



Sri Lanka Institute Of Information Technology

Predicting Coupon Redemption Data Warehouse

Data Warehouse & Business Intelligence

Assignment 1

Kumaran.S

IT20211400

1.Data Preparation

This dataset gives us an idea about the Predicting Coupon Redemption.

Data Set: Predicting Coupon Redemption

Site: Kaggle

Source Link:

<https://www.kaggle.com/datasets/vasudeva009/predicting-coupon-redemption-pca>

Modifications were done accordingly to the data set derived from the source This data set reflects combinations between customer transactions and promotion campaigns. Customer-specific details involved in transactions, Items customers are keen to purchase, and customer participation in promotion campaigns are some of the key details included in the data set.

ER Diagram

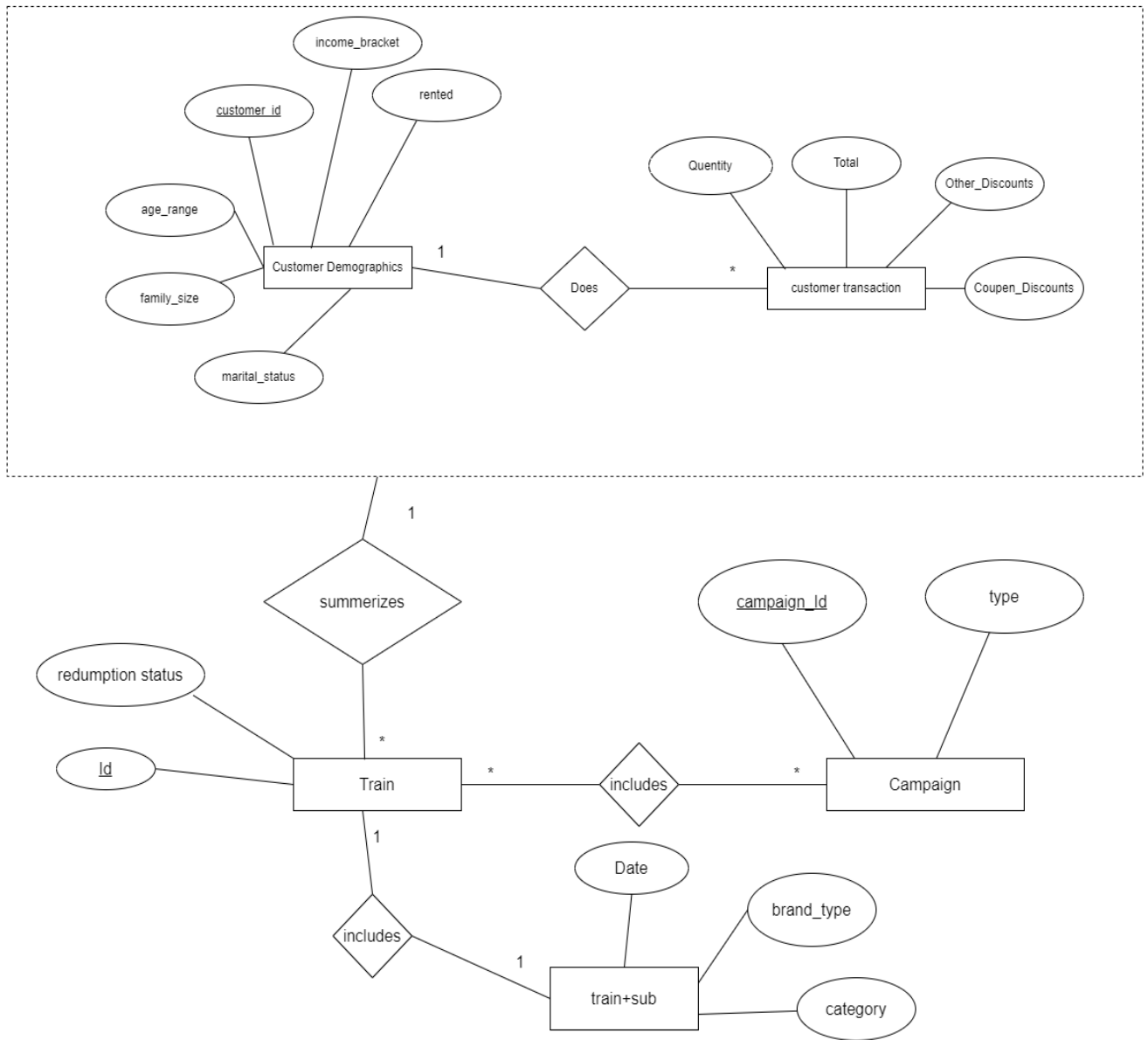







Figure 1: ER Diagram

2. Preparation of Data Sources

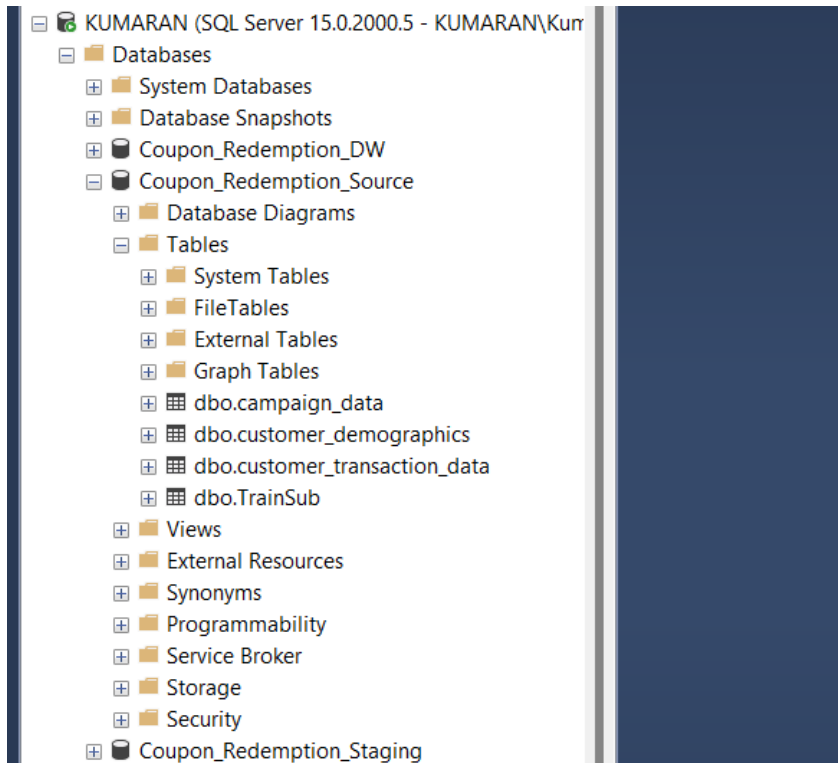
There are 5 source tables in two formats (CSV & Text formats)

| | | | |
|---|--------------------|-----------------------|----------|
|  campaign_data.csv | 5/14/2022 1:21 PM | Microsoft Excel Co... | 1 KB |
|  customer_demographics.csv | 5/14/2022 11:53 AM | Microsoft Excel Co... | 50 KB |
|  customer_transaction_data.csv | 5/14/2022 11:51 AM | Microsoft Excel Co... | 39 KB |
|  Train.txt | 5/14/2022 2:40 PM | Text Document | 1,227 KB |
|  TrainSub.csv | 5/15/2022 11:59 AM | Microsoft Excel Co... | 3,041 KB |

Files were imported to the SQL source database:

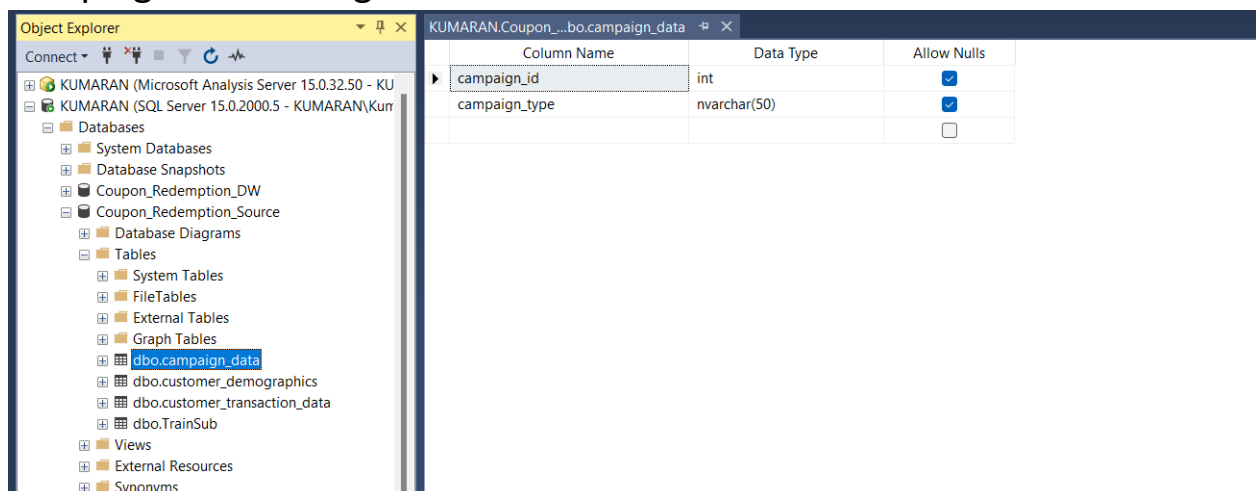
- Campaign_data
- Customer_demographics
- Customer_transaction_data
- TrainSub

