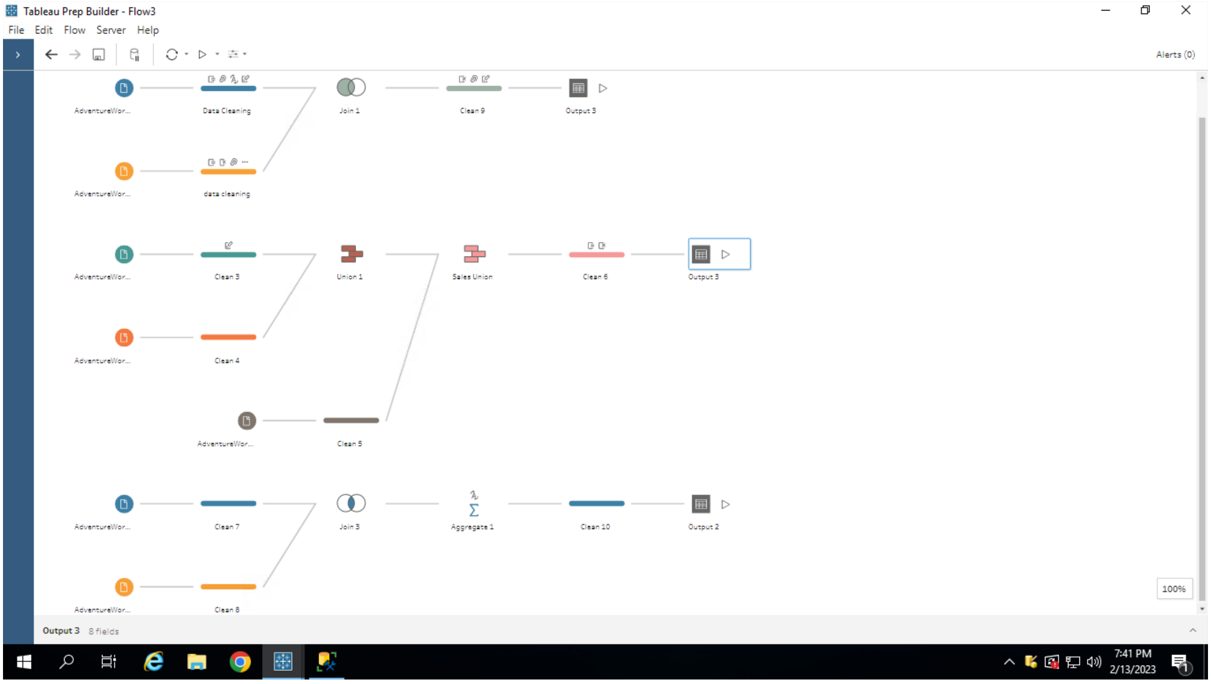
FUNDAMENTALS OF DATA ENGINEERING.

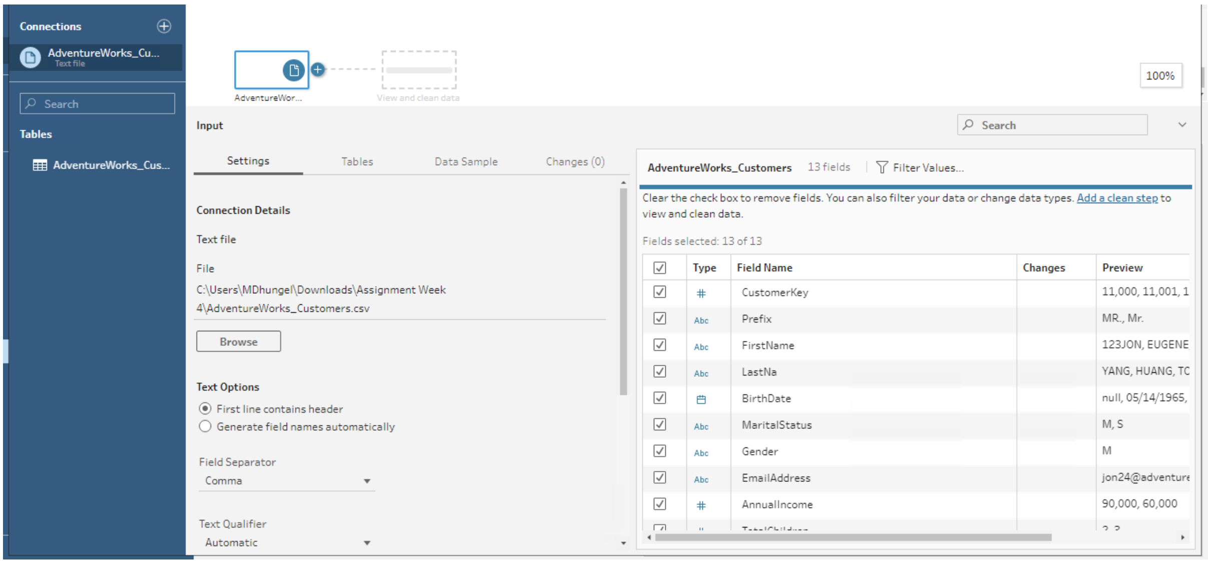
MAHIM DHUNGEL

Overview: ETL Processes in Tableau



Phase 1

1. Connect to Customer file



Input

Connection Details

Text file

File
C:\Users\MDhungel\Downloads\Assignment Week 4\AdventureWorks_Customers.csv

Browse

Text Options

● First line contains header
○ Generate field names automatically

Field Separator
Comma

Text Qualifier
Automatic

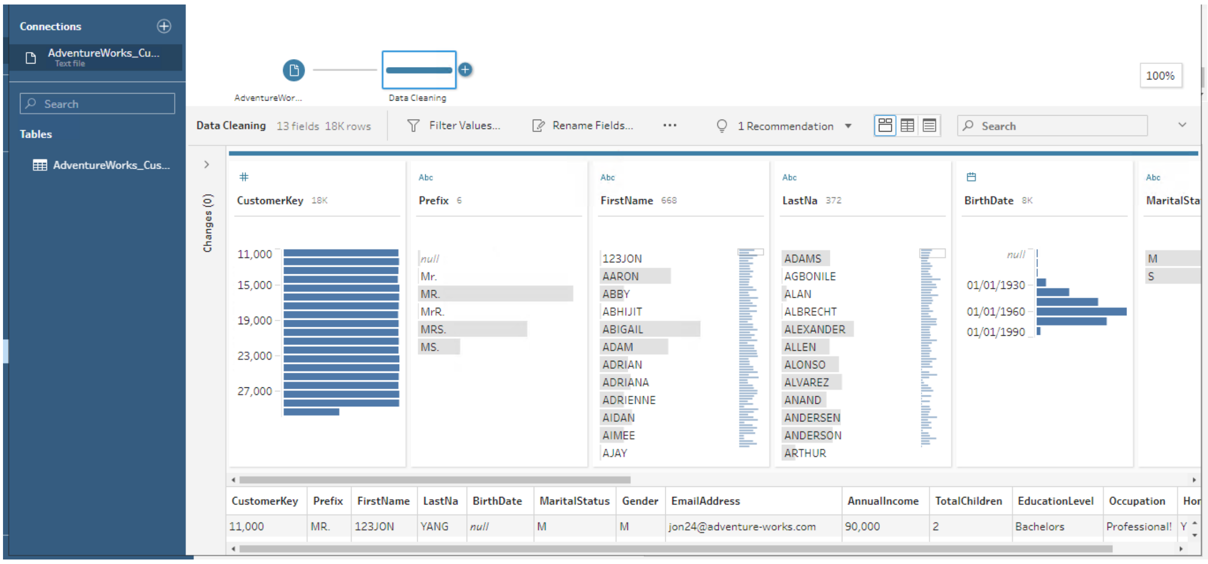
AdventureWorks_Customers 13 fields | Filter Values...

