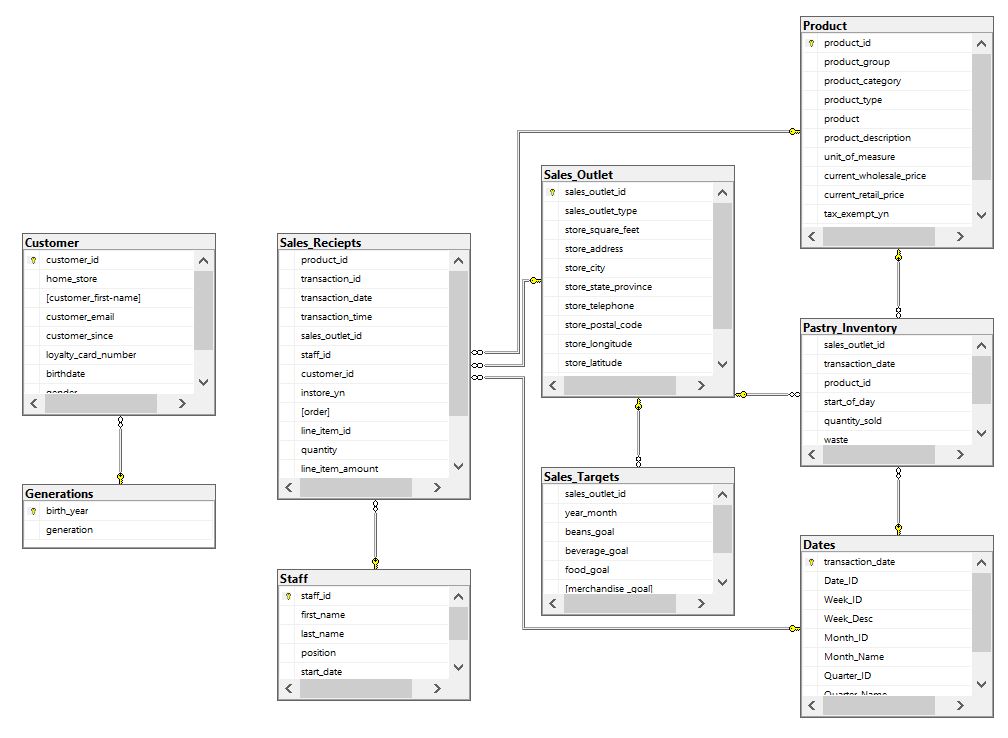
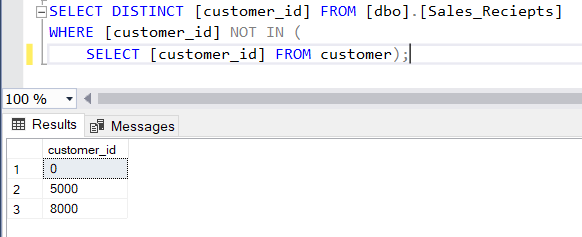
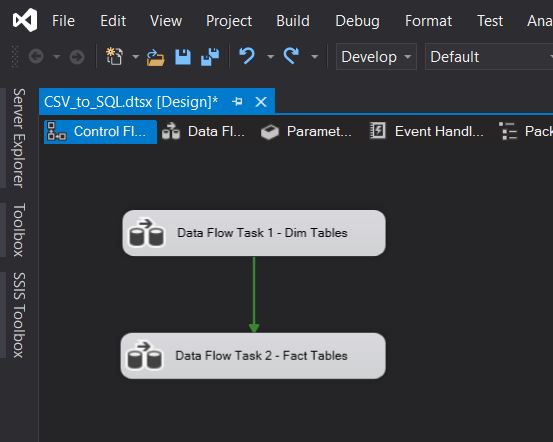
* Design and create data warehouse tables schema (design star schema), (use database MySQL, MSSQL, PostgreSQL)



