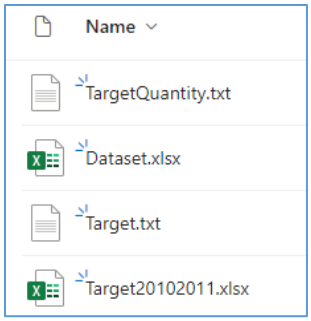
Challenge #1 – Prepare the data

Adventure Works Cycles is a large, multinational manufacturing company. The company manufactures and sells metal and composite bicycles to North American, European and Asian commercial markets. While its base operation is located in Bothell, Washington with 290 employees, several regional sales teams are located throughout their market base.

In 2009, Adventure Works Cycles bought a small manufacturing plant, Importadores Neptuno, located in Mexico. Importadores Neptuno manufactures several critical subcomponents for the Adventure Works Cycles product line. These subcomponents are shipped to the Bothell location for final product assembly. In 2010, Importadores Neptuno, became the sole manufacturer and distributor of the touring bicycle product group.

Coming off a successful fiscal year, Adventure Works Cycles is looking to broaden its market share by targeting their sales to their best customers, extending their product availability through an external Web site, and reducing their cost of sales through lower production costs.

1. Product, Customer, Sales and Targets are stored in Excel workbooks, csv/txt files. Extract the ZIP file shared by instructor.



2. Dataset.xls workbook has information about the customers, their geography, products and sales.

Table	Contains this kind of content	Comment
DimProduct	Information about each product sold by Adventure Works Cycles or used to manufacture Adventure Works Cycles bicycles and bicycle components.	The **FinishedGoodsFlag** column indicates whether a product is sold. Products that are not sold are components of a product that is sold.
DimProductCategory	The most general classification of products. For example, bike or accessory.	
DimProductSubCategory	Subcategories of product categories. For example, Mountain, Road, and Touring are subcategories of the category Bike.	
DimGeography	State, city, pin code of each country	

Challenge #1 – Prepare the data

DimCustomer	Customer details including date of joining, annual income, family	
FactInternetSales	Contains individual customer Internet sales order and detail data.	

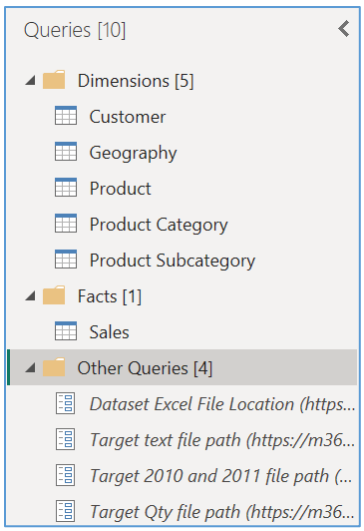
3. Create four parameters to store the location of these files.

4. Import the following tables from Dataset workbook (DON'T import DimDate)
    - DimProduct
    - DimProductSubCategory
    - DimProductCategory
    - FactInternetSales
    - DimCustomer
    - DimGeography

5. Rename the queries
    - DimProduct to Product
    - DimProductSubCategory to Product Subcategory,
    - DimProductCategory to Product Category,
    - FactInternetSales to Sales
    - DimCustomer to Customer
    - DimGeography to Geography

6. Group the queries based on their purpose

🛑 Make sure the query and parameter names match the figure shown below.



Challenge #1 – Prepare the data

7. Apply the following transformations to Customer query
    a. The following columns should be removed.
        i. CustomerAlternateKey
        ii. Title
        iii. NameStyle
        iv. Suffix
        v. EmailAddress
        vi. NumberChildrenAtHome
        vii. SpanishEducation
        viii. FrenchEducation
        ix. SpanishOccupation
        x. FrenchOccupation
        xi. Phone
        xii. CommuteDistance
    b. Combine FirstName, MiddleName, and LastName into the Name column.
    c. Replace values
        i. Marital Label: M = Married, S = Single
        ii. Gender: M for Male, F for Female
        iii. HouseOwnerFlag: 1 indicates Yes, 0 indicates No.
    d. Rename columns
        i. EnglishEducation to Education
        ii. EnglishOccupation to Occupation