Clear the check box to remove fields. You can also filter your data or change data types. Add a clean step to view and clean data.

Fields selected: 13 of 13

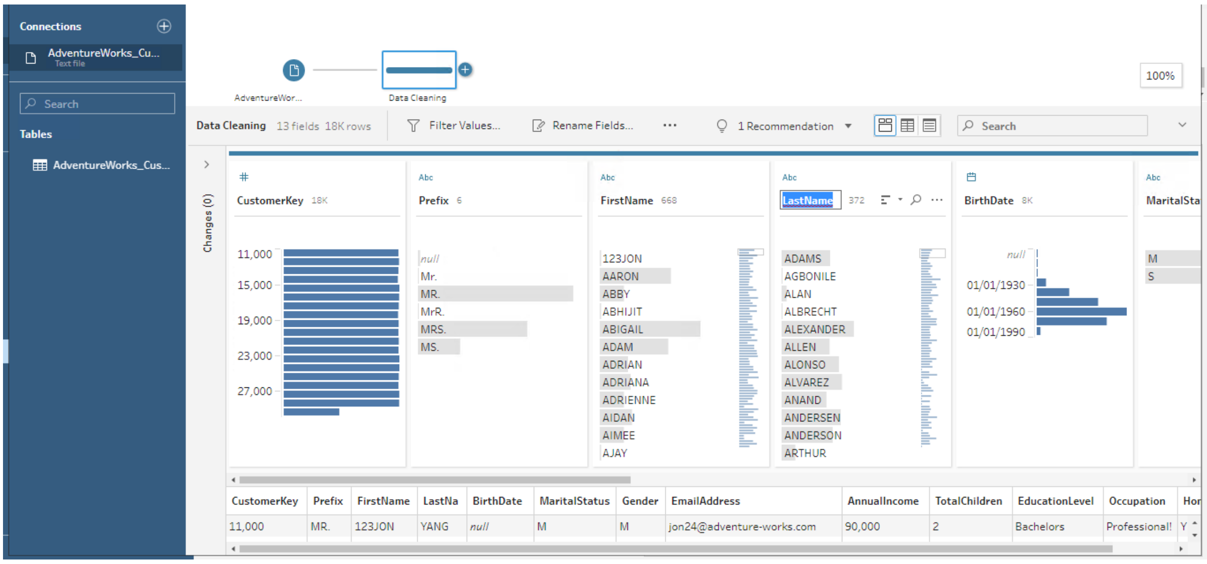
☑	Type	Field Name	Changes	Preview
☑	#	CustomerKey		11,000, 11,001, 1
☑	Abc	Prefix		MR., Mr.
☑	Abc	FirstName		123JON, EUGENE
☑	Abc	LastNa		YANG, HUANG, TC
☑	📅	BirthDate		null, 05/14/1965,
☑	Abc	MaritalStatus		M, S
☑	Abc	Gender		M
☑	Abc	EmailAddress		jon24@adventure
☑	#	AnnualIncome		90,000, 60,000

2. Add a “Clean Step



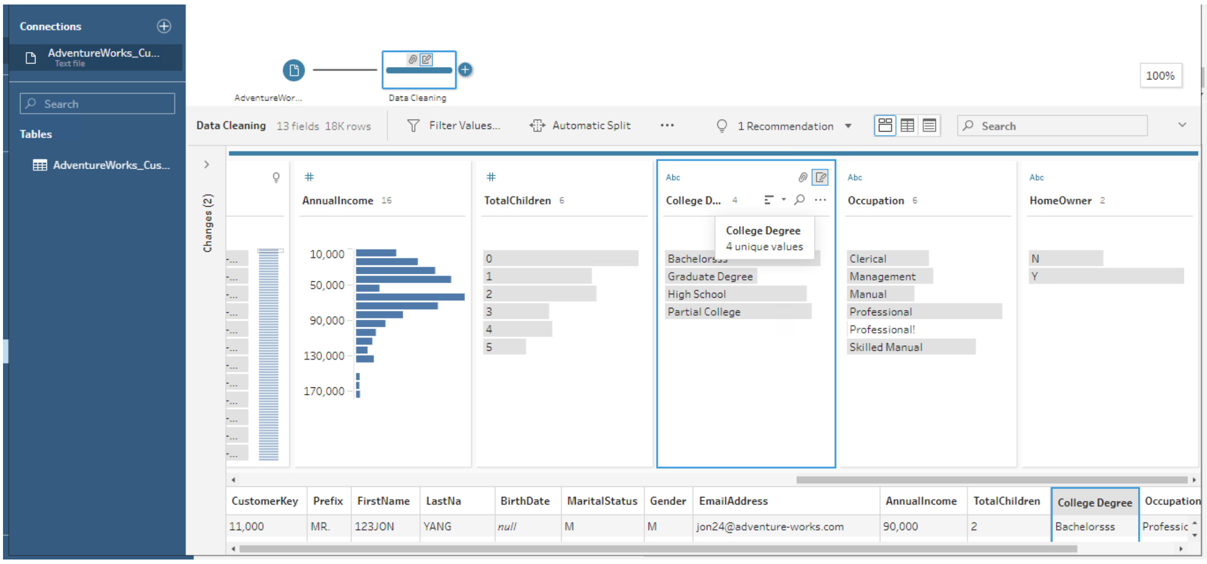
CustomerKey	Prefix	FirstName	LastNa	BirthDate	MaritalStatus	Gender	EmailAddress	AnnualIncome	TotalChildren	EducationLevel	Occupation	Hor
11,000	MR.	123JON	YANG	*null*	M	M	jon24@adventure-works.com	90,000	2	Bachelors	Professional!	Y

