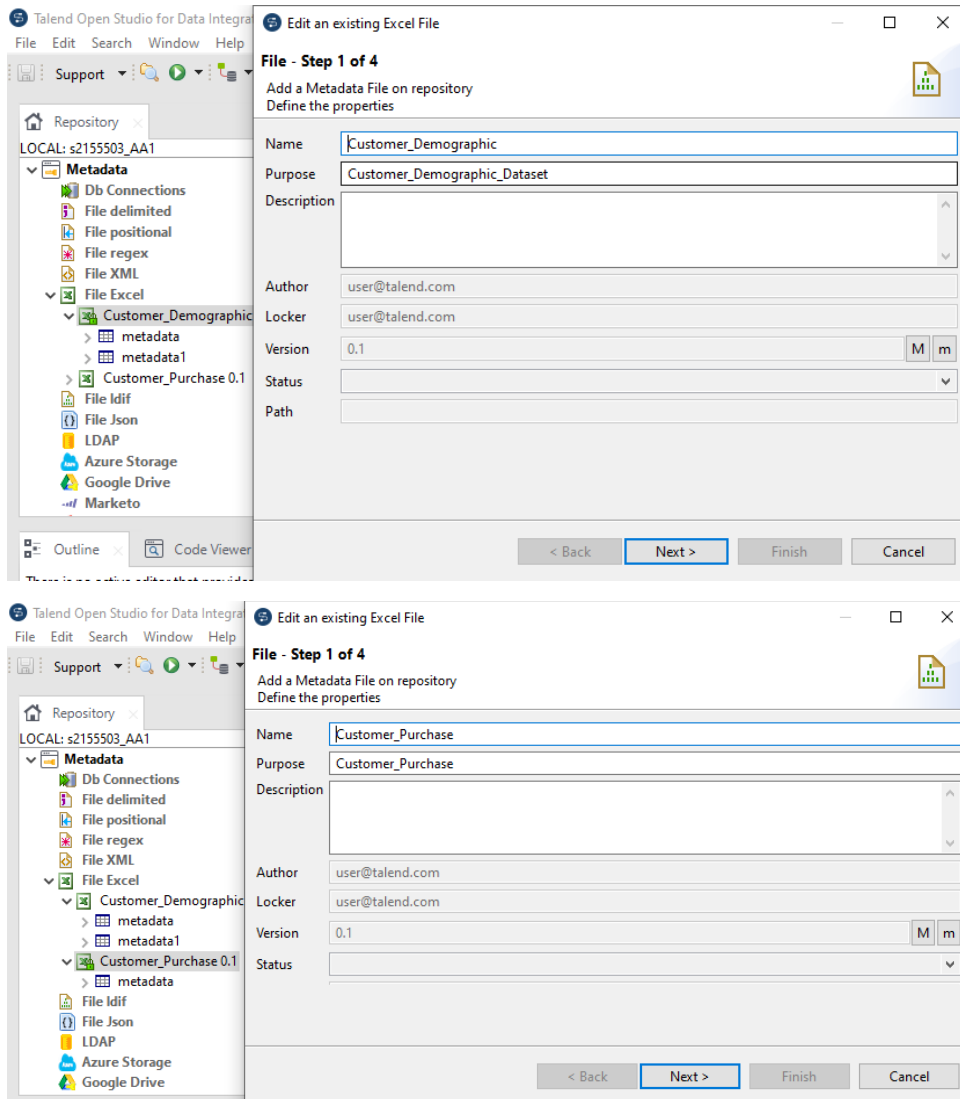


Data Integration

I utilized the Talend data integration tool to carefully combine these datasets, ensuring the seamless merging of information while preserving accuracy and integrity.

Metadata creation for both Excel datasets: -



Edit Scheme and identify correct Data Type

Schema

Click Guess button to update the schema below according to your settings

Guess

Description of the Schema

Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pattern (Ctrl+S...	Length	Precision	Default	Comment
CustomerID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
Age	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
Gender	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
Location	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
MembershipLevel	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
Occupation	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
WebsiteVisitFrequency	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		