Description of the data set



Design of Data_Source

- campaign table design



- Customer demographics table design

The screenshot shows the SQL Server Enterprise Manager interface. On the left, the Object Explorer displays the database structure for KUMARAN (Microsoft Analysis Server 15.0.32.50 - KU) and KUMARAN (SQL Server 15.0.2000.5 - KUMARAN\Kum). The right pane shows the design of the KUMARAN.Coupon_R...er_demographics table.

| Column Name | Data Type | Allow Nulls |
|----------------|--------------|-------------------------------------|
| customer_id | int | <input checked="" type="checkbox"/> |
| age_range | nvarchar(50) | <input checked="" type="checkbox"/> |
| marital_status | nvarchar(50) | <input checked="" type="checkbox"/> |
| rented | int | <input checked="" type="checkbox"/> |
| family_size | nvarchar(50) | <input checked="" type="checkbox"/> |
| income_bracket | int | <input checked="" type="checkbox"/> |

- Customer transaction table design

The screenshot shows the SQL Server Enterprise Manager interface. On the left, the Object Explorer displays the database structure for KUMARAN (Microsoft Analysis Server 15.0.32.50 - KU) and KUMARAN (SQL Server 15.0.2000.5 - KUMARAN\Kum). The right pane shows the design of the KUMARAN.Coupon_R...transaction_data table.

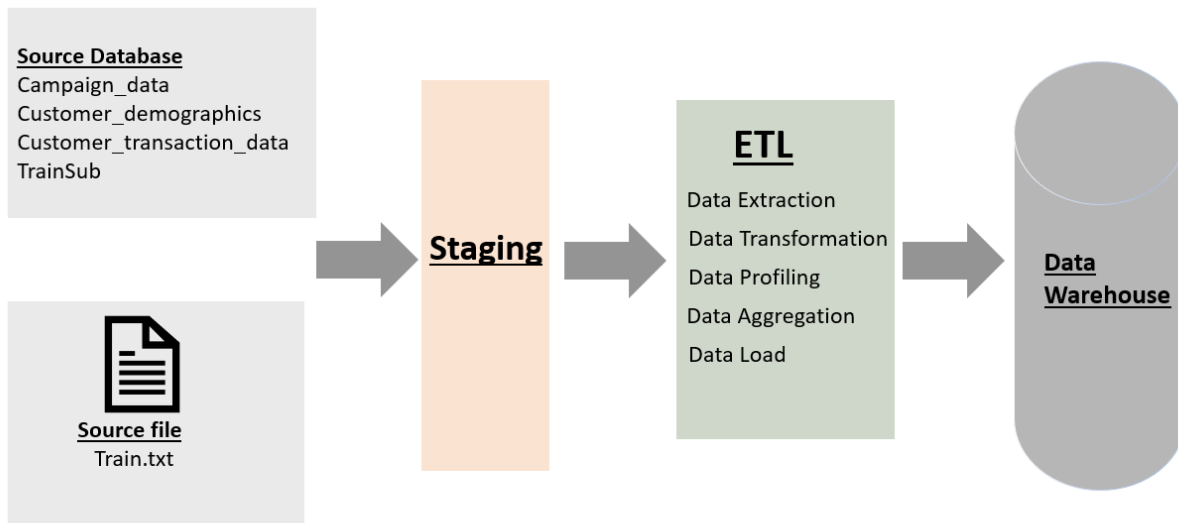
| Column Name | Data Type | Allow Nulls |
|-----------------|-----------|-------------------------------------|
| customer_id | int | <input checked="" type="checkbox"/> |
| quantity | int | <input checked="" type="checkbox"/> |
| Total | money | <input checked="" type="checkbox"/> |
| other_discount | money | <input checked="" type="checkbox"/> |
| coupon_discount | money | <input checked="" type="checkbox"/> |

- Train sub table design

The screenshot shows the SQL Server Enterprise Manager interface. On the left, the Object Explorer displays the database structure for KUMARAN (Microsoft Analysis Server 15.0.32.50 - KU) and KUMARAN (SQL Server 15.0.2000.5 - KUMARAN\Kum). The right pane shows the design of the KUMARAN.Coupon_R...e - dbo.TrainSub table.

| Column Name | Data Type | Allow Nulls |
|-------------|---------------|-------------------------------------|
| id | int | <input checked="" type="checkbox"/> |
| date | datetime | <input checked="" type="checkbox"/> |
| category | nvarchar(500) | <input checked="" type="checkbox"/> |
| brand_type | nvarchar(500) | <input checked="" type="checkbox"/> |

Solution Architecture

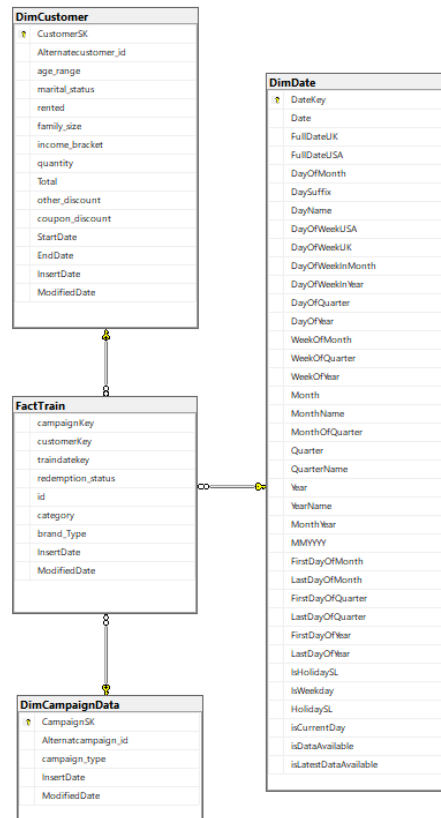


The first step is staging the source data, after staging tables are created, there are:

- StgCampaignData
- StgCustomerDemographics
- StgCustomerTransactionData
- StgTrain
- StgTrainSub

Next staged tables are profiled and aggregated. Then data is transformed and loaded. Finally completing the stages and all data is tested and validated then the Data Warehouse is created.

Data warehouse design and development



The snowflake schema is used to design the Datawarehouse design. There is one fact table as transactions and 3 dimensions.






Assumptions.

Customer Details were considered as a slowly changing dimension.

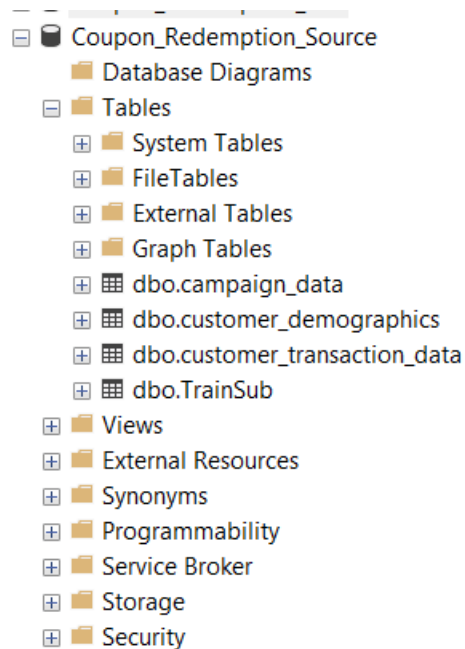
ETL development process

Steps with Screenshots will be displayed in this category.

