



## Mapping data flow customer and dimension tables

1. In the last lab we created the fact table and loaded the table using the mapping data flow. But in this lab, we will create our dimensions tables which are product and customer tables then load the data in them using mapping data flows.
2. So, first go to Synapse Studio and create your product and dimension table using the below command. Also, delete the data from the fact sales table.

```
1 CREATE TABLE dimProduct
2 (
3     ProductID int NOT NULL,
4     ProductNumber varchar(100) NOT NULL,
5     Color varchar(20) NOT NULL,
6     ProductCategoryID int NOT NULL,
7     ProductCategoryName varchar(200) NOT NULL
8
9 )
```

```
1 CREATE TABLE dimCustomer
2 (
3     CustomerID int NOT NULL,
4     CompanyName varchar(500) NOT NULL
5 )
```

3. Also, we are going to follow the below script to add and remove the desired columns.

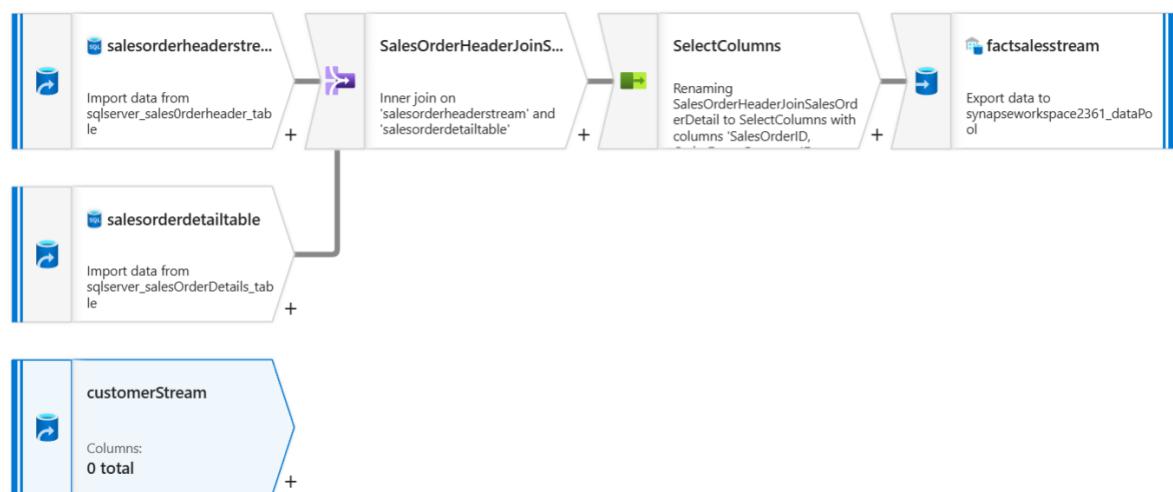
```
-- For the Product dimension table

SELECT pd.[ProductID],pd.[ProductNumber],pd.[Color],pd.[ProductCategoryID],
ct.[Name] AS 'ProductCategoryName'
FROM [SalesLT].[Product] pd INNER JOIN [SalesLT].[ProductCategory] ct
ON pd.[ProductCategoryID]=ct.[ProductCategoryID]
WHERE pd.[Color] IS NOT NULL

-- For the Customer dimension table

SELECT [CustomerID],[CompanyName] FROM [SalesLT].[Customer];
```

- From the same data flow add a new source for the customer dimension table.



- Choose to add a new dataset for the customer table from the SQL database.

Source settings    Source options    Projection    Optimize    Inspect    Data preview

---

Output stream name *	<input type="text" value="customerStream"/>	<a href="#">Learn more</a>
Description	<input type="text" value="Add source dataset"/>	Reset
Source type *	Dataset     Inline	
Dataset *	<input type="button" value="Select..."/>	New
Options	<input checked="" type="checkbox"/> Allow schema drift ⓘ <input type="checkbox"/> Infer drifted column types ⓘ <input type="checkbox"/> Validate schema ⓘ	
Sampling * ⓘ	<input type="radio"/> Enable <input checked="" type="radio"/> Disable	

6. Then just choose your link service and click on OK.

## Set properties

Name	<input type="text" value="sqlserver_customer_table"/>
Linked service *	<input type="text" value="sqlserver120"/>
Table name	<input type="button" value="Select..."/>
	<input type="checkbox"/> Enter manually
Import schema	<input checked="" type="radio"/> From connection/store <input checked="" type="radio"/> None
<a href="#">Advanced</a>	