Schema

Click Guess button to update the schema below according to your settings

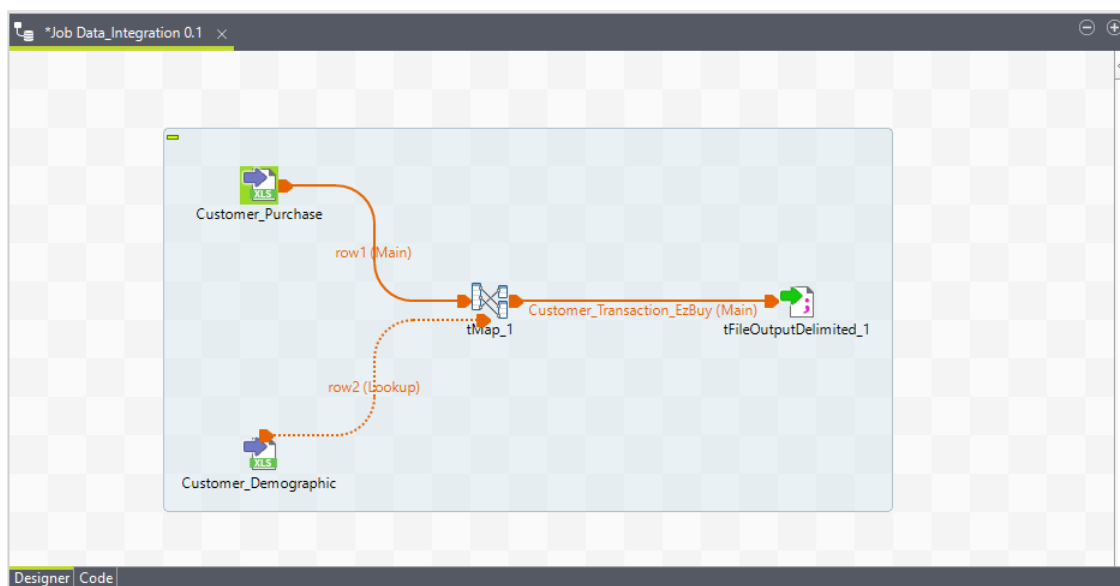
Guess

Description of the Schema

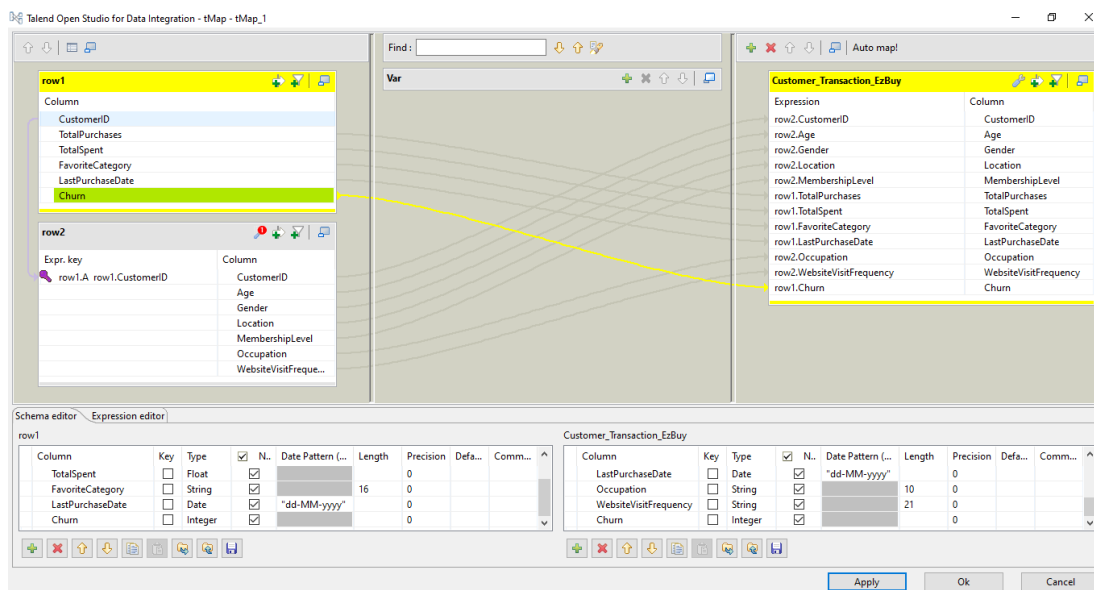
Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pattern (Ctrl+S...	Length	Precision	Default	Comment
CustomerID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
TotalPurchases	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>				0		
FavoriteCategory	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			16	0		
LastPurchaseDate	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>		"dd-MM-yyyy"		0		
Churn	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		



