

ETL AND TEXTUAL ANALYTICS

Data Analytics for Professional Accountants
(ACCTG 522)

Class 3 | MPAcc class of 2026



Welcome!

Class Agenda

- Review and ETL Overview
- ETL Cases in the FSB Remote Labs
- Conclusion and look ahead

Review: ETL Overview

Results - Input Data (1) - Output

3 of 3 Fields | Cell Viewer | 597 records displayed

Record	EmployeeCode	FirstName	LastName
14	Canada342-9W	Sean	Morgan
15	US476-10W	Julie	Johnson
16	MEX93-5M	Samuel	Perry
17	ENG442-9Mo	Jarrold	Beckett
18	Canada17-7W	Lillian	Nelson
19	ARG559-10W	Delfina	Brandon
20	ENG389-10W	Shaun	Jacobs
21	MEX121-11W	Theresa	Barnes
22	ENG489-12W	Stephen	Powell
23	AUS88-10W	Queenie	Thibodeau
24	AUS455-3W	Juan	King
25	AUS491-13M	Laureen	Varney
26	ENG193-10Mo	Anthony	Griffin
27	ARG508-14W	Olga	Rowe

5 new output fields again

Results - Output Data (4) - Input

8 of 8 Fields | Cell Viewer | 597 records displayed

Record	EmployeeCode	FirstName	LastName	Location	EmpID	PlantID	PayPeriod	LocationPlantID
14	Canada342-9W	Sean	Morgan	CAN	342	9	W	CAN-9
15	US476-10W	Julie	Johnson	USA	476	10	W	USA-10
16	MEX93-5M	Samuel	Perry	MEX	93	5	M	MEX-5
17	ENG442-9Mo	Jarrold	Beckett	ENG	442	9	M	ENG-9
18	Canada17-7W	Lillian	Nelson	CAN	17	7	W	CAN-7
19	ARG559-10W	Delfina	Brandon	ARG	559	10	W	ARG-10
20	ENG389-10W	Shaun	Jacobs	ENG	389	10	W	ENG-10
21	MEX121-11W	Theresa	Barnes	MEX	121	11	W	MEX-11
22	ENG489-12W	Stephen	Powell	ENG	489	12	W	ENG-12
23	AUS88-10W	Queenie	Thibodeau	AUS	88	10	W	AUS-10
24	AUS455-3W	Juan	King	AUS	455	3	W	AUS-3
25	AUS491-13M	Laureen	Varney	AUS	491	13	M	AUS-13
26	ENG193-10Mo	Anthony	Griffin	ENG	193	10	M	ENG-10
27	ARG508-14W	Olga	Rowe	ARG	508	14	W	ARG-14

W True/False: Our regular expression `(\u*)(\d*)-(\d*)(\u*)` only worked because of an Alteryx-specific feature?

0

True: Alteryx treats `\u`, `\l`, and `\w` the same

False: `\d` is any digit

True: Alteryx has an option to ignore case

False: This would work in all cases

None of the above

W True/False: Our regular expression `(\u*)(\d*)-(\d*)(\u*)` only worked because of an Alteryx-specific feature?

0

True: Alteryx treats `\u`, `\l`, and `\w` the same

0%

False: `\d` is any digit

0%

True: Alteryx has an option to ignore case

0%

False: This would work in all cases

0%

None of the above

0%

W True/False: Our regular expression `(\u*)(\d*)-(\d*)(\u*)` only worked because of an Alteryx-specific feature?

0

True: Alteryx treats `\u`, `\l`, and `\w` the same

0%

False: `\d` is any digit

0%

True: Alteryx has an option to ignore case

0%

False: This would work in all cases

0%

None of the above

0%

Extension

Using RegEx101 tool attempt to Transform these similar EmployeeCodes:

Canada1234-1Mo
Mexico0001234-121Monthly
Australia45789-13Annual
AUS123-13M
FRA334455-145weekly
GRE67 1a

Take aways

Regular expressions:

[A-Za-z] lists of acceptable terms (alphabetic)

[A-Za-z0-9] as above, alphanumeric

*,+ zero, one or more

^,\$ beginning, end of the line

| or

\d any digit, \d{N} a specific “N” number of digits

. Any character, \. A full stop (escaping)

Extension Exercise

`^[A-Za-z]+\d+(-|\s)\d+[MAWmaw]`

`^[A-Za-z]+\d+[-\s]\d+[MAWmaw]`

After Class Solution:

`^([A-Za-z]+)(\d+)[- \s](\d+)([MAWmaw])[A-Za-z]*$`

Captured (line 1, 4 separate outputs):

Canada,1234,1,M

Discarded (line 1):

-,0

```
REGULAR EXPRESSION
: / ^([A-Za-z]+)(\d+)[- \s](\d+)([MAWmaw])[A-Za-z]*$

TEST STRING
Canada1234-1Mo
Mexico0001234-121Monthly
Australia45789-13Annual
AUS123-13M
FRA334455-145weekly
GRE67-1a
```

Extra RegEx examples

Dates:

- Invoice Date: 03/15/2025
- Payment Due: 04/14/2025

`^\d*\s*(\S*)`

`\d*\d*\d*`

Invoices:

- INV-100234 `^(?!Reference :)INV-\d+$`
- INV-567890
- Reference: INV-456

`^([A-Z]+)-(\d*)`

`INV-(\d{6})`

Extra RegEx examples

Firm Identifiers:

- CIK: 0000320193
- Ticker: TSLA
- Ticker: AAPL

`[CIK|Ticker]+: (\d+ | [A-Z]+)`

Matching dollar amounts and accounts:

- Total: \$12,450.00 `^Total:\s([\d+,\*\d+.\*\d*)$`
- Late Fee: \$25
- Refund: (\$1,200)
- `.+: (.+)`

Extra RegEx examples

Phone numbers, email, credit card numbers:

- 206-123-4567
- `^\d{3}-\d{3}-\d{4}$`
- real_person@example.com
- Potential fraud detected: 4532-9876-1234-5678

`(\d+-\d+-\d+)(\d+)|([A-Za-z]+)([@_.]) ([A-Za-z]+)([@_.]) ([A-Za-z]+)([@_.]) ([A-Za-z]+)`

ETL Cases in the FSB

Remote Labs

ETL Cases, Continued

Today we will examine “Joining Data”:

- Joining data is another **fundamental step** in the ETL process, as in most cases, developing insight will require data from different databases.
- Joining data can sometimes be referred to as merging data, blending data, or combining data.

Why is joining data important?

Joining data is part of the transforming step in ETL.

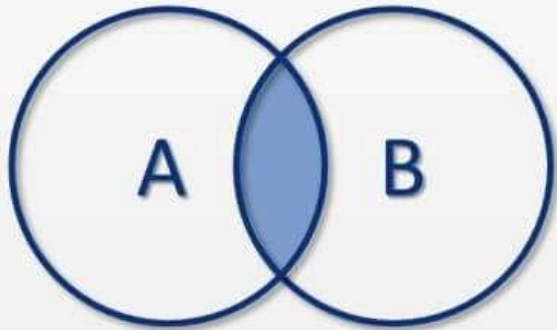
- Incorrect joins are a **data quality issue**, incorrect joins can:
 - **crash** (or really slow down) our programs!
 - have us **lose** important data.
 - Have us generate **incorrect conclusions/insight** from our analysis.

Joining using unique keys

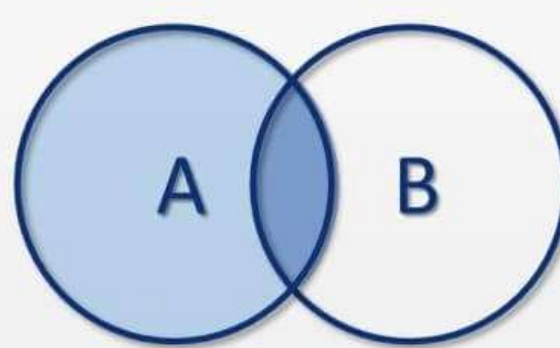
In all cases we will use unique keys to join the data.

- All the unique keys are described in the case
- One of them was tricky, can you remember which one?

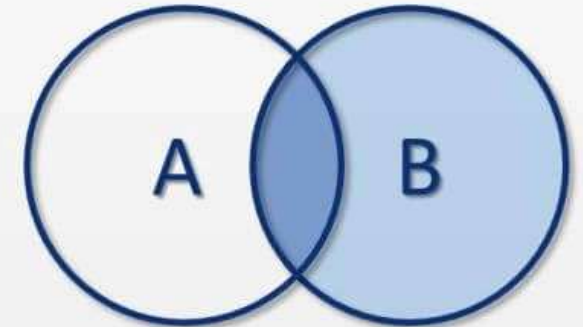
INNER JOIN



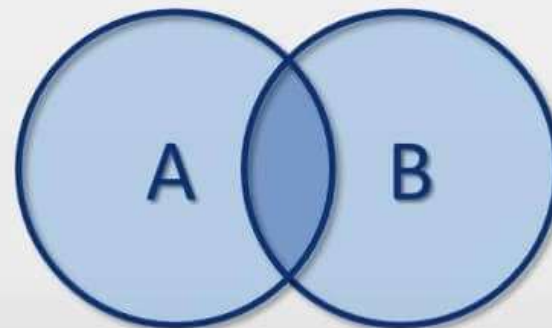
OUTER JOIN



LEFT

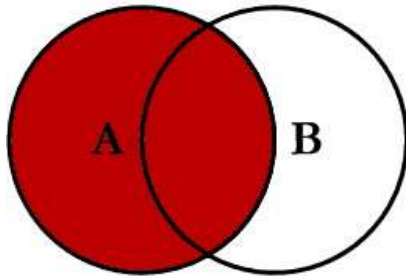


RIGHT

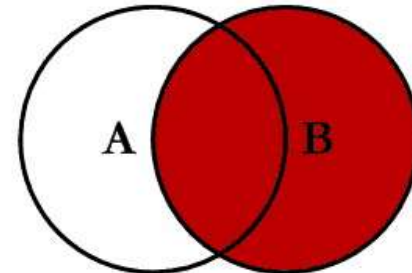


FULL

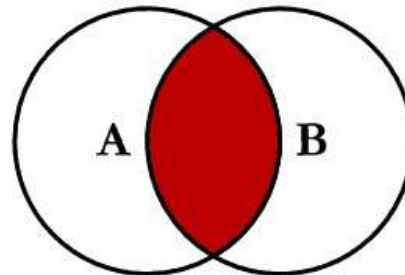
SQL JOINS



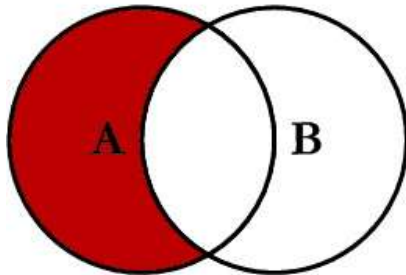
```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```



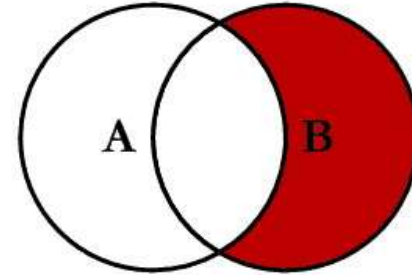
```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```



