Global Superstore

The growth of supermarkets in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of supermarket company which has recorded in many different branches for months data. Predictive data analytics methods are easy to apply with this dataset.

A random sample was selected from the data between 2012 and 2015. The data reflects supermarket data in several countries around the world.

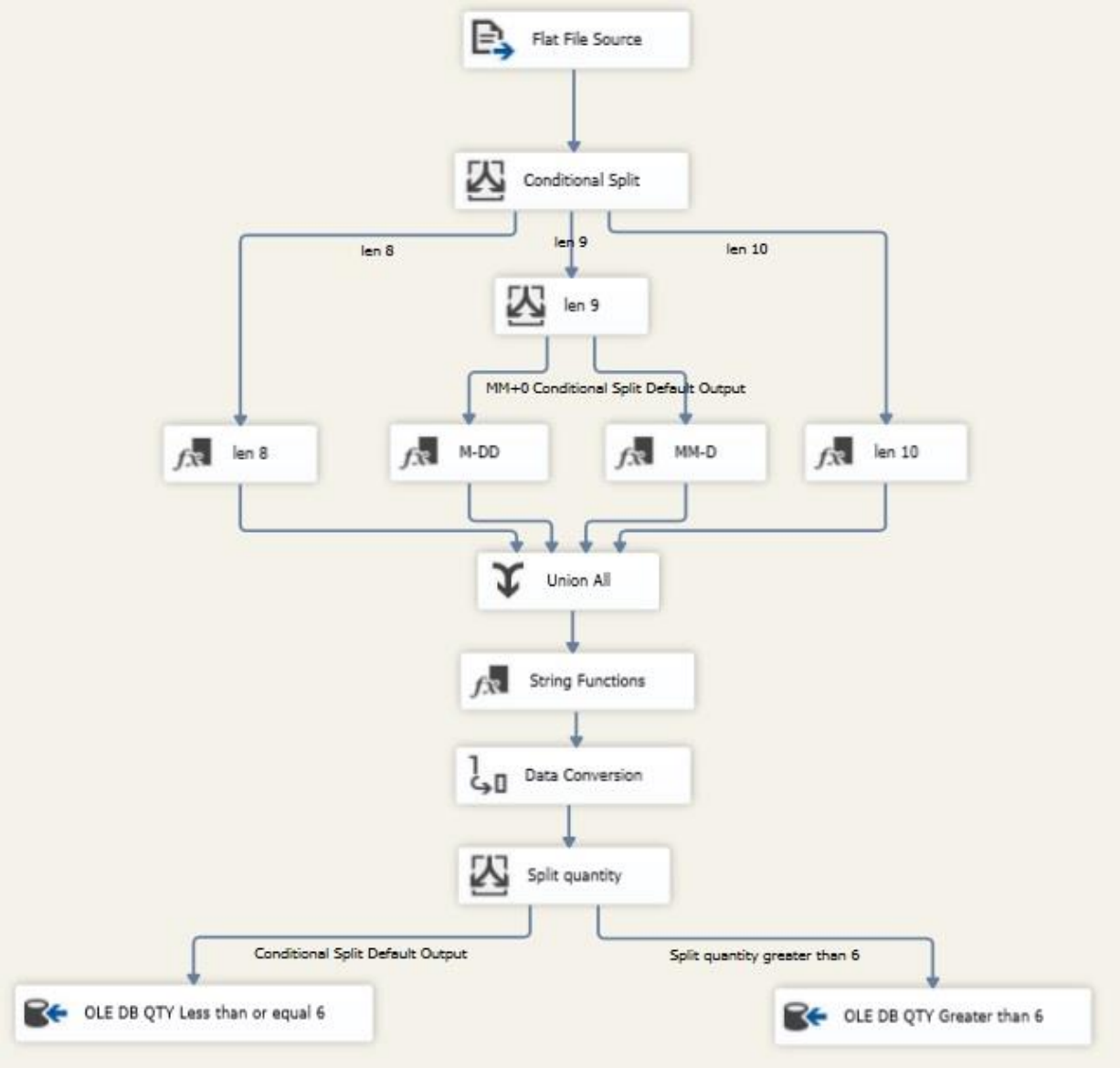
The main columns in the dataset file are:

- Row ID
- Order ID
- Order Date
- Ship Date
- Ship Mode
- Customer ID
- Customer Name
- Segment
- Postal Code
- City
- State
- Country
- Region
- Market
- Product ID
- Category
- Sub-Category
- Product Name
- Sales Quantity
- Discount
- Profit
- Shipping Cost
- Order Priority

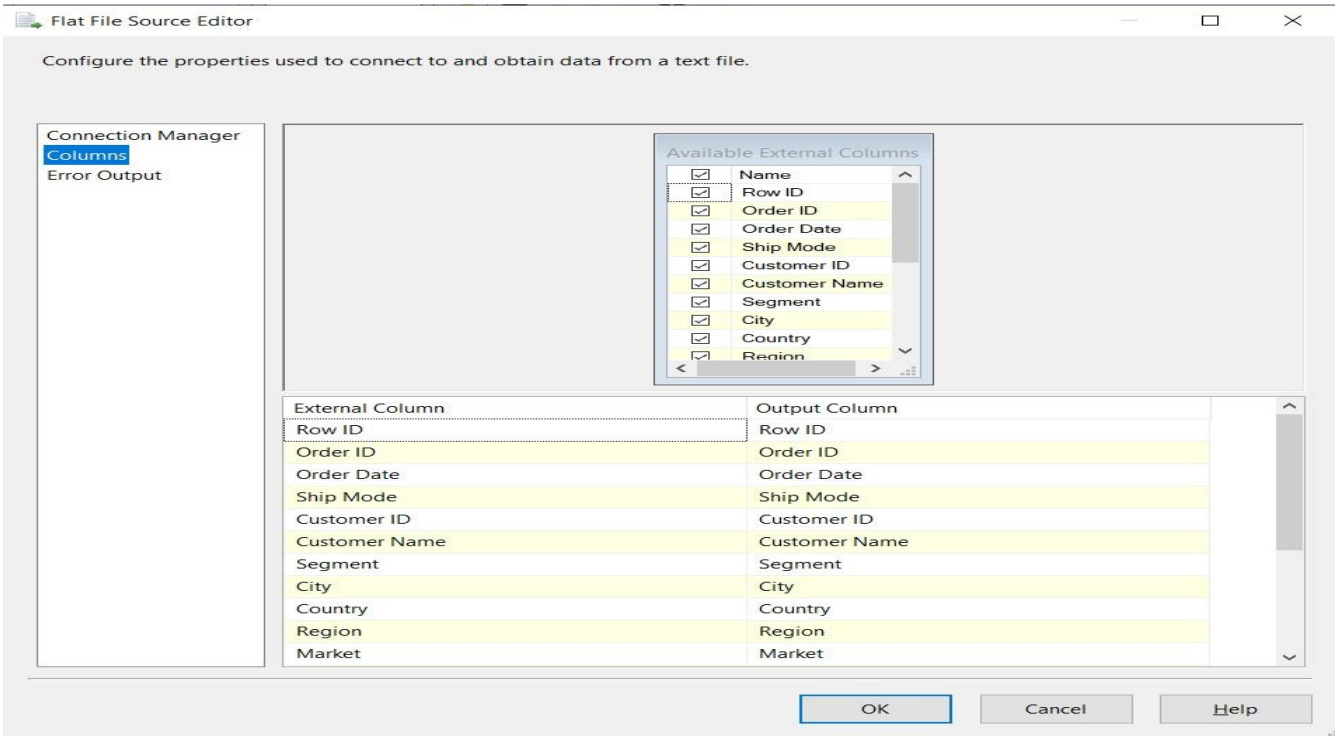
A screenshot from Dataset:

Flow of components

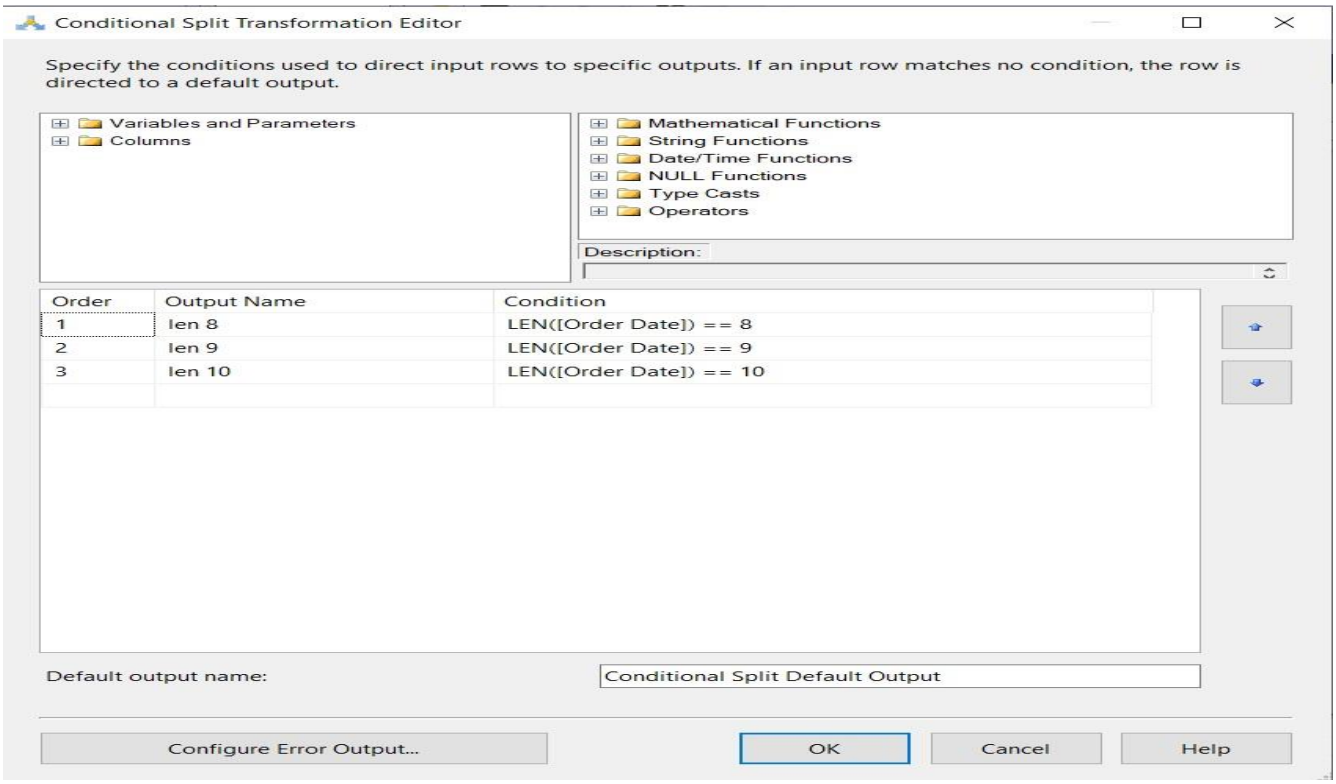
1- Flat File:



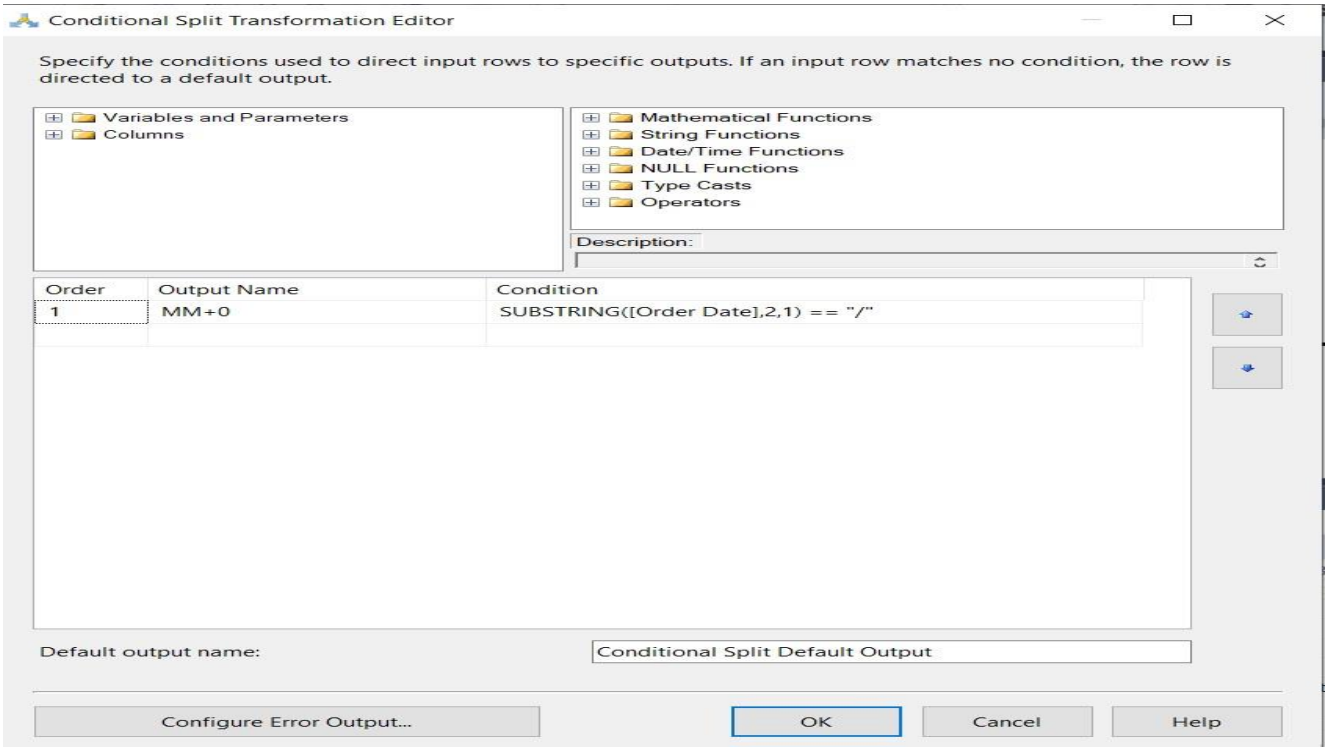
Flat File source :