- Step 01: Setting up the Environment

| | | | |
|---|--------------------|-----------------------|----------|
|  campaign_data.csv | 5/14/2022 1:21 PM | Microsoft Excel Co... | 1 KB |
|  customer_demographics.csv | 5/14/2022 11:53 AM | Microsoft Excel Co... | 50 KB |
|  customer_transaction_data.csv | 5/14/2022 11:51 AM | Microsoft Excel Co... | 39 KB |
|  Train.txt | 5/14/2022 2:40 PM | Text Document | 1,227 KB |
|  TrainSub.csv | 5/15/2022 11:59 AM | Microsoft Excel Co... | 3,041 KB |

Text and Excel File



SourceDB in sql

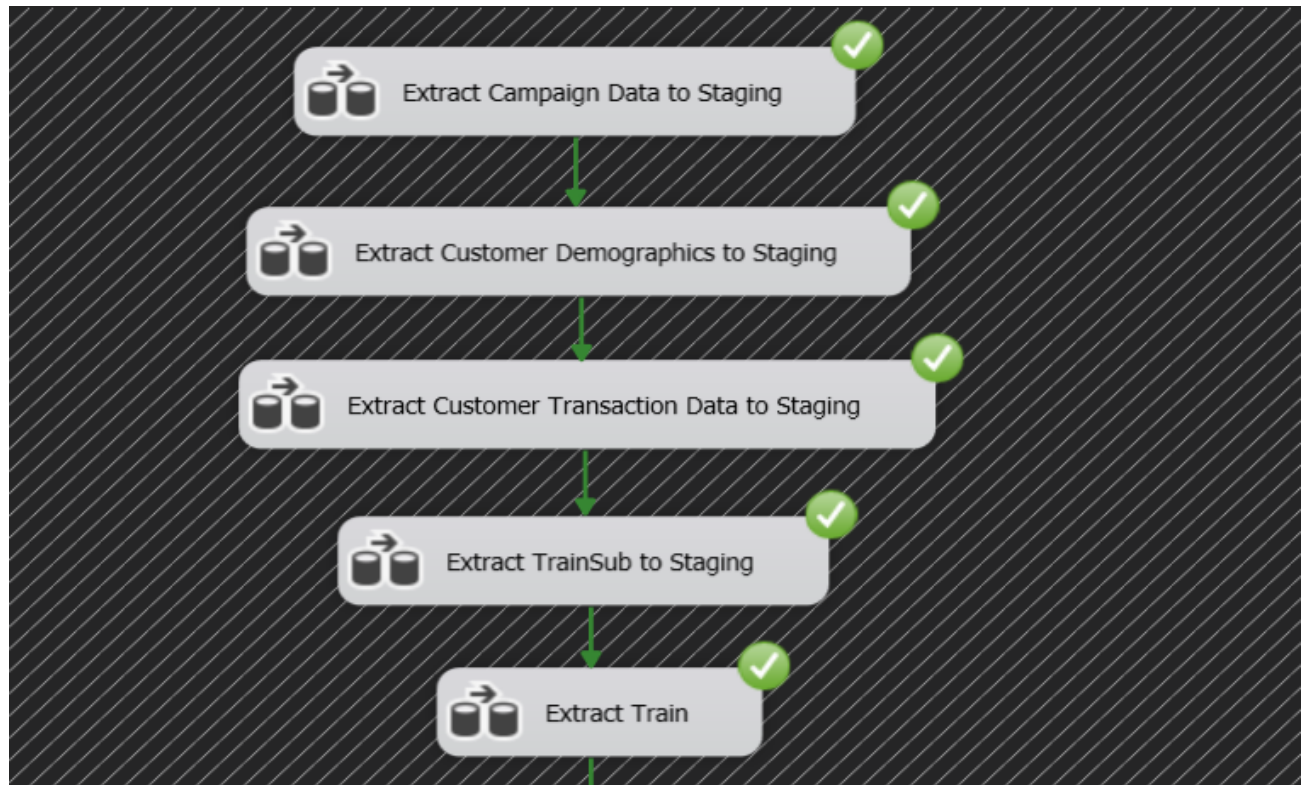
- [-] [DB Icon] Coupon_Redemption_Staging
 - [+] [Folder Icon] Database Diagrams
 - [-] [Folder Icon] Tables
 - [+] [Folder Icon] System Tables
 - [+] [Folder Icon] FileTables
 - [+] [Folder Icon] External Tables
 - [+] [Folder Icon] Graph Tables
 - [+] [Table Icon] dbo.StgCampaignData
 - [+] [Table Icon] dbo.StgCustomerDemographics
 - [+] [Table Icon] dbo.StgCustomerTransactionData
 - [+] [Table Icon] dbo.StgTrain
 - [+] [Table Icon] dbo.StgTrainSub
 - [+] [Folder Icon] Views
 - [+] [Folder Icon] External Resources
 - [+] [Folder Icon] Synonyms
 - [+] [Folder Icon] Programmability
 - [+] [Folder Icon] Service Broker
 - [+] [Folder Icon] Storage
 - [+] [Folder Icon] Security

Staging

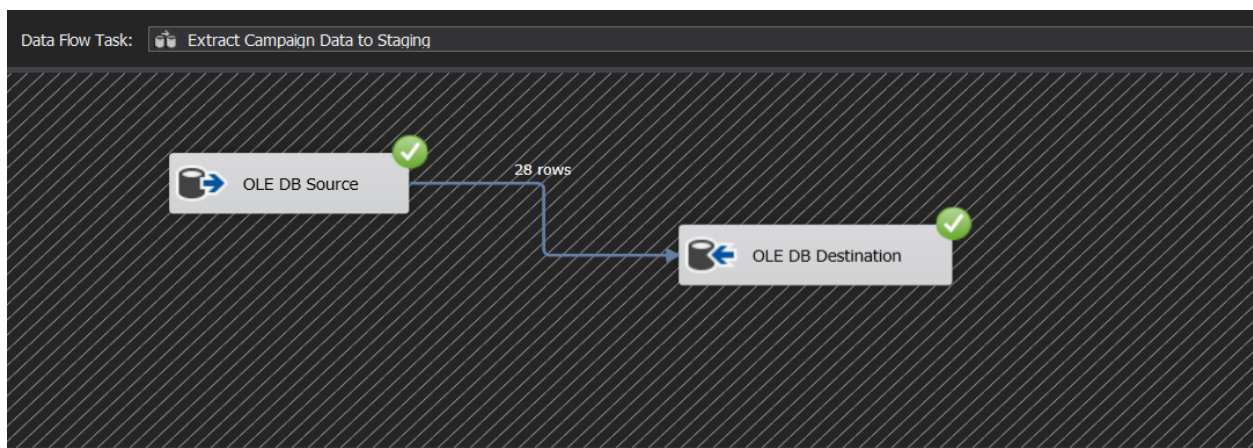
- [-] [DB Icon] Coupon_Redemption_DW
 - [+] [Folder Icon] Database Diagrams
 - [-] [Folder Icon] Tables
 - [+] [Folder Icon] System Tables
 - [+] [Folder Icon] FileTables
 - [+] [Folder Icon] External Tables
 - [+] [Folder Icon] Graph Tables
 - [+] [Table Icon] dbo.DimCampaignData
 - [+] [Table Icon] dbo.DimCustomer
 - [+] [Table Icon] dbo.DimDate
 - [+] [Table Icon] dbo.FactTrain
 - [+] [Folder Icon] Views
 - [+] [Folder Icon] External Resources
 - [+] [Folder Icon] Synonyms
 - [+] [Folder Icon] Programmability
 - [+] [Folder Icon] Service Broker
 - [+] [Folder Icon] Storage
 - [+] [Folder Icon] Security

Data Warehouse

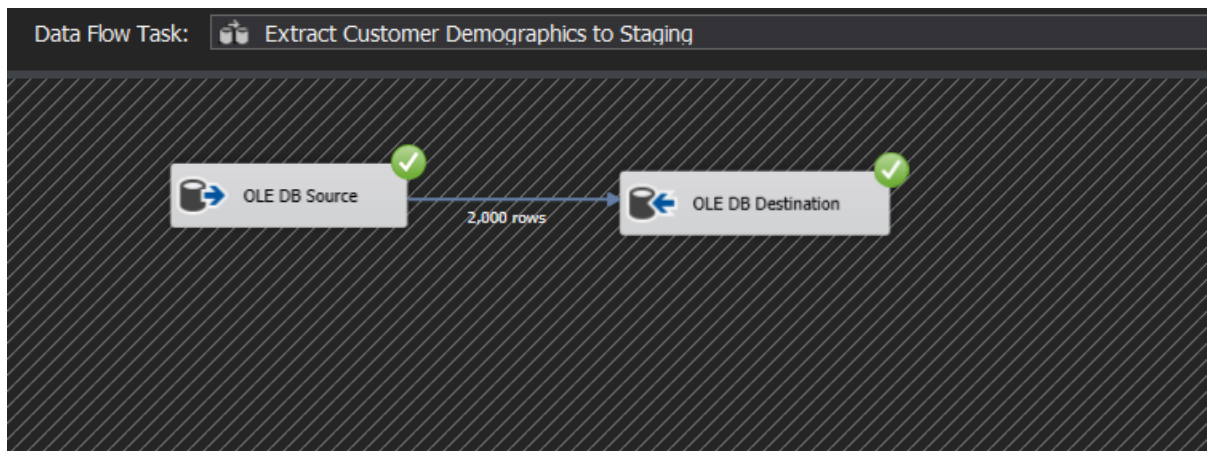
- Step 02: Data Extracting from source to staging tables