7. Now dataset is added click on open to open the dataset and choose your customer table.

Source settings    Source options    Projection    Optimize    Inspect    Data preview

---

**Output stream name \*** customerStream [Learn more](#)

**Description** Import data from sqlserver\_customer\_table [Reset](#)

**Source type \*** [Dataset](#) [Inline](#)

**Dataset \*** [sqlserver\\_customer\\_table](#) [Test connection](#) [Open](#) [New](#)

**Options**

- Allow schema drift ①
- Infer drifted column types ①
- Validate schema ①

**Sampling \*** ① [Enable](#) [Disable](#)

8. Choose the customer table, go to Schema, and import the schema.

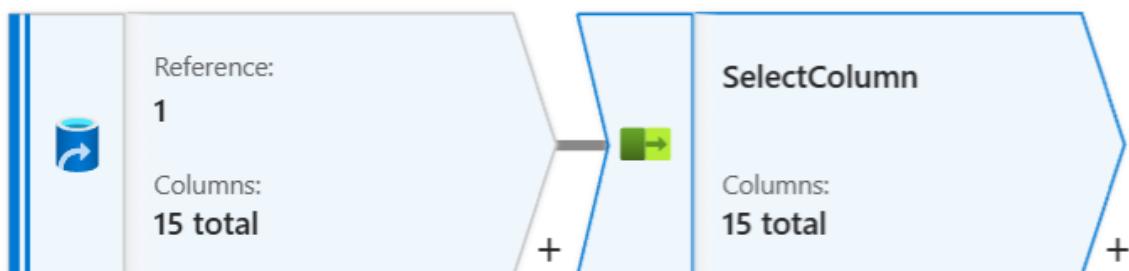
Connection    Schema    Parameters

---

**Linked service \*** [sqlserver120](#) [Test connection](#) [Edit](#) [New](#) [Learn more](#)

**Table** SalesLT.Customer [Refresh](#) [Preview data](#) [Enter manually](#)

9. Then we will create a new transformation which is **Select** transformation for our dim customer table and choose the columns of our interest.



10. Now from the settings delete all the unnecessary columns just choose customer ID and company Name.

Select settings   Optimize   Inspect   Data preview   [← Previous](#) [Next →](#)

**Output stream name \***  [Learn more](#)

**Description**  [Reset](#)

**Incoming stream \***

**Options**

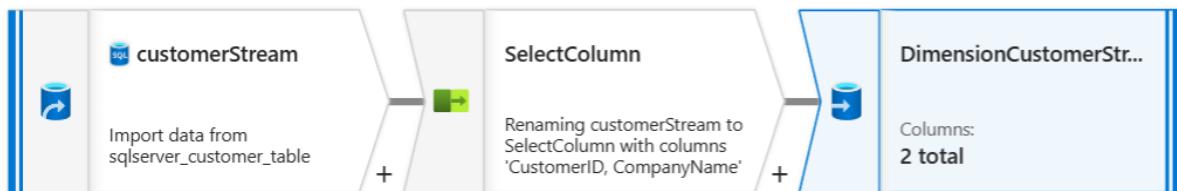
- Skip duplicate input columns [i](#)
- Skip duplicate output columns [i](#)

**Input columns \***

<input type="checkbox"/> customerStream's column	<a href="#">Y</a>	Name as	<a href="#">Y</a>
<input type="checkbox"/> CustomerID	<a href="#">→</a>	<input type="text" value="CustomerID"/>	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> CompanyName	<a href="#">→</a>	<input type="text" value="CompanyName"/>	<a href="#">+</a> <a href="#">Delete</a>

2 mappings: 13 column(s) from the inputs left unmapped [i](#)

11. Then create the sink from the customer stream,



12. Click on new to create a new dataset for it.

Sink   Settings   Errors   Mapping   Optimize   Inspect   Data preview

**Output stream name \***  [Learn more](#)

**Description**  [Reset](#)

**Incoming stream \***

**Sink type \***

Dataset	Inline	Cache
---------	--------	-------

**Dataset \***  [New](#)

**Options**

- Allow schema drift [i](#)
- Validate schema [i](#)

13. Choose the same option as shown below, here we are going to use the dedicated SQL Pool to choose our table.

## Set properties

Name

sqlserver\_dimcustomer\_Pooldb

Linked service \*

synapseworkspace2361



Select from existing table  New table

Table name

Select...