3. Rename Field LastNa to LastName



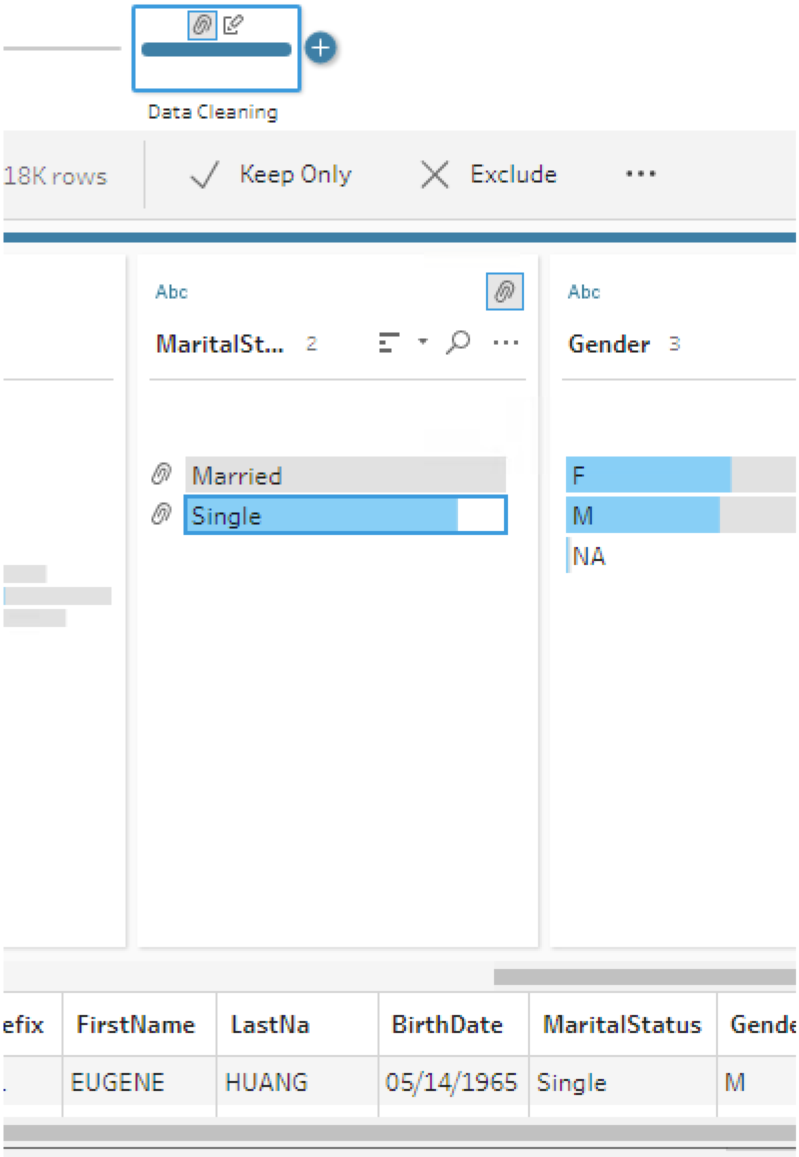
CustomerKey	Prefix	FirstName	LastNa	BirthDate	MaritalStatus	Gender	EmailAddress	AnnualIncome	TotalChildren	EducationLevel	Occupation	Hor
11,000	MR.	123JON	YANG	*null*	M	M	jon24@adventure-works.com	90,000	2	Bachelors	Professional!	Y

4. Group Education and Replace by College Degree



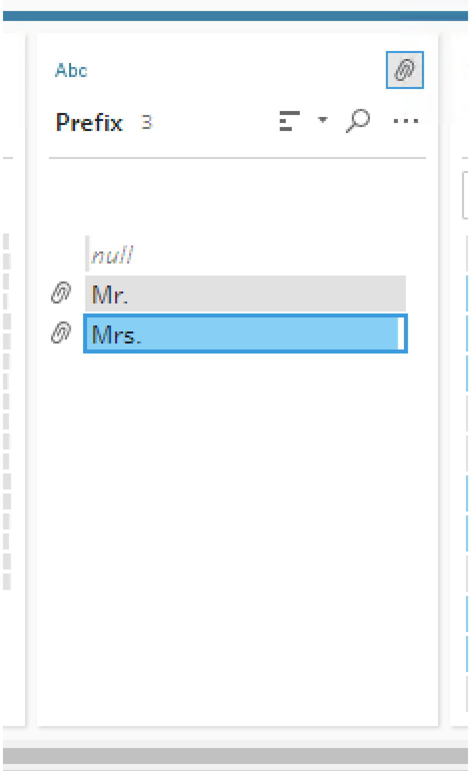
CustomerKey	Prefix	FirstName	LastNa	BirthDate	MaritalStatus	Gender	EmailAddress	AnnualIncome	TotalChildren	College Degree	Occupation
11,000	MR.	123JON	YANG	*null*	M	M	jon24@adventure-works.com	90,000	2	Bachelorsss	Professic

5. Change marital status M and S to “Married” and “Single”

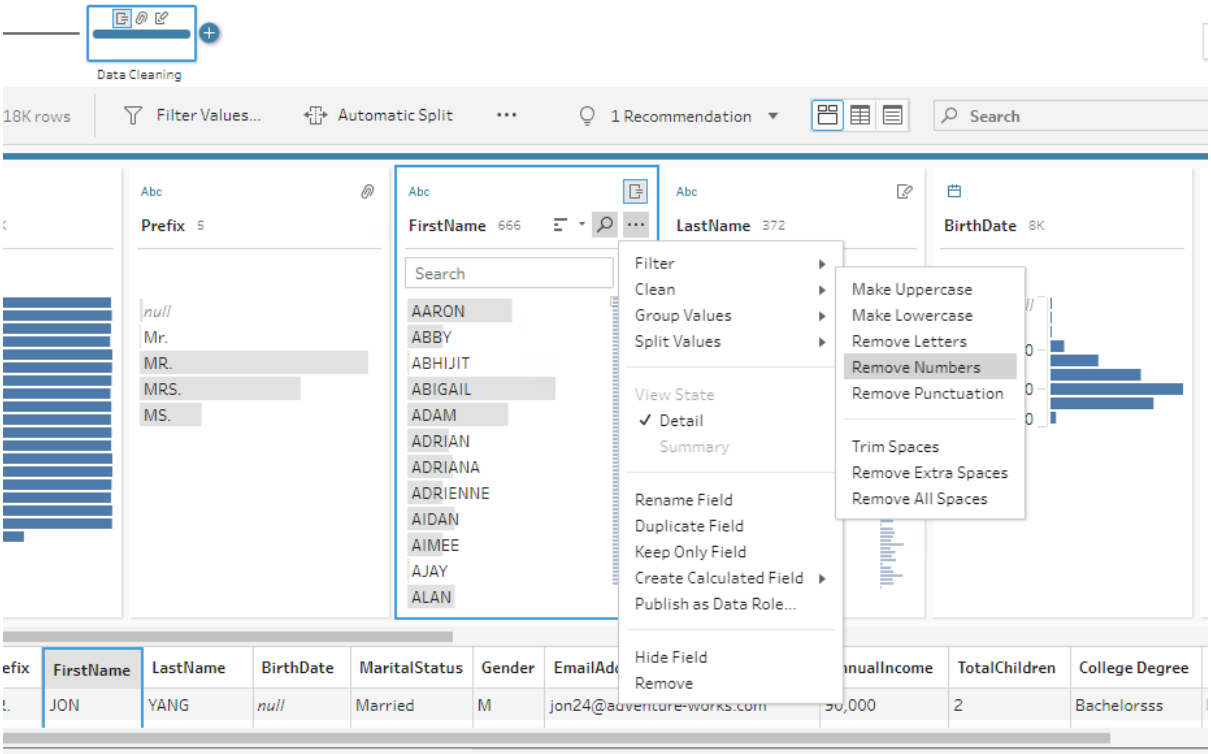


efix	FirstName	LastNa	BirthDate	MaritalStatus	Gende
.	EUGENE	HUANG	05/14/1965	Single	M

6. Clean Prefix MrR to MR

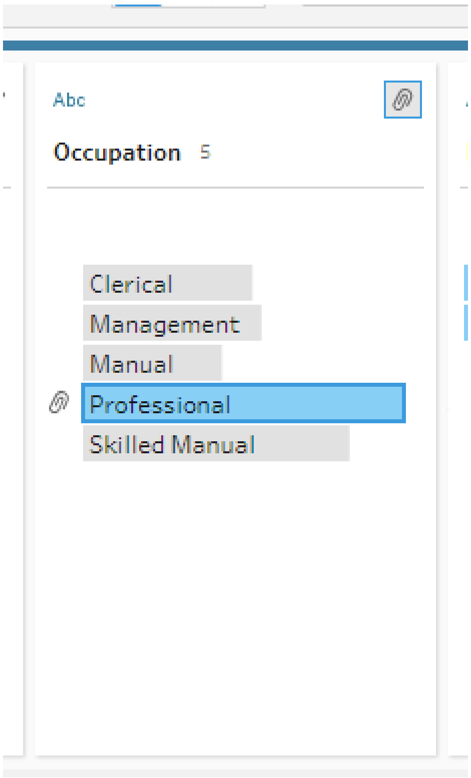


7. Remove Numbers from FirstName



efix	FirstName	LastName	BirthDate	MaritalStatus	Gender	EmailAdd	nualIncome	TotalChildren	College Degree
.	JON	YANG	null	Married	M	jon24@adventure-works.com	90,000	2	Bachelorsss