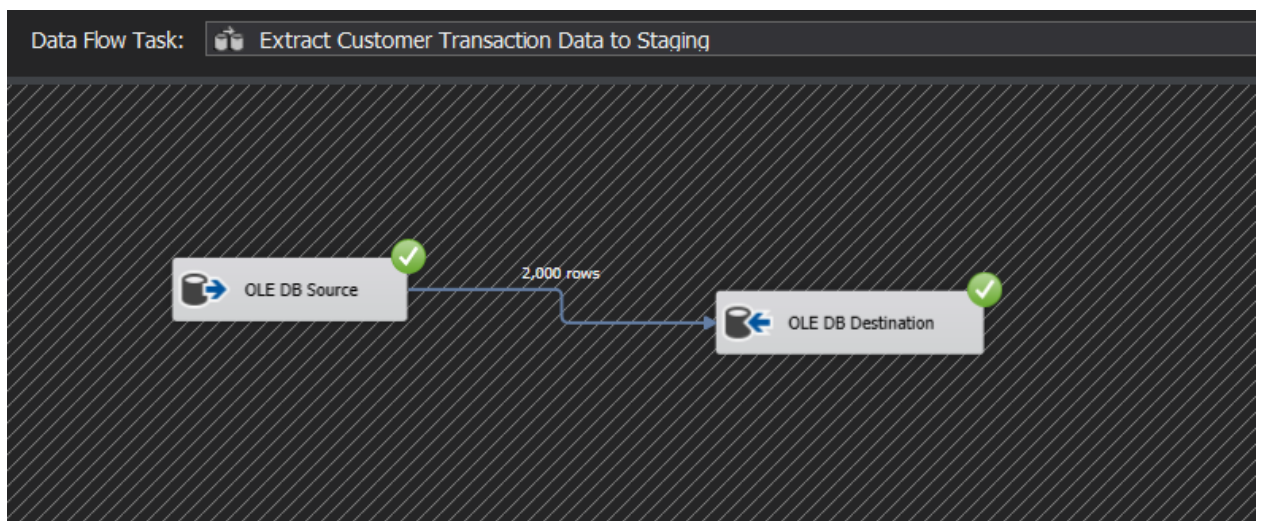
Extracting data



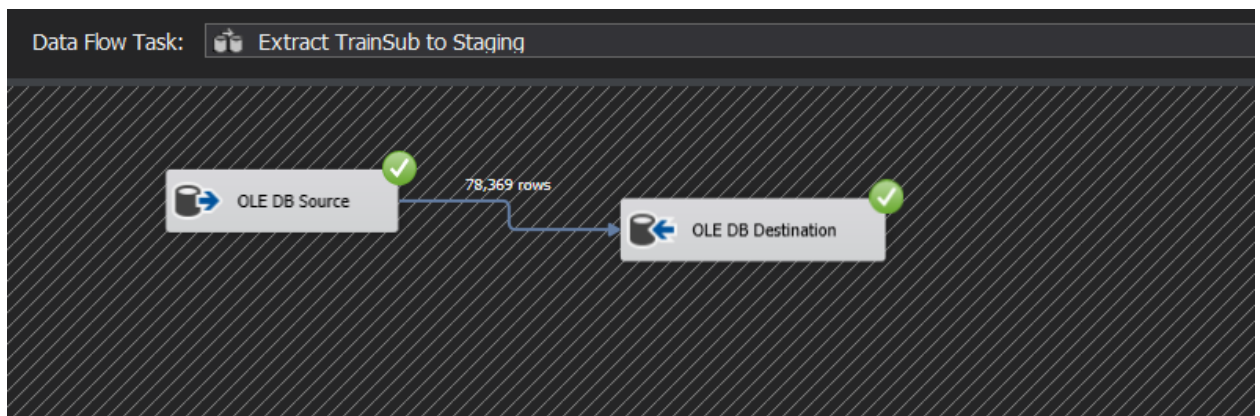
Extracting Campaign data



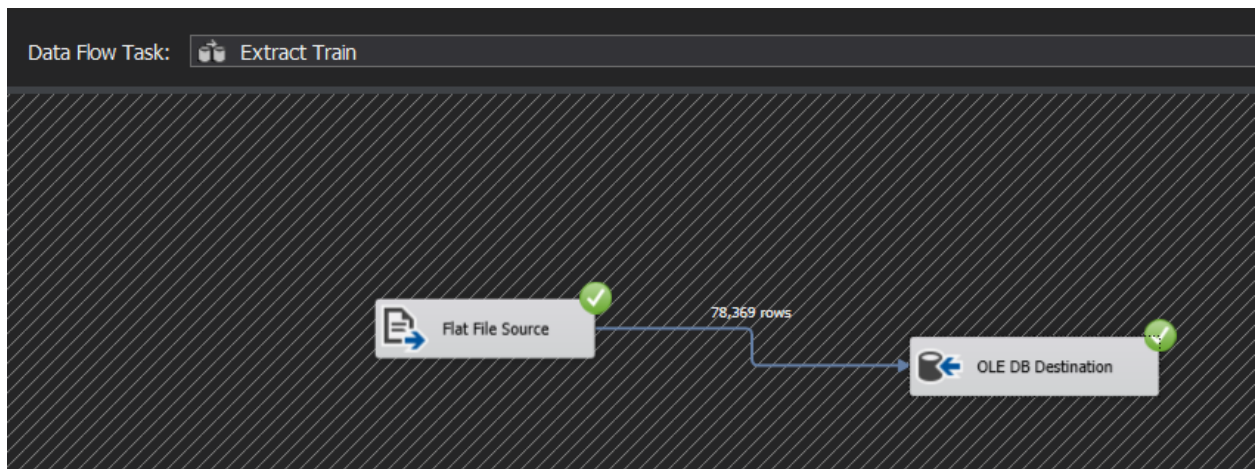
Extracting customer demographics



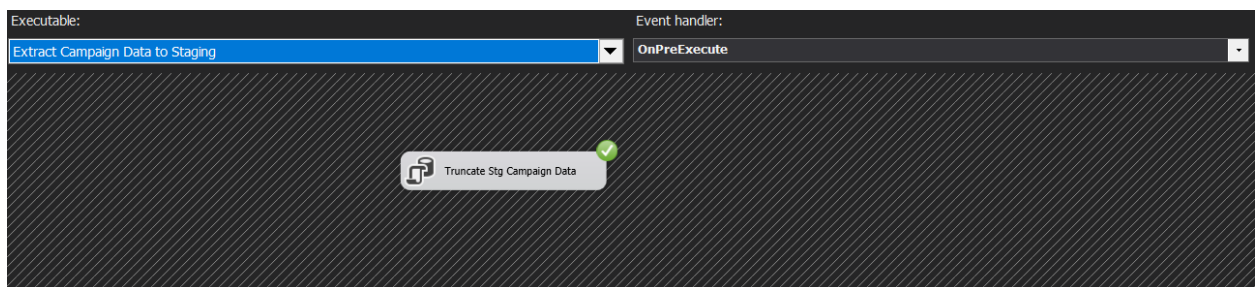
Extracting customer transaction data



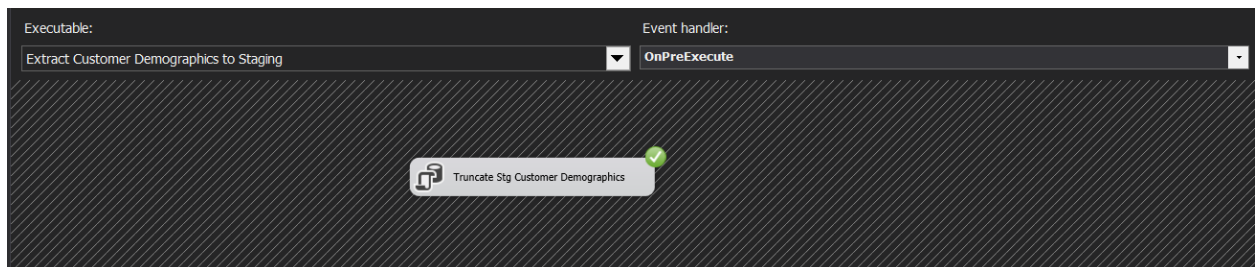
Extracting Train sub data