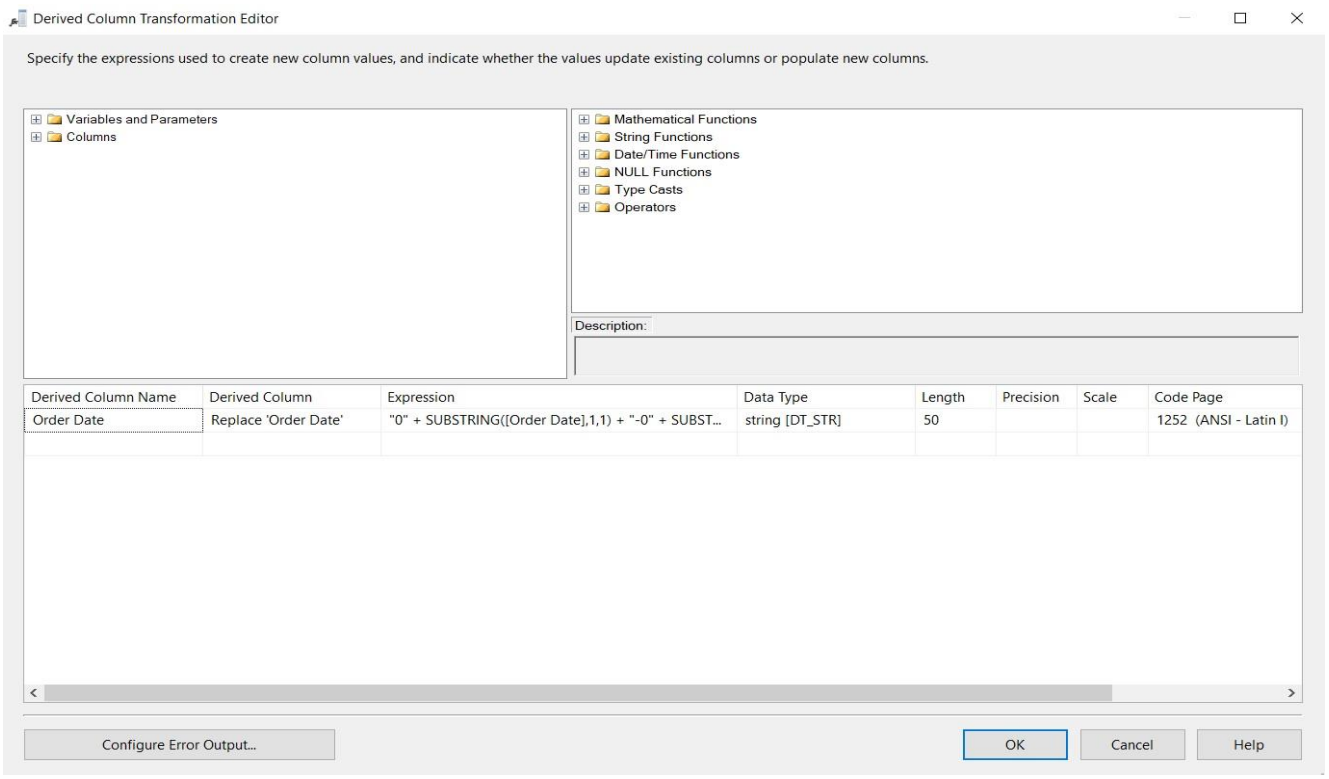
Conditional split :



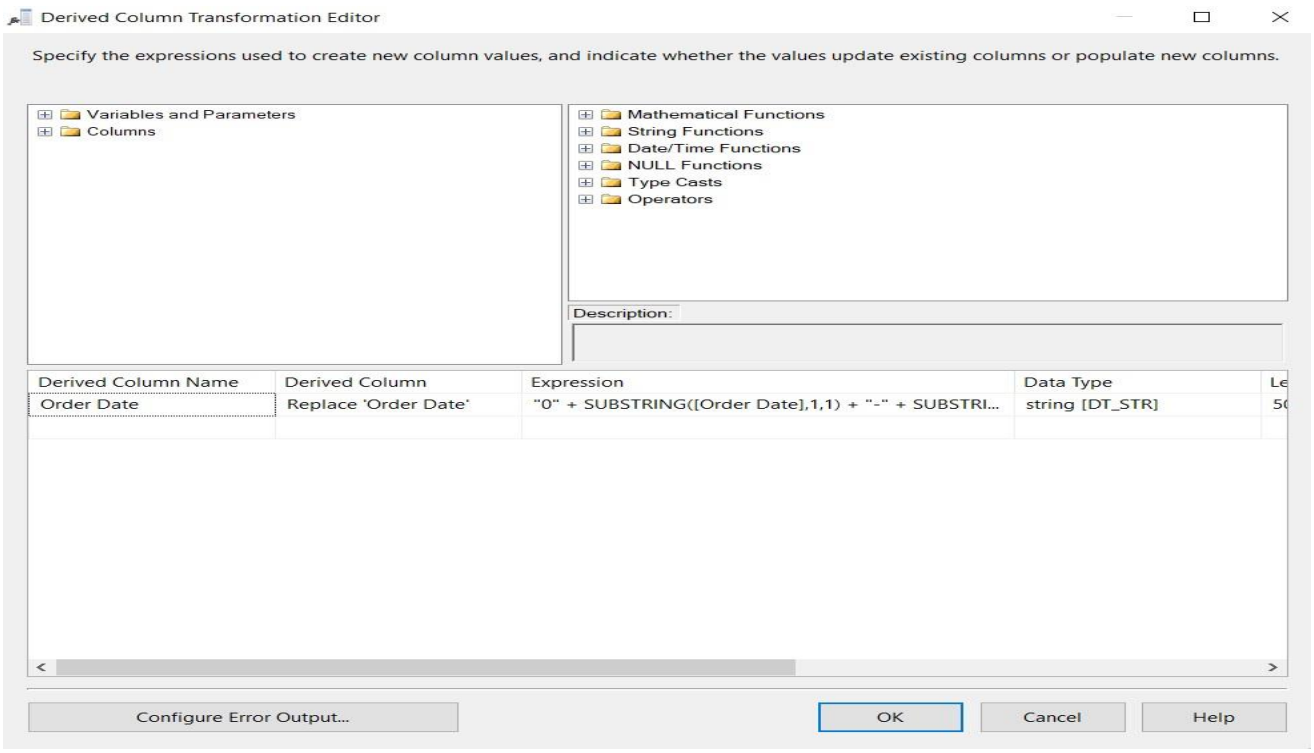
Len 9:


Conditional Split Transformation Editor
Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Order
Output Name
Condition
1
MM+0
SUBSTRING([Order Date],2,1) == "/"
Default output name:
Conditional Split Default Output
Configure Error Output...
OK
Cancel
Help

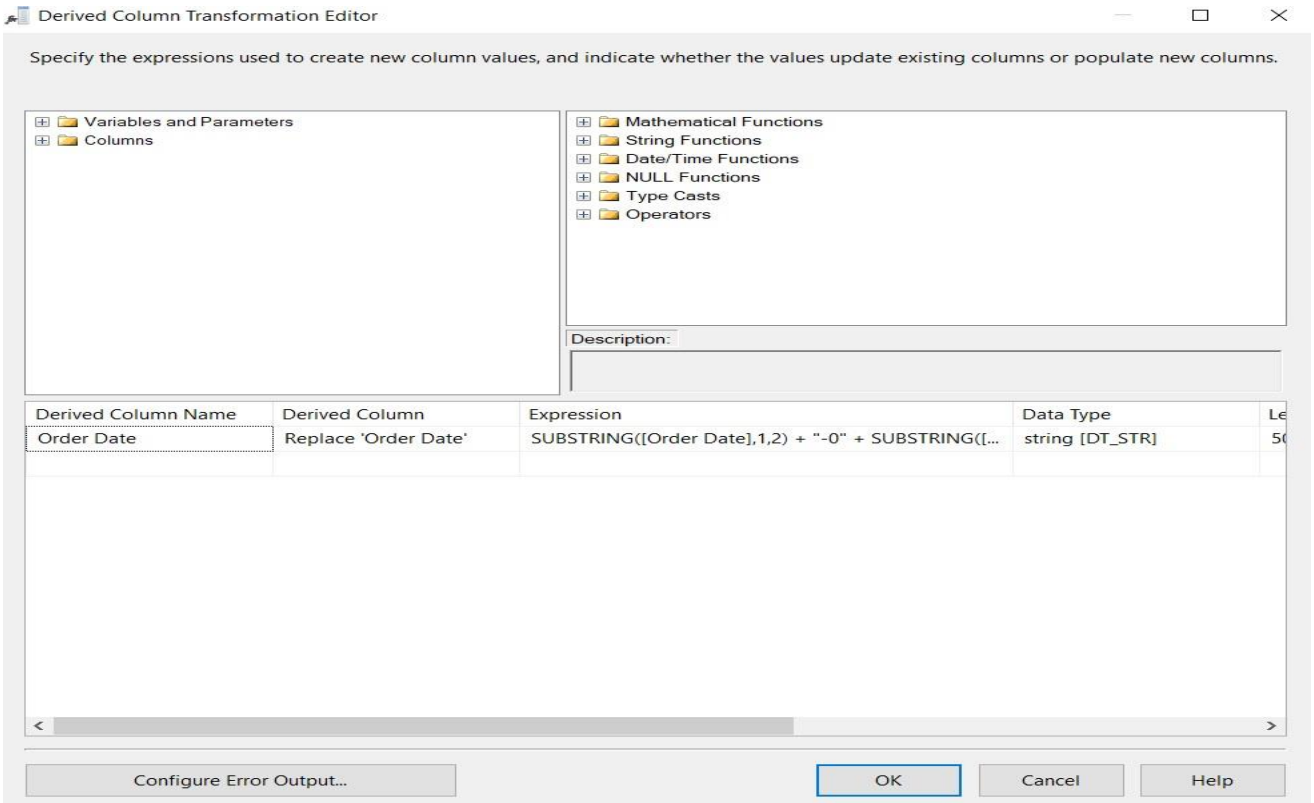
Len 8:


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
Length
Precision
Scale
Code Page
Order Date
Replace 'Order Date'
"0" + SUBSTRING([Order Date],1,1) + "-0" + SUBST...
string [DT_STR]
50
1252 (ANSI - Latin I)
Configure Error Output...
OK
Cancel
Help

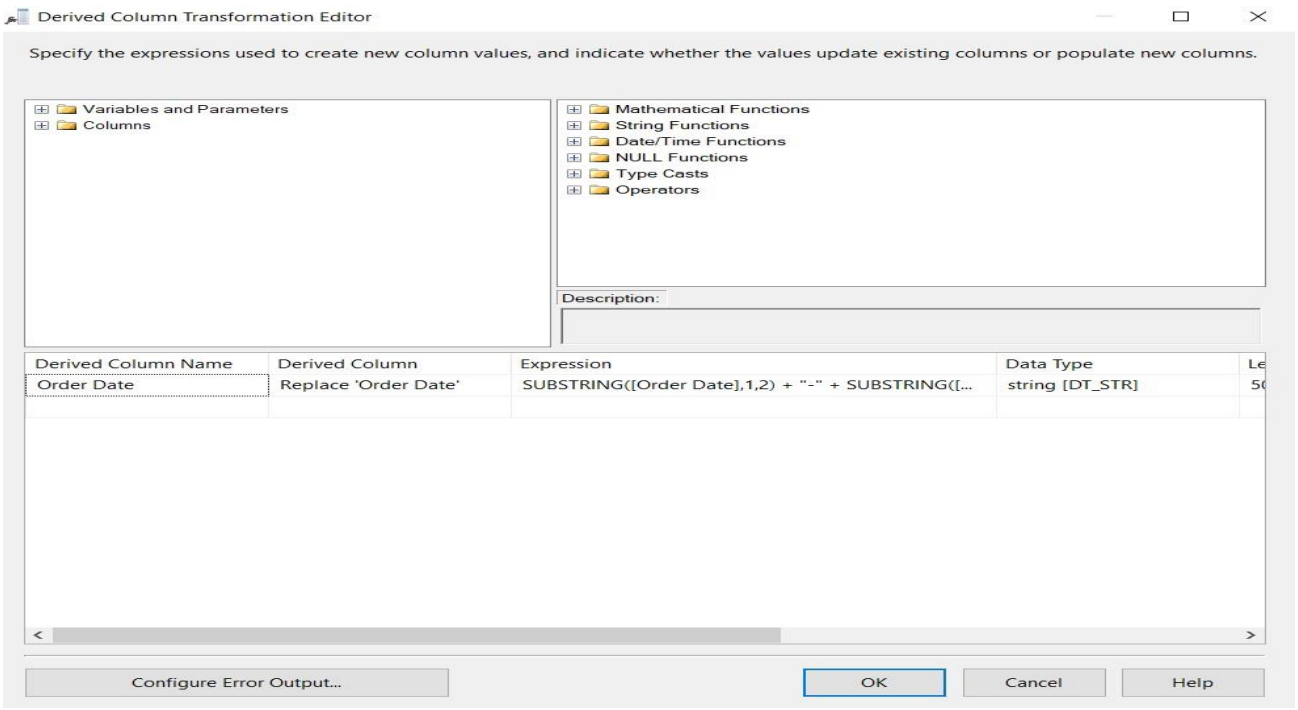
M-DD :


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
Order Date
Replace 'Order Date'
"0" + SUBSTRING([Order Date],1,1) + "-" + SUBSTRI...
string [DT_STR]
Configure Error Output...
OK
Cancel
Help

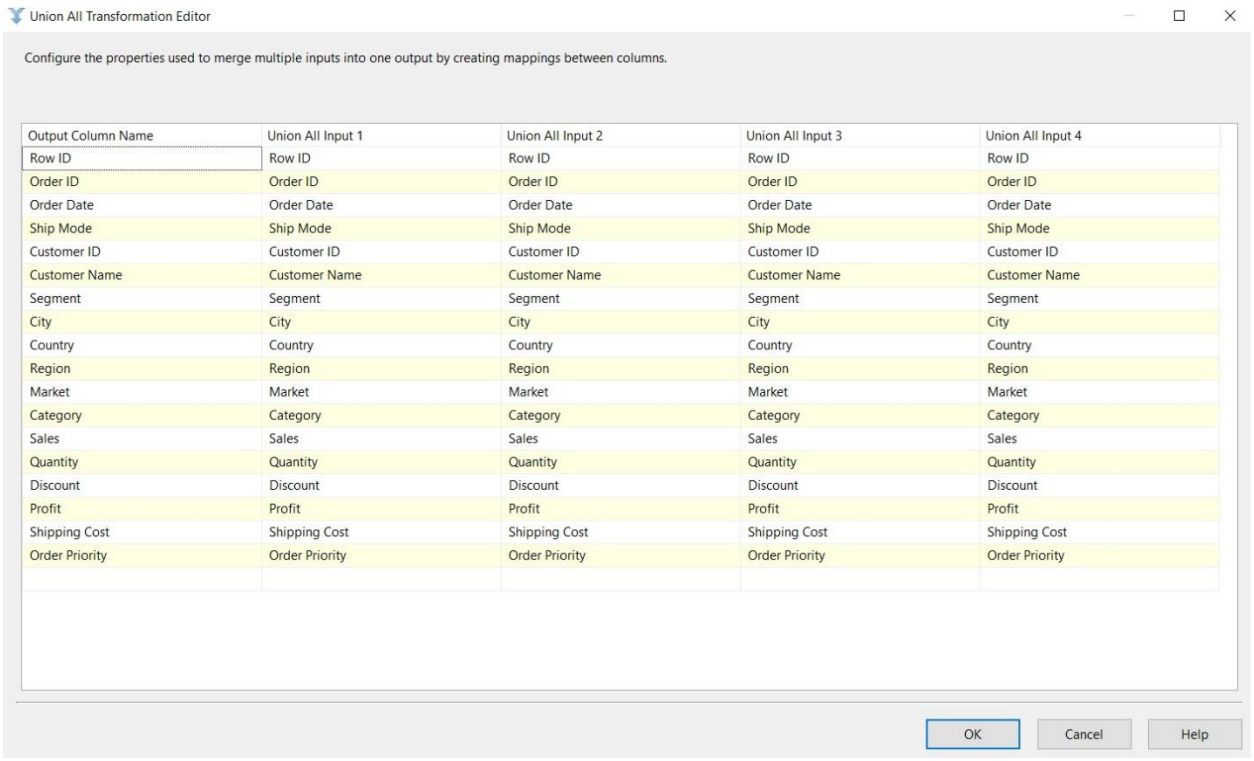
MM-D:


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
Order Date
Replace 'Order Date'
SUBSTRING([Order Date],1,2) + "-0" + SUBSTRING([...
string [DT_STR]
Configure Error Output...
OK
Cancel
Help

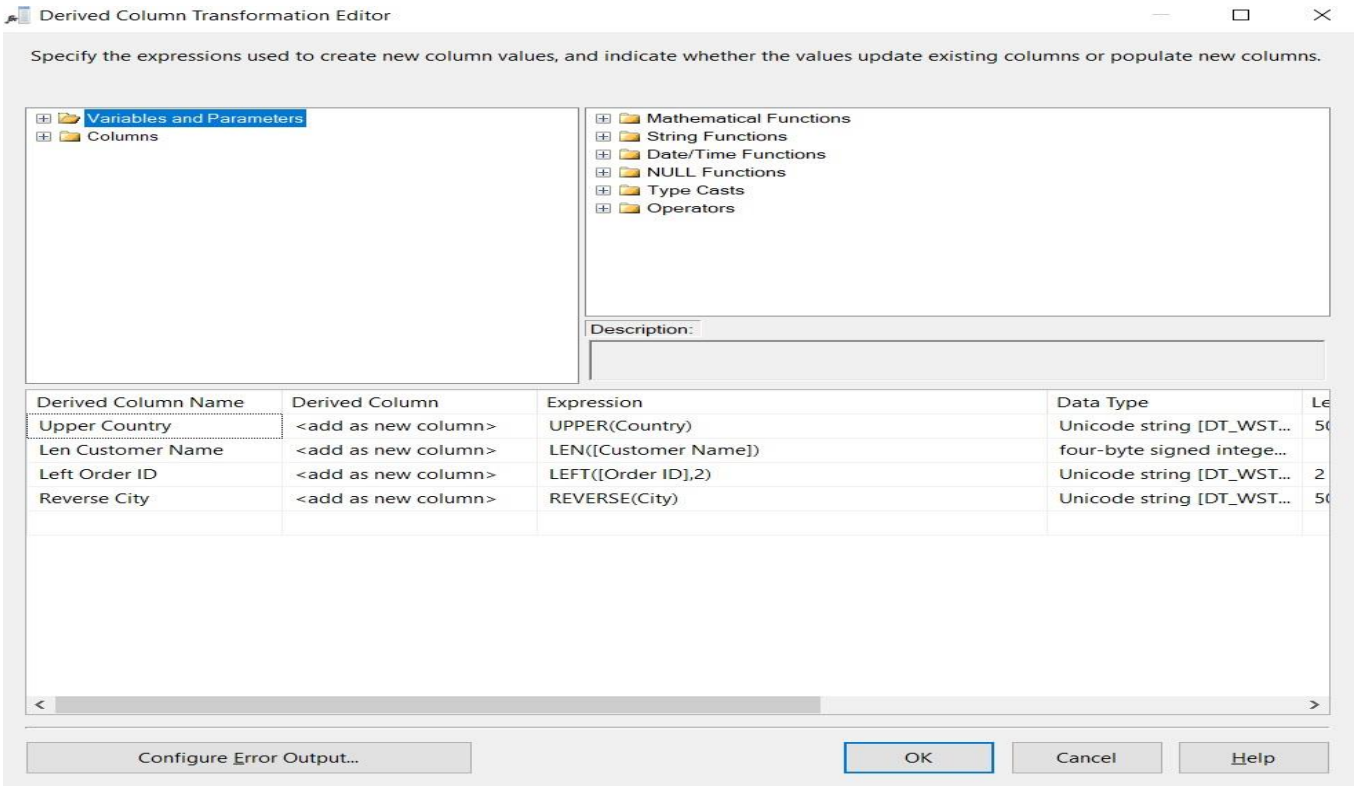
Len 10:



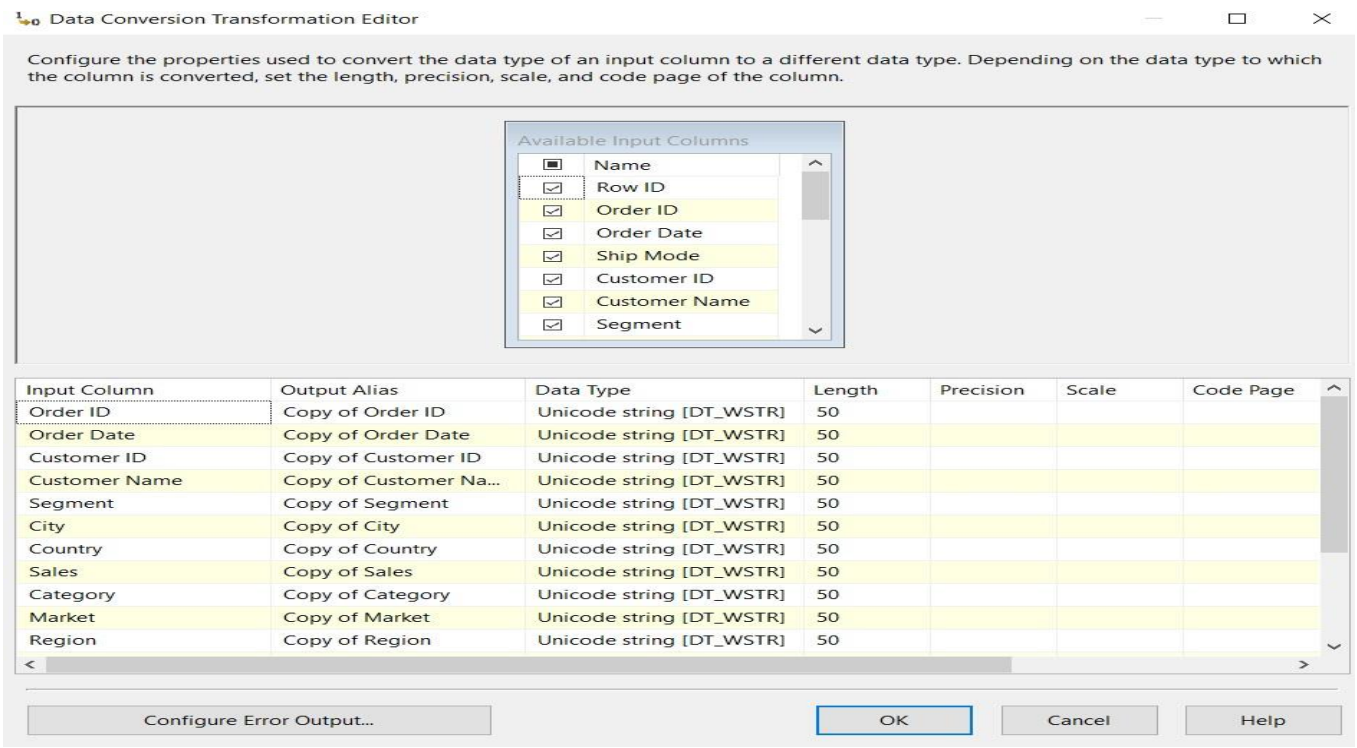
Union all:



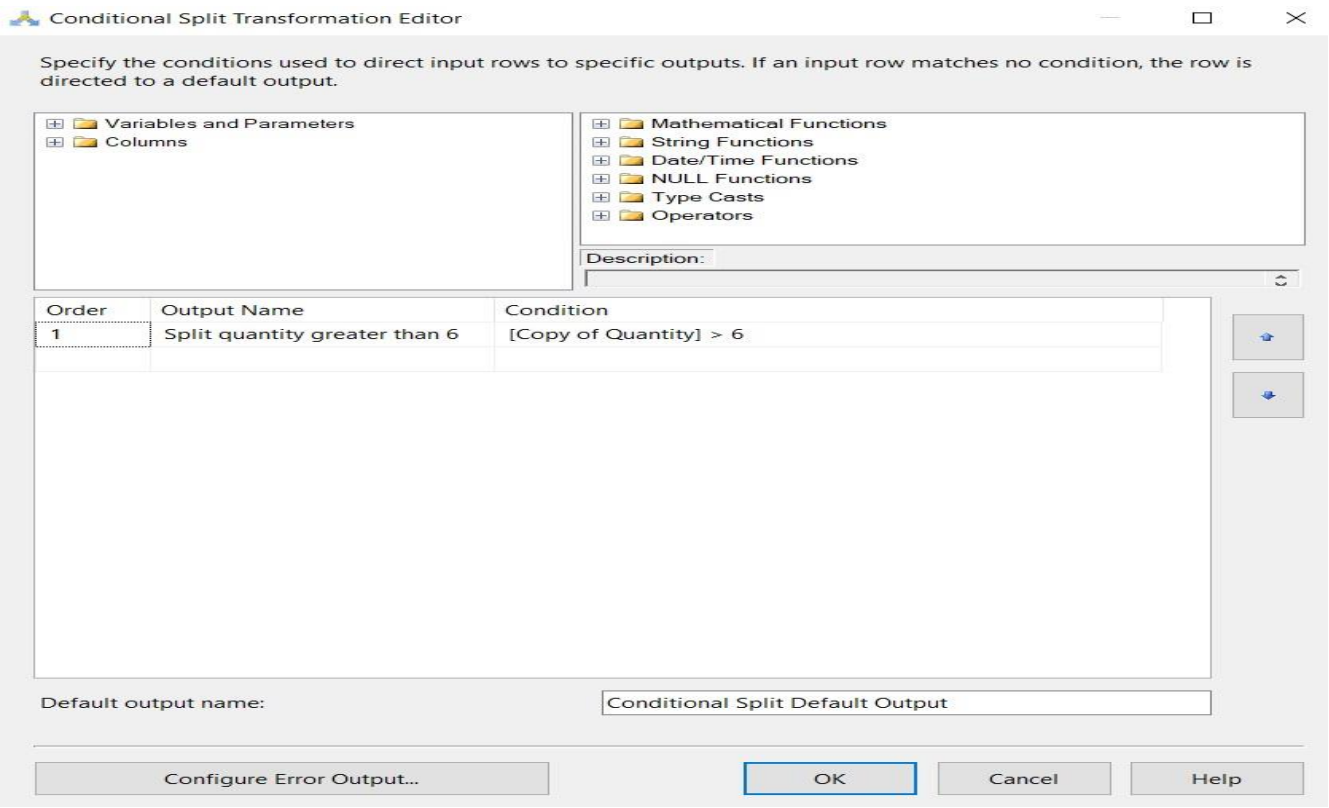
String Functions:



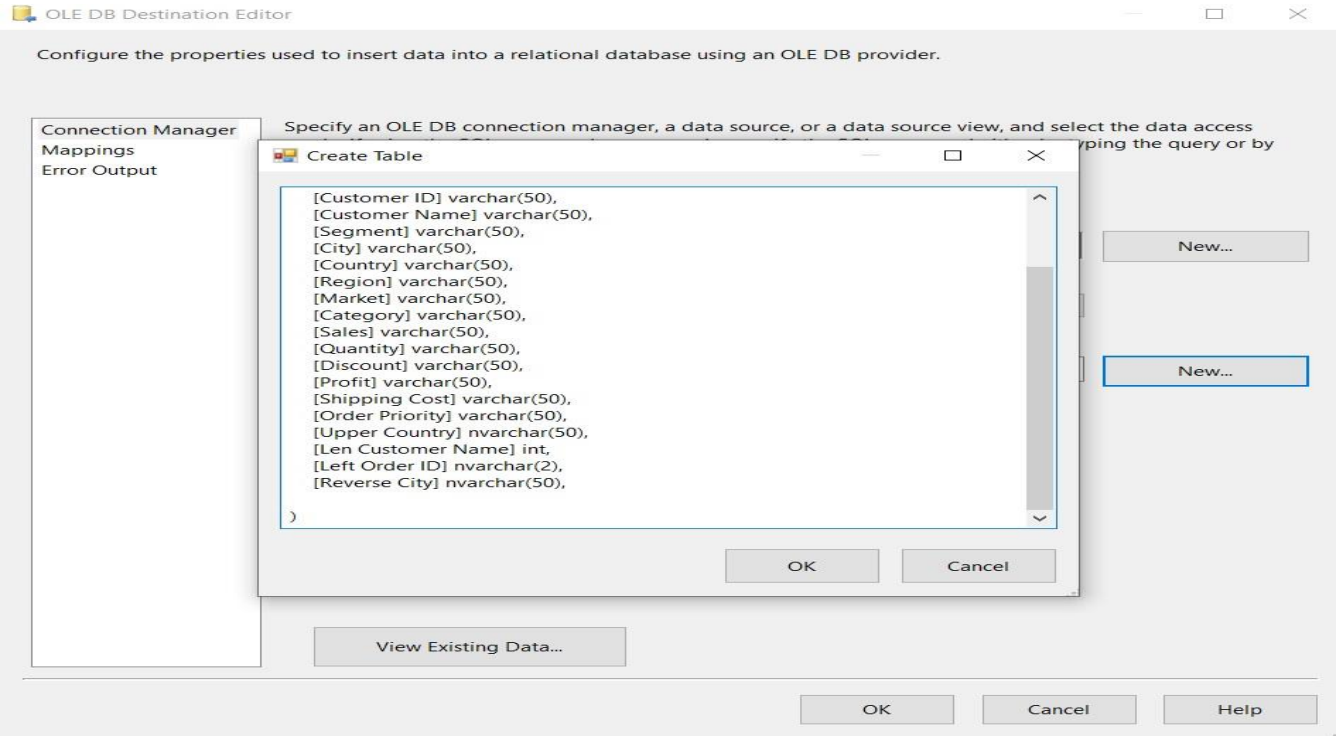
Data Conversion:

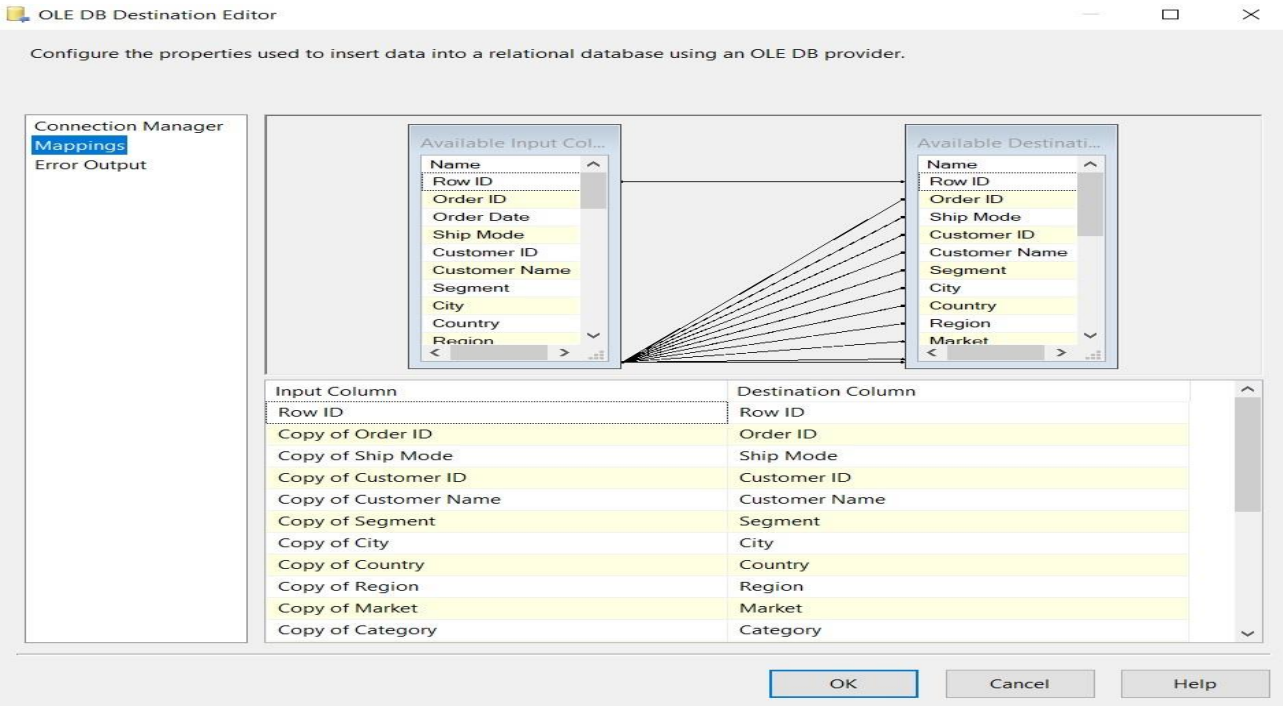


Split Quantity:

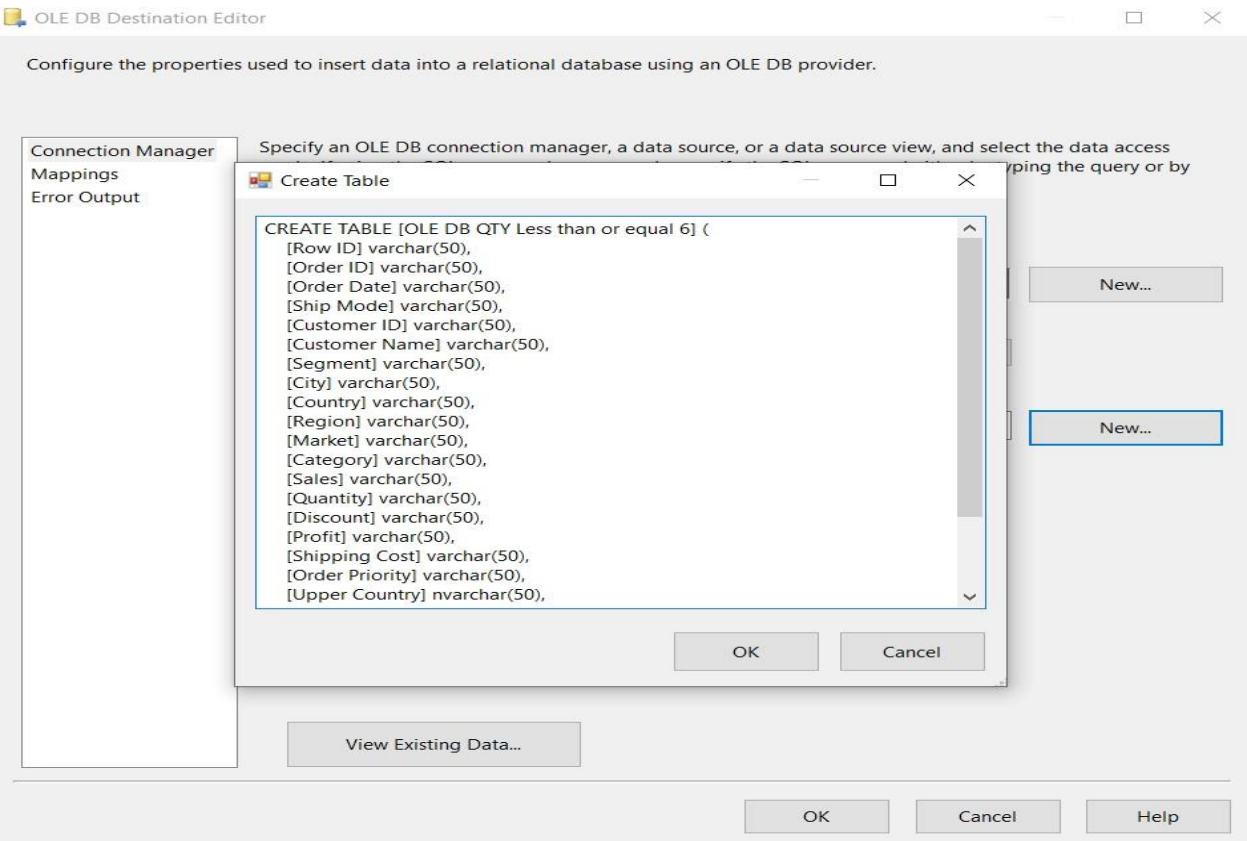


OLE DB QTY Greater than 6:





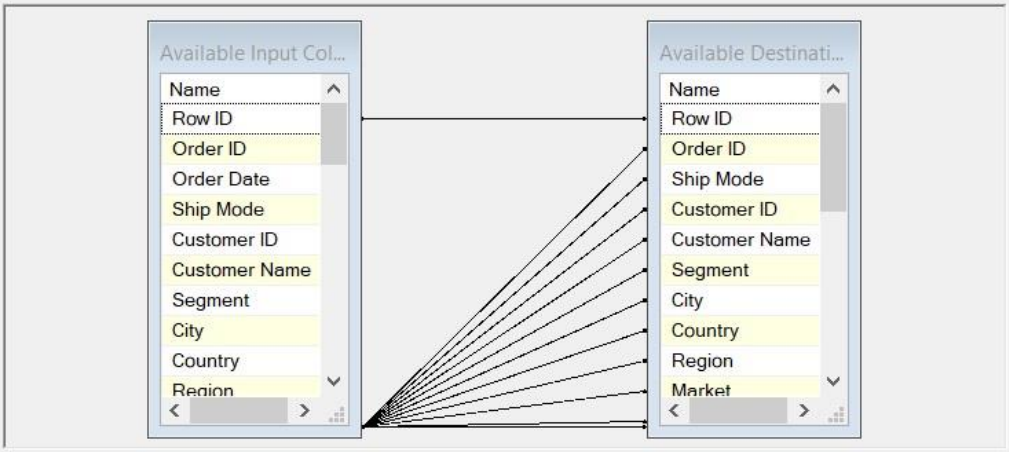
OLE DB QTY Less than or equal 6:



OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

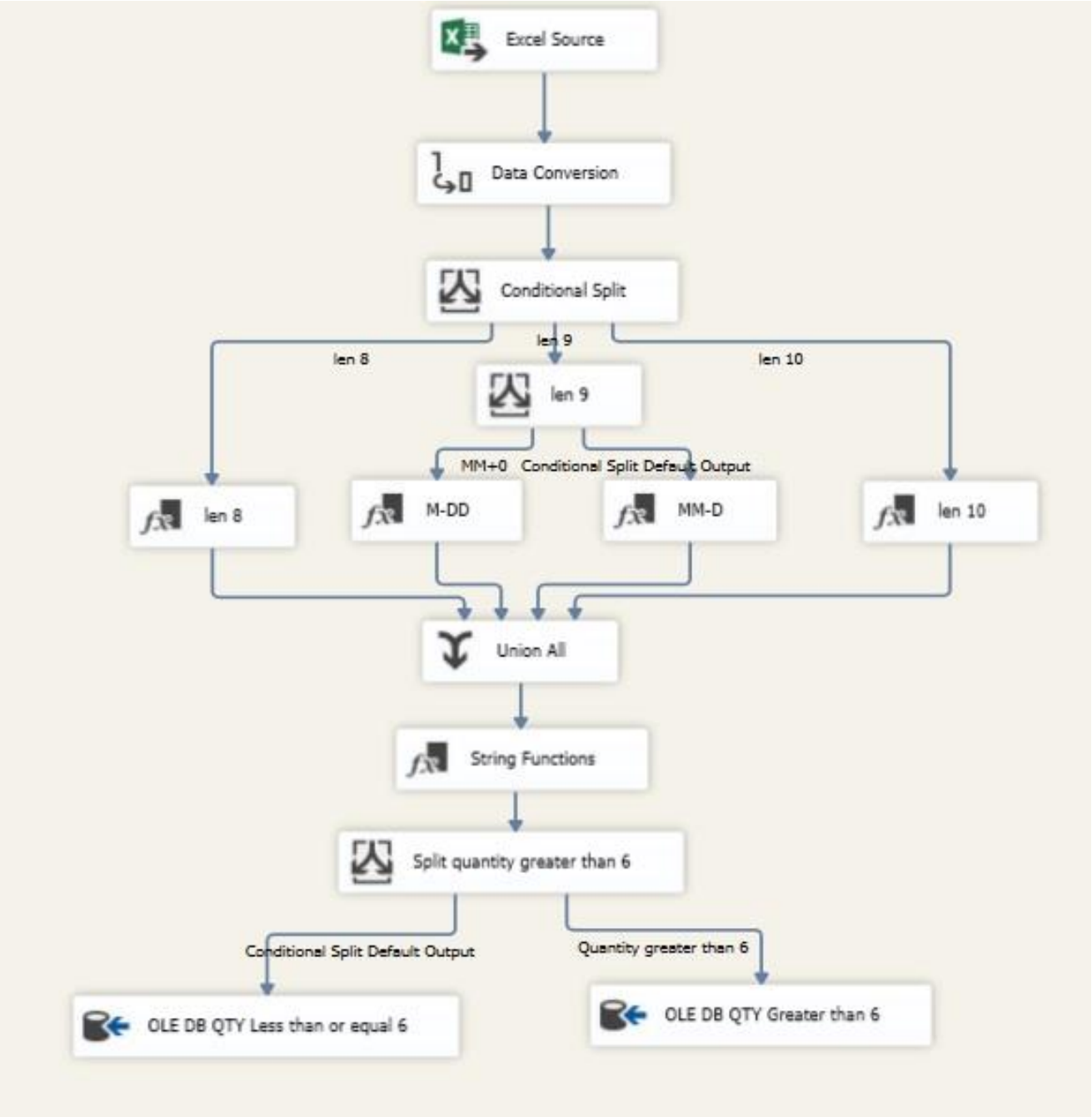
- Connection Manager
- Mappings
- Error Output



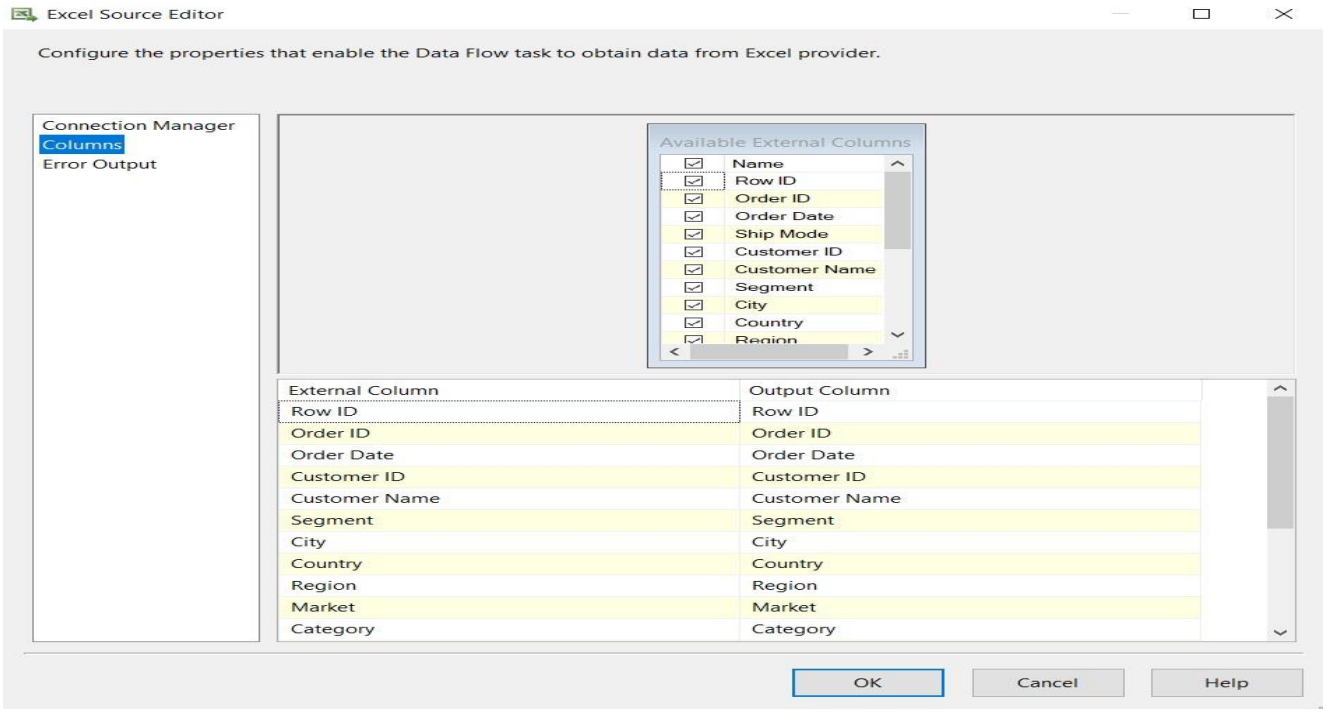
Input Column	Destination Column
Row ID	Row ID
Copy of Order ID	Order ID
Copy of Ship Mode	Ship Mode
Copy of Customer ID	Customer ID
Copy of Customer Name	Customer Name
Copy of Segment	Segment
Copy of City	City
Copy of Country	Country
Copy of Region	Region
Copy of Market	Market
Copy of Category	Category

