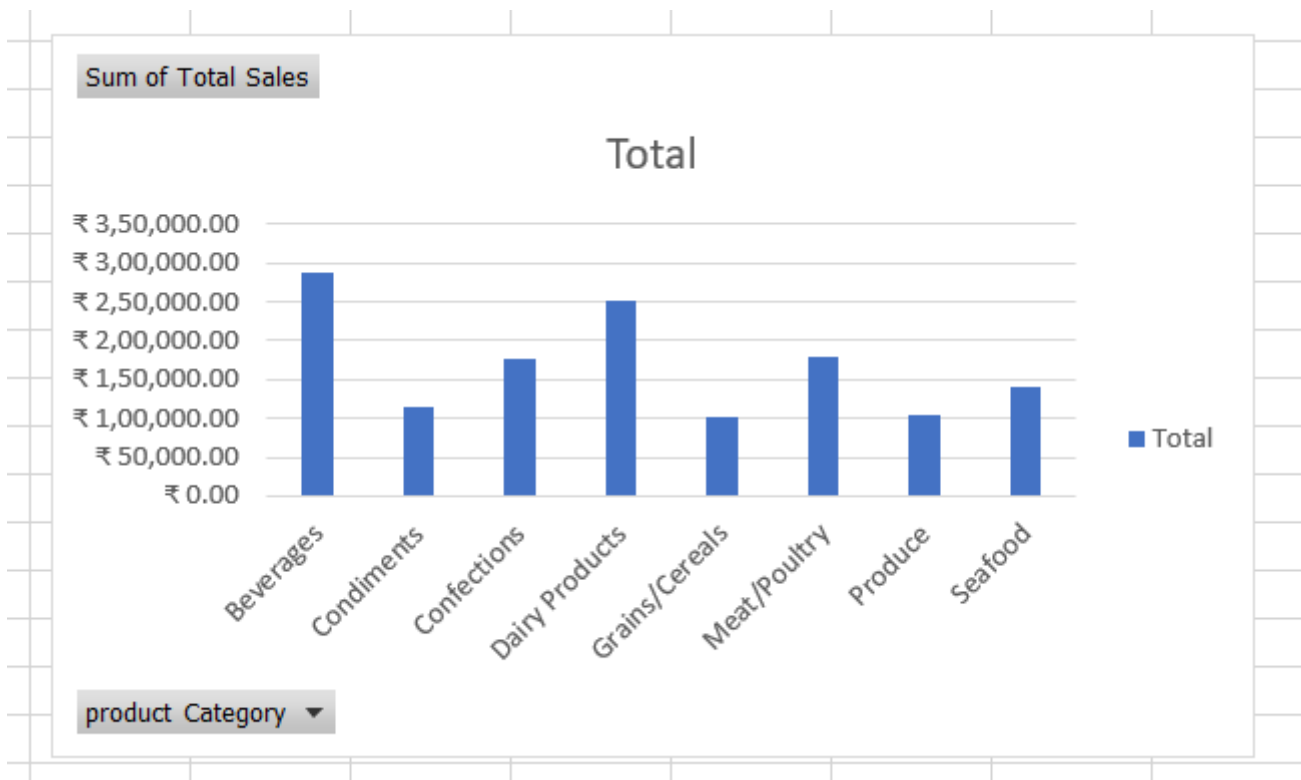
Count of Sales Amount

order month

- April
- August
- December
- February
- January
- July
- June

orderID

10248 10285 10322 10359 10396 10433 10470 10507 10544 10581 10618 10655 10692 10729 10766 10803 10840 10877 10914 10951 10988 11025 11062



Row Labels	Sum of Total Sales
Beverages	₹ 2,86,526.95
Condiments	₹ 1,13,694.75
Confections	₹ 1,77,099.10
Dairy Products	₹ 2,51,330.50
Grains/Cereals	₹ 1,00,726.80
Meat/Poultry	₹ 1,78,188.80
Produce	₹ 1,05,268.60
Seafood	₹ 1,41,623.09
<b>Grand Total</b>	<b>₹ 13,54,458.59</b>

## Result:

Thus, a Data Warehouse and a ETL Data Pipeline is created using the Excel Power Query for Data Warehouse Design for E-commerce Environments.