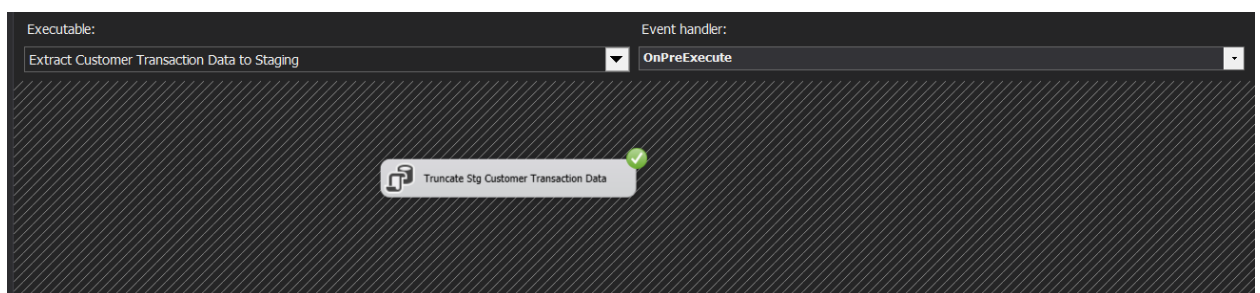
Extract train data



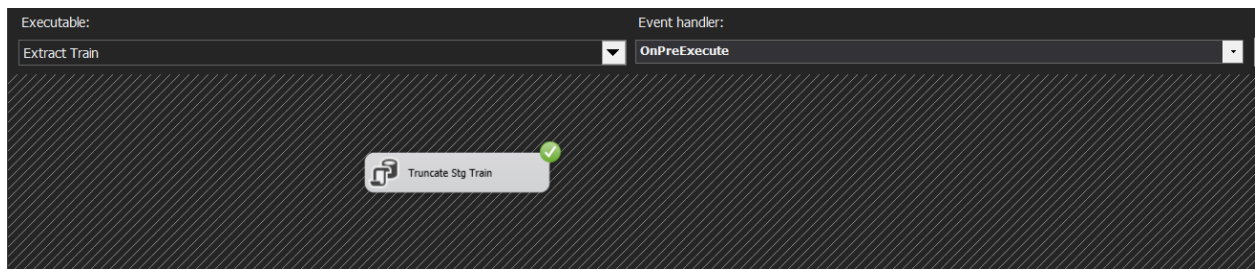
Truncate campaign data



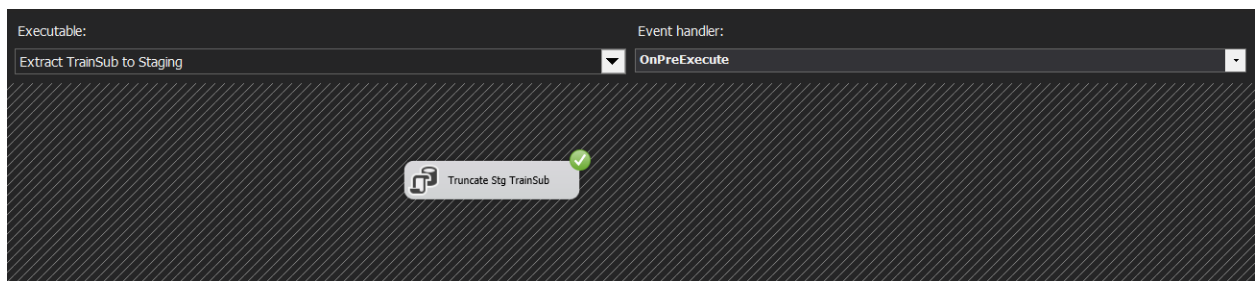
Truncate customer data



Truncate customer Transaction data

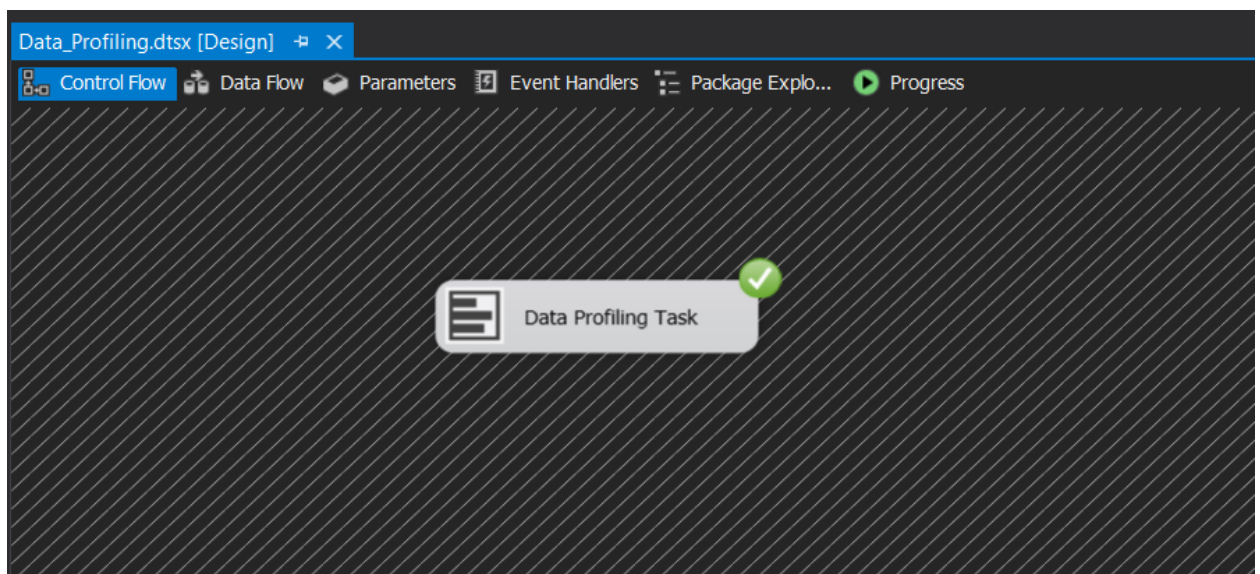


Truncate Train data

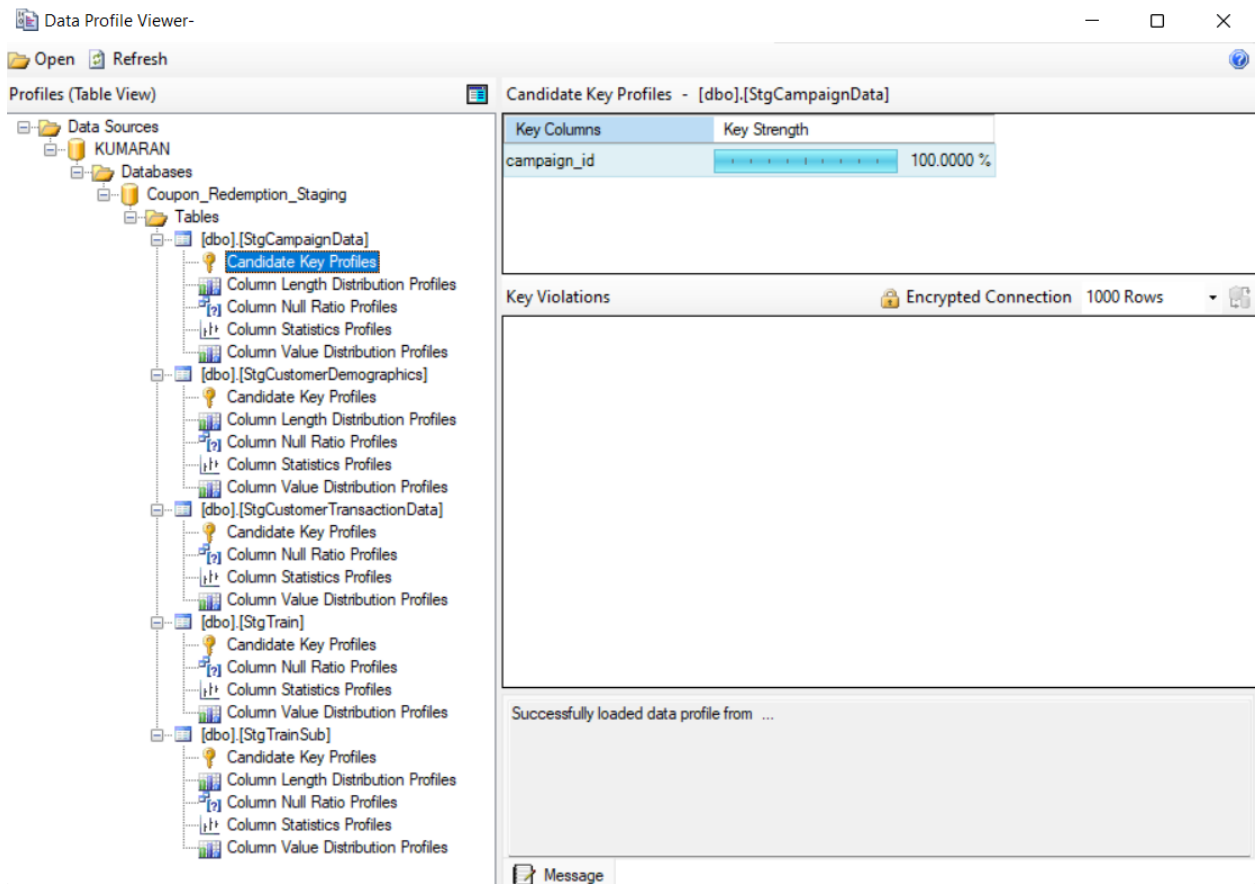