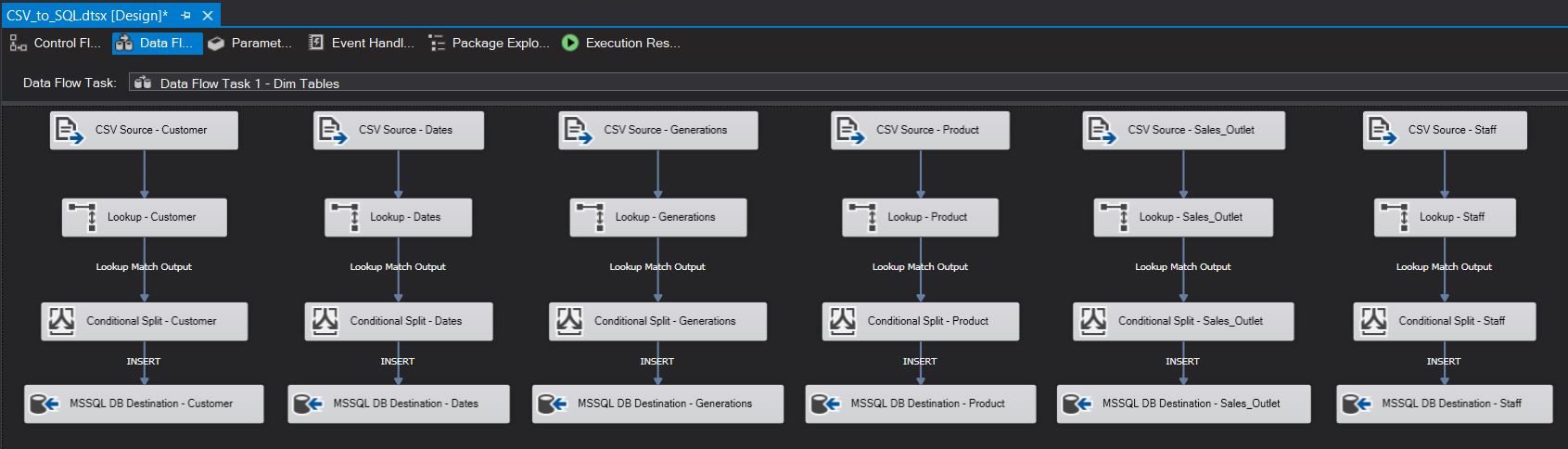
--Customer id values 0, 5000, 8000 in Sales\_Reciepts does not exist in Customer table and because of this we can not create foreign key relation between these two tables but we could create this relation on Power BI manually by filtering these customer id’s.



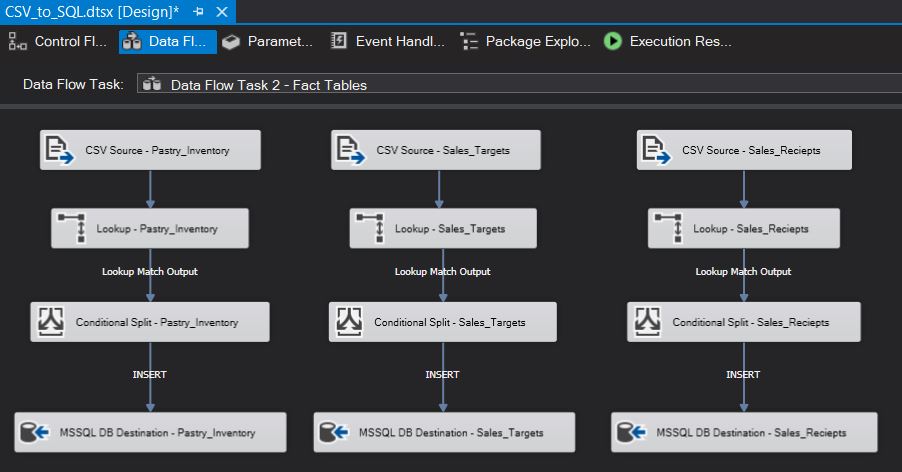
* Design and implement ETL to extract data from CSV into the data warehouse tables (better list the ETL options: nodeJS, SSIS, Talend, Apache NiFi, Informatica)



SSIS Workflows Screen Shot



SSIS Load Tasks For Dimension Tables



SSIS Load Tasks For Fact Tables

* What are the customers propensity to buy according to their generation?

-- Sold Quantities with Product Group by Generations

SELECT GNR.[generation] AS [Generation], PRD.[product\_group] AS [ProductGroup], SUM(SRC.[quantity]) AS [TotalQuantity]

FROM [dbo].[Sales\_Reciepts] SRC

INNER JOIN [dbo].[Customer] CST ON CST.[customer\_id] = SRC.[customer\_id]

LEFT JOIN [dbo].[Generations] GNR ON CST.[birth\_year] = GNR.[birth\_year]

LEFT JOIN [dbo].[Product] PRD ON SRC.[product\_id] = PRD.[product\_id]

GROUP BY GNR.[generation], PRD.[product\_group]

ORDER BY GNR.[generation]



* What is the most 3 seller products according to weeks?