Enter manually

Import schema

From connection/store  None

> Advanced

14. Once your dataset is added click on open.

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name \* DimensionCustomerStream [Learn more](#)

Description Export data to sqlserver\_dimcustomer\_Pooldb [Reset](#)

Incoming stream \* SelectColumn

Sink type \* [Dataset](#) [Inline](#) [Cache](#)

Dataset \* [sqlserver\\_dimcustomer\\_Pooldb](#) [Test connection](#) [Open](#) [New](#)

Options  Allow schema drift [①](#)  Validate schema [①](#)

15. Choose your customer table, go to Schema and import the schema.

Connection Schema Parameters

Linked service \* [synapseworkspace2361](#) [Down arrow](#)

[Test connection](#) [Edit](#) [New](#) [Learn more](#)

Table [dbo.dimCustomerTable](#) [Down arrow](#) [Refresh](#) [Preview data](#)

Enter manually

16.Go onto mappings and disable auto mappings.

Sink Settings Errors **Mapping** Optimize Inspect Data preview

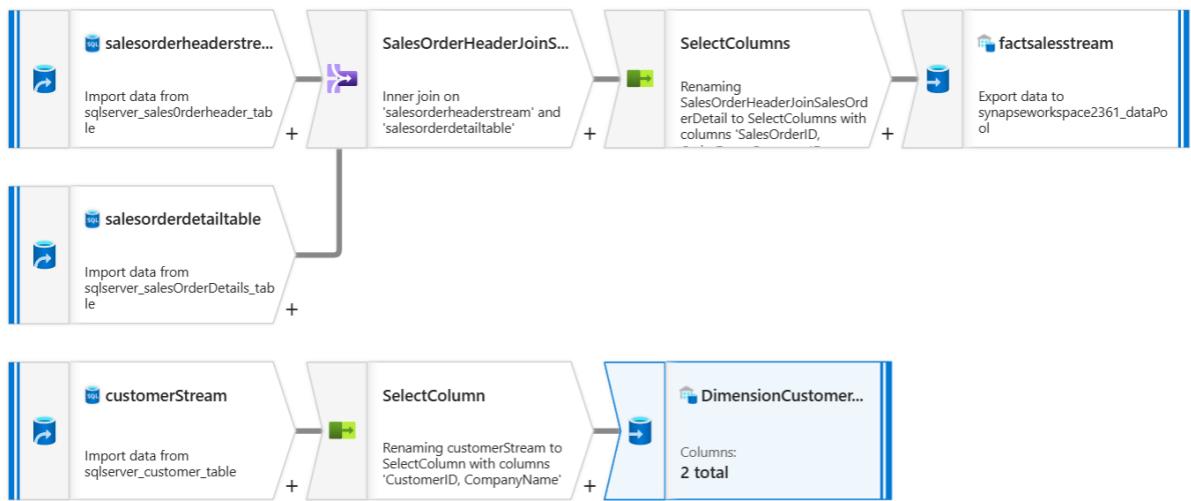
Options

Auto mapping [\(1\)](#)  Skip duplicate input columns [\(1\)](#)  
 Skip duplicate output columns [\(1\)](#)

[Reset](#) [Add mapping](#) [Delete](#) [Output format](#)

Input columns	Output columns
123 CustomerID	123 CustomerID
abc CompanyName	abc CompanyName

17.Once everything is done then click on publish all.



18.Move to the pipeline and trigger the pipeline but before that use the delete query to delete the data from your fact sales tables or the data will be copied into the table.

## Product category and product table

19.Now start with the product table and first create a new resource. Then click on new to create a new dataset for the product table.

**pRODUCTStream**

Columns: 0 total

**Source settings**    Source options    Projection    Optimize    Inspect    Data preview

**Output stream name \*** pRODUCTStream [Learn more](#)

**Description** Add source dataset [Reset](#)

**Source type \*** Dataset [Inline](#)

**Dataset \*** Select... [New](#)

**Options**  Allow schema drift [ⓘ](#)  Infer drifted column types [ⓘ](#)

20. Then give a name to your data set and click on OK. Then open your dataset.

## Set properties

**Name**  
sqlserver\_PRODUCT\_TABLE

**Linked service \***  
sqlserver120 [Edit](#)

**Table name**  
Select... [Reset](#) [Edit](#)  
 Enter manually

**Import schema**  
 From connection/store  None  
[Advanced](#)

21. Choose your product table from the database and import the schema.

Connection Schema Parameters

Linked service \*

[Test connection](#) [Edit](#) [New](#) [Learn more](#)

