Contoso Retail Dataset Cheat Sheet

Welcome to the Contoso Database Cheat sheet!
This guide provides an overview of the Contoso Retail dataset, a popular sample database for learning and testing in data analytics. Designed for beginners and aspiring BI professionals, the cheat sheet aims to build both **business and technical knowledge** essential for senior Business Intelligence roles.

Created: Nov 2024

BrainAnalytix
LinkedIn

Arkadiusz Wos

1 Business Model Overview

What is Contoso?

Contoso is a fictional company created by Microsoft to help people learn and practice data analysis. It's designed as a retail business that sells a variety of products to simulate real-world business data.
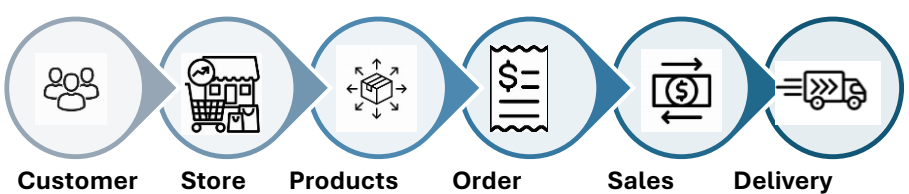Overview of Contoso Corporation - Microsoft 365 Enterprise | Microsoft Learn
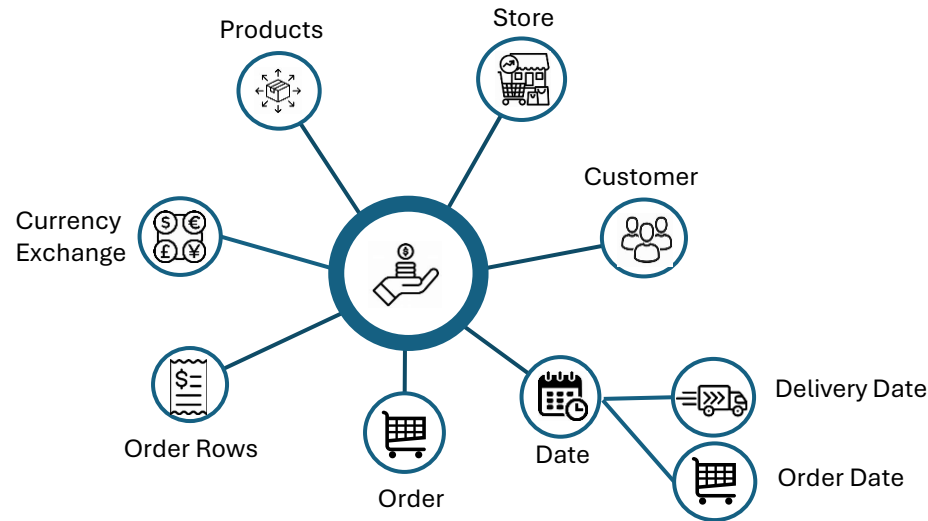
Retail Business Model

Manufactures → Products → Stores → Orders → Customers → Sales

Order Fulfilment Process

Customer → Store → Products → Order → Sales → Delivery

Sales Transaction Data Model

2 Database Access

You can download the Contoso dataset from GitHub, where it's available in multiple formats and sizes to suit different needs.
Release Ready to use data - sql-bi/Contoso-Data-Generator-V2-Data · GitHub

File Format

Format	
bak	SQL Server database backup, used for restoring the Contoso database in SQL Server environments.
pbix	Power BI Desktop file, directly usable in Power BI for data analysis and reporting.
pbit	Power BI Desktop template file, used to import the SQL Server database structure into Power BI.
csv	Standard CSV file format, suitable for quick data viewing, import into Excel, or lightweight analysis tools.
delta	Delta Table format, optimized for large-scale data analysis in cloud environments, especially with Apache Spark or Databricks.
parquet	Parquet file format, ideal for use in data lakes or with big data platforms like Amazon S3, Google BigQuery, and Apache Hadoop. It's optimized for efficient storage and fast query performance.

Suggested Data Format: .bak file (SQL Server backup) is recommended, as it supports both Import Mode and Direct Query Mode in Power BI. This format preserves all relationships, allowing you to view the full data model and providing flexibility for in-depth analysis and data exploration in Power BI.