-- TOP 3 Most Seller Products According to Weeks

SELECT [Product], [YearWeek], [TotalQuantity] FROM (

SELECT

ROW\_NUMBER() OVER(PARTITION BY CONCAT(DTE.[Year\_ID],DTE.[Week\_ID]) ORDER BY SUM(SRC.[quantity]) DESC) AS [RowNumber]

,PRD.[product] AS [Product]

,CONCAT(DTE.[Year\_ID],DTE.[Week\_ID]) AS [YearWeek]

,SUM(SRC.[quantity]) AS [TotalQuantity]

FROM [dbo].[Sales\_Reciepts] SRC

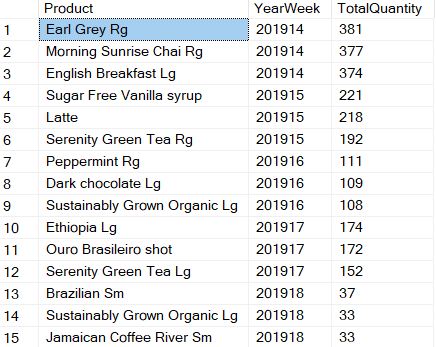
INNER JOIN [dbo].[Customer] CST ON CST.[customer\_id] = SRC.[customer\_id]

LEFT JOIN [dbo].[Product] PRD ON SRC.[product\_id] = PRD.[product\_id]

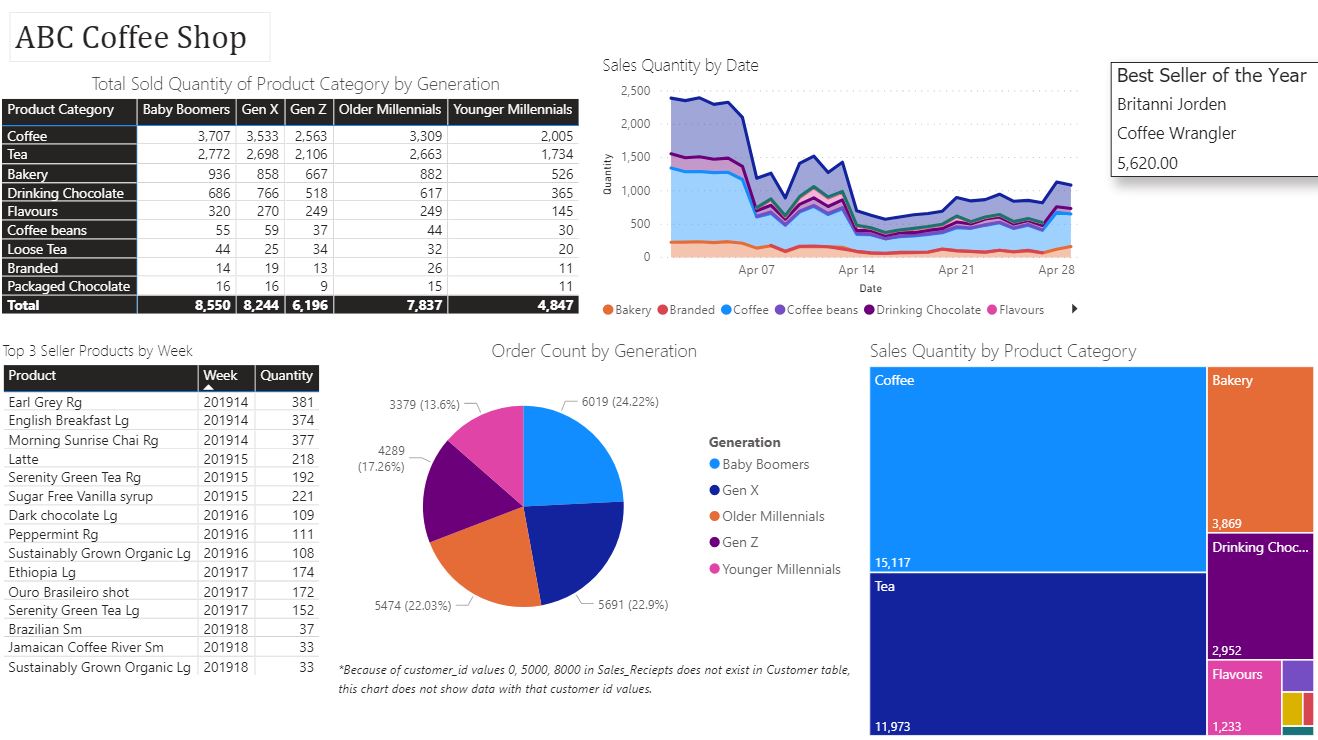
LEFT JOIN [dbo].[Dates] DTE ON SRC.[transaction\_date] = DTE.[transaction\_date]

GROUP BY CONCAT(DTE.[Year\_ID],DTE.[Week\_ID]), PRD.[product] ) TBL

WHERE [RowNumber] IN (1, 2, 3)



* Implement dashboard for data analysis (on Tableau, Power BI, Qlikview)



* Implement incremental load ETL job that takes new inserted sales in the source (folder contains CSV files) and add only those new sells in the database