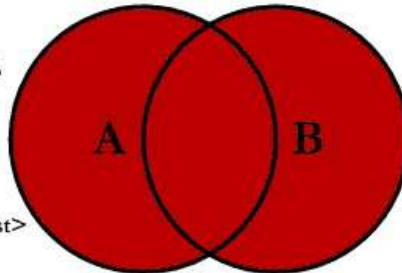
```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```



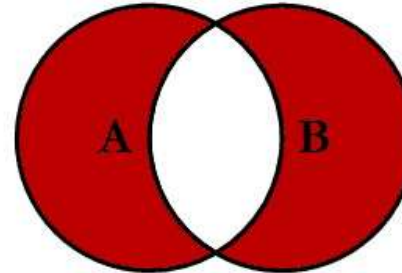
```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```



```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```



```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

Joining using unique keys

In all joins we will use “one to many” matches.

- Joining the large information source (“JELineItems”) to the smaller set of information in the other data sheets.

Why join across multiple software?

Because joining data is so fundamental, we want to work towards becoming software agnostic.

Working with cutting-edge software allows you to:

- Become adaptable and resilient to **new software**.
- Impress recruiters with practical discussions of cutting-edge software.

Joining across different tools:

Some terminology:

- Alteryx: drag and drop “Join” and “Union” tools
- PowerBI: data relationships tab
- Python: code-level with package “pandas”
- SQL: base code-level
- Tableau: Both relationships and joins.

Alteryx Joins:

Join (1) - Configuration

☐ Join by Record Position

☒ Join by Specific Fields

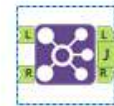
Left Right

* [v] [v]

L J R

Options | ↑ ↓ TIP: To reorder multiple rows: select, right-click and drag.

	Input	Field	Type	Size	Rename	Description
▶	☒	*Unknown	Unknown	0		Dynamic or Unknown Fields



SQL/Python:

```
SELECT
    JEL.*,
    BU.*
FROM JELineitems AS JEL
LEFT JOIN BusinessUnits AS BU
    ON JEL.BusinessUnitID = BU.BusinessUnitID;
```

```
SELECT
    JEL.*,
    BU.*
FROM JELineitems AS JEL
RIGHT JOIN BusinessUnits AS BU
    ON JEL.BusinessUnitID = BU.BusinessUnitID;
```

```
import pandas as pd
```

```
# Example: read from Excel file (each sheet as
DataFrame)
file_path = "your_file.xlsx"
```

```
# Load the two sheets
JELineitems = pd.read_excel(file_path,
    sheet_name="JELineitems")
BusinessUnits = pd.read_excel(file_path,
    sheet_name="BusinessUnits")
```

```
# Perform SQL-style join on BusinessUnitID
# Equivalent to: SELECT * FROM JELineitems
#             INNER JOIN BusinessUnits
#             ON JELineitems.BusinessUnitID =
BusinessUnits.BusinessUnitID;
merged_df = JELineitems.merge(
    BusinessUnits,
    on="BusinessUnitID",
    how="inner" # can also use 'left', 'right', or 'outer'
)
```

Lab Agenda

1. Open files in excel / open case and identify the keys
2. Compare Tableau and Power BI

Tableau Relationships

📁 JELineItems+ (Analytics_mindset_case_studies_ET...