File Sizes

Size	Description
10K	About 10,000 orders; impacts only **Orders**, **Order Item**, or **Sales** view
100K	About 100,000 orders; impacts only **Orders**, **Order Item**, or **Sales** view
1M	About 1 million orders; impacts only **Orders**, **Order Item**, or **Sales** view
10M	About 10 million orders; impacts only **Orders**, **Order Item**, or **Sales** view
100M	About 100 million orders (split into sub-files); impacts only **Orders**, **Order Item**, or **Sales** views

Suggested Data Size: 1M (around 1 million orders) is manageable for testing on typical desktop setups and large enough to reveal real-world performance challenges, helping users optimize for scalability.

3 Basic KPIs

Total Cost:
Calculates the total cost of products sold by multiplying the quantity sold by the unit cost for each sale. This KPI is essential for understanding the total expenses associated with sales.
DAX Measure:
Total Cost = SUMX (Sales, Sales[Quantity] * Sales[Unit Cost])

Total Quantity:
Sums up the total quantity of items sold. This KPI provides insight into sales volume and product demand.
DAX Measure:
Total Quantity = SUM (Sales[Quantity])

Sales Amount:
Calculates the total revenue generated from sales by multiplying the quantity sold by the net price per unit. It reflects the gross revenue before any deductions for cost.
DAX Measure:
Sales Amount = SUMX (Sales, Sales[Quantity] * Sales[Net Price])

Margin:
Represents the absolute profit by subtracting the total cost from the total sales amount. It shows the monetary profit made on sales.
DAX Measure:
Margin = [Sales Amount] - [Total Cost]

Margin %:
Calculates the profit margin as a percentage of the sales revenue, indicating profitability. It helps measure how much of the revenue is retained as profit after costs are covered.
DAX Measure:
Margin % = DIVIDE ([Margin], [Sales Amount])

4 Tables in Data Model

OrderRows
Table captures line-item details for orders, recording each product in an order and its price and quantity. This table enables detailed sales analysis, e.g. product basket analysis.

Column Name	Description
OrderKey	Unique identifier for each order.
Line Number	Line item number within an order.
ProductKey	Identifier for the product.
Quantity	Number of units ordered.
Unit Price	Price per unit of the product.
Net Price	Total price after adjustments.
Unit Cost	COGS (Cost of Good Sold) per unit of the product.

Orders
Table provides summary-level details of customer orders, including dates, currency, and related customer and store keys.

Column Name	Description
OrderKey	Unique identifier for each order.
CustomerKey	Identifier for the customer placing the order.
StoreKey	Identifier for the store fulfilling the order.
Order Date	Date when the order was placed.
Delivery Date	Expected or actual delivery date.
Currency Code	Currency code used in the order.

CurrencyExchange
Table stores exchange rate information, allowing accurate conversion of amounts across different currencies. It is essential for businesses operating in multiple currencies to ensure that sales and expenses are recorded at the correct rates.

Column Name	Description
Date	Date of the exchange rate.
FromCurrency	Currency code being converted from.
ToCurrency	Currency code being converted to.
Exchange	Conversion rate between the two currencies.

Customer
Table stores information about customers, including demographics and location. It helps in understanding customer profiles and behaviors for targeted marketing and sales analysis.

Column Name	Description
CustomerKey	Unique identifier for each customer.
Gender	Gender of the customer.
Title	Customer's title (e.g., Mr., Ms.).
Name	Full name of the customer.
Address	Full address of the customer.
GivenName	Customer's first name.
MiddleInitial	Middle initial of the customer.
City	City where the customer resides.
State Code	State code of the customer's address.
Surname	Last name of the customer.
StreetAddress	Street address of the customer.
State	State where the customer resides.
ZipCode	Postal code of the customer's address.
Country Code	Country code of the customer's location.
Country	Full country name of the customer's location.
Birthday	Date of birth of the customer.
Age	Age of the customer.
Occupation	Job or occupation of the customer.
Company	Company where the customer works.
Vehicle	Type of vehicle owned by the customer.
Latitude	Latitude of the customer's location.
Longitude	Longitude of the customer's location.
Continent	Continent of the customer's address.

Date
table provides calendar-based information, supporting analysis by different time dimensions such as year, quarter, month, and day of the week. This table is useful for time-based reporting.

