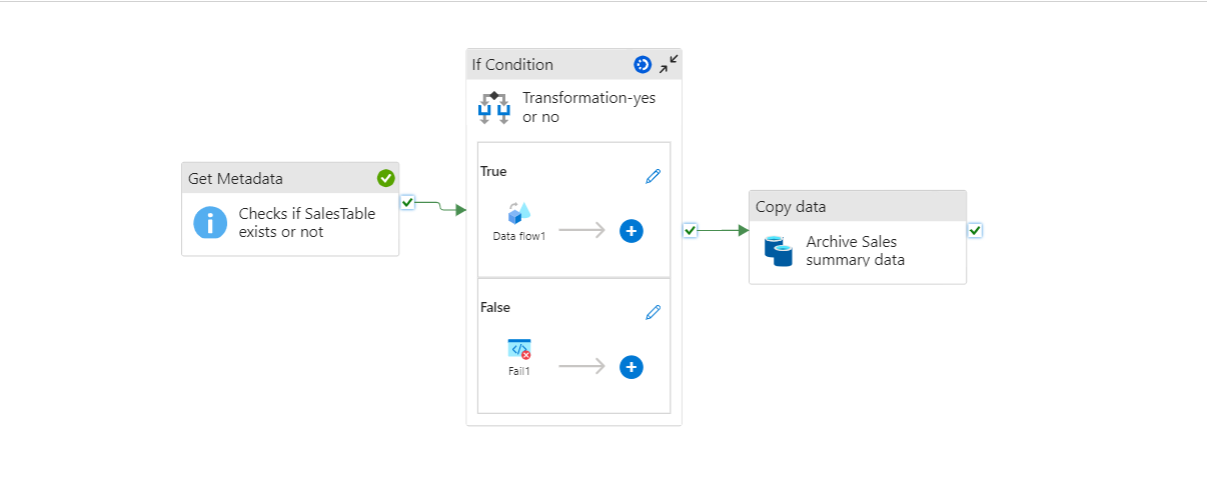
# PROJECT 4 – Create an Azure data factory Pipeline with data flow activity

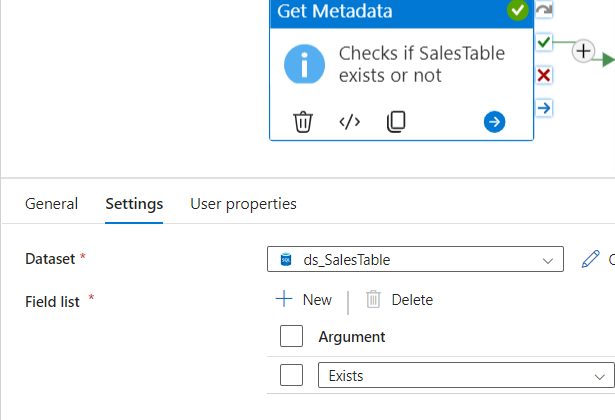
## **Scenario/Problem:** To obtain average and total sales of the products in year 2023 using sales and product data.

## **Solution:** Create a **ADF** pipeline that extract the sales data, transforms it and load the sales summary data at the sink. For that we need a set of activities.

* Source: Azure SQL DB tables(sales and product tables)
* Sink: Azure SQL DB table( sales summary table)

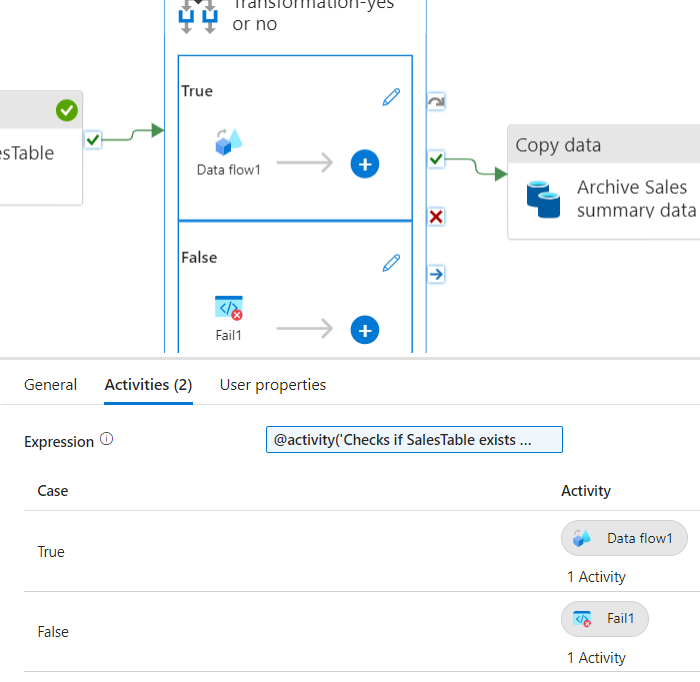


1. Get Metadata Activity:

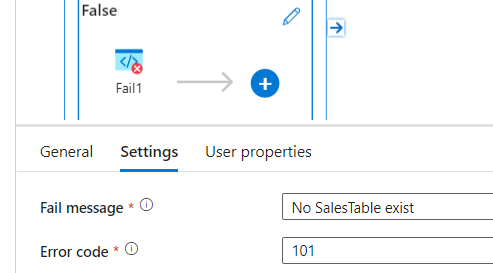


* Dataset: SalesTable
* Argument: Exists - This field will return a Boolean value indicating whether the item exists or not
* True – item exists
* False – item doesn’t exists