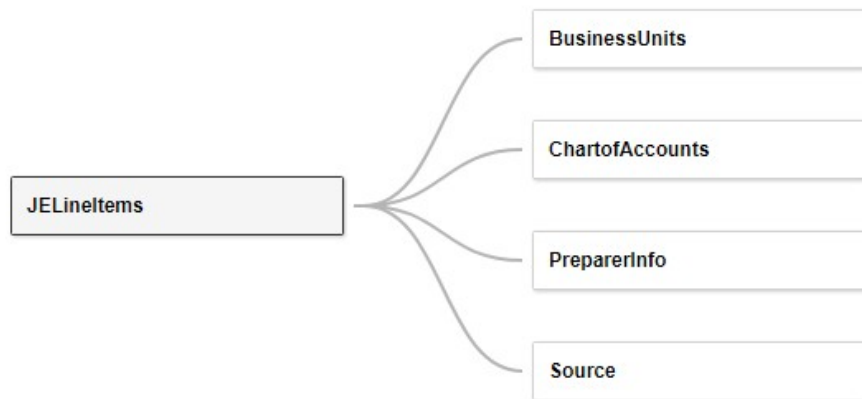
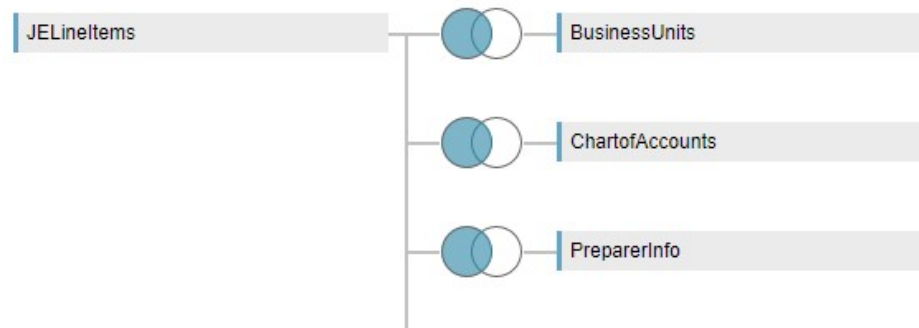
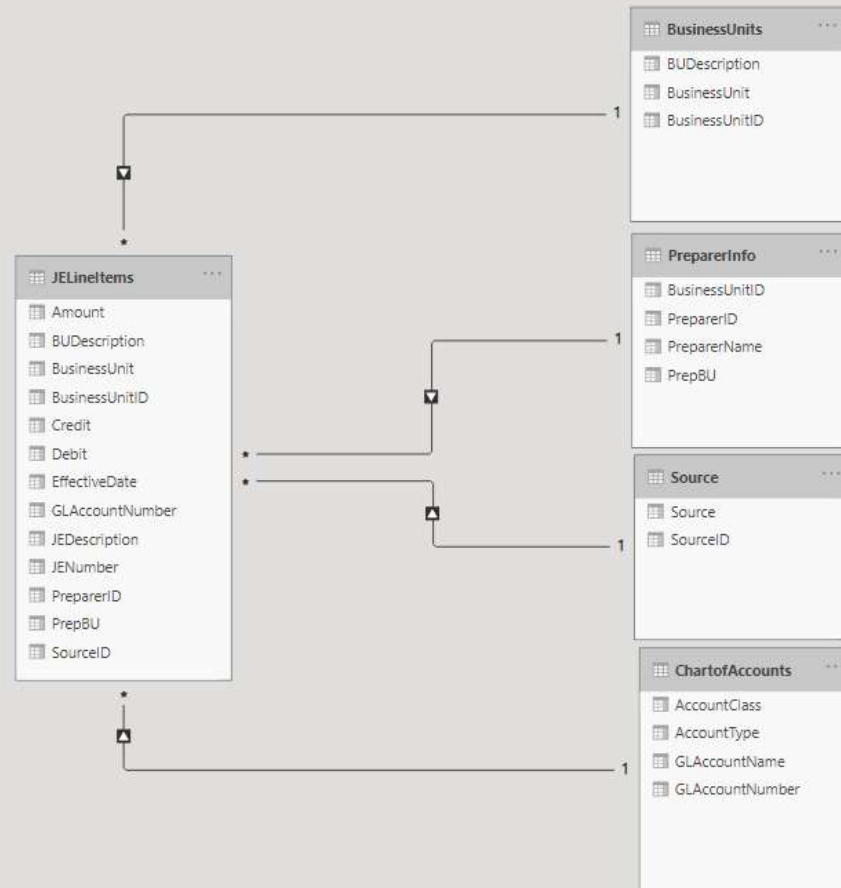


Tableau Joins

JELineItems is made of 5 tables. ⓘ



Power BI Data Relationships



To FSB remote labs

Again I recommend **saving screenshots** of your work in Alteryx (I've added blank slides if you want to save them in this ppt document, ctrl+m adds more slides)

Two Potential Issues Explained:

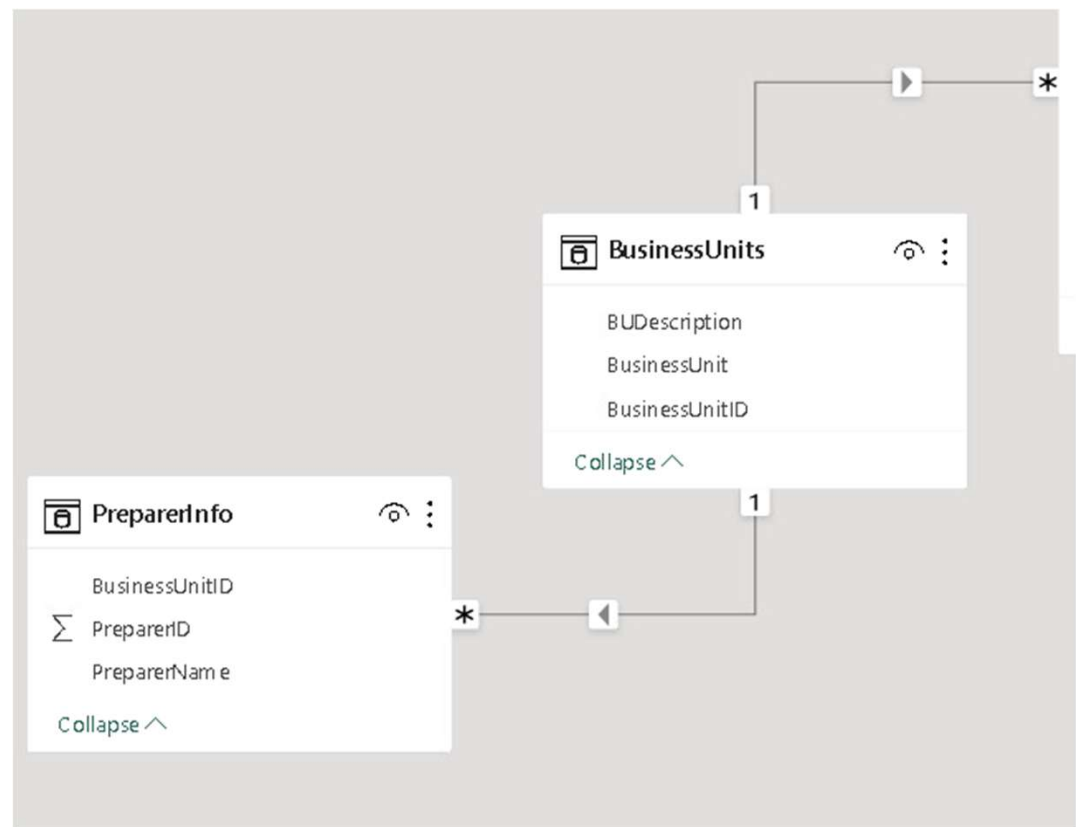
Issue 1 In the underlying dataset PreparerID is **not unique**:

For example ID 101 appears 6 times, repeating once for each BusinessUnit

PreparerName	PreparerID	BusinessUnitID
Carolyn Slater Food Operations Manager	101	1
Alexander Knox Assistant Manager	101	2
Trevor Roberts Housekeeping Supervisor	101	4
Gavin McGrath Dining Commons Manager	101	3
Rodney Wilkins Maintenance Supervisor	101	6

Two Potential Issues Explained:

PowerBI: maps the multiple prepareIDs to a unique BusinessUnitID, allowing it to join both datasets to JELineItems via BusinessUnitID.



Two Potential Issues Explained:

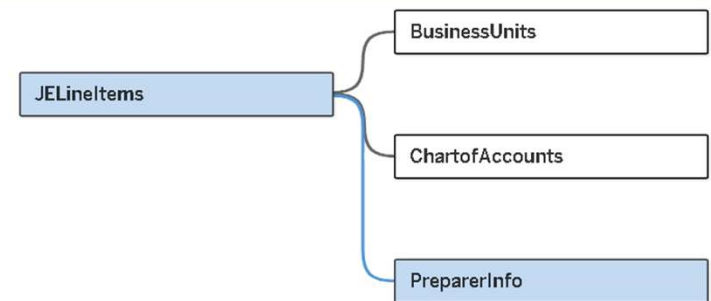
Tableau: we can resolve the alert by filling out the keys which are

BusinessUnitID= BusinessUnitID

AND

PreparerID = PreparerID.

Which means it has to match both keys.



The screenshot shows the Tableau relationship editor interface. At the top, a dropdown menu shows 'JELineItems — PreparerIn...'. Below this, a table displays the relationship configuration. The table has three columns: 'JELineItems', 'Operator', and 'PreparerInfo'. The first row shows a join on 'Business Unit ID' with an equals sign operator. The second row shows a join on 'Preparer ID' with an equals sign operator. The 'PreparerID (Prej)' field in the second row is highlighted with a yellow border. At the bottom, there is a button labeled '+ Add more fields'. On the right side, there is a vertical list of names: 'Abc', 'Prepi', 'Prep', 'Carc', 'Alex', 'Trev', and 'Gavi'.

JELineItems	Operator	PreparerInfo
# Business Unit ID	=	# BusinessUnitID
# Preparer ID	=	# PreparerID (Prej)

+ Add more fields

Tableau Completed

Add the Source dataset and the **data relationship** is complete.

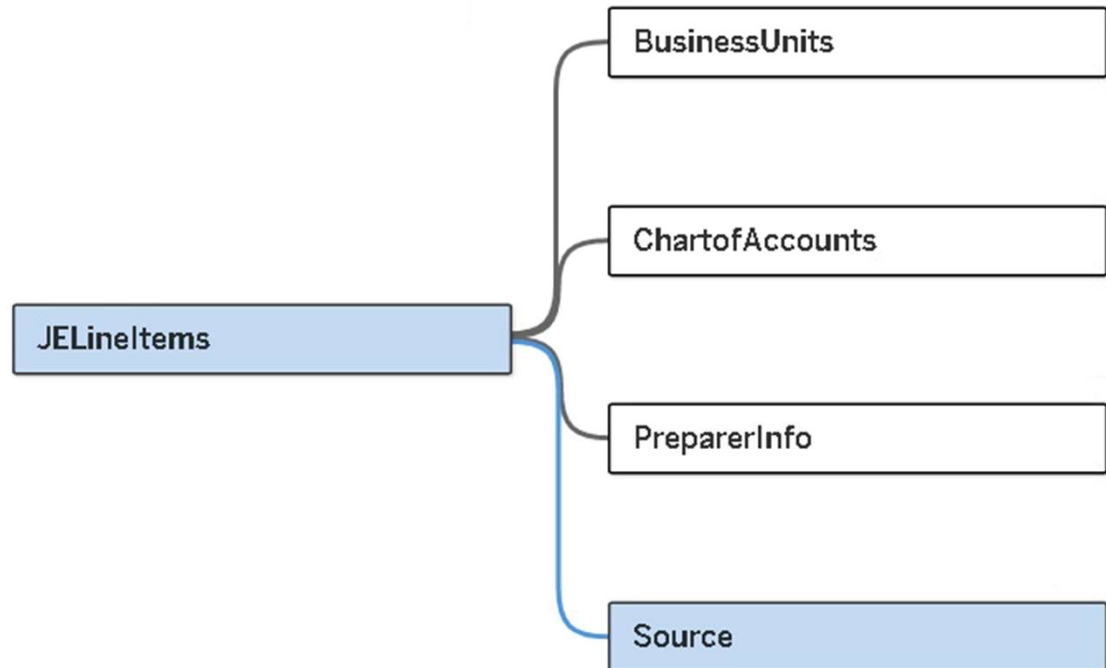


Tableau Join Version

Defaults to inner joins, we are expecting 789 line items (rows), but get 639 rows of data (missing 150 rows / JE's).