Truncate Train sub data

- Step 03: Data Profiling

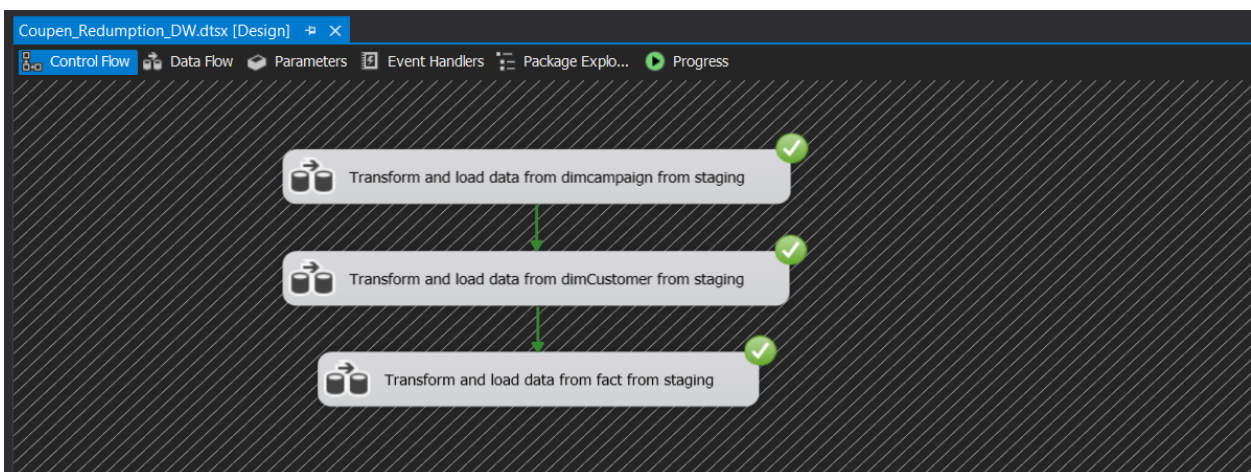


Data profiling

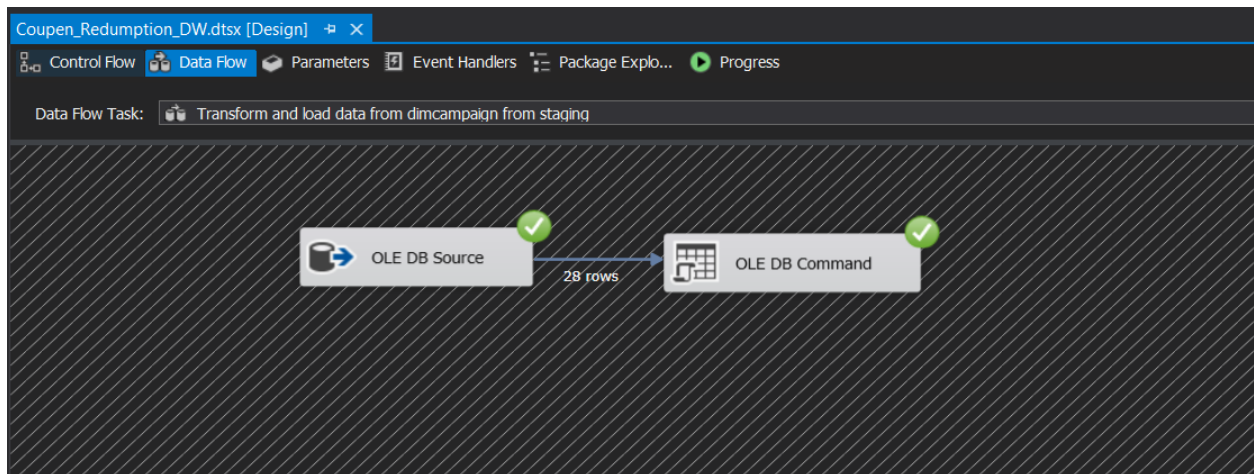


profile viewer

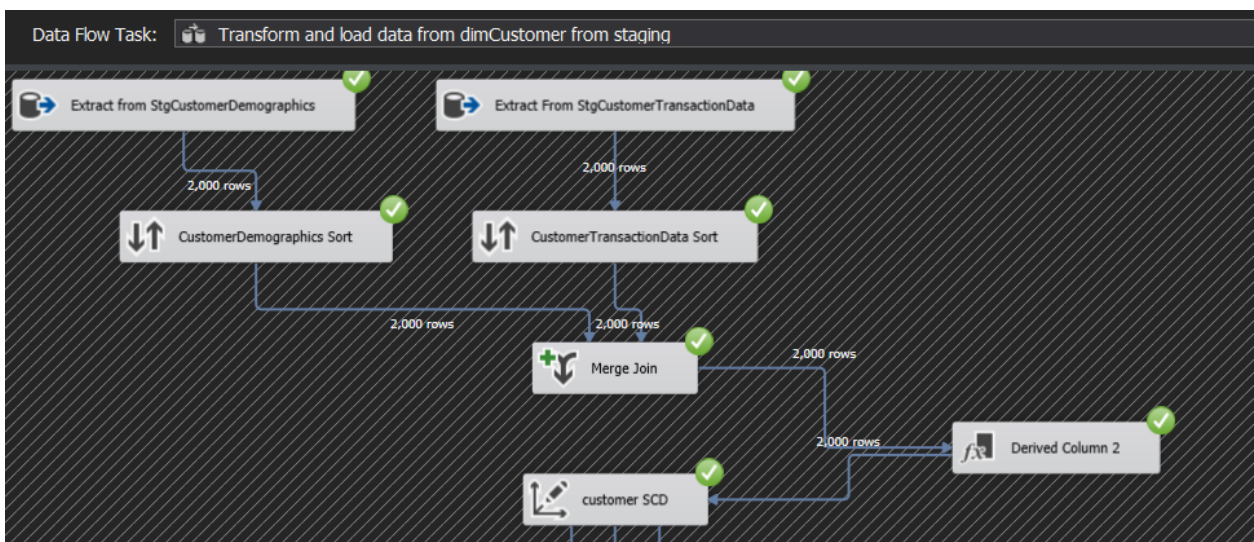
- Step 04: Data Transformation from staging to warehouse

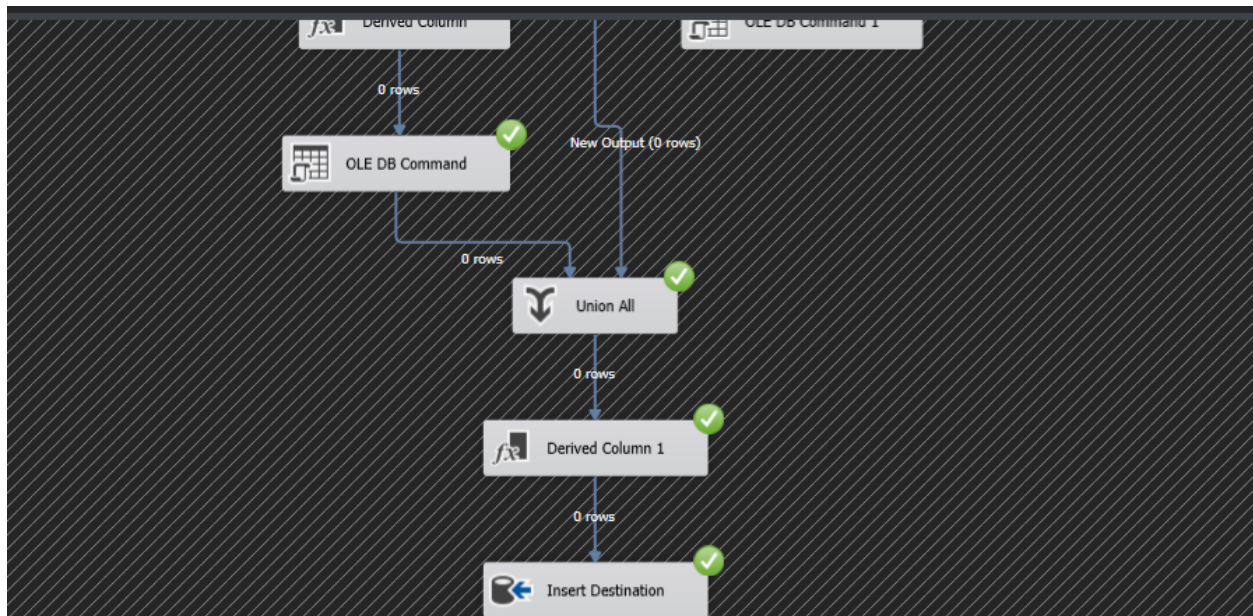
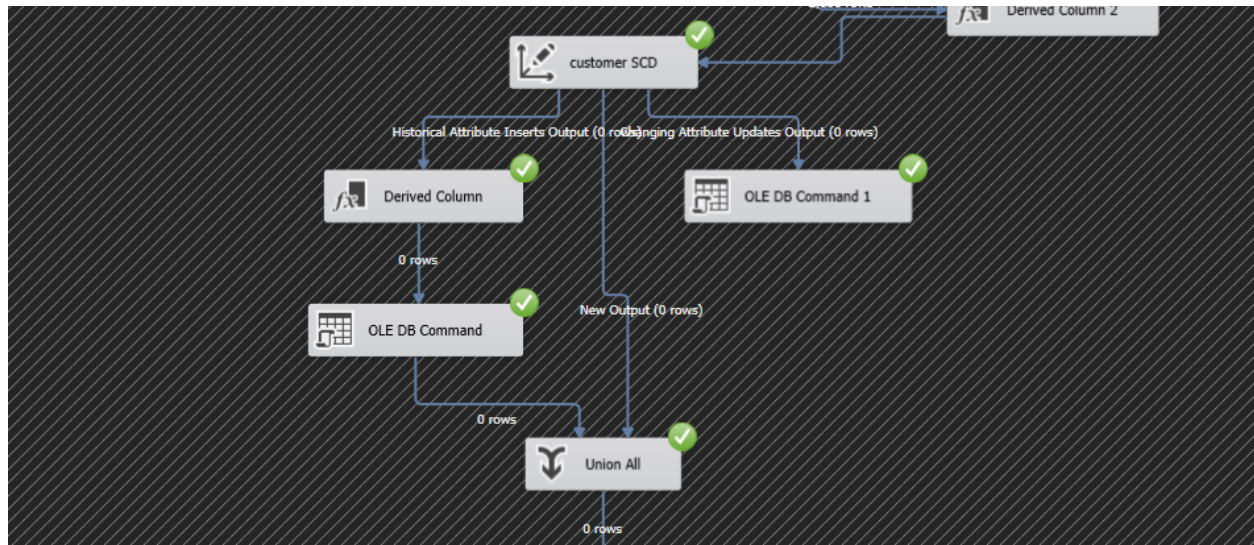