Table  [Refresh](#) [Preview data](#)

Enter manually

22. Then for the other table that is the product category choose to add a new source.

ProductCategory  
Columns: 0 total

Source settings Source options Projection Optimize Inspect Data preview

Output stream name \*  [Learn more](#)

Description  [Reset](#)

Source type \*  Dataset  Inline

Dataset \*  [New](#)

## Set properties

Name

Linked service \*

Table name

  Enter manually

Import schema

 From connection/store  None

> Advanced

23. Click on Open the dataset and choose product category table and import the schema.

Source settings Source options Projection Optimize Inspect Data preview

Output stream name \*  Learn more 

Description  

Source type \*  Dataset  Inline

Dataset \*   Test connection  Open 

Options  Allow schema drift   
 Infer drifted column types   
 Validate schema 

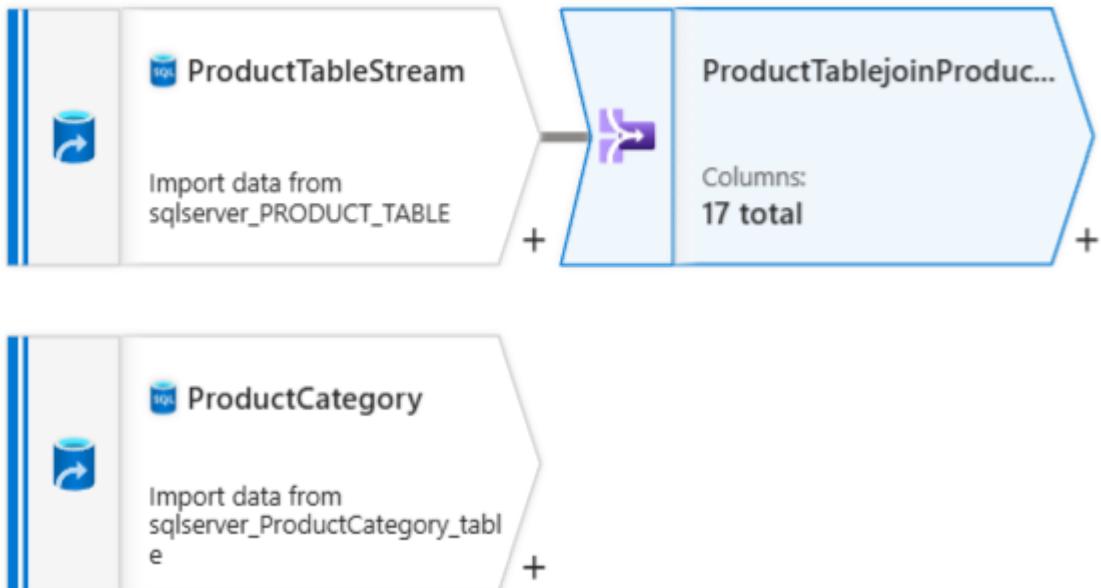
Sampling \*   Enable  Disable

Connection Schema Parameters

Linked service \*   Test connection  Edit  New Learn more

Table   Refresh  Preview data  
 Enter manually

24. As we have two tables so we need to perform a join. Form a product table and perform a join.



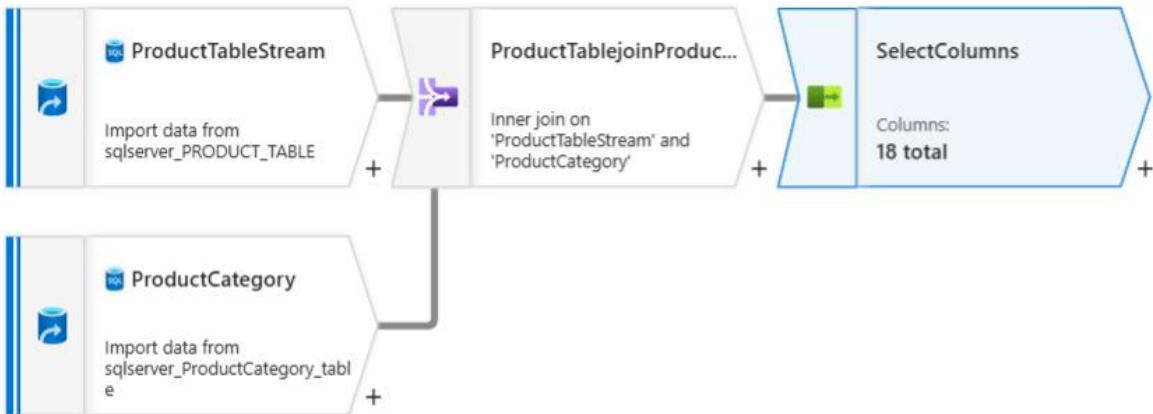
25. Choose the same options as you see below.