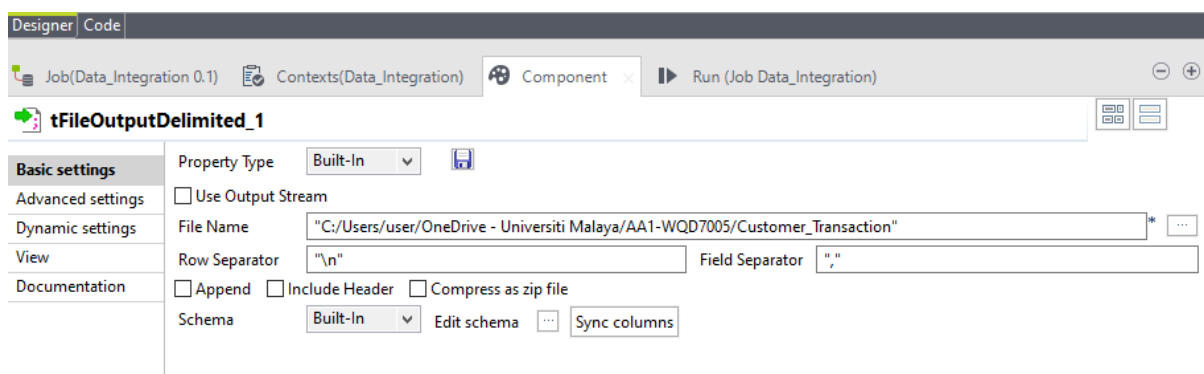
Set Up component's diagram.



In this process, I used four key components. First, source ExcelFileInput dragged from both metadata. Second component called TMap to connect the input data to the output data. Primarily serving the purpose of mapping input data to output data. This involves the transformation of one schema into another, ensuring seamless data integration. Lastly, TOutputDelimited facilitates the outputting of the processed data to a delimited file, adhering to the defined schema.



The component TMap figure above shows how was used to map both datasets, using a unique common key, the customerID from both datasets, to bring all the records together.



To ensure that TOutputDelimited functions as intended, it's important to follow a series of steps. First, we must verify the designated output path, making sure it's correctly specified.

Secondly, we should set the field separator to a comma (","), which defines how the data is separated in the output file.

Schema of tFileOutputDelimited_1

Customer_Transaction_EzBuy (Input)