8. Remove punctuations from Occupation

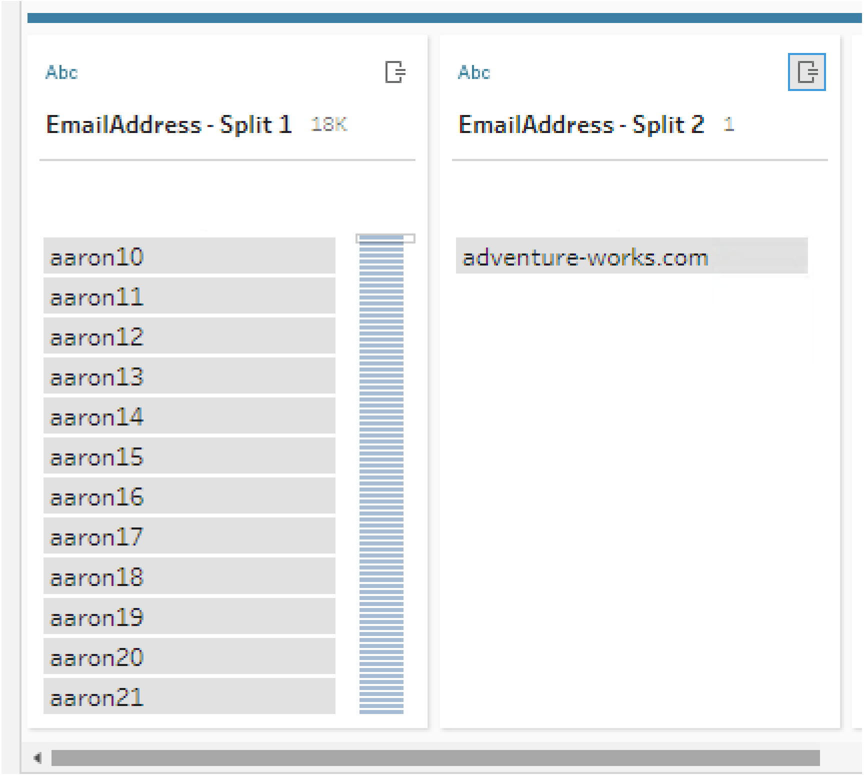


Abc

Occupation 5

Clerical
Management
Manual
Professional
Skilled Manual

9. Split email address and remove the first letters prior to @



Abc

EmailAddress - Split 1 18K

aaron10
aaron11
aaron12
aaron13
aaron14
aaron15
aaron16
aaron17
aaron18
aaron19
aaron20
aaron21

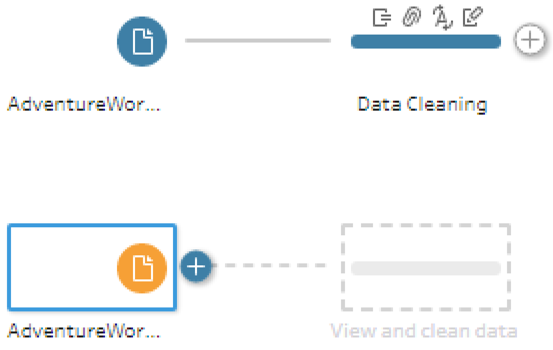
Abc

EmailAddress - Split 2 1

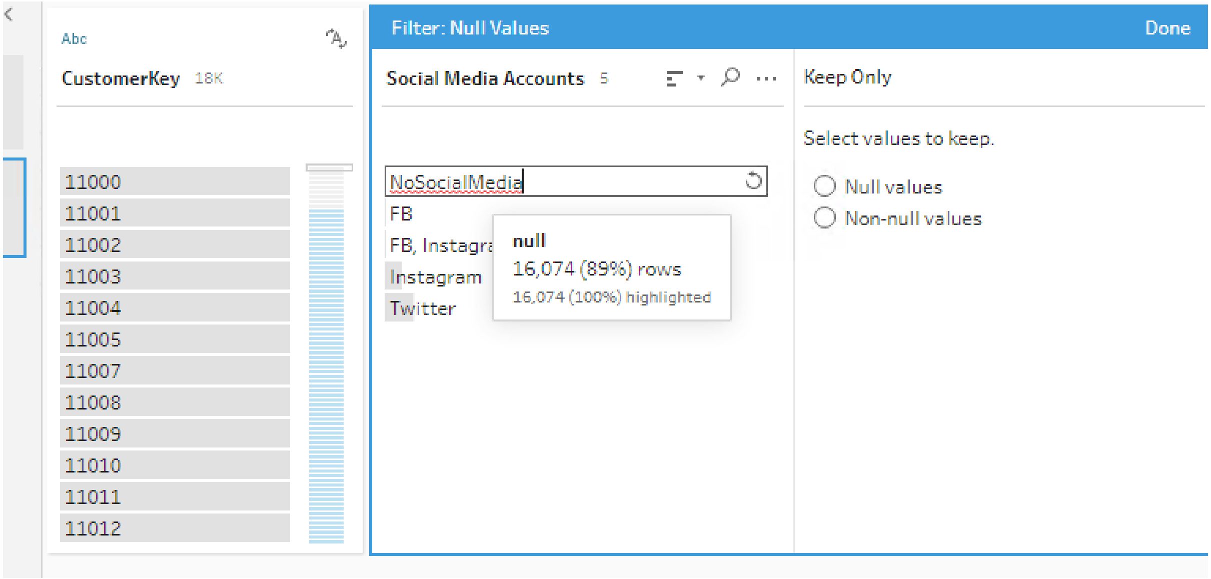
adventure-works.com

Phase 2

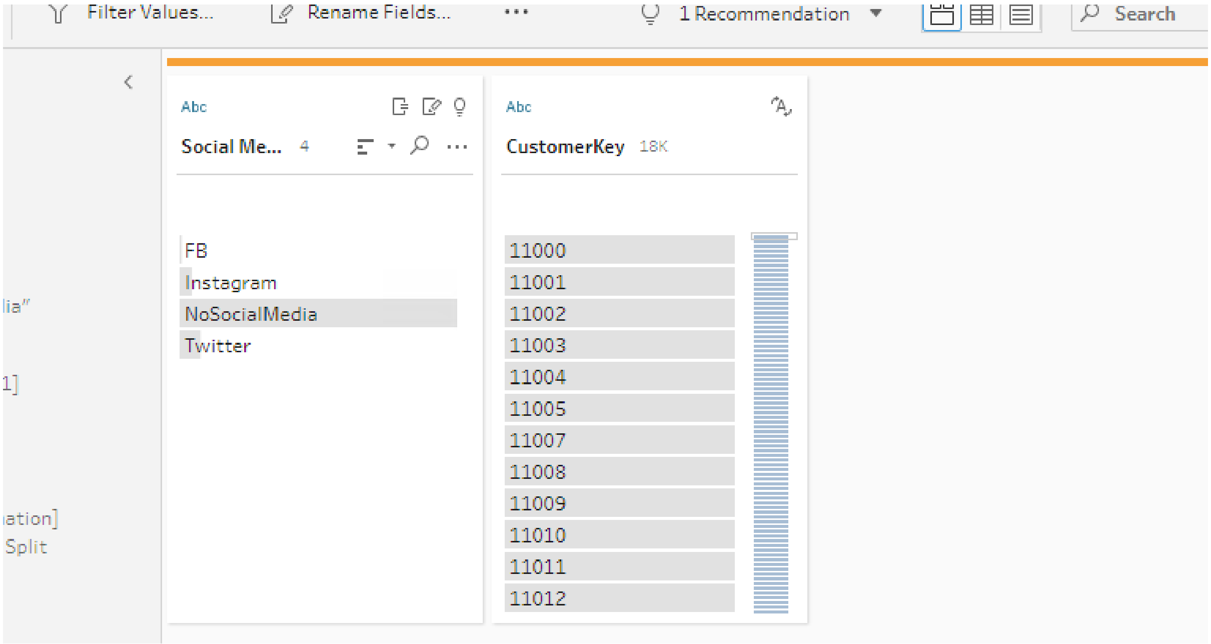
10. Add Adventure Works Customer New file