Name	TypeName	Kind	Row Count
CustomerKey	Int64.Type	number	*18484*
GeographyKey	Int64.Type	number	*18484*
Name	Text.Type	text	*18484*
BirthDate	Date.Type	date	*18484*
Gender	Text.Type	text	*18484*
YearlyIncome	Currency.Type	number	*18484*
TotalChildren	Int64.Type	number	*18484*
NumberChildrenAtHome	Int64.Type	number	*18484*
Education	Text.Type	text	*18484*
Occupation	Text.Type	text	*18484*
HouseOwnerFlag	Text.Type	text	*18484*
NumberCarsOwned	Int64.Type	number	*18484*
AddressLine1	Text.Type	text	*18484*
AddressLine2	Text.Type	text	*18484*
DateFirstPurchase	Date.Type	date	*18484*
Marital Label	Text.Type	text	*18484*

8. Apply the following transformations to Sales query
    a. Keep the following columns
        i. Product Key

Challenge #1 – Prepare the data

ii. Order Date
iii. Due Date
iv. Ship Date
v. Customer Key
vi. Sales Order Number
vii. Order Quantity
viii. Product Standard Cost
ix. Tax Amount
x. Sales Amount

Name	TypeName	Kind	Row Count
ProductKey	Int64.Type	number	60398
CustomerKey	Int64.Type	number	60398
SalesOrderNumber	Text.Type	text	60398
OrderQuantity	Int64.Type	number	60398
ProductStandardCost	Currency.Type	number	60398
SalesAmount	Currency.Type	number	60398
TaxAmt	Currency.Type	number	60398
OrderDate	Date.Type	date	60398
DueDate	Date.Type	date	60398
ShipDate	Date.Type	date	60398

9. Create a Date query
   i. Full Date [Min(OrderDate) … Max(OrderDate,DueDate, ShipDate)]
   ii. Day Of Week
   iii. Day Name
   iv. Day
   v. Day Of Year
   vi. Week Of Year
   vii. Month Name
   viii. Month
   ix. Quarter of Year
   x. Year

Name	TypeName	Kind	Row Count
Full Date	Date.Type	date	1139
Day of Week	Int64.Type	number	1139
Day Name	Text.Type	text	1139
Day	Int64.Type	number	1139
Day of Year	Int64.Type	number	1139
Week of Year	Int64.Type	number	1139
Month Name	Text.Type	text	1139
Month	Int64.Type	number	1139
Quarter	Int64.Type	number	1139
Year	Int64.Type	number	1139

Challenge #1 – Prepare the data

10. Apply the following transformations to Geography query
    a. The following columns should be removed.
        i. StateProvinceCode
        ii. CountryRegionCode
        iii. SpanishCountryRegionName
        iv. FrenchCountryRegionName
        v. SalesTerritoryKey
        vi. IpAddressLocator

Name	TypeName	Kind	Row Count
GeographyKey	Int64.Type	number	655
City	Text.Type	text	655
State	Text.Type	text	655
Country	Text.Type	text	655
PostalCode	Text.Type	text	655

11. Apply the following transformations to Product Category query
    a. The following columns should be removed.
        i. ProductCategoryAlternateKey
        ii. SpanishProductCategoryName
        iii. FrenchProductCategoryName
    b. Rename columns
        i. EnglishProductCategoryName to Product Category Name

Name	TypeName	Kind	Row Count
ProductCategoryKey	Int64.Type	number	4
Product Category Name	Text.Type	text	4

12. Apply the following transformations to Product Subcategory query
    a. The following columns should be removed.
        i. ProductSubcategoryAlternateKey
        ii. SpanishProductSubcategoryName
        iii. FrenchProductSubcategoryName
    b. Rename columns
        i. EnglishProductSubcategoryName to Product Subcategory Name

Name	TypeName	Kind	Row Count
ProductSubcategoryKey	Int64.Type	number	37
ProductCategoryKey	Int64.Type	number	37
Product Subcategory Name	Text.Type	text	37