Transforming and loading

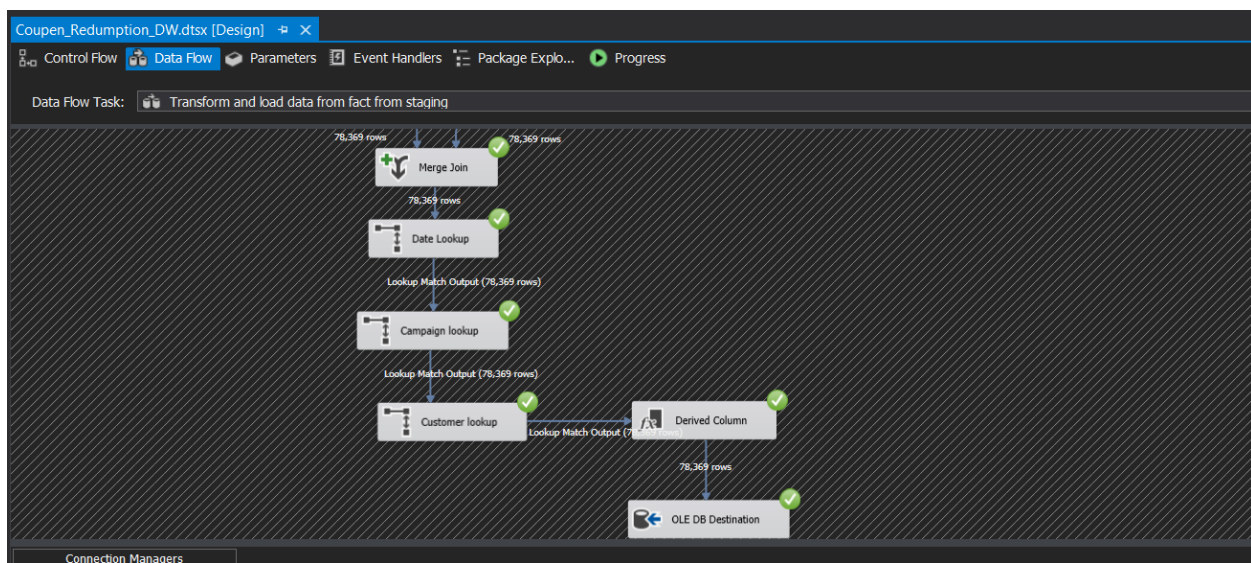
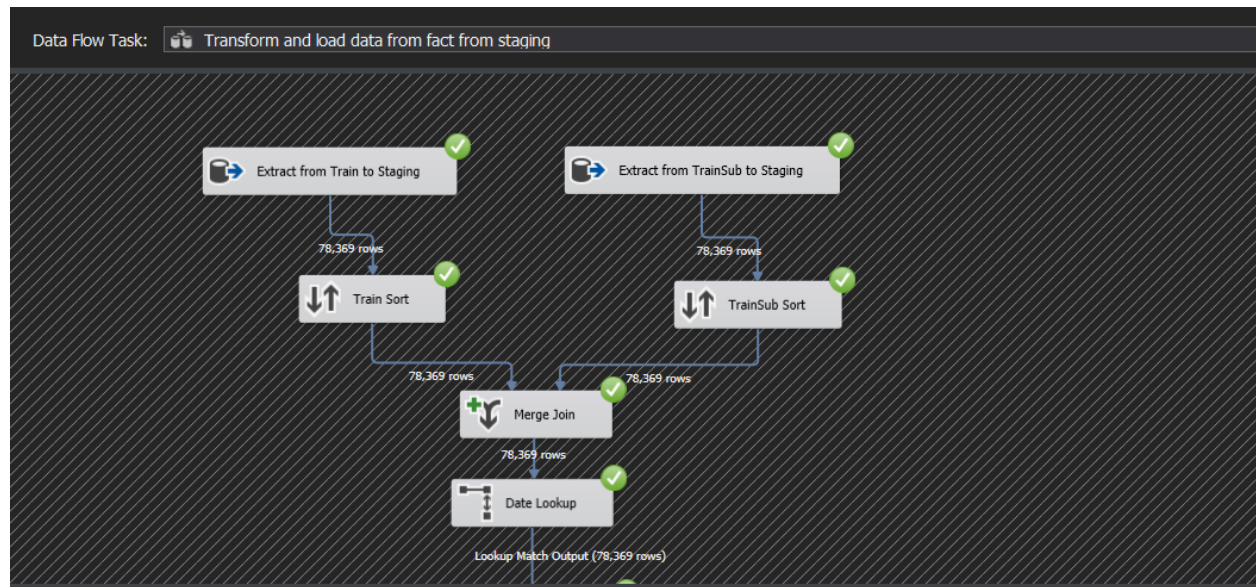


Transform and load dim campaign





Transform and load dim customer



Transform from fact train

Procedures

```
1
2 CREATE PROCEDURE dbo.UpdateDimCampaignData
3     @campaign_id int,
4     @campaign_type nvarchar(50)
5
6 AS
7 BEGIN
8     if not exists (select CampaignSK
9                     from dbo.DimCampaignData
10                    where Alternatcampaign_id = @campaign_id)
11     BEGIN
12         insert into dbo.DimCampaignData
13             (Alternatcampaign_id, campaign_type, InsertDate, ModifiedDate)
14             values
15             (@campaign_id, @campaign_type, GETDATE(), GETDATE())
16     END;
17     if exists (select CampaignSK
18               from dbo.DimCampaignData
19               where Alternatcampaign_id = @campaign_id)
20     BEGIN
21         update DimCampaignData
22             set campaign_type = @campaign_type,
23                 ModifiedDate = GETDATE()
24             where Alternatcampaign_id = @campaign_id
25     END;
26 END;
```

Dim campaign procedure

Ware house tables with values

| Results | | Messages | | | |
|---------|------------|---------------------|---------------|-------------------------|-------------------------|
| | CampaignSK | Alternatcampaign_id | campaign_type | InsertDate | ModifiedDate |
| 1 | 1 | 24 | Y | 2022-05-14 15:49:18.277 | 2022-05-17 20:38:22.480 |
| 2 | 2 | 25 | Y | 2022-05-14 15:49:18.280 | 2022-05-17 20:38:22.480 |
| 3 | 3 | 20 | Y | 2022-05-14 15:49:18.280 | 2022-05-17 20:38:22.483 |
| 4 | 4 | 23 | Y | 2022-05-14 15:49:18.280 | 2022-05-17 20:38:22.483 |
| 5 | 5 | 21 | Y | 2022-05-14 15:49:18.280 | 2022-05-17 20:38:22.483 |
| 6 | 6 | 22 | X | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.483 |
| 7 | 7 | 18 | X | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.483 |
| 8 | 8 | 19 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.487 |
| 9 | 9 | 17 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.487 |
| 10 | 10 | 16 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.487 |
| 11 | 11 | 13 | X | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.487 |
| 12 | 12 | 11 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.490 |
| 13 | 13 | 12 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.490 |
| 14 | 14 | 10 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.490 |
| 15 | 15 | 9 | Y | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.490 |
| 16 | 16 | 8 | X | 2022-05-14 15:49:18.283 | 2022-05-17 20:38:22.490 |