11. Replace Social Media Account Nulls with NoSocialMedia



12. Split Social Media Fields



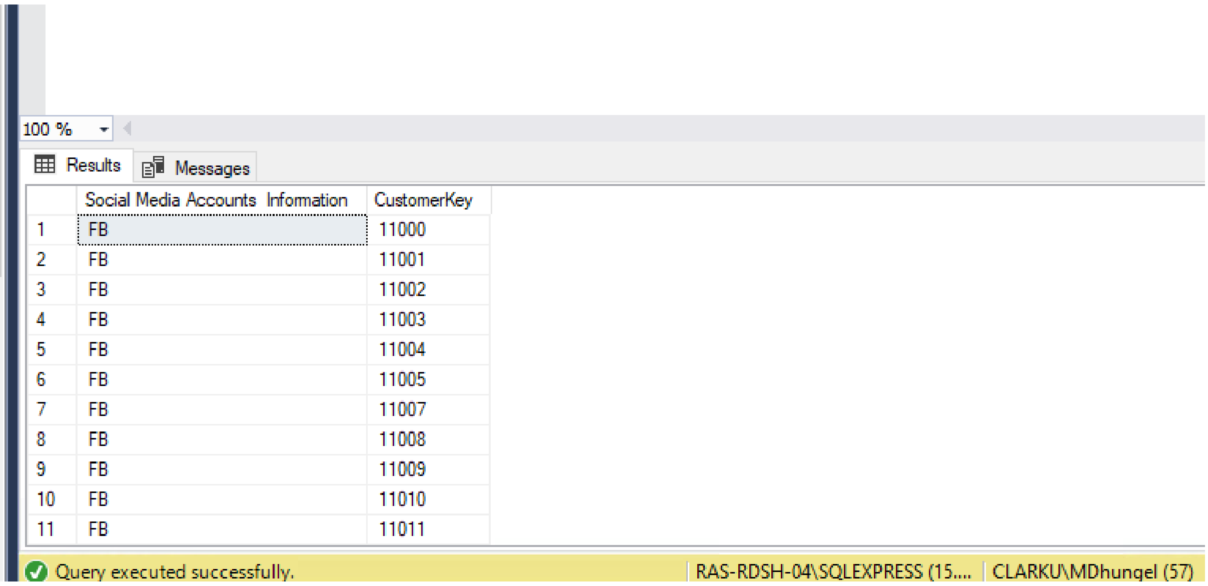
Abc
Social Me... 4

FB
Instagram
NoSocialMedia
Twitter

Abc
CustomerKey 18K

11000
11001
11002
11003
11004
11005
11007
11008
11009
11010
11011
11012

13. Load Data toDimCustomer Tables

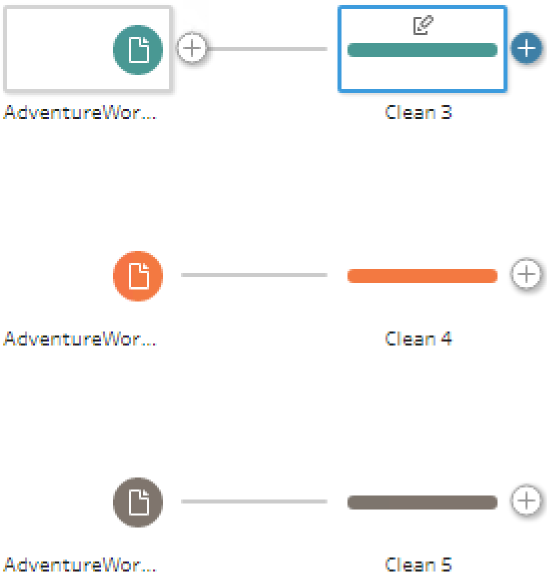


	Social Media Accounts Information	CustomerKey
1	FB	11000
2	FB	11001
3	FB	11002
4	FB	11003
5	FB	11004
6	FB	11005
7	FB	11007
8	FB	11008
9	FB	11009
10	FB	11010
11	FB	11011

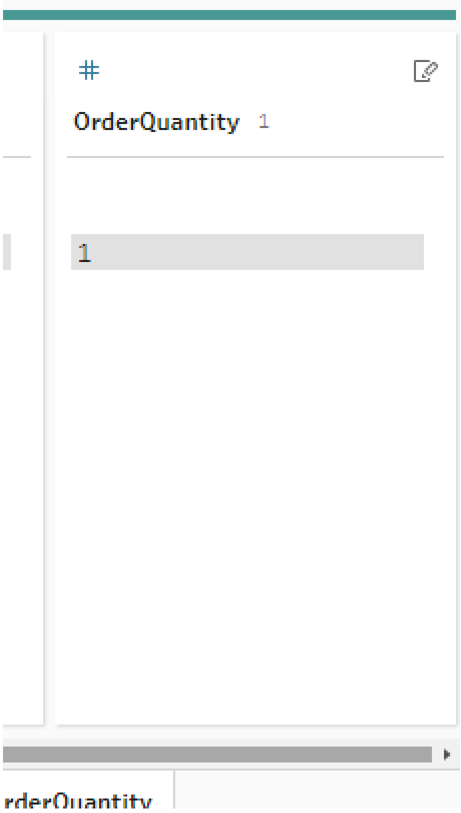
✓ Query executed successfully. | RAS-RDSH-04\SQLEXPRESS (15.... | CLARKU\MDhungel (57)