OK Cancel Help

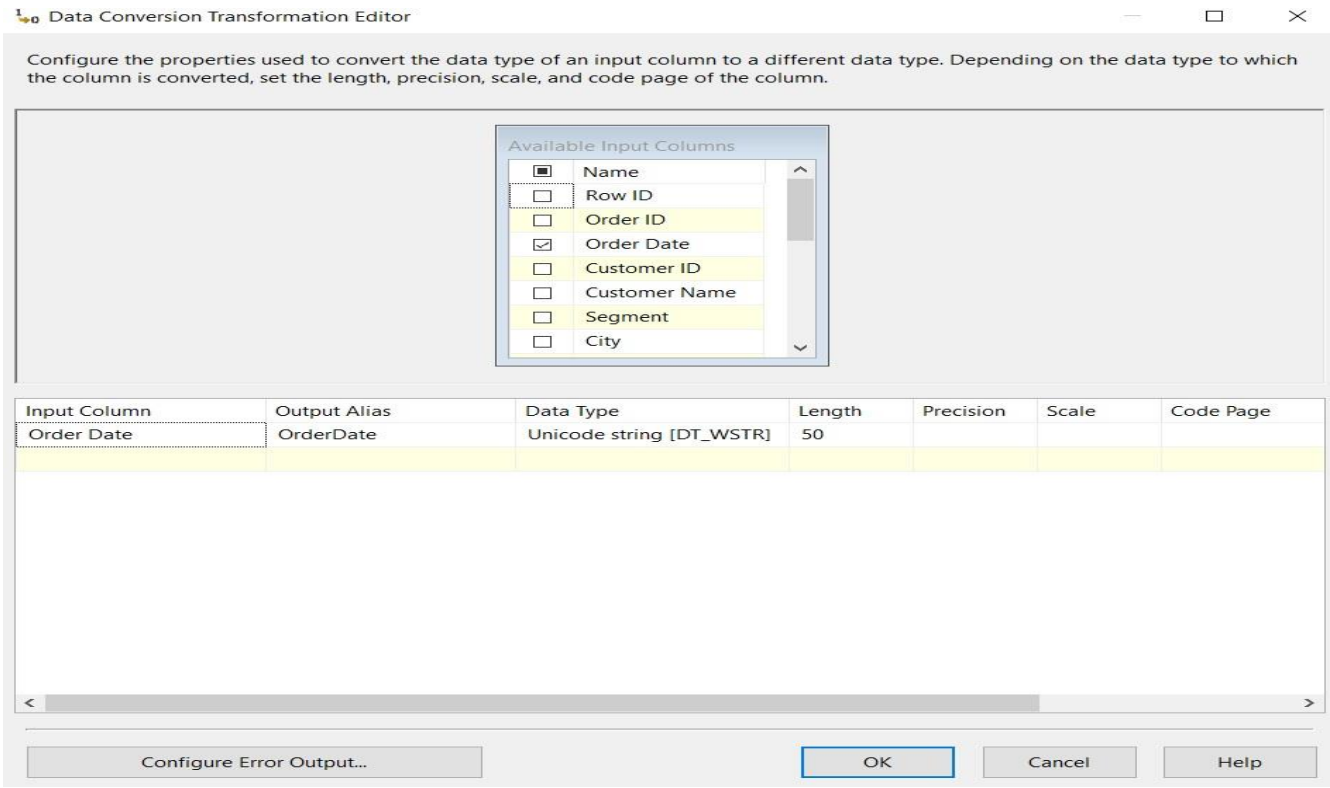
2- Excel :


Excel Source
Data Conversion
Conditional Split
len 8
len 9
len 10
len 9
MM+0
Conditional Split Default Output
len 8
M-DD
MM-D
len 10
Union All
String Functions
Split quantity greater than 6
Conditional Split Default Output
Quantity greater than 6
OLE DB QTY Less than or equal 6
OLE DB QTY Greater than 6

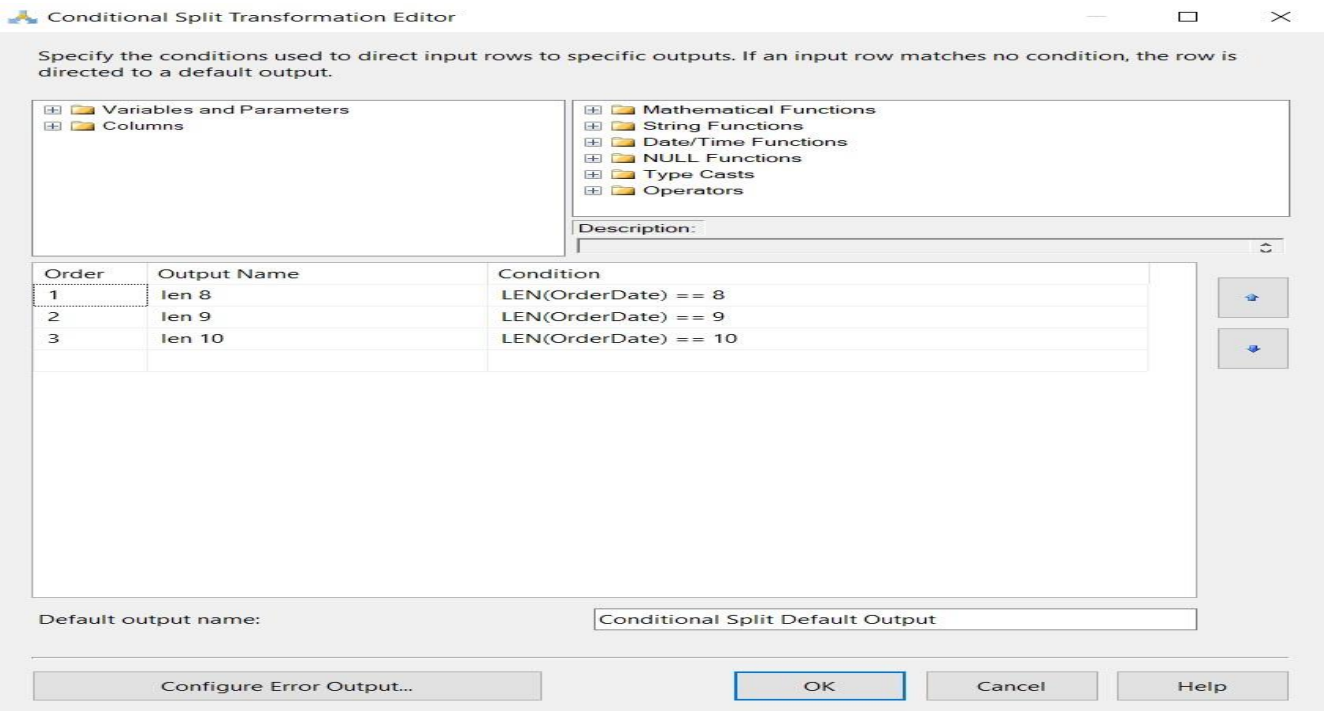
Excel Source:



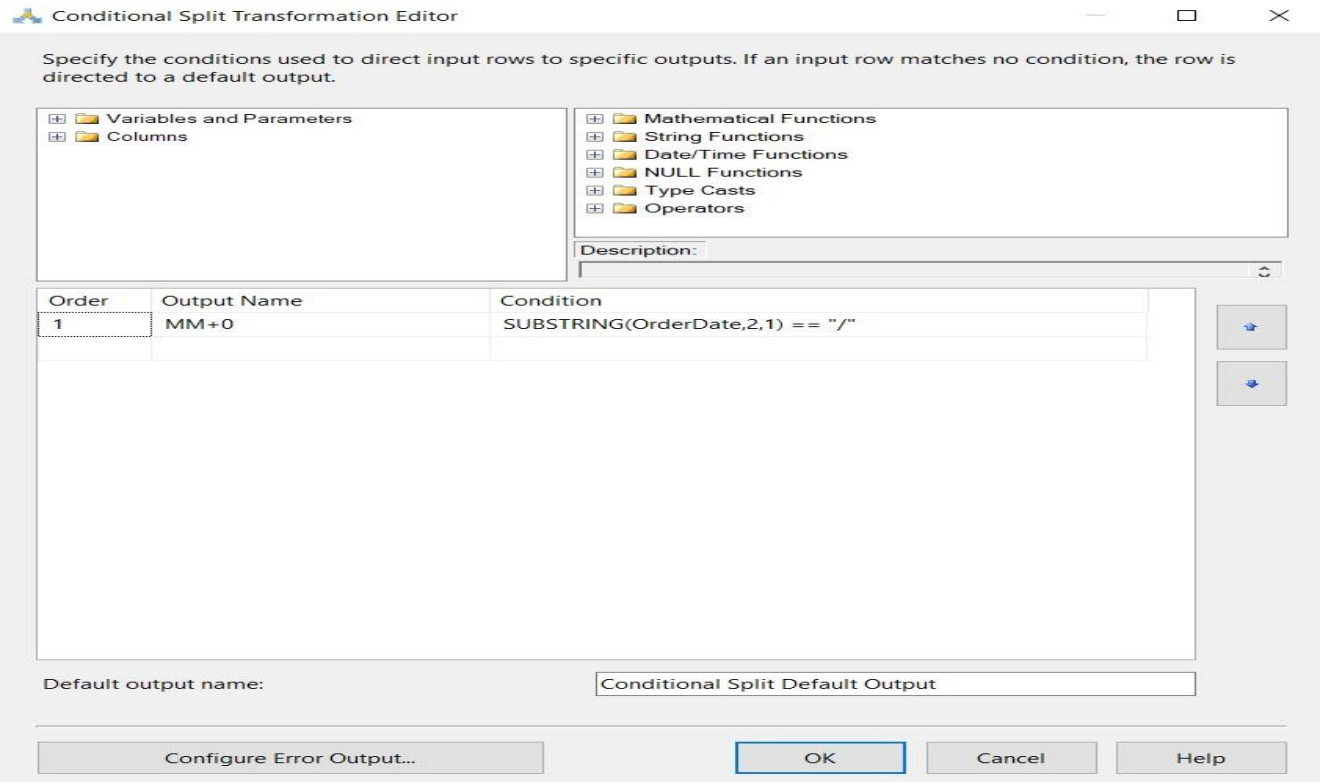
Data Conversion:



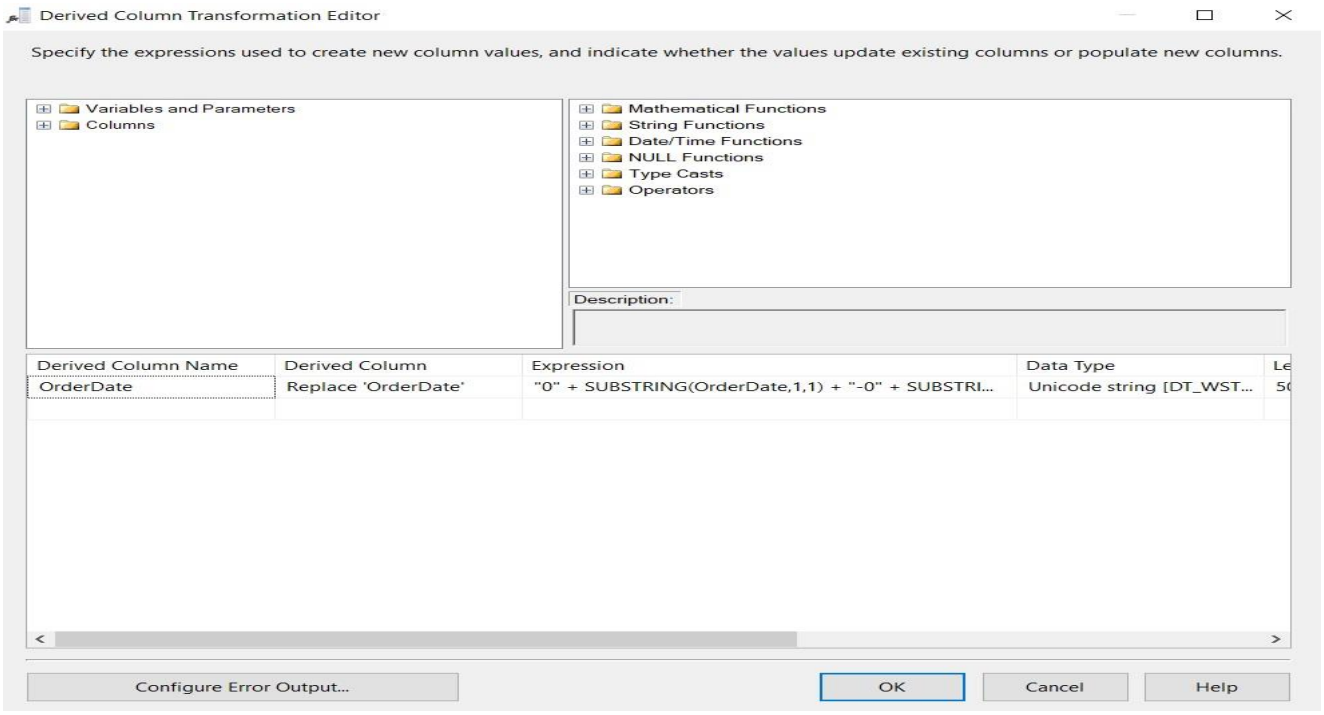
Conditional Split:



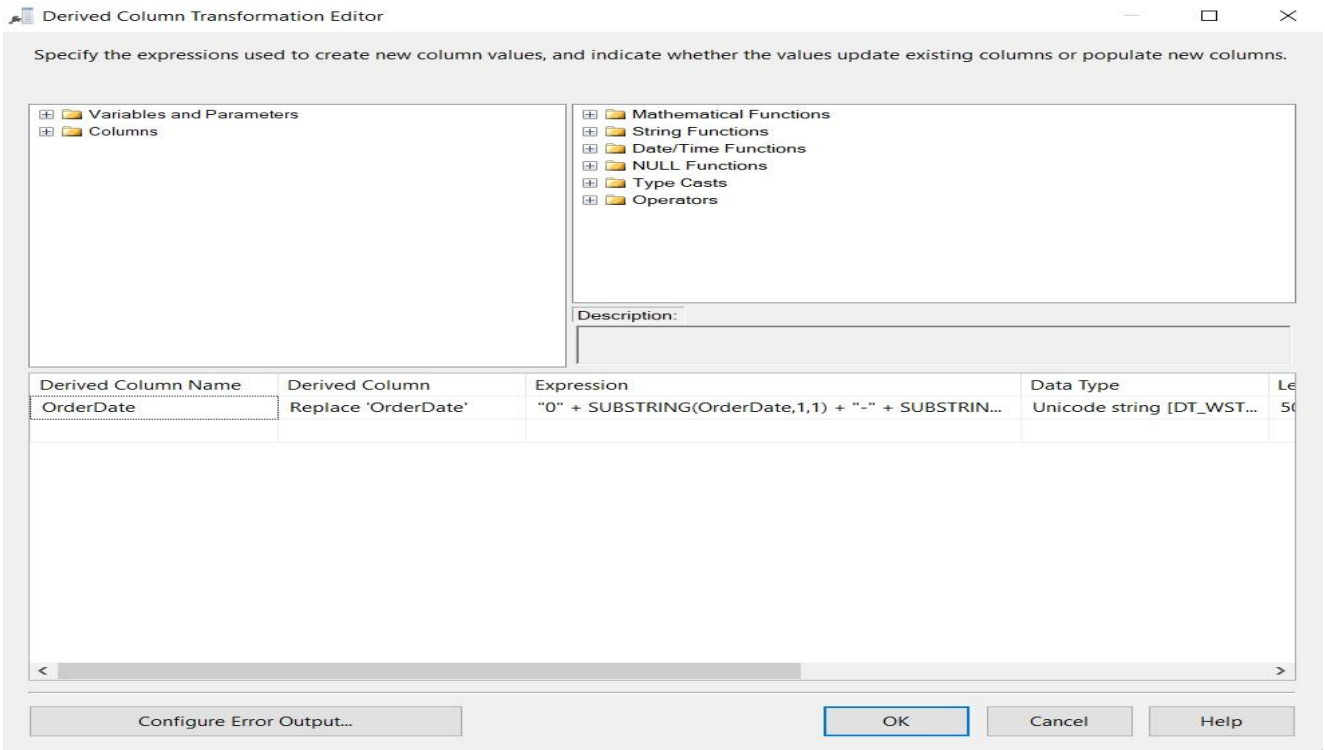
Len 9:



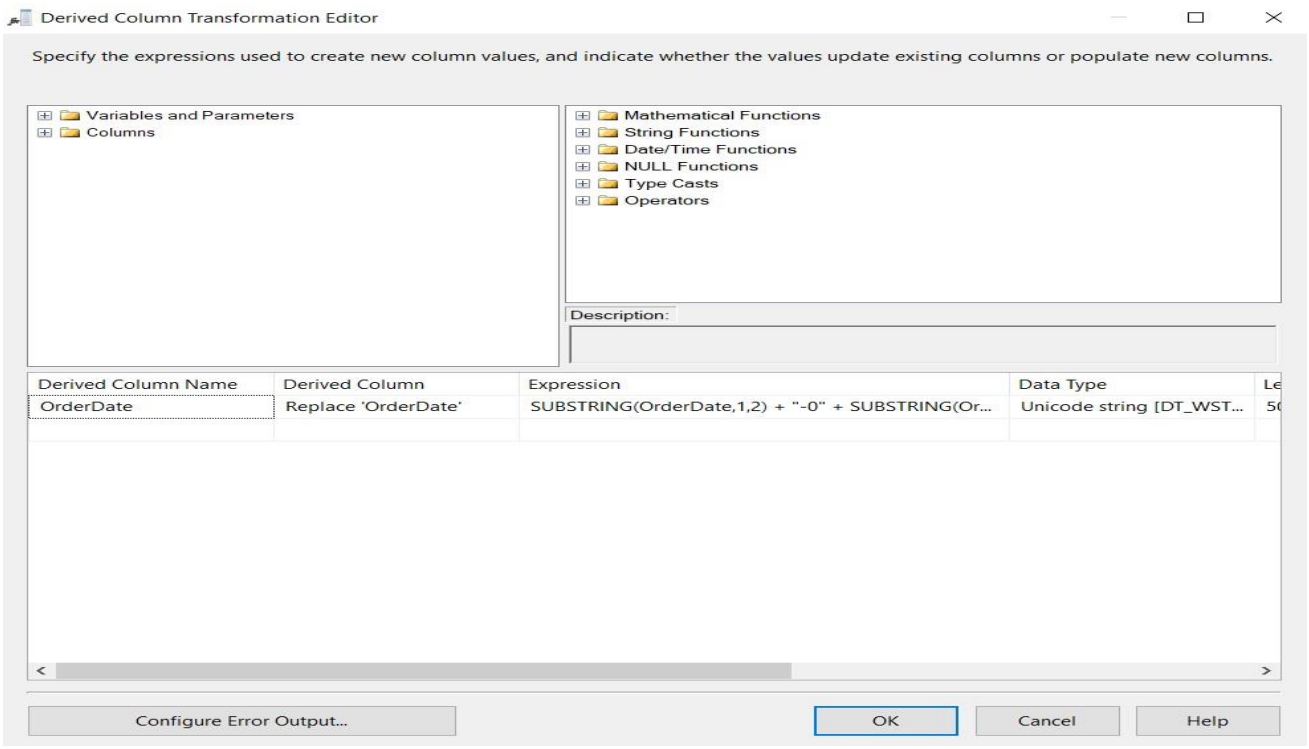
Len 8:


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
OrderDate
Replace 'OrderDate'
"0" + SUBSTRING(OrderDate,1,1) + "-0" + SUBSTRI...
Unicode string [DT_WST...
Configure Error Output...
OK
Cancel
Help

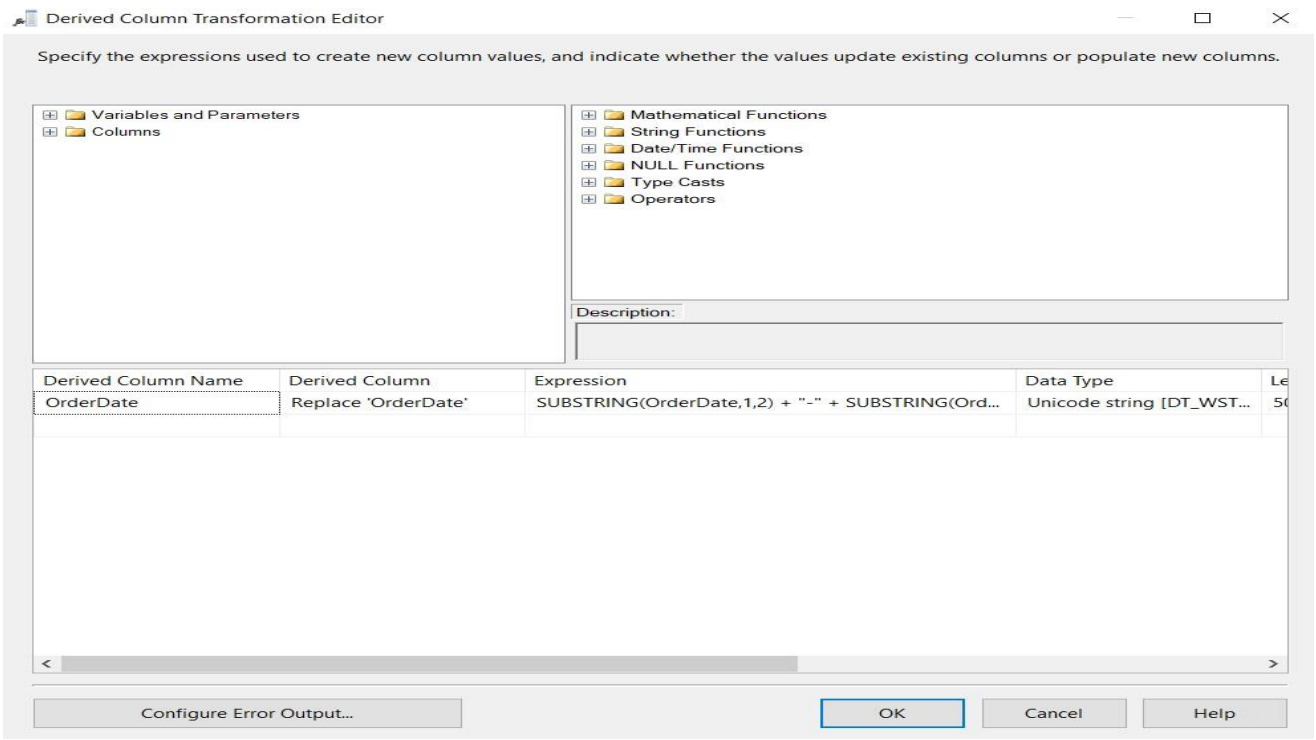
M-DD:


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
OrderDate
Replace 'OrderDate'
"0" + SUBSTRING(OrderDate,1,1) + "-" + SUBSTRIN...
Unicode string [DT_WST...
Configure Error Output...
OK
Cancel
Help

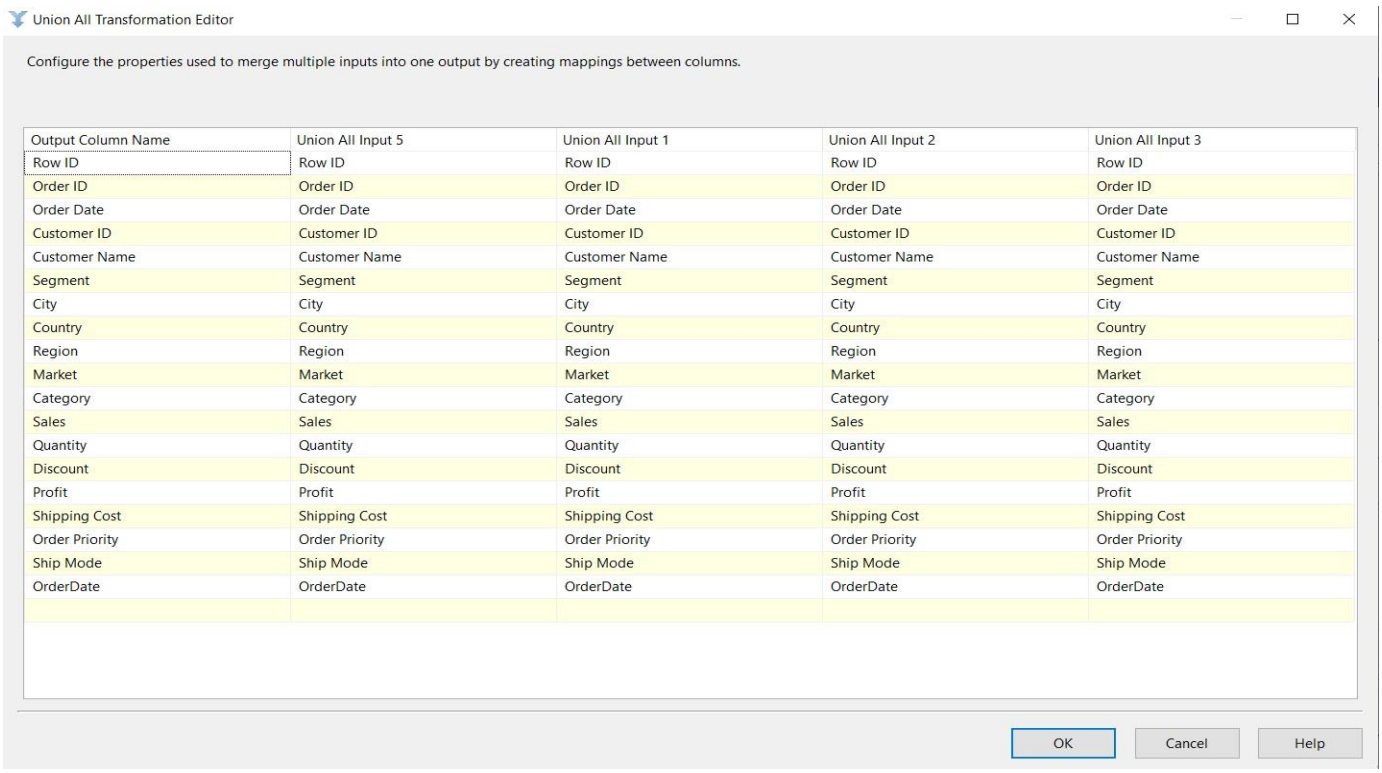
MM-D:


Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
OrderDate
Replace 'OrderDate'
SUBSTRING(OrderDate,1,2) + "-0" + SUBSTRING(Or...
Unicode string [DT_WST...
Configure Error Output...
OK
Cancel
Help

Len 10:

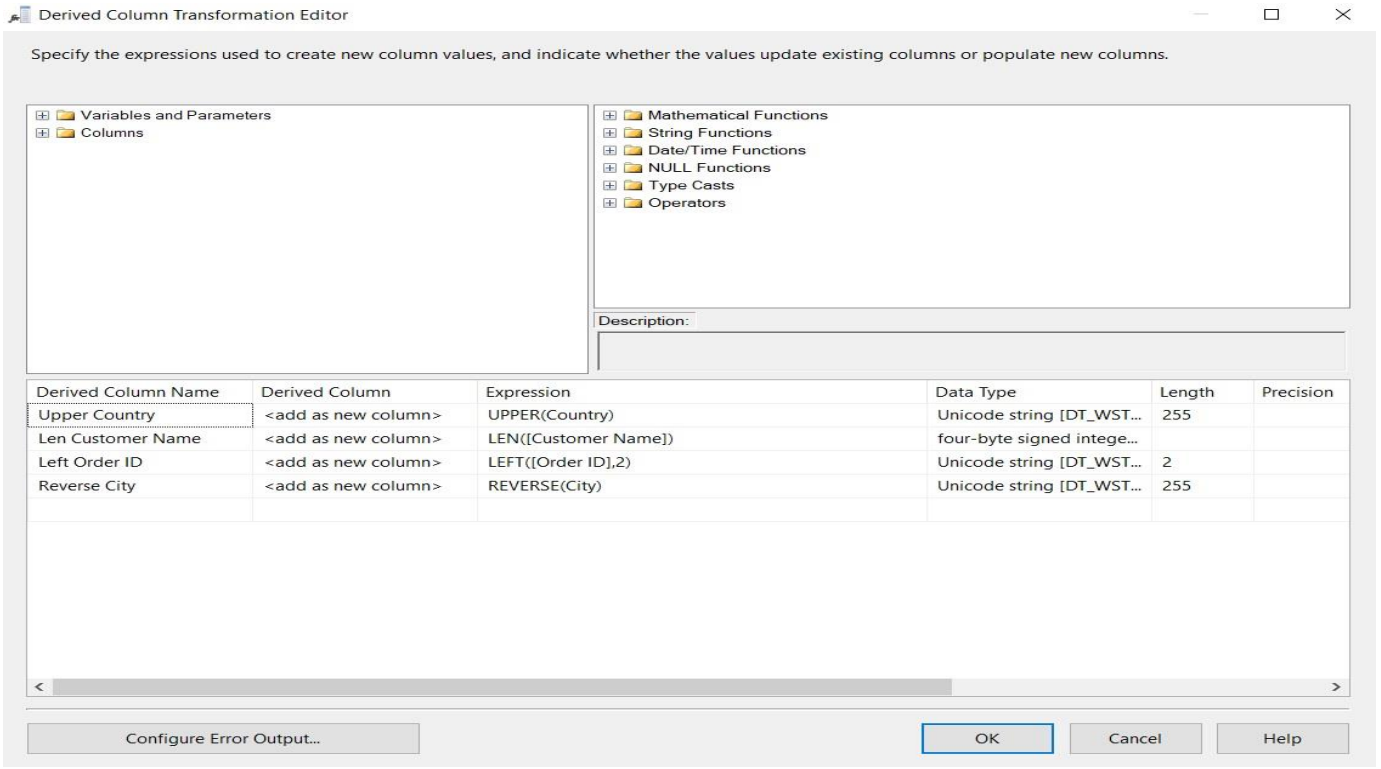

Derived Column Transformation Editor
Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.
Variables and Parameters
Columns
Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators
Description:
Derived Column Name
Derived Column
Expression
Data Type
OrderDate
Replace 'OrderDate'
SUBSTRING(OrderDate,1,2) + "-" + SUBSTRING(Ord...
Unicode string [DT_WST...
Configure Error Output...
OK
Cancel
Help