JELineItems is made of 5 tables. ⓘ

- JELineItems
- BusinessUnits
- ChartofAccounts
- PreparerInfo
- Source

Join

Data Source			PreparerInfo	
Business Unit ID	=		BusinessUnitID ...	
Preparer ID	=		PreparerID ...	
Add new join clause				

Edit the Preparer Info Join to join on both preparerID and BusinessUnit

Tableau Join Version

Change to left join,
confirm that 641 rows exists

**This occurs when we left join
to the ChartofAccounts tab,
implying that 2 JELineItems
have a GLAccountNumber not
in the summary
ChartOfAccounts dataset.

The screenshot shows the Tableau interface with three data sources: JELineItems, BusinessUnits, and ChartofAccounts. A join dialog box is open, showing the 'Left' join option selected for the relationship between JELineItems and ChartofAccounts. The join is based on the 'GL Account Number' field. The dialog box also shows the 'Inner', 'Right', and 'Full Outer' join options. The 'Data Source' is 'JELineItems' and the 'ChartofAccounts' is 'ChartofAccounts'. The join is based on the 'GL Account Number' field. The dialog box also shows the 'Inner', 'Right', and 'Full Outer' join options. The 'Data Source' is 'JELineItems' and the 'ChartofAccounts' is 'ChartofAccounts'. The join is based on the 'GL Account Number' field.

At the bottom of the Tableau interface, the status bar shows 'JELineItems' selected, with '22 fields' and '641 rows' displayed. The '641 rows' is highlighted in yellow.

Tableau Join Version

Change to left join,
confirm that 789 rows exists

**This occurs when we left join
to the PreparerInfo tab, implying
that 148 JELineItems have a
preparerID not in the summary
PreparerInfo dataset.

JELineItems is made of 5 tables. ①

Diagram illustrating the data source structure for JELineItems, which is composed of 5 tables. The tables are joined to JELineItems using various join types (Inner, Left, Right, Full Outer). The PreparerInfo table is highlighted in yellow, indicating it is the focus of the join operation.

Join configuration window showing the join type (Left) and the join condition (Source ID = SourceID (Source)).

Join Type	Data Source	Source
Inner		
Left		
Right		
Full Outer		

Source ID = SourceID (Source)

Add new join clause

JELineItems 22 fields 789 rows

Pages

Columns

Rows

JE Number

GLAccountNumber (Cha..

Business Unit ID

Preparer ID

Preparer Name

Prep Filter

Preparer Name: Null

Marks

Automatic

Color

Size

Text

Detail

Tooltip

JE Number	GLAccount..	Business U..	Preparer ID	Preparer N..
2033532	11845	1	106	Null
	43310	1	106	Null
2033558	11845	1	106	Null
	43310	1	106	Null
2033572	11845	1	106	Null
	43440	1	106	Null
2033573	11845	1	106	Null
	43310	1	106	Null
2033574	11845	1	106	Null
	43310	1	106	Null
2033581	11845	1	106	Null
	43440	1	106	Null
2033589	11845	1	106	Null
	43310	1	106	Null
2033590	11845	1	106	Null
	43310	1	106	Null
2033596	11845	1	106	Null
	43440	1	106	Null
2033597	22110	6	105	Null
	51402	6	105	Null
2033599	22110	4	105	Null
	51610	4	105	Null
2033603	22110	1	105	Null
	51610	1	105	Null
2033606	22110	6	105	Null
	51400	6	105	Null
2033607	11845	6	105	Null
	51140	6	105	Null
2033608	22110	4	105	Null
	51402	4	105	Null
2033613	22110	1	105	Null
	51402	1	105	Null
2033615	22110	1	105	Null
	51610	1	105	Null
2033616	22110	6	105	Null
	51510	6	105	Null
2033621	22110	4	105	Null
	51400	4	105	Null
2033623	22110	4	105	Null

Filtering the data we find 148 rows with businessUnit-PreparerID pairs that are not in the PreparerInfo dataset (and so we do not know the preparer name, which is "Null").

And 2 with missing GLNumbers/Names

Pages

Columns

Rows

JE Number

GLAccountNumber (Cha..

GL Account Name

GLFilter

GLAccountNumber (C..

Marks

Automatic

Color

Size

Text

JE Number	GLAccount..	GL Account..	Business U..	Preparer ID	Preparer Name
2033847	Null	Null	3	101	Gavin McGrath Dining Co..
2033851	Null	Null	1	101	Carolyn Slater Food Oper..

Two Potential Issues Explained:

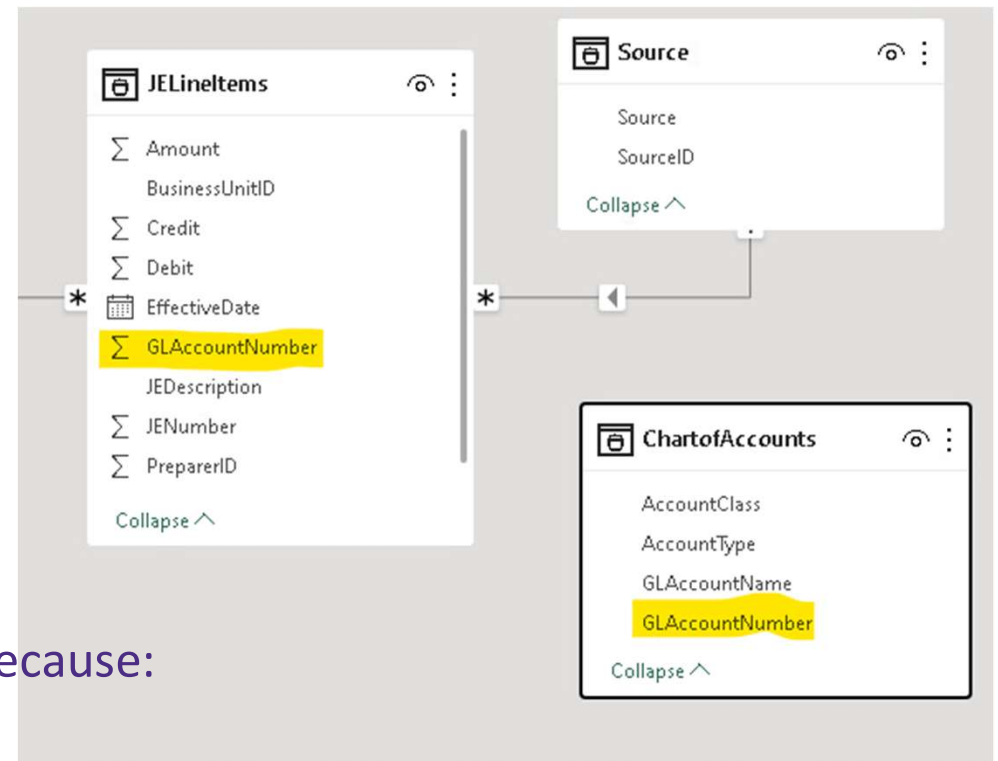
Issue 2 GLAccounts not mapped in PowerBI:

PowerBI read GLAccountNumbers as numbers instead of string format (warned us with 2 read errors)

PowerBI Workflow

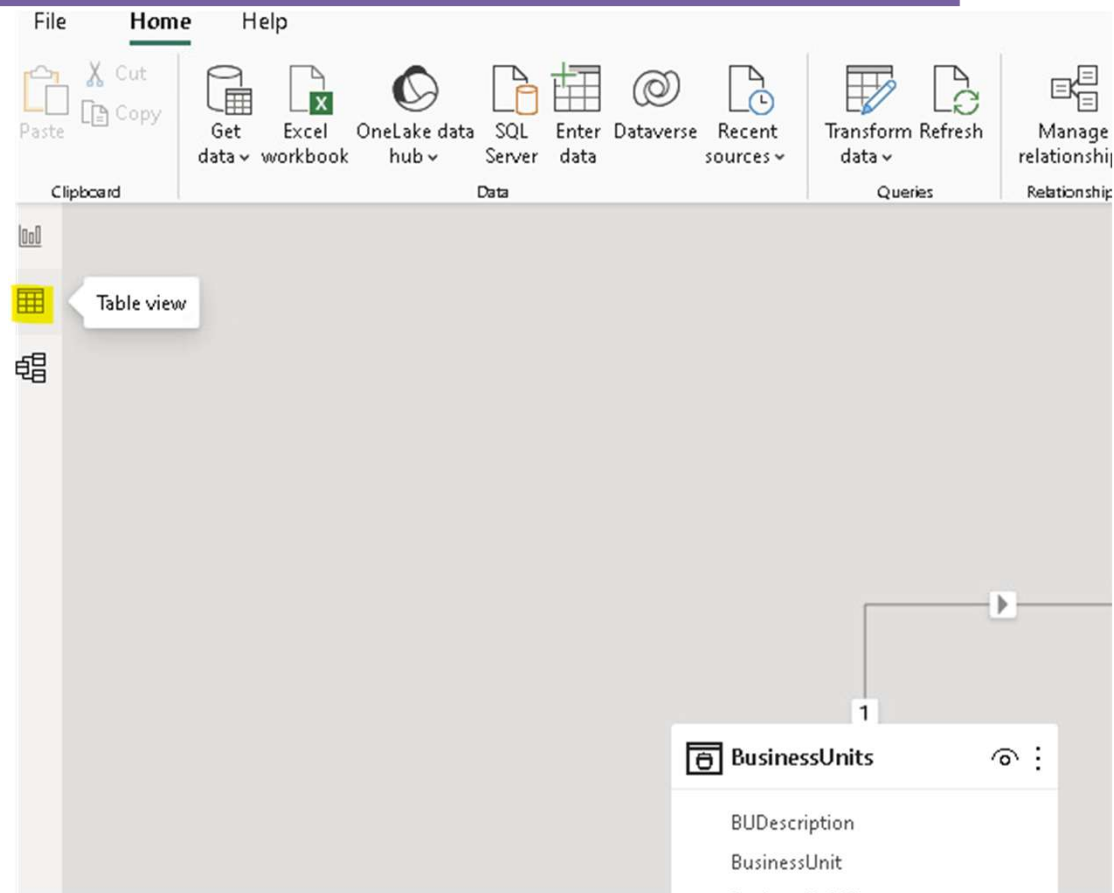
Notice that GLACccountNumber is summed (numeric) in JELineItems but not in Chart of accounts (means it is string).