Phase 3

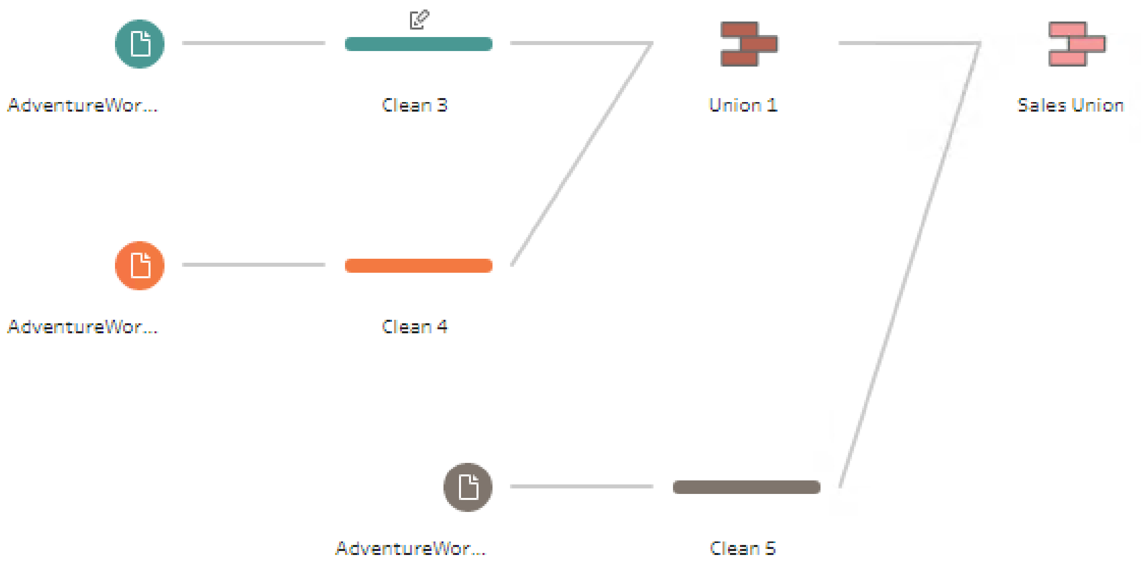
- Add AdventureWorks_Sales_2015
- Add AdventureWorks_Sales_2016
- Add AdventureWorks_Sales_2017



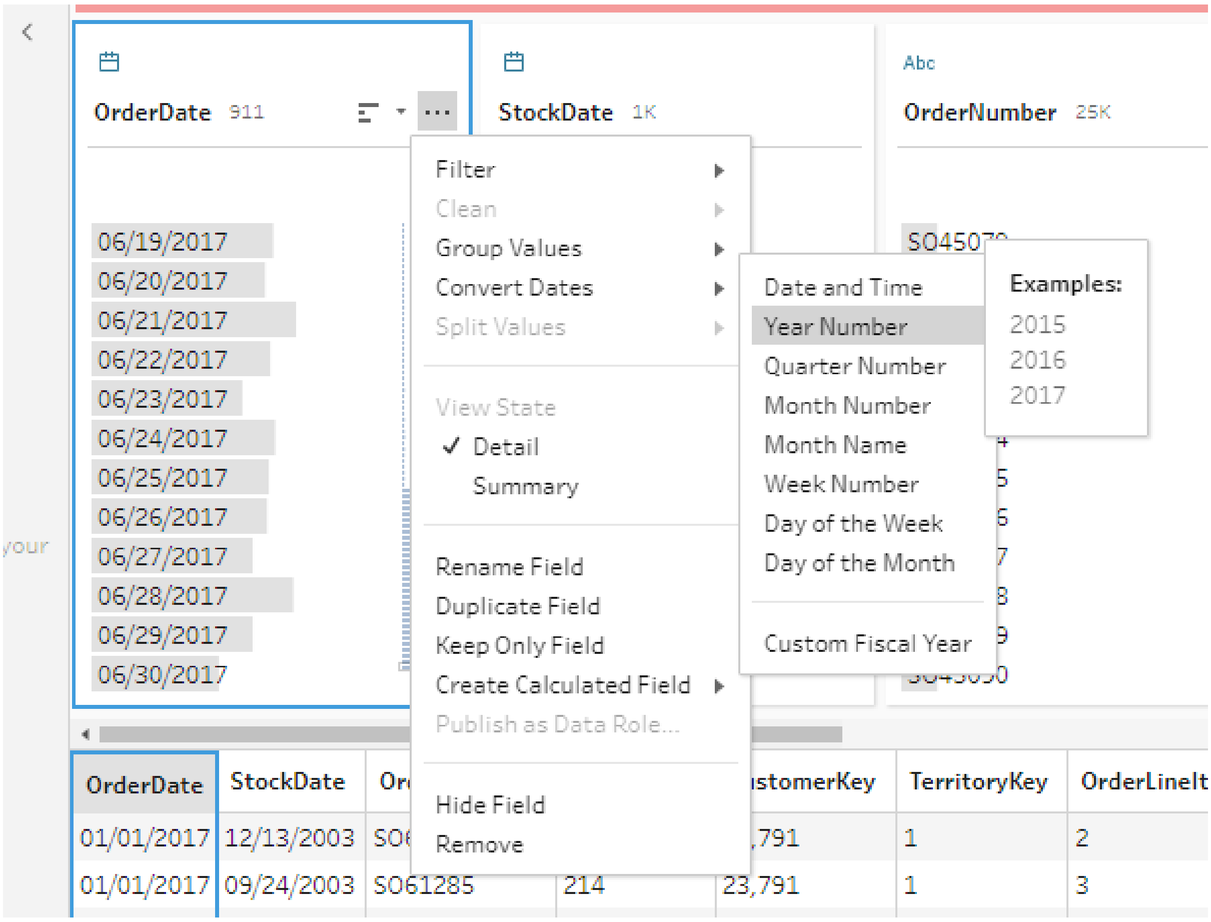
14. Clean OrderQuantity from “Quantite de Ventes” to “OrderQuantity”



15. Union all three Files

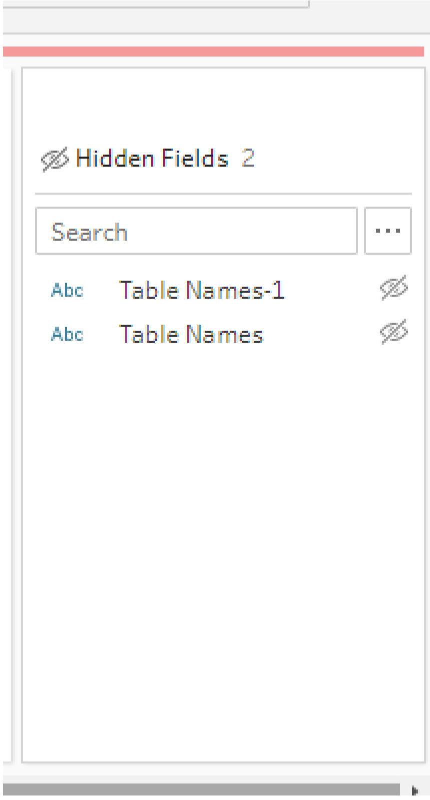


16. Add a calculation to extract “Year” from “Order Date”



OrderDate	StockDate	Or		stomerKey	TerritoryKey	OrderLineIt
01/01/2017	12/13/2003	SO6		,791	1	2
01/01/2017	09/24/2003	SO61285	214	23,791	1	3