Column Name	Description
Date	Specific date.
DateKey	Unique identifier for the date.
Year	Year component of the date.
Year Quarter	Year and quarter combination.
Year Quarter Number	Numeric representation of the quarter.
Quarter	Quarter component (Q1-Q4).
Year Month	Year and month combination.
Year Month Short	Abbreviated year and month.
Year Month Number	Numeric representation of the year and month.
Month	Full month name.
Month Short	Abbreviated month name.
Month Number	Numeric representation of the month.
Day of Week	Full name of the day of the week.
Day of Week Short	Abbreviated day name.
Day of Week Number	Numeric representation of the day.
Working Day	Indicates if it's a working day.
Working Day Number	Numeric code for working day.

Product
Table holds information about each product, including identifiers, descriptions, prices, and category details. This is vital for product-level reporting.

Column Name	Description
ProductKey	Unique identifier for each product.
Product Code	Code representing the product.
Product Name	Descriptive name of the product.
Manufacturer	Company producing the product.
Brand	Brand name of the product.
Color	Color description of the product.
Weight Unit Measure	Measurement unit for weight.
Weight	Weight of the product.
Unit Cost	Cost to produce or acquire the product.
Unit Price	Price at which the product is sold.
Subcategory Code	Code for the product's subcategory.
Subcategory	Name of the subcategory.
Category Code	Code for the product's category.
Category	Name of the product category.

Store
Table contains information about each retail store, including its location, size, and operational status. This table is useful for analyzing store performance, tracking locations, and understanding factors such as store size and regional distribution.

Column Name	Description
StoreKey	Unique identifier for each store.
Store Code	Code assigned to the store for identification.
Country	Country where the store is located.
State	State where the store is located.
Name	Name of the store.
Square Meters	Size of the store in square meters.
Open Date	Date when the store was opened.
Close Date	Date when the store was closed (if applicable).
Status	Current operational status of the store (e.g., Open, Closed).

5 View in Data model

Sales
View includes transactional details of each order, including order date, customer, store, and product information, which are key to analyzing sales performance, financial numbers are exchange to $ USD.

```
CREATE VIEW dbo.Sales AS
SELECT
    Orders.OrderKey AS [Order Number],
    OrderRows.[Line Number] AS [Line Number],
    Orders.[Order Date],
    Orders.[Delivery Date],
    Orders.CustomerKey,
    Orders.StoreKey,
    OrderRows.ProductKey,
    OrderRows.Quantity,
    OrderRows.[Unit Price],
    OrderRows.[Net Price],
    OrderRows.[Unit Cost],
    Orders.[Currency Code],
    [CurrencyExchange].Exchange AS [Exchange Rate]
FROM
    [Data].Orders
    LEFT OUTER JOIN [Data].OrderRows
        ON Orders.OrderKey = OrderRows.OrderKey
    LEFT OUTER JOIN [Data].CurrencyExchange
        ON [CurrencyExchange].Date = Orders.[Order Date]
        AND [CurrencyExchange].[ToCurrency] = Orders.[Currency Code]
        AND [CurrencyExchange].[FromCurrency] = 'USD'
```

Column Name	Description
Order Number	Unique identifier for each order.
Line Number	Line item number within an order.
Order Date	Date the order was placed.
Delivery Date	Date the order was delivered.
CustomerKey	Identifier for the customer who placed the order.
StoreKey	Identifier for the store fulfilling the order.
ProductKey	Identifier for the product ordered.
Quantity	Number of units ordered.
Unit Price	Price per unit of the product.
Net Price	Total price after any discounts.
Unit Cost	Cost per unit of the product.
Currency Code	Currency code used in the transaction.
Exchange Rate	Exchange rate applicable to the order.

Currency Exchange
View provides exchange rate information between different currencies. It's essential for calculating sales in various currencies accurately.

```
CREATE VIEW dbo.[Currency Exchange] AS
SELECT
    [Date],
    [FromCurrency],
    [ToCurrency],
    [Exchange]
FROM [Data].[CurrencyExchange]
```

Column Name	Description
Date	Date of the currency exchange rate.
FromCurrency	Currency code of the currency being converted from.
ToCurrency	Currency code of the currency being converted to.
Exchange	Exchange rate between the two currencies.