1. If Condition Activity:

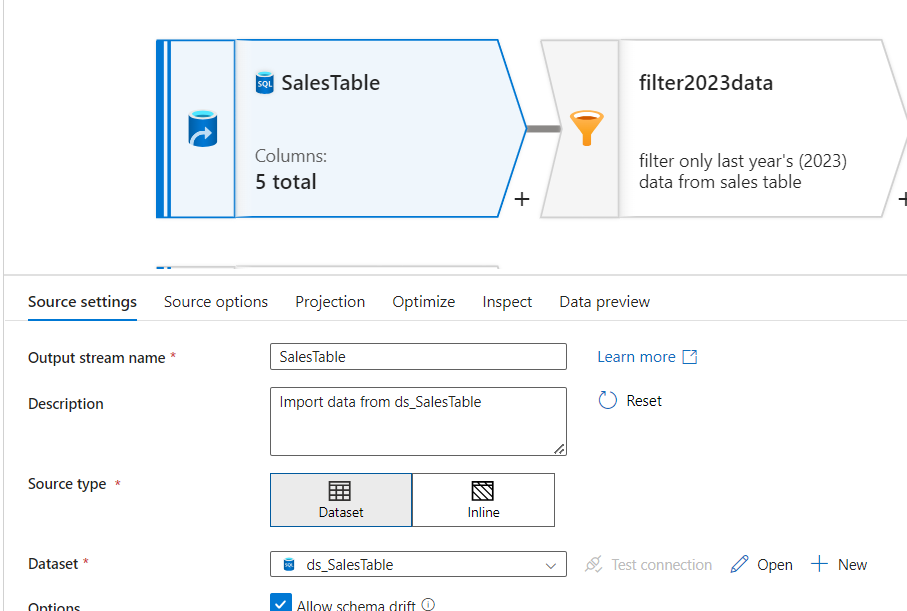


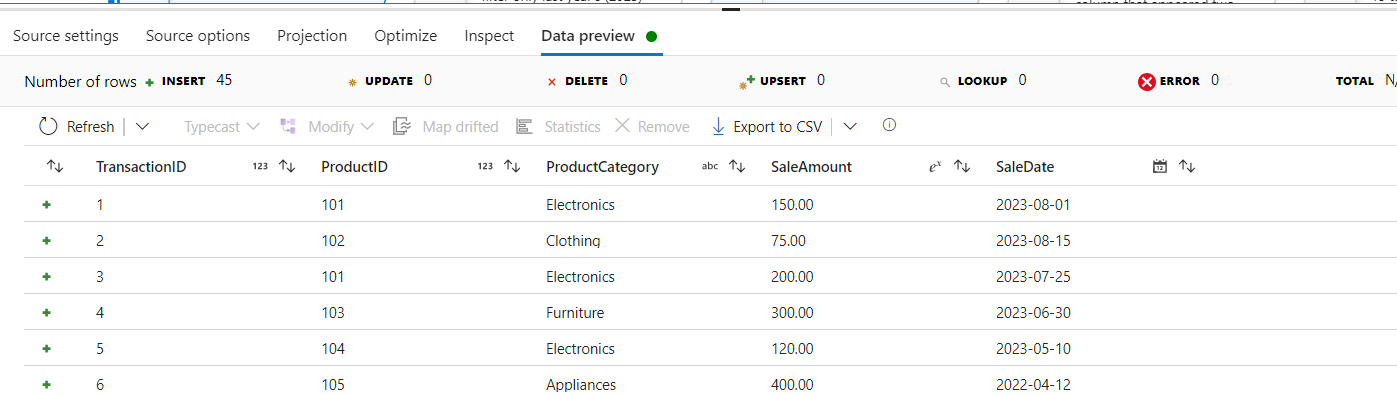
* There are 2 activities inside If Condition Activity:
* If condition is true >> Dataflow activity
* If condition is False >> Fail activity: which says no SalesTable Exists with an error code.



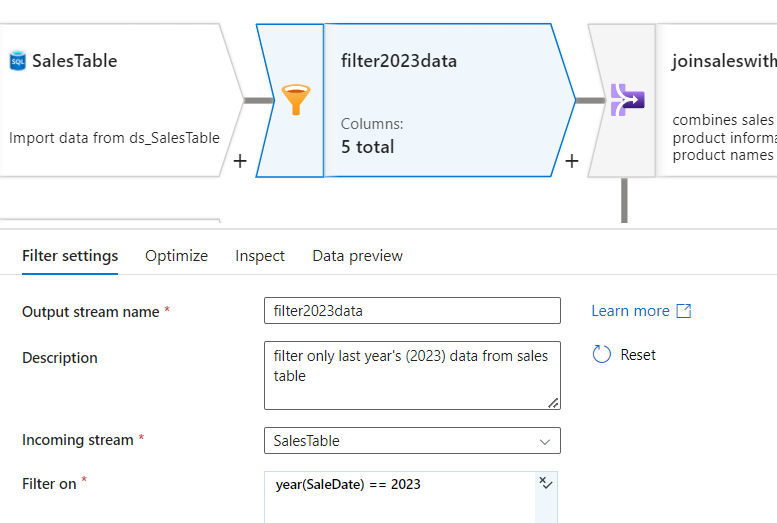
**Dataflow activity : To obtain total and avg sales of products based on product category**