17. Remove Table Names

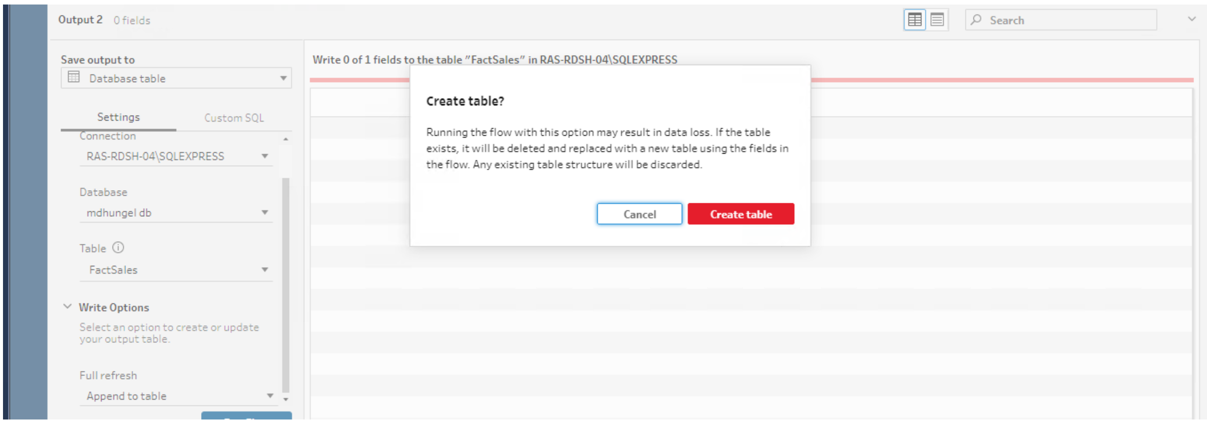


Hidden Fields 2

Abc Table Names-1

Abc Table Names

18. Load Data into FactSales table



Save output to: Database table

Connection: RAS-RDSH-04\SQLEXPRESS

Database: mdhungel db

Table: FactSales

Write Options

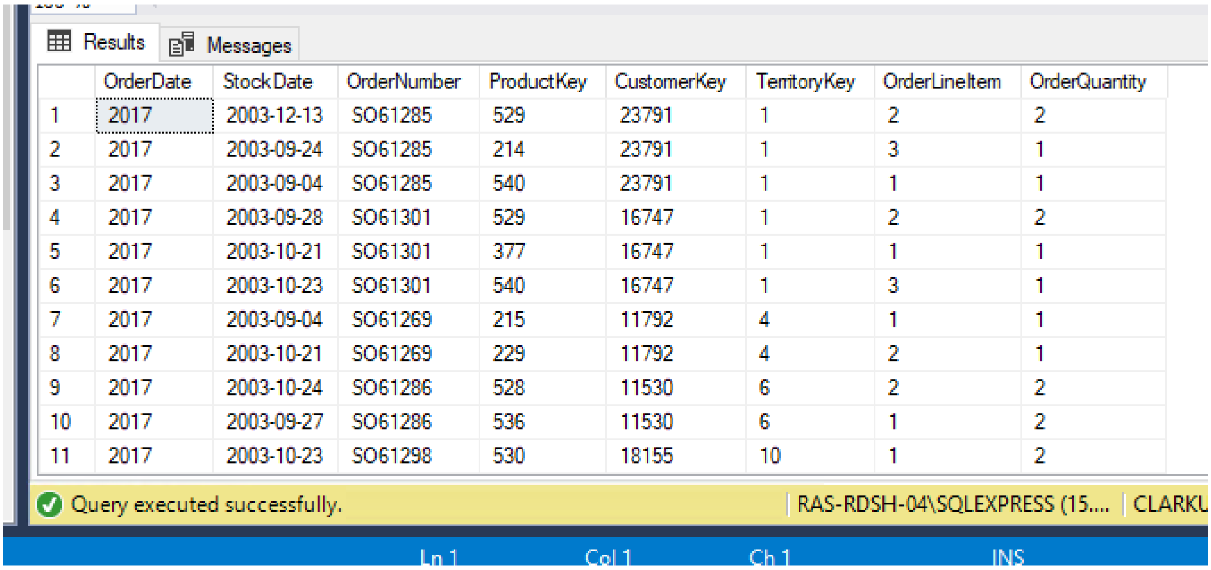
Select an option to create or update your output table.

Full refresh: Append to table

Write 0 of 1 fields to the table "FactSales" in RAS-RDSH-04\SQLEXPRESS

Create table?

Running the flow with this option may result in data loss. If the table exists, it will be deleted and replaced with a new table using the fields in the flow. Any existing table structure will be discarded.

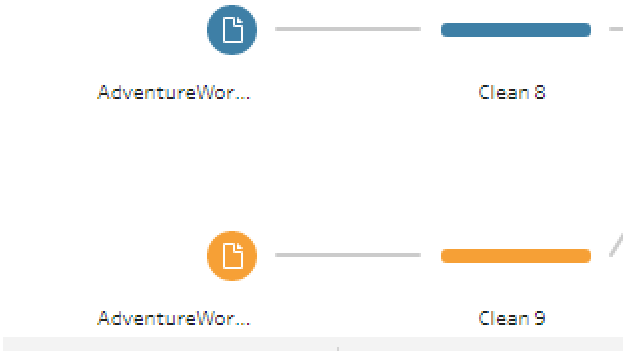


	OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	OrderLineItem	OrderQuantity
1	2017	2003-12-13	SO61285	529	23791	1	2	2
2	2017	2003-09-24	SO61285	214	23791	1	3	1
3	2017	2003-09-04	SO61285	540	23791	1	1	1
4	2017	2003-09-28	SO61301	529	16747	1	2	2
5	2017	2003-10-21	SO61301	377	16747	1	1	1
6	2017	2003-10-23	SO61301	540	16747	1	3	1
7	2017	2003-09-04	SO61269	215	11792	4	1	1
8	2017	2003-10-21	SO61269	229	11792	4	2	1
9	2017	2003-10-24	SO61286	528	11530	6	2	2
10	2017	2003-09-27	SO61286	536	11530	6	1	2
11	2017	2003-10-23	SO61298	530	18155	10	1	2

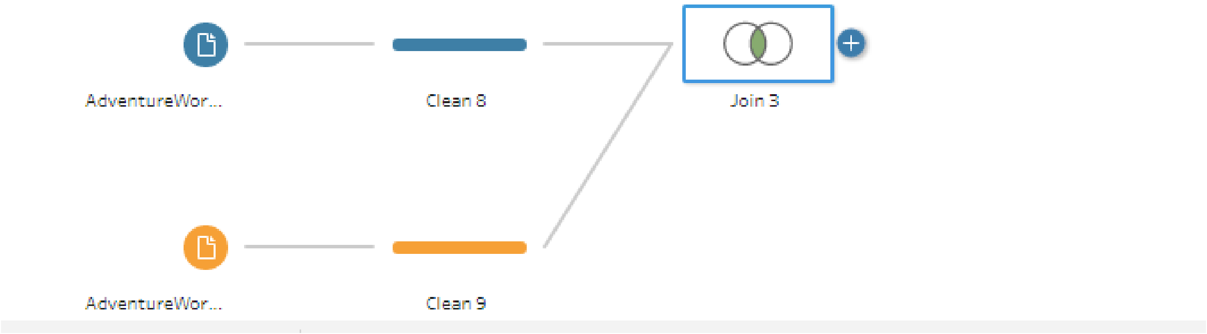
Query executed successfully. | RAS-RDSH-04\SQLEXPRESS (15.... | CLARKU

Phase 4

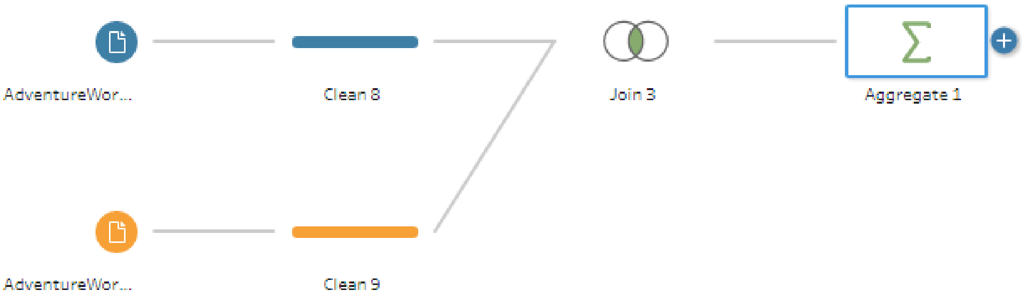
19. Add Returns table and Add Product Table