Join settings   [Optimize](#)   [Inspect](#)   [Data preview](#)

---

Output stream name *	<input type="text" value="ProductTablejoinProductCategory"/>	<a href="#">Learn more</a>
Description	<input type="text" value="Inner join on 'ProductTableStream' and 'ProductCategory'"/>	
Left stream *	<input type="text" value="ProductTableStream"/>	
Right stream *	<input type="text" value="ProductCategory"/>	
Join type *	<input checked="" type="radio"/> Full outer <input checked="" type="radio"/> Inner <input type="radio"/> Left outer <input type="radio"/> Right outer <input type="radio"/> Custom (cross)	
Use fuzzy matching	<input type="checkbox"/>	
Join conditions *	Left: ProductTableStream's column      Right: ProductCategory's column <div style="border: 1px solid #ccc; padding: 5px; display: flex; align-items: center;"> <div style="flex-grow: 1; margin-right: 10px;"> <input type="text" value="123 ProductCategoryID"/> </div> <div style="margin-right: 10px;">==</div> <div style="flex-grow: 1; margin-right: 10px;"> <input type="text" value="123 ProductCategoryID"/> </div> <div style="margin-right: 10px;">+</div> <div style="border: 1px solid #ccc; padding: 2px 5px;">Delete</div> </div>	

26. After the join we need to perform the select columns.



27. Below are the columns of interest and the rest we have removed.

Input column	Name as
123 ProductID	ProductID
abc ProductNumber	ProductNumber
abc Color	Color
123 ProductTableStream@ProductCategoryId	ProductCategoryID
abc ProductCategory@Name	ProductCategoryName

28. Also, we don't want the null values. We will choose a filter transformation to filter out the null values.



29. Open the expression builder from the filter on option.

Filter settings    Optimize    Inspect    Data preview

---

Output stream name * <b>filteNullColorRows</b>  Description  Incoming stream * <b>SelectColumnsDimProduct</b>  <b>Filter on *</b>  <div style="border: 1px solid #ccc; padding: 5px; height: 100px; width: 100%;"><p>Enter filter...</p></div> <div style="text-align: center; margin-top: -10px;"><a href="#">Open expression builder</a></div>	<a href="#">Learn more</a>  <a href="#">Reset</a>
--	---

30. We need to write this expression. This says that only if the color is not null then filter those values.

**!(isNull(Color))**

### Expression

```
!(isNull(color))
```

Filter settings    Optimize    Inspect    Data preview

---

Output stream name * <b>filteNullColorRows</b>  Description  Incoming stream * <b>SelectColumnsDimProduct</b>  <b>Filter on *</b>  <div style="border: 1px solid #ccc; padding: 5px; height: 100px; width: 100%;"><p>!(isNull(Color))</p></div>	<a href="#">Learn more</a>  <a href="#">Reset</a>
---	---

31. Once this is done then we can add our sink respectively.



32. Give a name and click on new to create a new dataset.

Sink   Settings   Errors   Mapping   Optimize   Inspect   Data preview

---

**Output stream name \*** dimproductstream [Learn more](#)

**Description** Add sink dataset [Reset](#)

**Incoming stream \*** filterNullColorRows

**Sink type \***  Dataset  Inline  Cache

**Dataset \*** Select... [New](#)

**Options**  Allow schema drift ⓘ  Validate schema ⓘ

## Set properties

### Name

### Linked service \*

Select from existing table  New table

#### Table name

  Enter manually

#### Import schema

From connection/store  None

> Advanced

33. Then click on the Open button choose your table and import schema.

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name \*  [Learn more](#) 