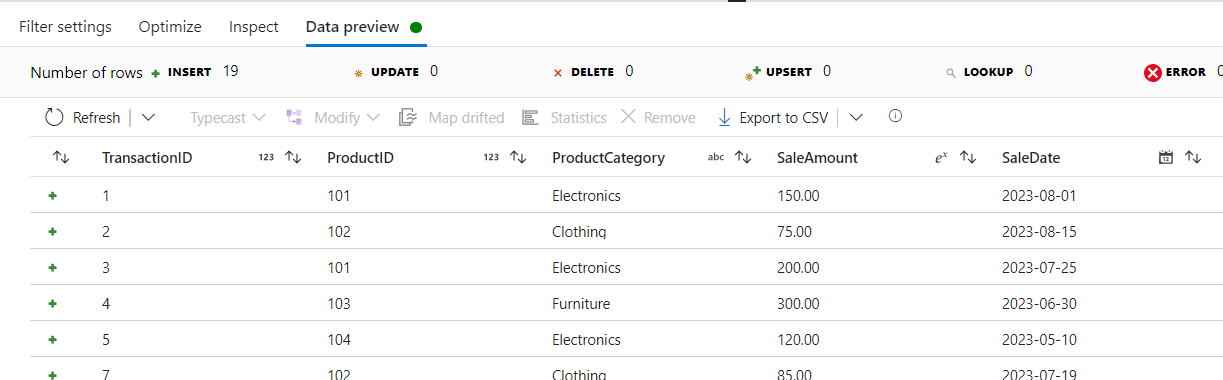
* Source:



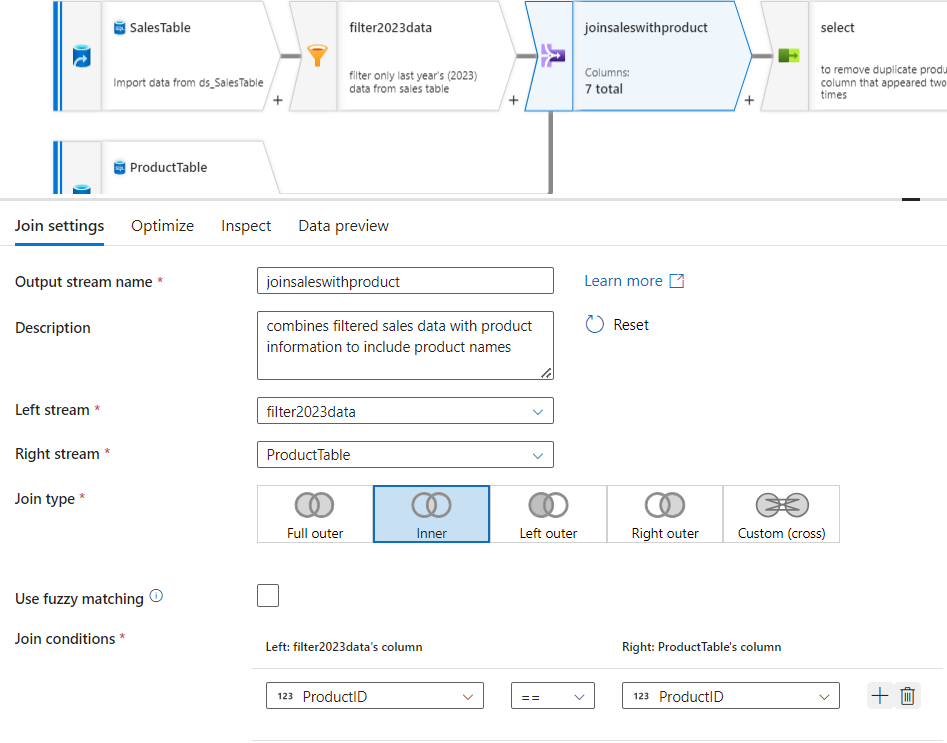


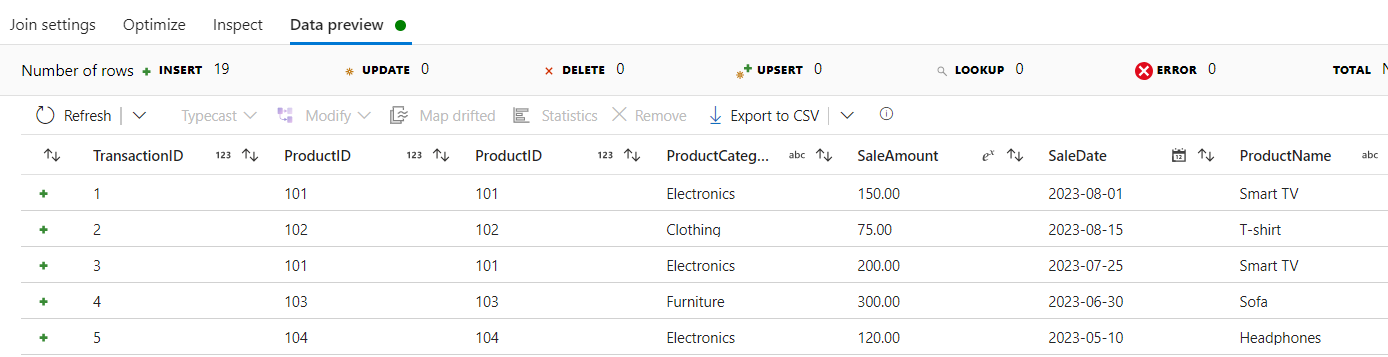
* Filter:

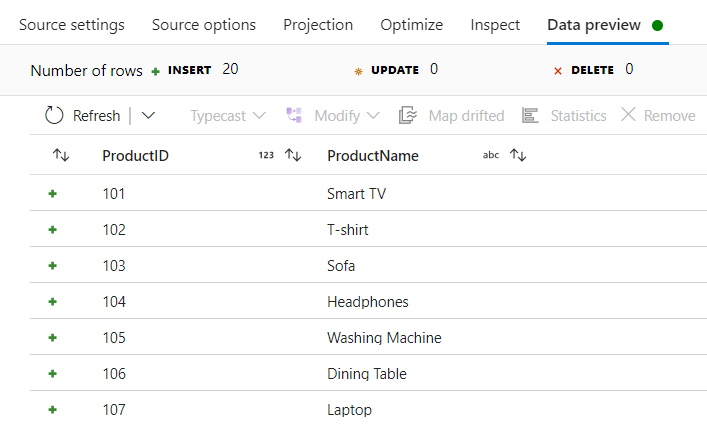




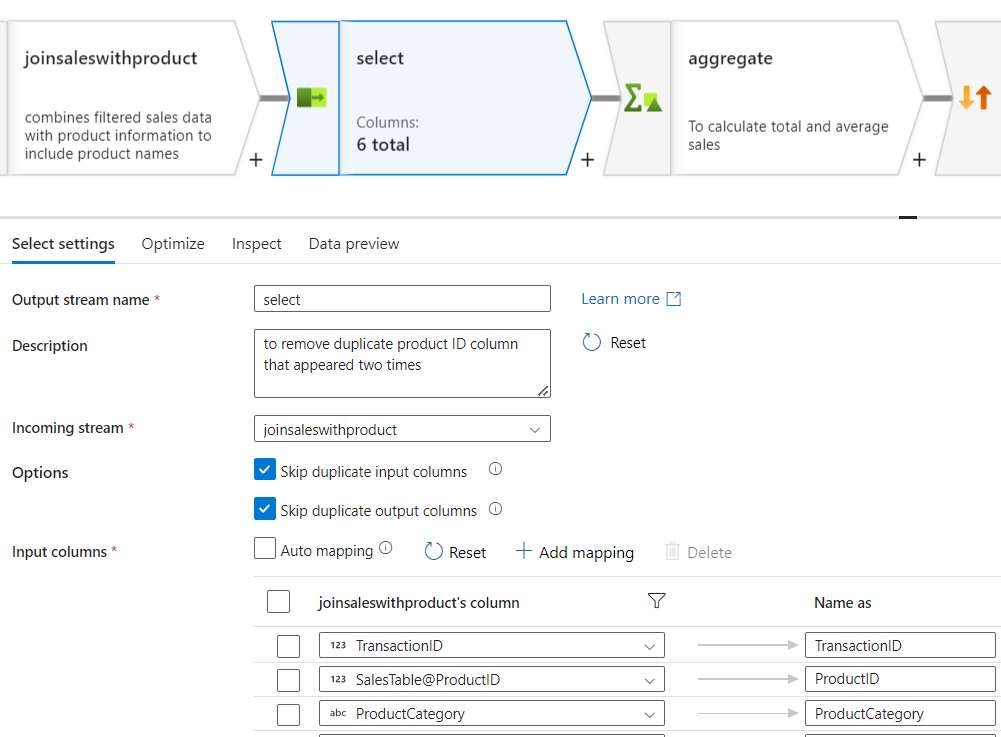
* Join:

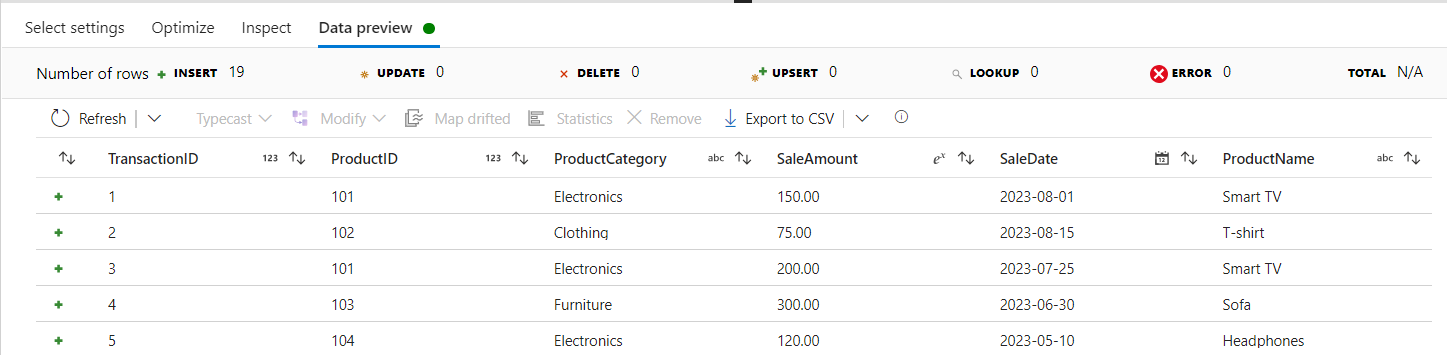




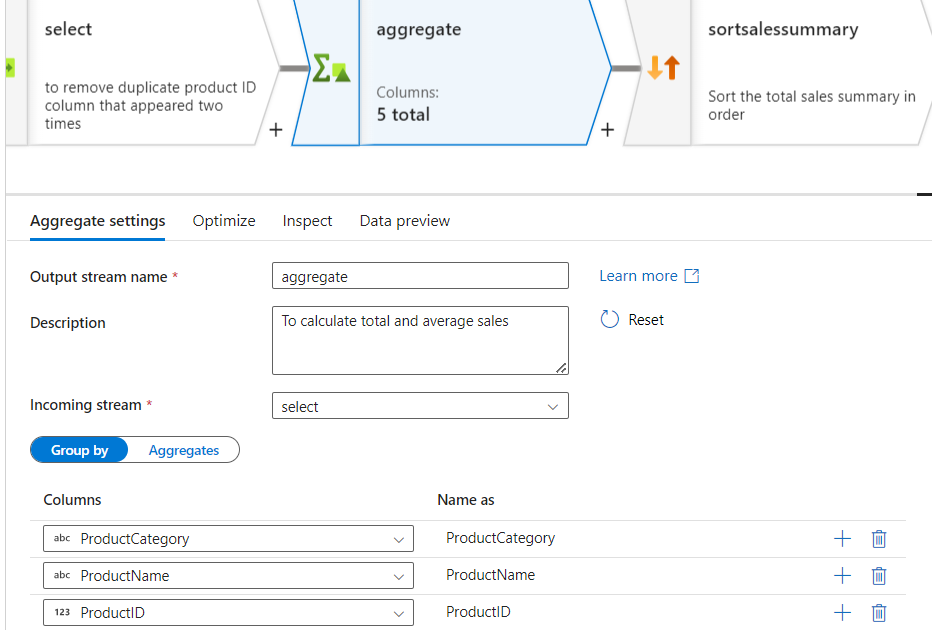
 : the productTable

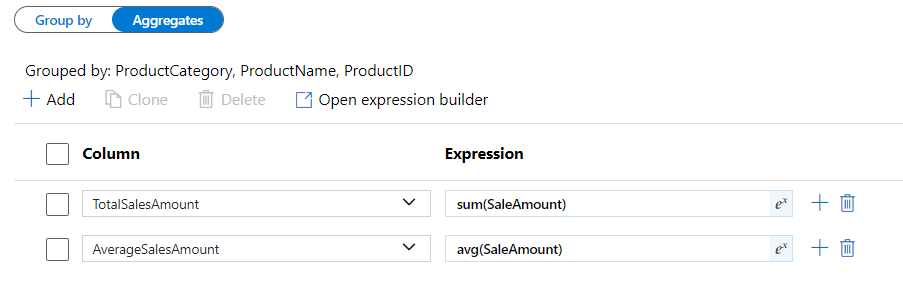
* Select:

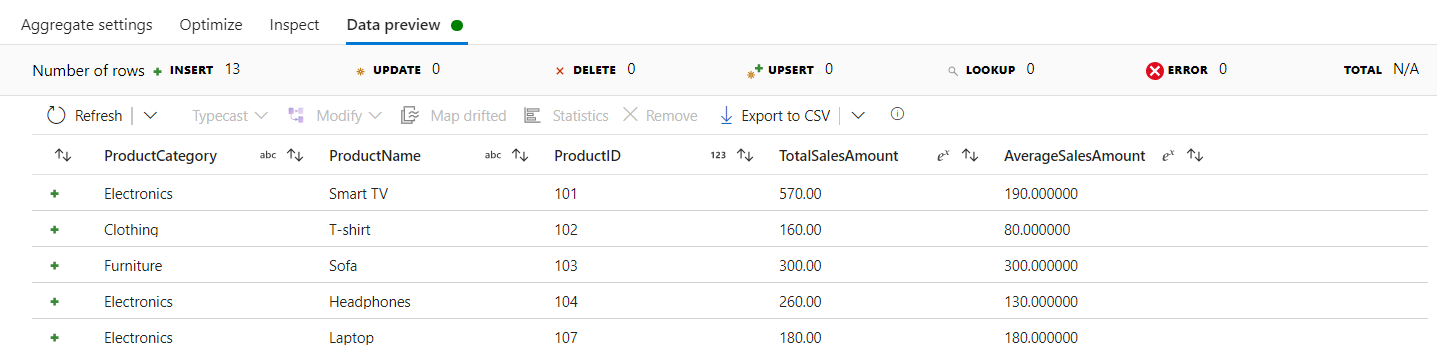




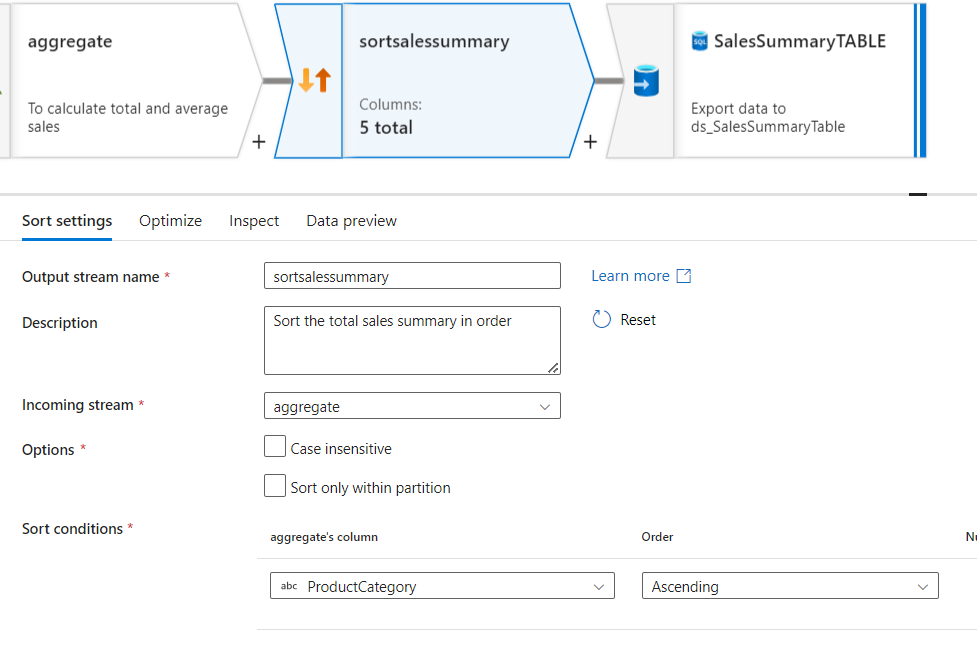
* Aggregate:

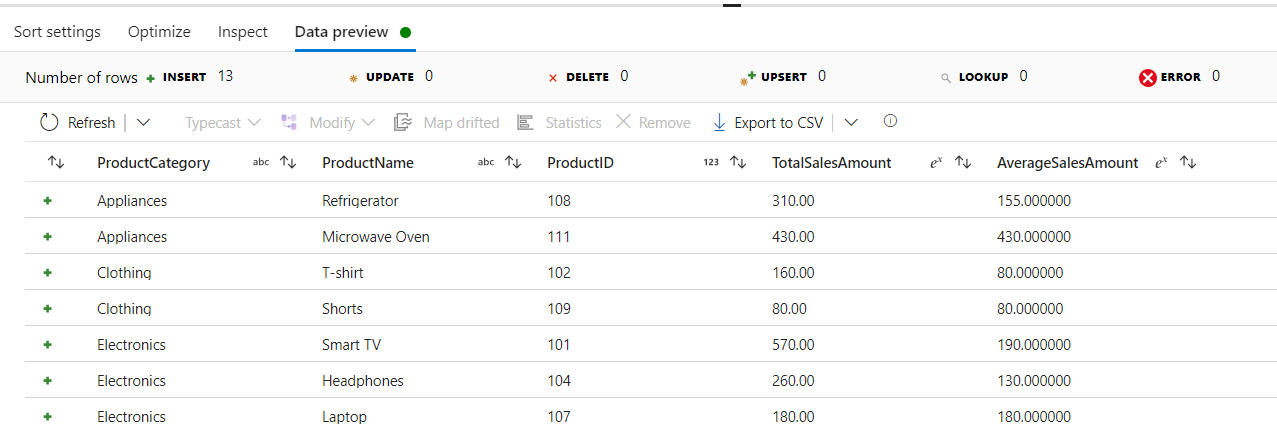




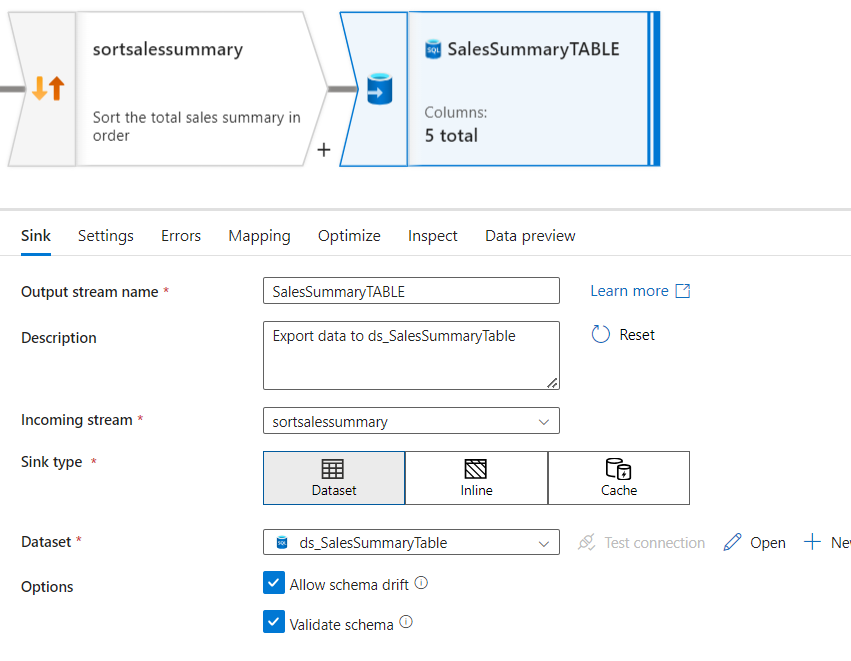


* Sort:

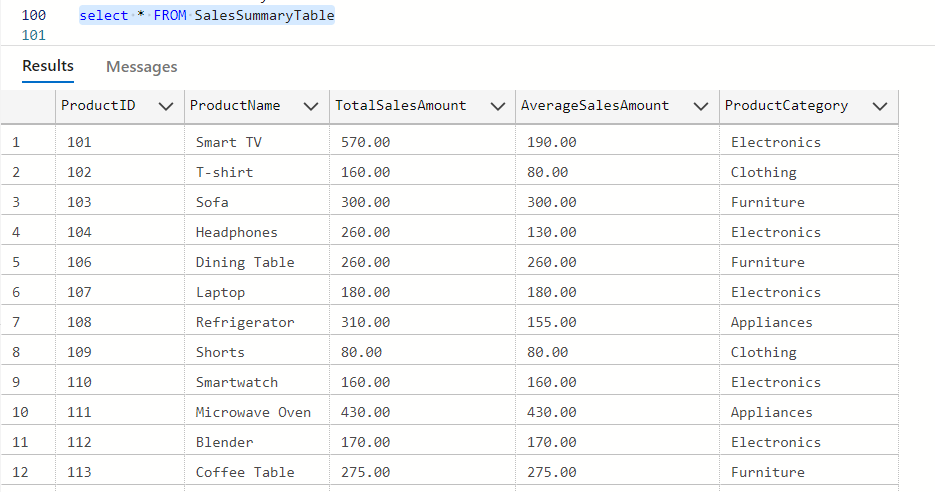




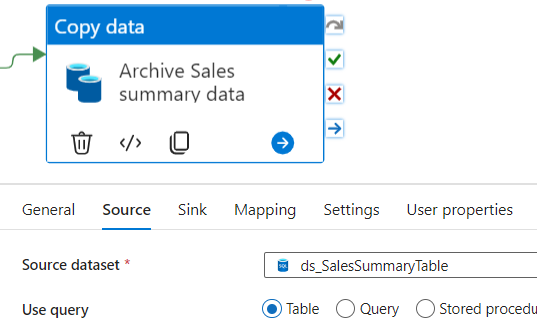
* Sink:

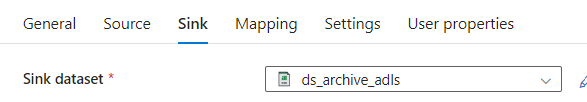


The final Sales Summary Table in Azure sql DB

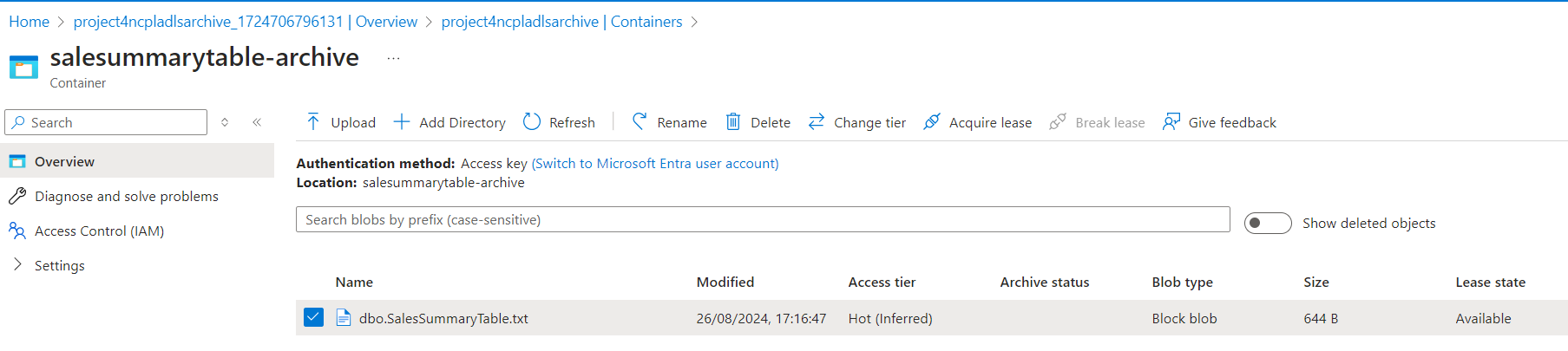


1. Copy Data Activity:





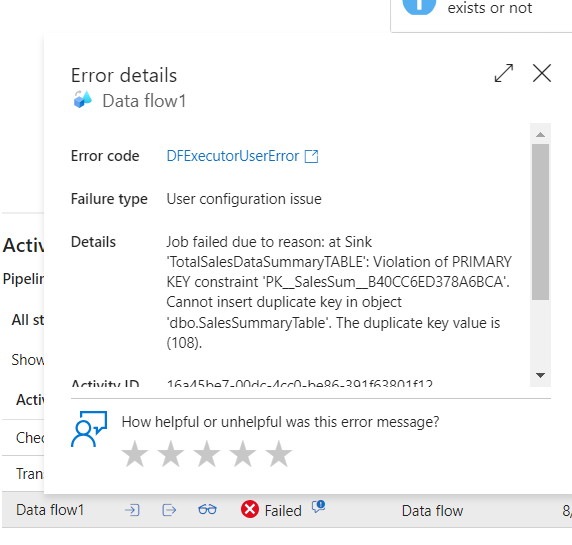
* To archive the Sales summary data we copy it to a container in ADLS .It is like creating back up of transformed data.