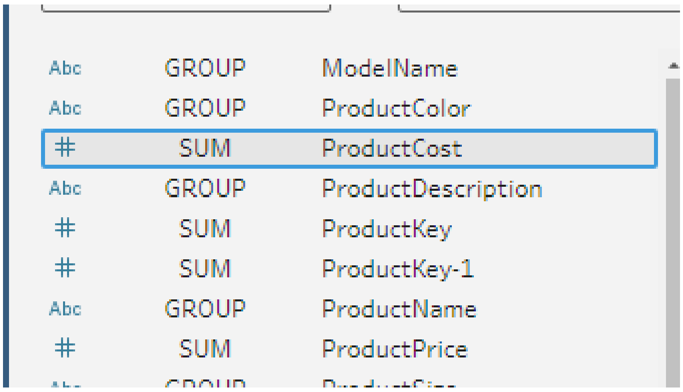
20. Inner Join on Product key



21. Add aggregation



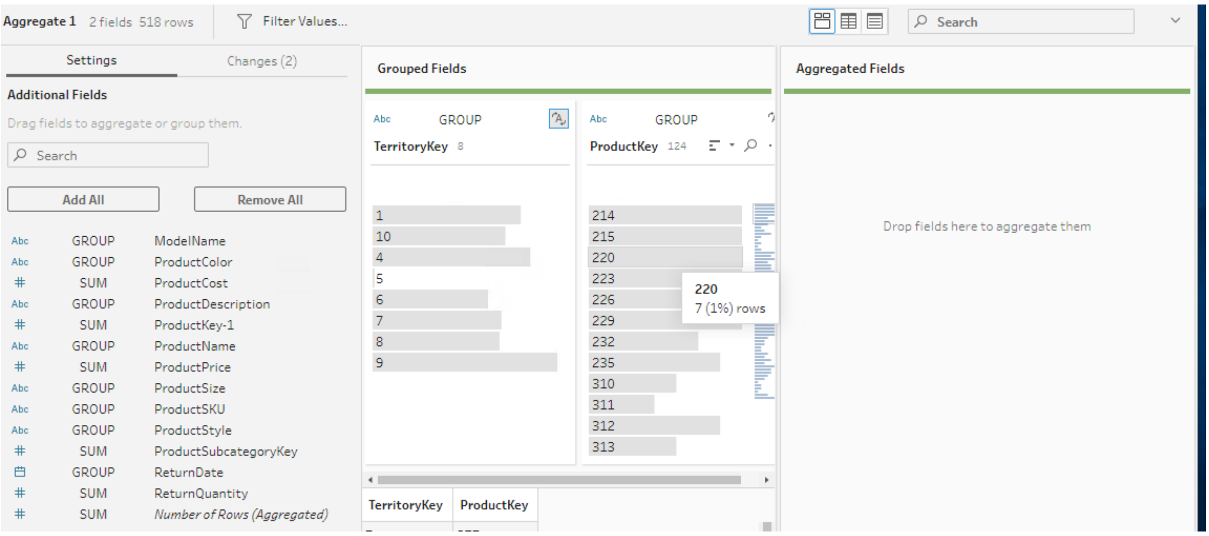
22. Sum of Prices for Product Returns



Abc	GROUP	ModelName
Abc	GROUP	ProductColor
#	SUM	ProductCost
Abc	GROUP	ProductDescription
#	SUM	ProductKey
#	SUM	ProductKey-1
Abc	GROUP	ProductName
#	SUM	ProductPrice

23. Create a new table with the following columns:

- Product Key
- Territory Key



Aggregate 1 2 fields 518 rows

Filter Values...

Search

Settings Changes (2)

Additional Fields

Drag fields to aggregate or group them.

Search

Add All Remove All

Abc	GROUP	ModelName
Abc	GROUP	ProductColor
#	SUM	ProductCost
Abc	GROUP	ProductDescription
#	SUM	ProductKey-1
Abc	GROUP	ProductName
#	SUM	ProductPrice
Abc	GROUP	ProductSize
Abc	GROUP	ProductSKU
Abc	GROUP	ProductStyle
#	SUM	ProductSubcategoryKey
📅	GROUP	ReturnDate
#	SUM	ReturnQuantity
#	SUM	*Number of Rows (Aggregated)*

Grouped Fields

Abc GROUP TerritoryKey 8

1, 10, 4, 5, 6, 7, 8, 9

Abc GROUP ProductKey 124

214, 215, 220, 223, 226, 229, 232, 235, 310, 311, 312, 313

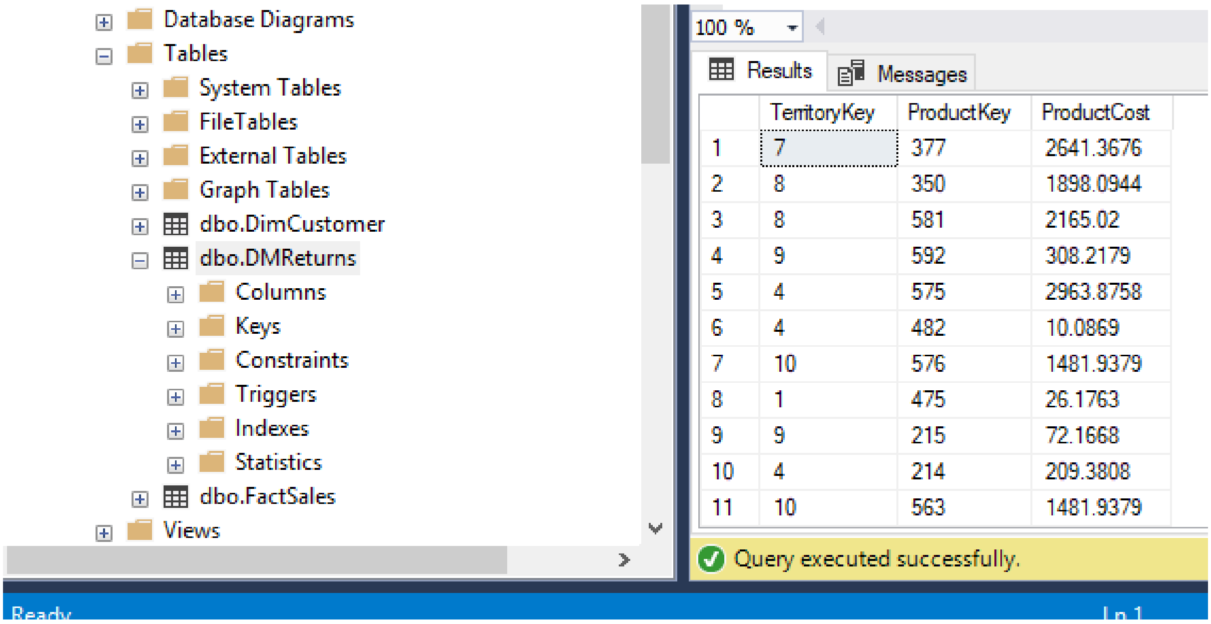
220
7 (1%) rows

Aggregated Fields

Drop fields here to aggregate them

TerritoryKey ProductKey

24. Create a new table in the database called DMReturns



	TerritoryKey	ProductKey	ProductCost
1	7	377	2641.3676
2	8	350	1898.0944
3	8	581	2165.02
4	9	592	308.2179
5	4	575	2963.8758
6	4	482	10.0869
7	10	576	1481.9379
8	1	475	26.1763
9	9	215	72.1668
10	4	214	209.3808
11	10	563	1481.9379

Query executed successfully.