Challenge #1 – Prepare the data

13. Apply the following transformations to Product query
    a. Retain the following columns
        i. ProductKey
        ii. ProductSubcategoryKey
        iii. EnglishProductName
        iv. StandardCost
        v. FinishedGoodsFlag
        vi. Color
        vii. ListPrice
        viii. Size
        ix. Weight
        x. Style
        xi. ModelName
        xii. EnglishDescription
    b. Rename columns
        i. EnglishProductName to Product Name
        ii. EnglishDescription to Description
    c. Only include finished products and remove FinishedGoodsFlag column.
    d. Replace values
        i. Replace empty values in Standard Cost with its average value
        ii. Replace empty values in List Price with its average value
        iii. Replace empty values in Size with 38
        iv. Replace empty values in Weight with 56
        v. Replace empty values in Style with U
        vi. Replace U with Unisex, W with Women and M with Men in Style column using custom functions

Name	TypeName	Kind	Row Count
ProductKey	Int64.Type	number	397
ProductSubcategoryKey	Int64.Type	number	397
Product Name	Text.Type	text	397
StandardCost	Currency.Type	number	397
Color	Text.Type	text	397
ListPrice	Currency.Type	number	397
Size	Text.Type	text	397
Weight	Number.Type	number	397
Style	Text.Type	text	397
ModelName	Text.Type	text	397
Description	Text.Type	text	397

14. Import [Target.txt] and apply the necessary transformations to match the specified output. Rename the query to Target.

Challenge #1 – Prepare the data

	Calendar Year	Category Name	Target
1	2012	Accessories	13,00,000.00
2	2012	Bikes	12,00,000.00
3	2012	Clothing	12,00,000.00
4	2013	Accessories	16,00,000.00
5	2013	Bikes	15,00,000.00
6	2013	Clothing	15,00,000.00
7	2014	Accessories	18,00,000.00
8	2014	Bikes	16,00,000.00
9	2014	Clothing	16,00,000.00
10	2015	Accessories	21,00,000.00
11	2015	Bikes	17,00,000.00
12	2015	Clothing	17,00,000.00

15. Import [Target20102011.xlsx] and transform to match the output below. Rename the query to Target 2010 – 2011.

	Calendar Year	Category Name	Target
1	2010	Accessories	20,80,000.00
2	2010	Bikes	16,50,000.00
3	2010	Clothing	16,60,000.00
4	2011	Accessories	19,70,000.00
5	2011	Bikes	14,90,000.00
6	2011	Clothing	15,20,000.00

16. Combine the Target 2010–2011 query with the Target query.

	Calendar Year	Category Name	Target
1	2010	Accessories	20,80,000.00
2	2010	Clothing	16,60,000.00
3	2010	Bikes	16,50,000.00
4	2011	Clothing	15,20,000.00
5	2011	Bikes	14,90,000.00
6	2011	Accessories	19,70,000.00
7	2012	Accessories	13,00,000.00
8	2012	Clothing	12,00,000.00
9	2012	Bikes	12,00,000.00
10	2013	Accessories	16,00,000.00
11	2013	Clothing	15,00,000.00
12	2013	Bikes	15,00,000.00
13	2014	Bikes	16,00,000.00
14	2014	Accessories	18,00,000.00
15	2014	Clothing	16,00,000.00
16	2015	Clothing	17,00,000.00
17	2015	Bikes	17,00,000.00
18	2015	Accessories	21,00,000.00

Challenge #1 – Prepare the data

17. Import [TargetQuantity.csv] and join it with Target on Calendar Year and Category Name.

	Calendar Year	Category Name	Target	Target Quantity
1	2010	Bikes	16,50,000.00	870000
2	2010	Clothing	16,60,000.00	820000
3	2010	Accessories	20,80,000.00	920000
4	2011	Bikes	14,90,000.00	860000
5	2011	Clothing	15,20,000.00	910000
6	2011	Accessories	19,70,000.00	1070000
7	2012	Clothing	12,00,000.00	820000
8	2012	Bikes	12,00,000.00	870000
9	2012	Accessories	13,00,000.00	920000
10	2013	Clothing	15,00,000.00	750000
11	2013	Bikes	15,00,000.00	800000
12	2013	Accessories	16,00,000.00	840000
13	2014	Accessories	18,00,000.00	920000
14	2014	Clothing	16,00,000.00	820000
15	2014	Bikes	16,00,000.00	870000
16	2015	Bikes	17,00,000.00	860000
17	2015	Clothing	17,00,000.00	910000
18	2015	Accessories	21,00,000.00	1070000