Union All:



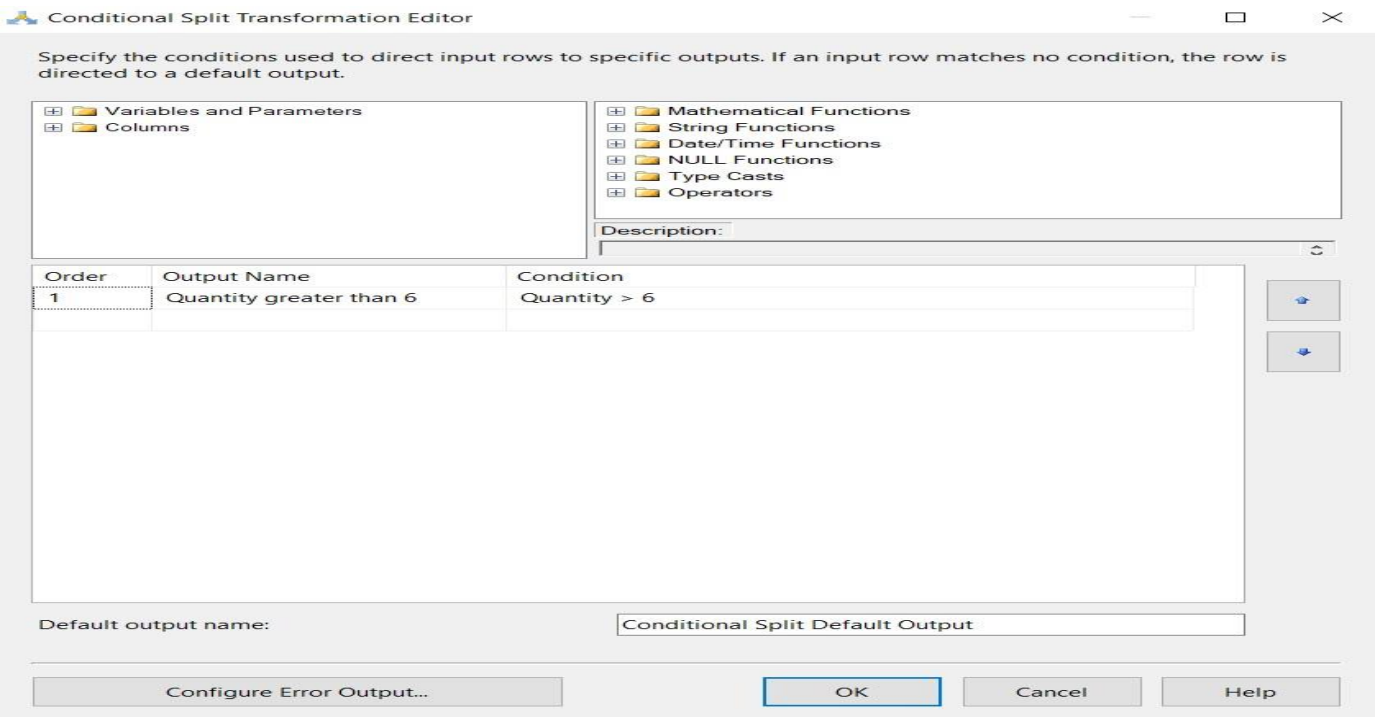
Output Column Name	Union All Input 5	Union All Input 1	Union All Input 2	Union All Input 3
Row ID	Row ID	Row ID	Row ID	Row ID
Order ID	Order ID	Order ID	Order ID	Order ID
Order Date	Order Date	Order Date	Order Date	Order Date
Customer ID	Customer ID	Customer ID	Customer ID	Customer ID
Customer Name	Customer Name	Customer Name	Customer Name	Customer Name
Segment	Segment	Segment	Segment	Segment
City	City	City	City	City
Country	Country	Country	Country	Country
Region	Region	Region	Region	Region
Market	Market	Market	Market	Market
Category	Category	Category	Category	Category
Sales	Sales	Sales	Sales	Sales
Quantity	Quantity	Quantity	Quantity	Quantity
Discount	Discount	Discount	Discount	Discount
Profit	Profit	Profit	Profit	Profit
Shipping Cost	Shipping Cost	Shipping Cost	Shipping Cost	Shipping Cost
Order Priority	Order Priority	Order Priority	Order Priority	Order Priority
Ship Mode	Ship Mode	Ship Mode	Ship Mode	Ship Mode
OrderDate	OrderDate	OrderDate	OrderDate	OrderDate

String Functions:

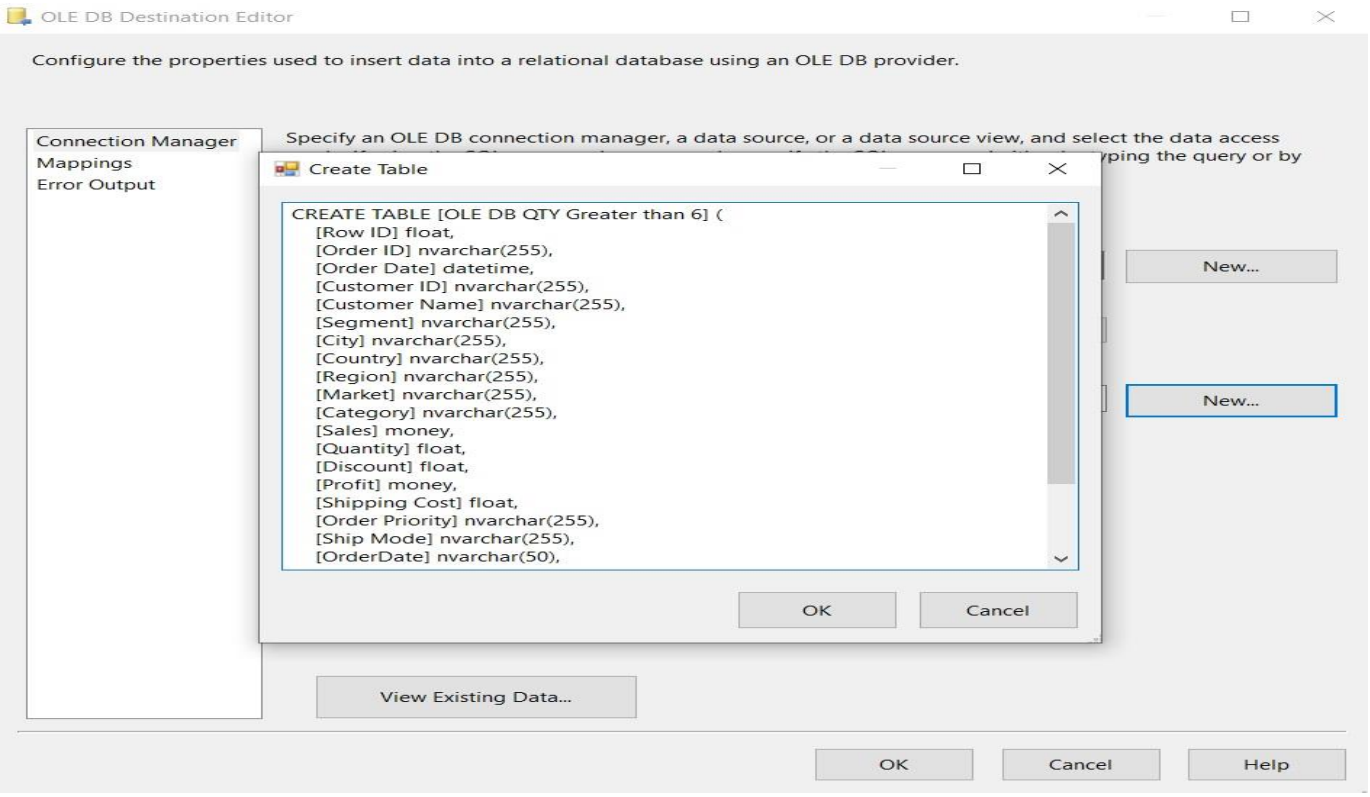


Derived Column Name	Derived Column	Expression	Data Type	Length	Precision
Upper Country	<add as new column>	UPPER(Country)	Unicode string [DT_WST...	255	
Len Customer Name	<add as new column>	LEN([Customer Name])	four-byte signed intege...		
Left Order ID	<add as new column>	LEFT([Order ID],2)	Unicode string [DT_WST...	2	
Reverse City	<add as new column>	REVERSE(City)	Unicode string [DT_WST...	255	

Split quantity greater than 6:



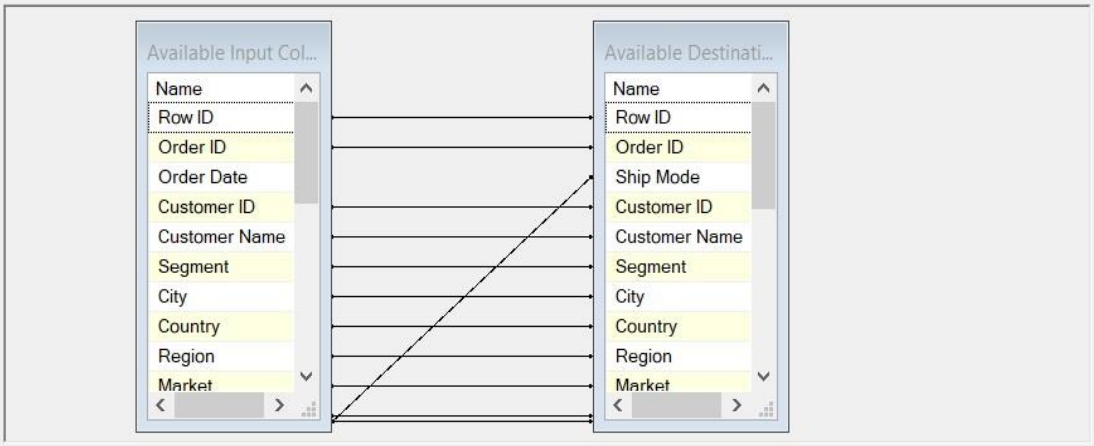
OLE DB QTY Greater than 6:



OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

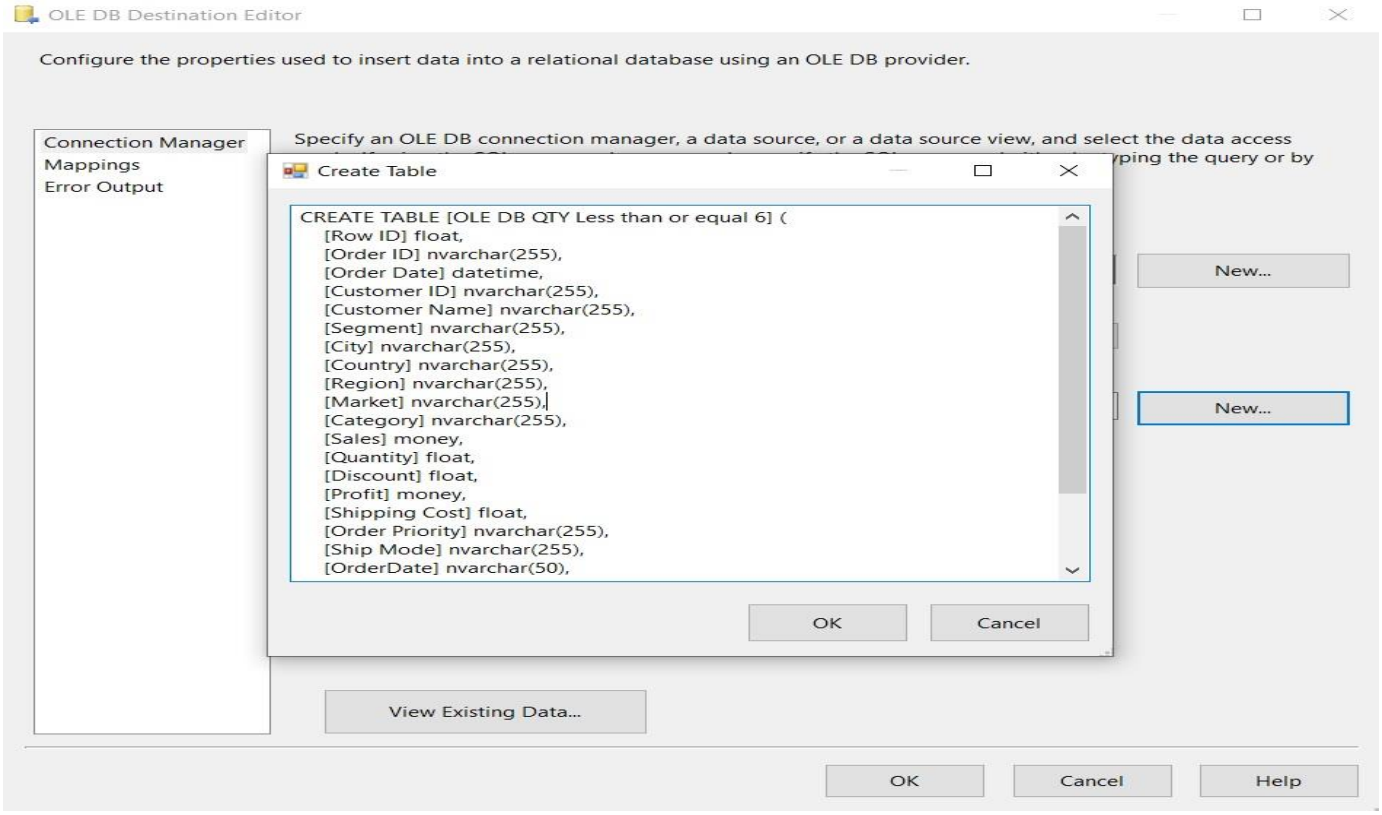
- Connection Manager
- Mappings
- Error Output