Customer
View contains information on each customer, including demographics and geographic location, which is useful for customer segmentation and targeted marketing.

```
CREATE VIEW dbo.Customer AS
SELECT
    [CustomerKey],
    [Gender],
    [GivenName] + ' ' + [Surname] AS [Name],
    [StreetAddress] AS [Address],
    [City],
    [State] AS [State Code],
    StateFull AS [State],
    ZipCode AS [Zip Code],
    [Country] AS [Country Code],
    [CountryFull] AS [Country],
    [Continent],
    [Birthday],
    [Age] AS [Age]
FROM
    [Data].Customer
```

Store
View provides information about each retail store, including location, size, and operational status, which is useful for geographic and operational analysis.

```
CREATE VIEW dbo.Store AS
SELECT
    StoreKey,
    [Store Code],
    [Country],
    [State],
    [Name],
    [Square Meters],
    [Open Date],
    [Close Date],
    [Status]
FROM
    [Data].Store
```

Column Name	Description
StoreKey	Unique identifier for each store.
Store Code	Code for identifying the store.
Country	Country where the store is located.
State	State where the store is located.
Name	Name of the store.
Square Meters	Size of the store in square meters.
Open Date	Date the store was opened.
Close Date	Date the store was closed, if applicable.
Status	Operational status of the store (e.g., Open).

Date
View provides calendar-based information for analyzing sales and events by various time dimensions, including year, quarter, month, and day of the week.

```
CREATE VIEW dbo.Date AS
SELECT
    [Date],
    [Year],
    [Year Quarter],
    [Year Quarter Number],
    [Quarter],
    [Year Month],
    [Year Month Short],
    [Month],
    [Month Short],
    [Month Number],
    [Day of Week],
    [Day of Week Short],
    [Day of Week Number],
    [Working Day],
    [Working Day Number]
FROM
    [Data].[Date]
```

Product
View provides detailed information on each product, including product identification, brand, category, and pricing, essential for inventory and sales analysis.

```
CREATE VIEW dbo.Product AS
SELECT
    ProductKey,
    [Product Code],
    [Product Name],
    [Manufacturer],
    [Brand],
    [Color],
    [Weight Unit Measure],
    [Weight],
    [Unit Cost],
    [Unit Price],
    [Subcategory Code],
    [Subcategory],
    [Category Code],
    Category
FROM
    [Data].Product
```

Column Name	Description
CustomerKey	Unique identifier for the customer.
Gender	Gender of the customer.
Title	Customer's title (e.g., Mr., Ms.).
Name	Full name of the customer.
MiddleInitial	Middle initial of the customer.
Surname	Last name of the customer.
Address	Address of the customer.
City	City of the customer's residence.
State Code	Code for the state where the customer resides.
State	Full state name.
StreetAddress	Street address of the customer.
Zip Code	Postal code of the customer's address.
Country Code	Code for the country where the customer resides.
Country	Full country name of the customer's residence.
Continent	Continent of the customer's address.
Birthday	Customer's date of birth.
Age	Age of the customer.
Occupation	Job or occupation of the customer.
Company	Company where the customer works.
Vehicle	Type of vehicle owned by the customer.
Latitude	Latitude of the customer's location.
Longitude	Longitude of the customer's location.

Column Name	Description
ProductKey	Unique identifier for each product.
Product Code	Code representing the product.
Product Name	Name of the product.
Manufacturer	Company producing the product.
Brand	Brand of the product.
Color	Color of the product.
Weight Unit Measure	Unit of measurement for the product's weight.
Weight	Weight of the product.
Unit Cost	Cost to produce or acquire the product.
Unit Price	Selling price of the product.
Subcategory Code	Code for the product subcategory.
Subcategory	Name of the subcategory.
Category Code	Code for the product category.
Category	Name of the product category.

Column Name	Description
Date	Specific date.
DateKey	Unique identifier for the date.
Year	Year component of the date.
Year Quarter	Year and quarter combined.
Year Quarter Number	Numeric quarter representation.
Quarter	Quarter of the year (e.g., Q1, Q2).
Year Month	Year and month combined.
Year Month Short	Abbreviated year and month.
Year Month Number	Numeric representation of the year and month.
Month	Name of the month.
Month Short	Abbreviated month name.
Month Number	Numeric representation of the month.
Day of Week	Day of the week.
Day of Week Short	Abbreviated day name.
Day of Week Number	Numeric representation of the day.
Working Day	Indicates if it's a working day.
Working Day Number	Numeric code for working days.

6 Entity Relationship Diagram

ERD Diagram for Tables

ERD Diagram for View