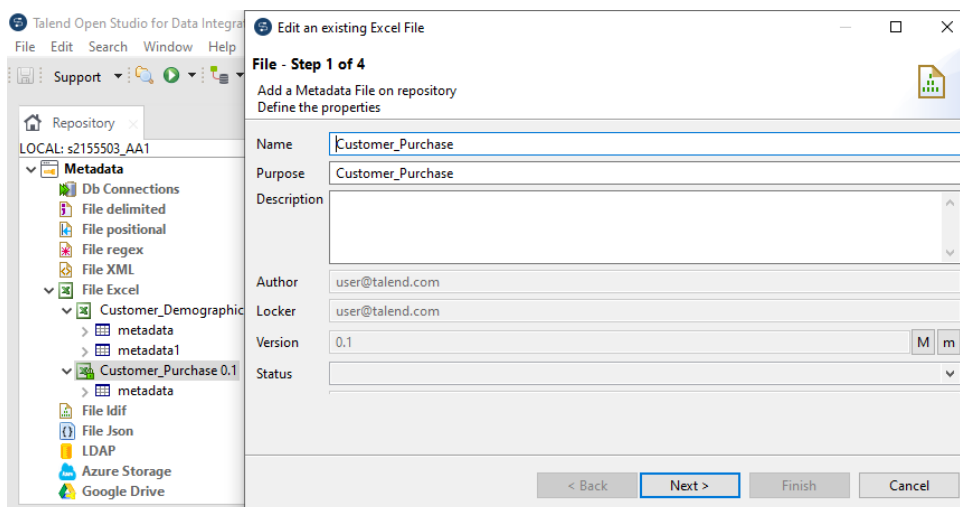
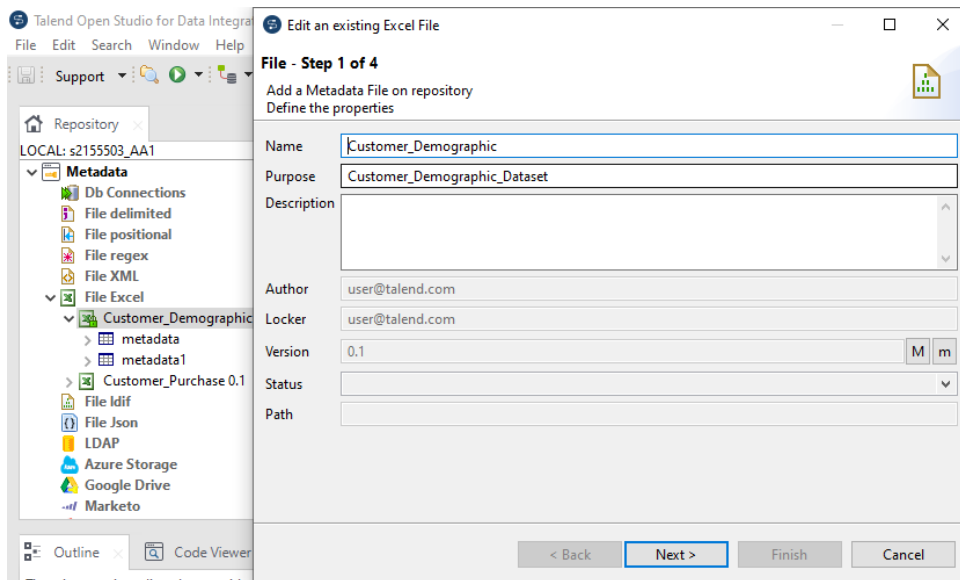
Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pa...	Len...	Pre...	D...	Co...
CustomerID	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			10	0		
Age	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			3	0		
Gender	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			6	0		
Location	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			8	0		
Membershi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			15	0		
TotalPurch...	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>				0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>				0		
FavoriteCat...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			16	0		
LastPurcha...	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>		"dd-M...		0		
Occupation	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			10	0		
WebsiteVisi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			21	0		
Churn	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>				0		

tFileOutputDelimited_1 (Output)

Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pa...	Len...	Pre...	D...	Co...
CustomerID	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			10	0		
Age	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			3	0		
Gender	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			6	0		
Location	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			8	0		
Membershi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			15	0		
TotalPurch...	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>				0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>				0		
FavoriteCat...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			16	0		
LastPurcha...	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>		"dd-M...		0		
Occupation	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			10	0		
WebsiteVisi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>			21	0		
Churn	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>				0		

Lastly, to ensure that all columns from the input schema are correctly copied over to the output schema. This process guarantees that the data is formatted and exported accurately in accordance with the defined criteria.

APPENDIX



Schema

Click Guess button to update the schema below according to your settings

Guess

Description of the Schema

Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pattern (Ctrl+S...	Length	Precision	Default	Comment
CustomerID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
Age	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
Gender	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
Location	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
MembershipLevel	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
Occupation	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		
WebsiteVisitFrequency	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			0	0		