Description  

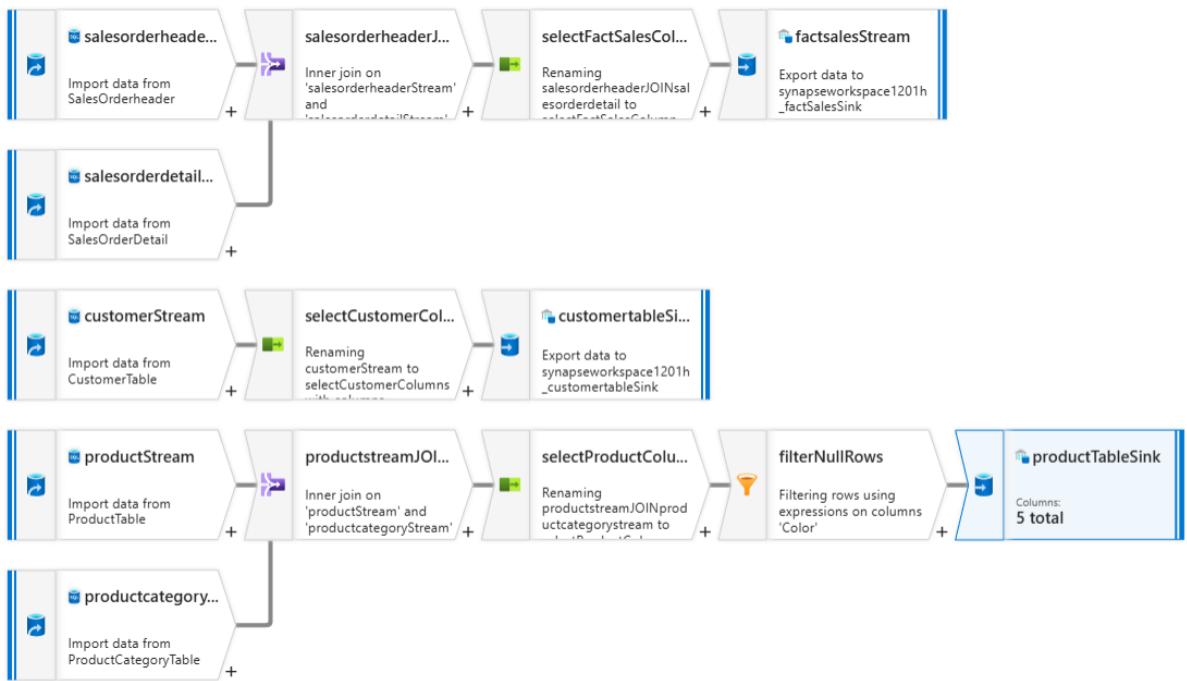
Incoming stream \*  

Sink type \*  Dataset  Inline  Cache

Dataset \*   Test connection  Open 

Options  Allow schema drift   Validate schema 

34. Disable auto mapping. Click on publish all to publish everything. This will be the data flow after completing everything.



35. After that you have to delete the data from the fact sales table and customer table to avoid data duplicity. Then trigger the pipeline you will see the data in your tables respectively.

16    `SELECT * FROM factSales`

Results    Messages

View    **Table**    Chart    Export results

Search

SalesOrderID	OrderDate	CustomerID	SubTotal	TaxAmt	Freight	TotalDue	OrderQty	ProductID
71774	2008-06-01T00:00:00.000Z	29847	880.3484	70.4279	22.0087	972.7850	1	836
71898	2008-06-01T00:00:00.000Z	29932	63980.9884	5118.4791	1599.5247	70698.9922	3	966
71774	2008-06-01T00:00:00.000Z	29847	880.3484	70.4279	22.0087	972.7850	1	822
71898	2008-06-01T00:00:00.000Z	29932	63980.9884	5118.4791	1599.5247	70698.9922	5	972
71776	2008-06-01T00:00:00.000Z	30072	78.8100	6.3048	1.9703	87.0851	1	907
71898	2008-06-01T00:00:00.000Z	29932	63980.9884	5118.4791	1599.5247	70698.9922	4	949

00:00:01 Query executed successfully.

```
10  SELECT * FROM dimProduct
```

Results Messages ^

View Table Chart Export results ▾

Search

ProductID	ProductNumber	Color	ProductCategoryID	ProductCategoryName
771	BK-M82S-38	Silver	5	Mountain Bikes
772	BK-M82S-42	Silver	5	Mountain Bikes
773	BK-M82S-44	Silver	5	Mountain Bikes
774	BK-M82S-48	Silver	5	Mountain Bikes
775	BK-M82B-38	Black	5	Mountain Bikes

00:00:01 Query executed successfully.

```
7  SELECT * FROM dimCustomer
```

Results Messages ^

View Table Chart Export results ▾

Search

CustomerID	CompanyName
1	A Bike Store
2	Progressive Sports
3	Advanced Bike Components
4	Modular Cycle Systems
5	Metropolitan Sports Supply
6	Aerobic Exercise Company
-	-

00:00:01 Query executed successfully.