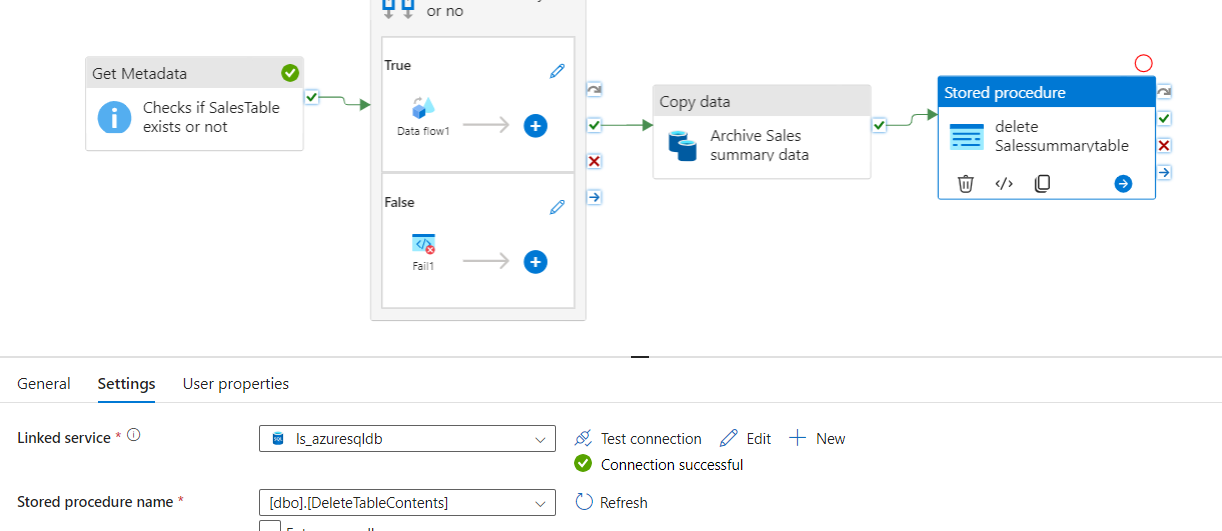


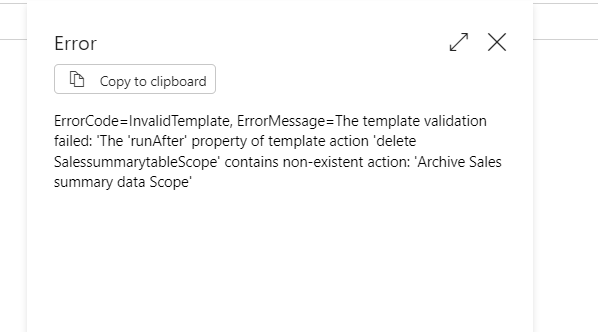
## ERRORS FACED:

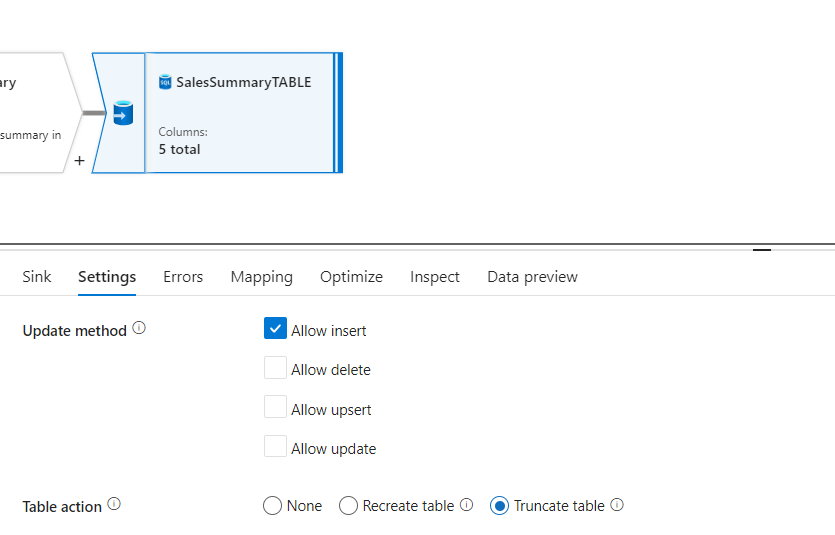
1. Pipeline failed when I ran it the second time:

* Issue:



* Solution: This happened data was already present in table due to previous run. So the old data has to be deleted to insert new records that have same primary key values.
* Tried using a stored procedure activity to delete the table contents after archiving it, but it showed some error. Will work on it.
* 





Another solution to delete the rows of Sales summary table is to Truncate it. Under settings of sink tab in data flow, we have option to truncate table while writing new records to it.