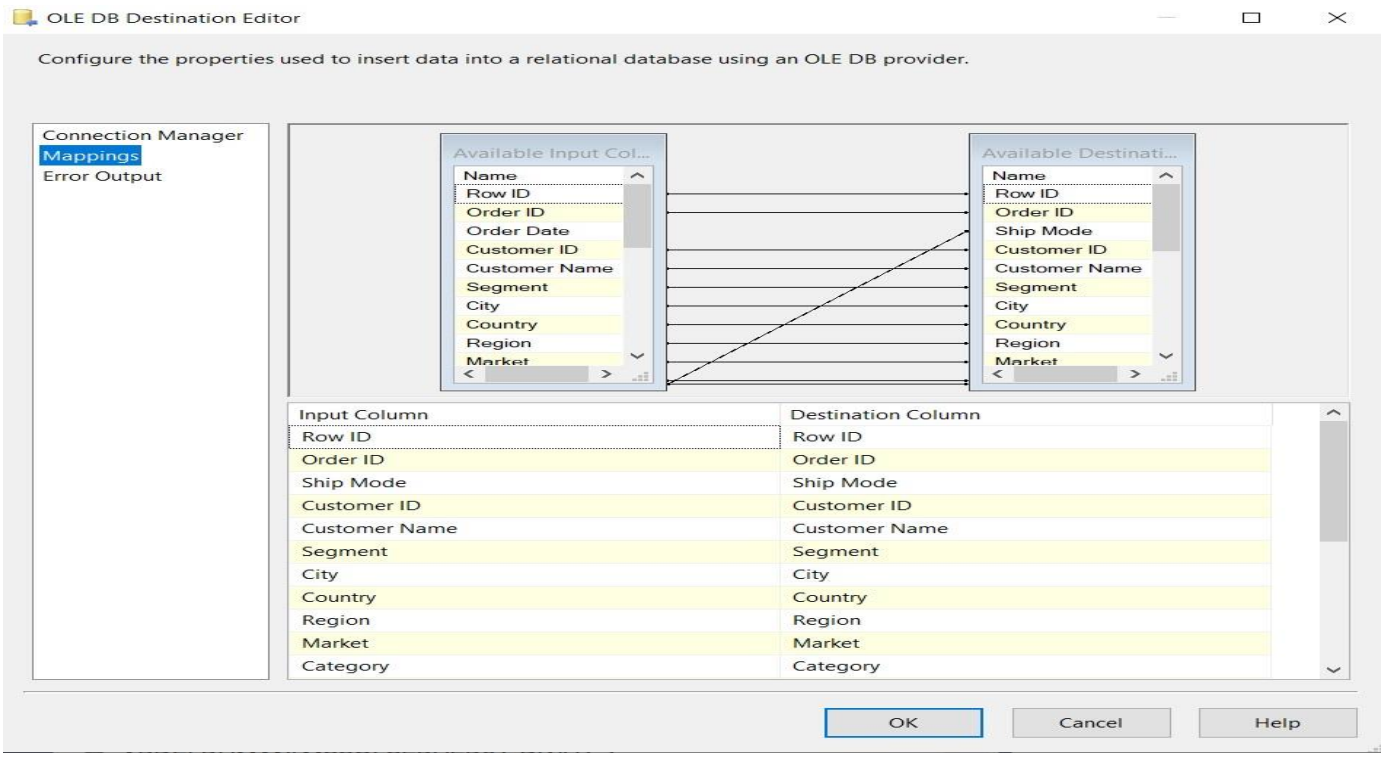


Input Column	Destination Column
Row ID	Row ID
Order ID	Order ID
Ship Mode	Ship Mode
Customer ID	Customer ID
Customer Name	Customer Name
Segment	Segment
City	City
Country	Country
Region	Region
Market	Market
Category	Category

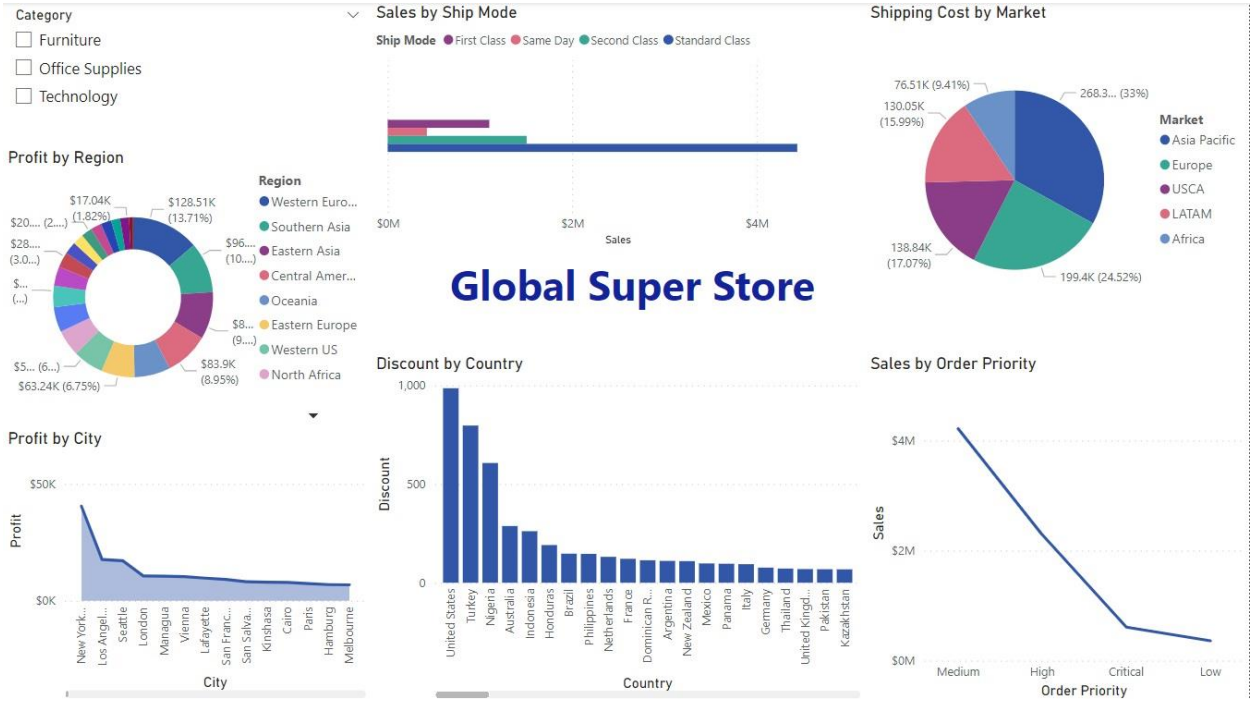
OK Cancel Help

OLE DB QTY Less than or equal 6:

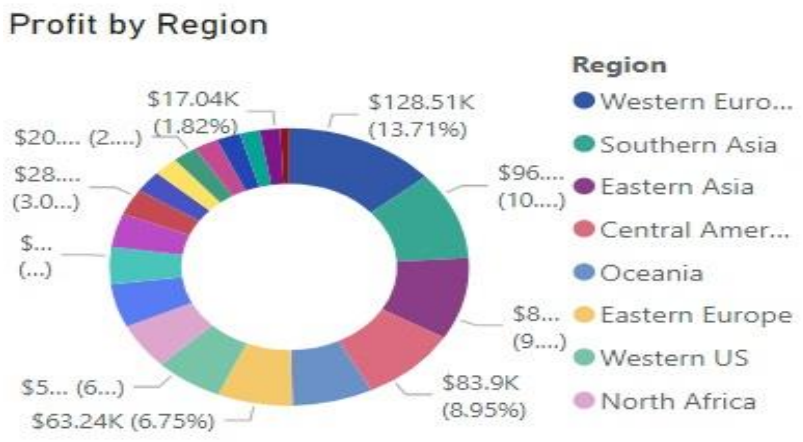




Power BI Dashboard



1-donut chart:

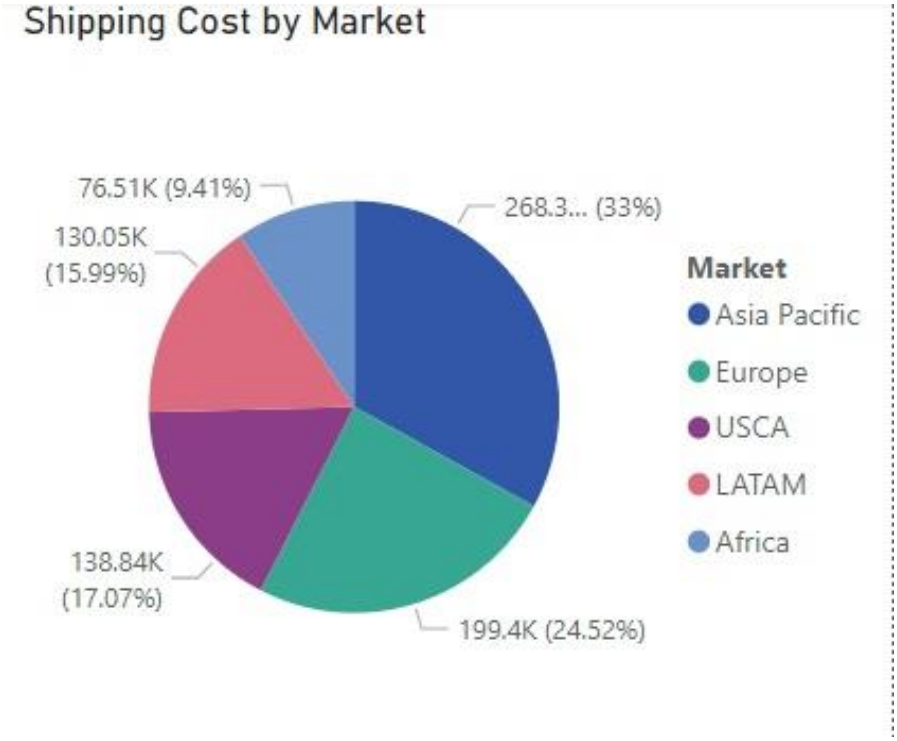


As shown in the previous donut chart, the highest profit in terms of regions is in Western Europe

2- Clustered Bar chart:

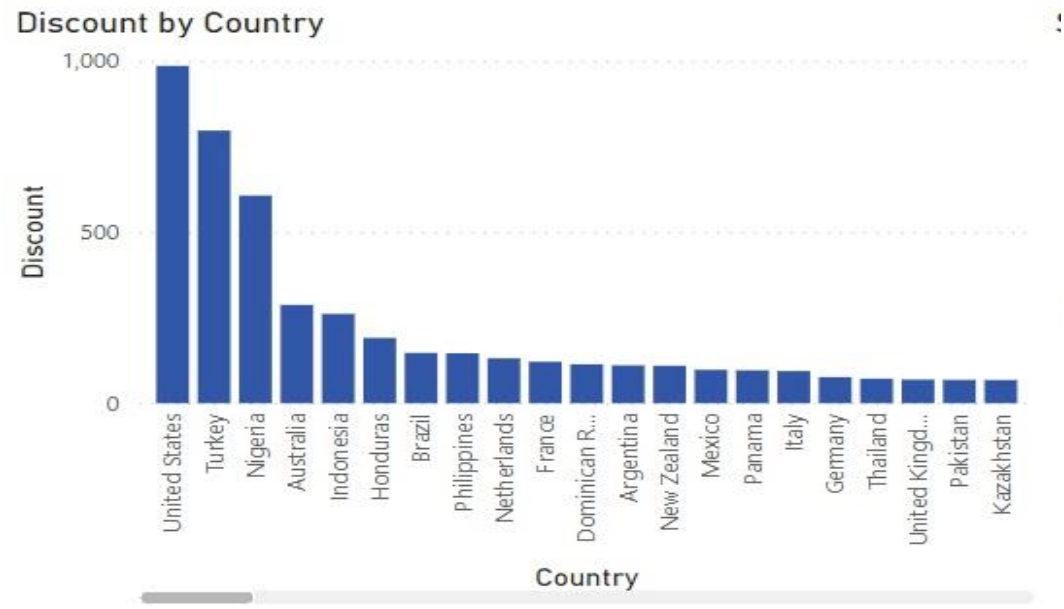
As shown in clustered bar chart, the bestselling ship mode is Standard Class

3- Pie chart:



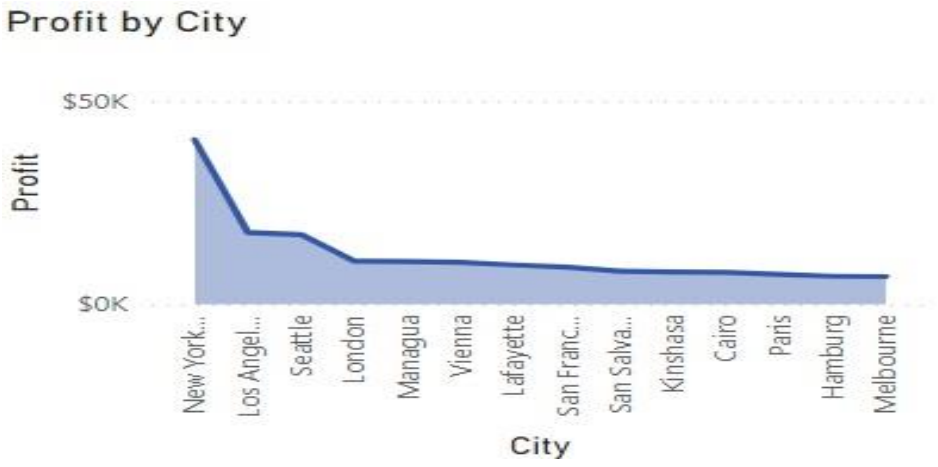
As shown in pie chart, the highest shipping cost in Asia Pacific market.

4- Clustered Column chart:



As shown in clustered column chart, the highest discount is USA

5- Area chart:



As shown in area chart, the highest profit is New York city in USA

6- Line chart:

As shown in line chart, the bestselling order that is in medium priority