Dim campaign table

| Results | | Messages | | | | | | | | | | | | | |
|---------|------------|----------------------|-----------|----------------|--------|-------------|----------------|----------|--------|----------------|-----------------|-------------------------|---------|-------------------------|-----------------|
| | CustomerSK | Alternatecustomer_id | age_range | marital_status | rented | family_size | income_bracket | quantity | Total | other_discount | coupon_discount | StartDate | EndDate | InsertDate | ModifiedDate |
| 1 | 1 | 1 | 70+ | Married | 0 | 2 | 4 | 1 | 35.26 | 10.69 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 2 | 2 | 2 | 46-55 | Married | 0 | 2 | 5 | 1 | 106.50 | 14.25 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 3 | 3 | 3 | 26-35 | Married | 0 | 3 | 3 | 3 | 32.06 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 4 | 4 | 4 | 26-35 | Married | 0 | 4 | 6 | 1 | 81.57 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 5 | 5 | 5 | 46-55 | Single | 0 | 1 | 5 | 1 | 63.76 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 6 | 6 | 6 | 70+ | Single | 0 | 2 | 1 | 1 | 44.52 | 12.11 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 7 | 7 | 7 | 46-55 | Married | 0 | 2 | 7 | 1 | 106.50 | 35.62 | 35.62 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 8 | 8 | 8 | 36-45 | Single | 0 | 1 | 2 | 2 | 336.25 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 9 | 9 | 9 | 26-35 | Married | 1 | 2 | 6 | 1 | 35.26 | 17.81 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 10 | 10 | 10 | 46-55 | Married | 0 | 2 | 6 | 2 | 426.73 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 11 | 11 | 11 | 36-45 | Married | 0 | 1 | 5 | 1 | 47.73 | 8.90 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 12 | 12 | 12 | 46-55 | Married | 0 | 1 | 3 | 1 | 45.95 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 13 | 13 | 13 | 36-45 | Single | 0 | 2 | 4 | 1 | 82.99 | 80.50 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 14 | 14 | 14 | 36-45 | Married | 0 | 2 | 8 | 1 | 89.05 | 17.45 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 15 | 15 | 15 | 46-55 | Married | 0 | 2 | 1 | 1 | 45.95 | 32.06 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |
| 16 | 16 | 16 | 70+ | Married | 0 | 1 | 5 | 1 | 320.22 | 0.00 | 0.00 | 2022-05-14 17:03:12.000 | NULL | 2022-05-14 17:03:21.483 | 2022-05-14 17:0 |

Dim customer table

| Results | | Messages | | | | | | | | | | | | | |
|---------|----------|-------------------------|------------|-------------|------------|-----------|-----------|--------------|-------------|------------------|-----------------|--------------|-----------|-------------|---------------|
| | DateKey | Date | FullDateUK | FullDateUSA | DayOfMonth | DaySuffix | DayName | DayOfWeekUSA | DayOfWeekUK | DayOfWeekInMonth | DayOfWeekInYear | DayOfQuarter | DayOfYear | WeekOfMonth | WeekOfQuarter |
| 1 | 19900101 | 1990-01-01 00:00:00.000 | 01/01/1990 | 01/01/1990 | 1 | 1st | Monday | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 19900102 | 1990-01-02 00:00:00.000 | 02/01/1990 | 01/02/1990 | 2 | 2nd | Tuesday | 3 | 2 | 1 | 1 | 1 | 2 | 1 | 1 |
| 3 | 19900103 | 1990-01-03 00:00:00.000 | 03/01/1990 | 01/03/1990 | 3 | 3rd | Wednesday | 4 | 3 | 1 | 1 | 1 | 3 | 1 | 1 |
| 4 | 19900104 | 1990-01-04 00:00:00.000 | 04/01/1990 | 01/04/1990 | 4 | 4th | Thursday | 5 | 4 | 1 | 1 | 1 | 4 | 1 | 1 |
| 5 | 19900105 | 1990-01-05 00:00:00.000 | 05/01/1990 | 01/05/1990 | 5 | 5th | Friday | 6 | 5 | 1 | 1 | 1 | 5 | 1 | 1 |
| 6 | 19900106 | 1990-01-06 00:00:00.000 | 06/01/1990 | 01/06/1990 | 6 | 6th | Saturday | 7 | 6 | 1 | 1 | 1 | 6 | 1 | 1 |
| 7 | 19900107 | 1990-01-07 00:00:00.000 | 07/01/1990 | 01/07/1990 | 7 | 7th | Sunday | 1 | 7 | 1 | 1 | 1 | 7 | 2 | 1 |
| 8 | 19900108 | 1990-01-08 00:00:00.000 | 08/01/1990 | 01/08/1990 | 8 | 8th | Monday | 2 | 1 | 2 | 2 | 2 | 8 | 2 | 2 |
| 9 | 19900109 | 1990-01-09 00:00:00.000 | 09/01/1990 | 01/09/1990 | 9 | 9th | Tuesday | 3 | 2 | 2 | 2 | 2 | 9 | 2 | 2 |
| 10 | 19900110 | 1990-01-10 00:00:00.000 | 10/01/1990 | 01/10/1990 | 10 | 10th | Wednesday | 4 | 3 | 2 | 2 | 2 | 10 | 2 | 2 |
| 11 | 19900111 | 1990-01-11 00:00:00.000 | 11/01/1990 | 01/11/1990 | 11 | 11th | Thursday | 5 | 4 | 2 | 2 | 2 | 11 | 2 | 2 |
| 12 | 19900112 | 1990-01-12 00:00:00.000 | 12/01/1990 | 01/12/1990 | 12 | 12th | Friday | 6 | 5 | 2 | 2 | 2 | 12 | 2 | 2 |
| 13 | 19900113 | 1990-01-13 00:00:00.000 | 13/01/1990 | 01/13/1990 | 13 | 13th | Saturday | 7 | 6 | 2 | 2 | 2 | 13 | 2 | 2 |
| 14 | 19900114 | 1990-01-14 00:00:00.000 | 14/01/1990 | 01/14/1990 | 14 | 14th | Sunday | 1 | 7 | 2 | 2 | 2 | 14 | 3 | 2 |
| 15 | 19900115 | 1990-01-15 00:00:00.000 | 15/01/1990 | 01/15/1990 | 15 | 15th | Monday | 2 | 1 | 3 | 3 | 3 | 15 | 3 | 3 |
| 16 | 19900116 | 1990-01-16 00:00:00.000 | 16/01/1990 | 01/16/1990 | 16 | 16th | Tuesday | 3 | 2 | 3 | 3 | 3 | 16 | 3 | 3 |

Dim date table

| Results Messages | | | | | | | | | |
|------------------|-------------|-------------|--------------|-------------------|----|----------------|-------------|-------------------------|-------------------------|
| | campaignKey | customerKey | traindatekey | redemption_status | id | category | brand_Type | InsertDate | ModifiedDate |
| 1 | 11 | 1053 | 20120102 | 0 | 1 | Grocery | Established | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 2 | 11 | 48 | 20120102 | 0 | 2 | Miscellaneous | Established | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 3 | 15 | 205 | 20120102 | 0 | 6 | Bakery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 4 | 11 | 1050 | 20120102 | 0 | 7 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 5 | 16 | 1489 | 20120102 | 0 | 9 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 6 | 12 | 793 | 20120102 | 0 | 11 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 7 | 15 | 590 | 20120102 | 0 | 14 | Pharmaceutical | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 8 | 25 | 368 | 20120102 | 0 | 15 | Bakery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 9 | 24 | 523 | 20120102 | 0 | 17 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 10 | 23 | 679 | 20120102 | 0 | 19 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 11 | 15 | 108 | 20120102 | 0 | 20 | Pharmaceutical | Established | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 12 | 11 | 89 | 20120102 | 0 | 21 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 13 | 15 | 712 | 20120102 | 0 | 22 | Grocery | Established | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 14 | 11 | 1067 | 20120102 | 0 | 23 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 15 | 20 | 761 | 20120102 | 0 | 24 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 16 | 11 | 248 | 20120102 | 0 | 25 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |
| 17 | 12 | 1464 | 20120102 | 0 | 26 | Grocery | Local | 2022-05-15 13:09:40.180 | 2022-05-15 13:09:40.180 |

Fact train table