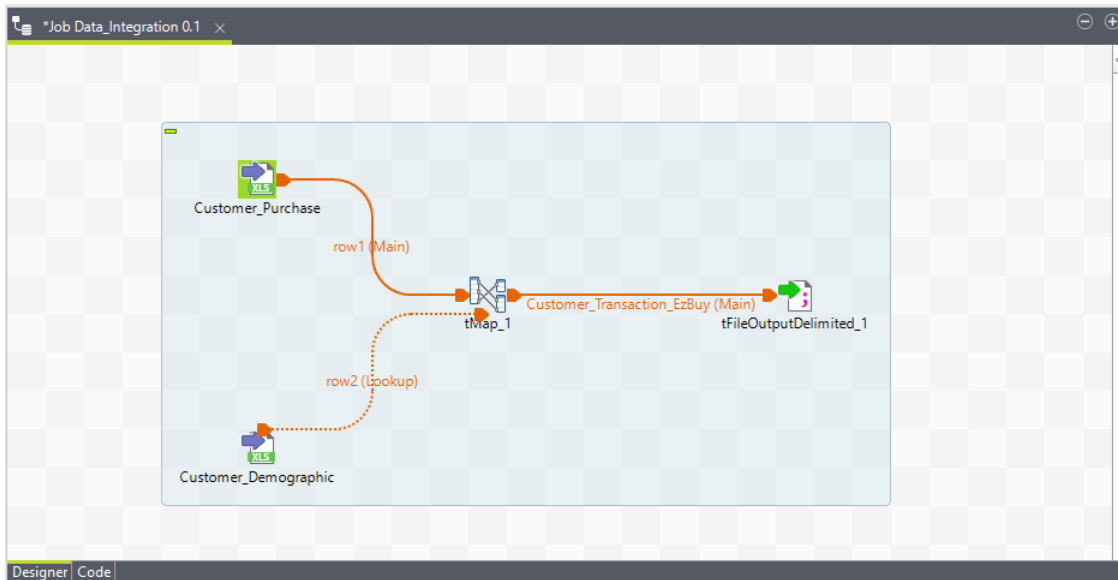
Schema

Click Guess button to update the schema below according to your settings

Guess

Description of the Schema

Column	Key	Type	<input checked="" type="checkbox"/>	N..	Date Pattern (Ctrl+S...	Length	Precision	Default	Comment
CustomerID	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
TotalPurchases	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>				0		
FavoriteCategory	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>			16	0		
LastPurchaseDate	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>		"dd-MM-yyyy"		0		
Churn	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>				0		



Talend Open Studio for Data Integration - tMap - tMap_1

Find:

Var

Auto map!

Customer_Transaction_EzBuy

Expression	Column
row2.CustomerID	CustomerID
row2.Age	Age
row2.Gender	Gender
row2.Location	Location
row2.MembershipLevel	MembershipLevel
row1.TotalPurchases	TotalPurchases
row1.TotalSpent	TotalSpent
row1.FavoriteCategory	FavoriteCategory
row1.LastPurchaseDate	LastPurchaseDate
row2.Occupation	Occupation
row2.WebsiteVisitFrequency	WebsiteVisitFrequency
row1.Churn	Churn

Schema editor Expression editor

row1

Column	Key	Type	<input checked="" type="checkbox"/> N.	Date Pattern (...)	Length	Precision	Defa...	Comm...
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>			0		
FavoriteCategory	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		16	0		
LastPurchaseDate	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-MM-yyyy"		0		
Churn	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>			0		

Customer_Transaction_EzBuy

Column	Key	Type	<input checked="" type="checkbox"/> N.	Date Pattern (...)	Length	Precision	Defa...	Comm...
LastPurchaseDate	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-MM-yyyy"	10	0		
Occupation	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		10	0		
WebsiteVisitFrequency	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		21	0		
Churn	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>			0		

Apply Ok Cancel

Designer Code

Job(Data_Integration 0.1) Contexts(Data_Integration) Component Run (Job Data_Integration)

tFileOutputDelimited_1

Basic settings Property Type Built-In

Advanced settings ☐ Use Output Stream

Dynamic settings File Name "C:/Users/user/OneDrive - Universiti Malaysia/AA1-WQD7005/Customr_Transaction" *

View Row Separator "\n" Field Separator ","

Documentation ☐ Append ☐ Include Header ☐ Compress as zip file

Schema Built-In Edit schema Sync columns

Schema of tFileOutputDelimited_1

Customer_Transaction_EzBuy (Input)

Column	Key	Type	<input checked="" type="checkbox"/> N.	Date Pa...	Len...	Pre...	D...	Co...
CustomerID	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		10	0		
Age	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>		3	0		
Gender	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		6	0		
Location	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		8	0		
Membershi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		15	0		
TotalPurch...	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>			0		
FavoriteCat...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		16	0		
LastPurcha...	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-M...		0		
Occupation	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		10	0		
WebsiteVisi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		21	0		
Churn	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			0		

tFileOutputDelimited_1 (Output)

Column	Key	Type	<input checked="" type="checkbox"/> N.	Date Pa...	Len...	Pre...	D...	Co...
CustomerID	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		10	0		
Age	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>		3	0		
Gender	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		6	0		
Location	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		8	0		
Membershi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		15	0		
TotalPurch...	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			0		
TotalSpent	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>			0		
FavoriteCat...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		16	0		
LastPurcha...	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"dd-M...		0		
Occupation	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		10	0		
WebsiteVisi...	<input type="checkbox"/>	Stri...	<input checked="" type="checkbox"/>		21	0		
Churn	<input type="checkbox"/>	Int...	<input checked="" type="checkbox"/>			0		