*Joins can't be between numbers and string because:
"1"≠1



PowerBI Workflow

Go to Table view on left index

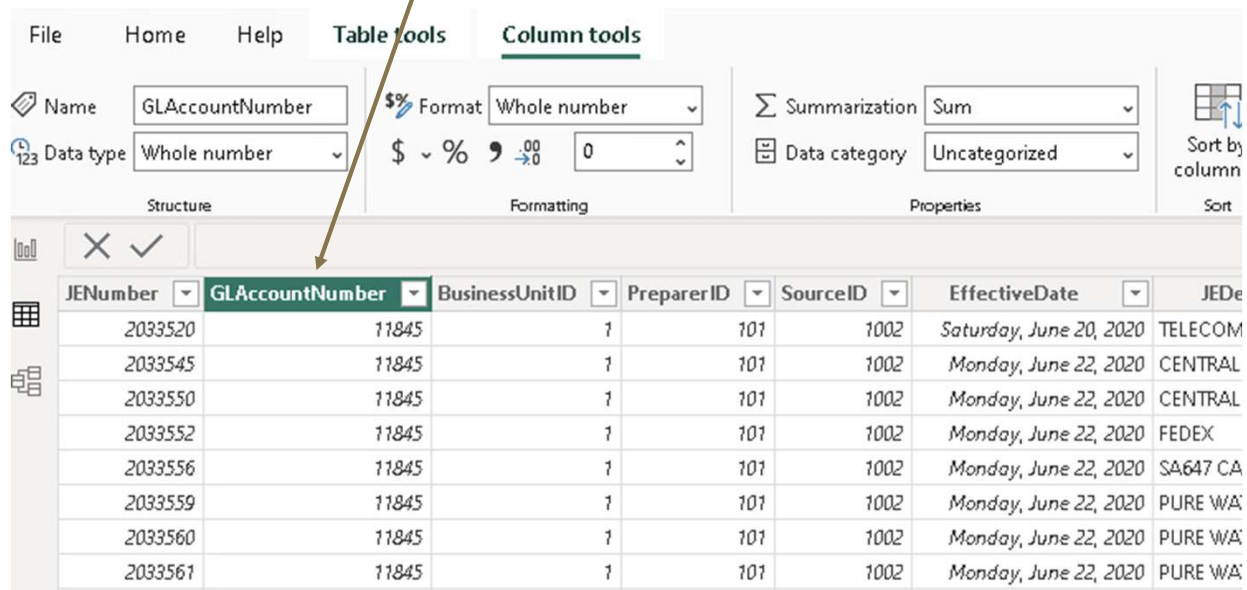


	-159.9	-159.9
	-159.9	-159.9

[illegible]

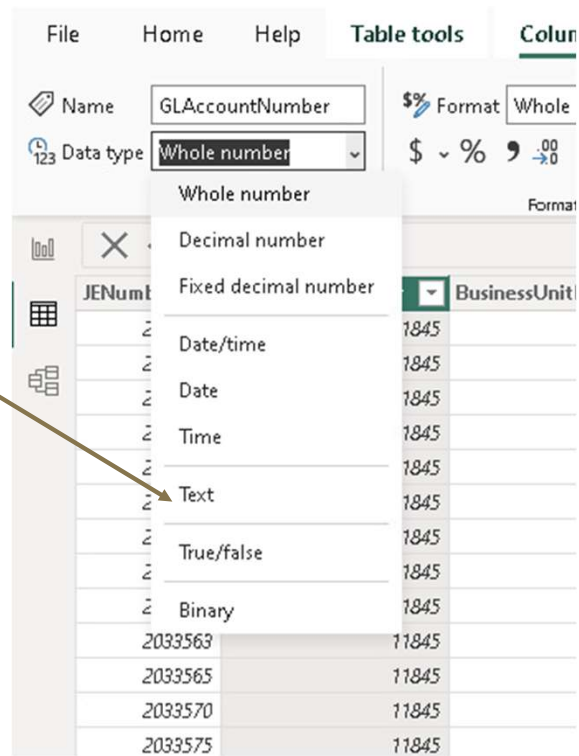
Note Dat type is whole number

Select the *GLAccountNumber* Field / Column



The screenshot displays the 'Column tools' ribbon in Microsoft Excel. The 'Name' field is set to 'GLAccountNumber' and the 'Data type' dropdown is set to 'Whole number'. The 'Format' dropdown is also set to 'Whole number'. The 'Summarization' dropdown is set to 'Sum' and the 'Data category' dropdown is set to 'Uncategorized'. The table below shows data for the 'GLAccountNumber' column.

JENumber	GLAccountNumber	BusinessUnitID	PreparerID	SourceID	EffectiveDate	JEDe
2033520	11845	1	101	1002	Saturday, June 20, 2020	TELECOM
2033545	11845	1	101	1002	Monday, June 22, 2020	CENTRAL
2033550	11845	1	101	1002	Monday, June 22, 2020	CENTRAL
2033552	11845	1	101	1002	Monday, June 22, 2020	FEDEX
2033556	11845	1	101	1002	Monday, June 22, 2020	SA647 CA
2033559	11845	1	101	1002	Monday, June 22, 2020	PURE WA
2033560	11845	1	101	1002	Monday, June 22, 2020	PURE WA
2033561	11845	1	101	1002	Monday, June 22, 2020	PURE WA



Confirm the change and how it affects format

Note format is just how the data is displayed, and it updates automatically to text (Data Type is how it is recognized by PowerBI).

File Home Help Table tools Column tools

Name GLAccountNumber

Data type Text

Format Text

Summarization Don't summarize

Data category Uncategorized

Sort by column

Data groups

JENumber	GLAccountNumber	BusinessUnitID	PreparerID	SourceID	EffectiveDate	JEDescription
2033520	11845	1	101	1002	Saturday, June 20, 2020	TELECOM GTA JUN 201
2033545	11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING
2033550	11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING
2033552	11845	1	101	1002	Monday, June 22, 2020	FEDEX
2033556	11845	1	101	1002	Monday, June 22, 2020	SA647 CARD JUN
2033559	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOI
2033560	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOI
2033561	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOI
2033562	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOI

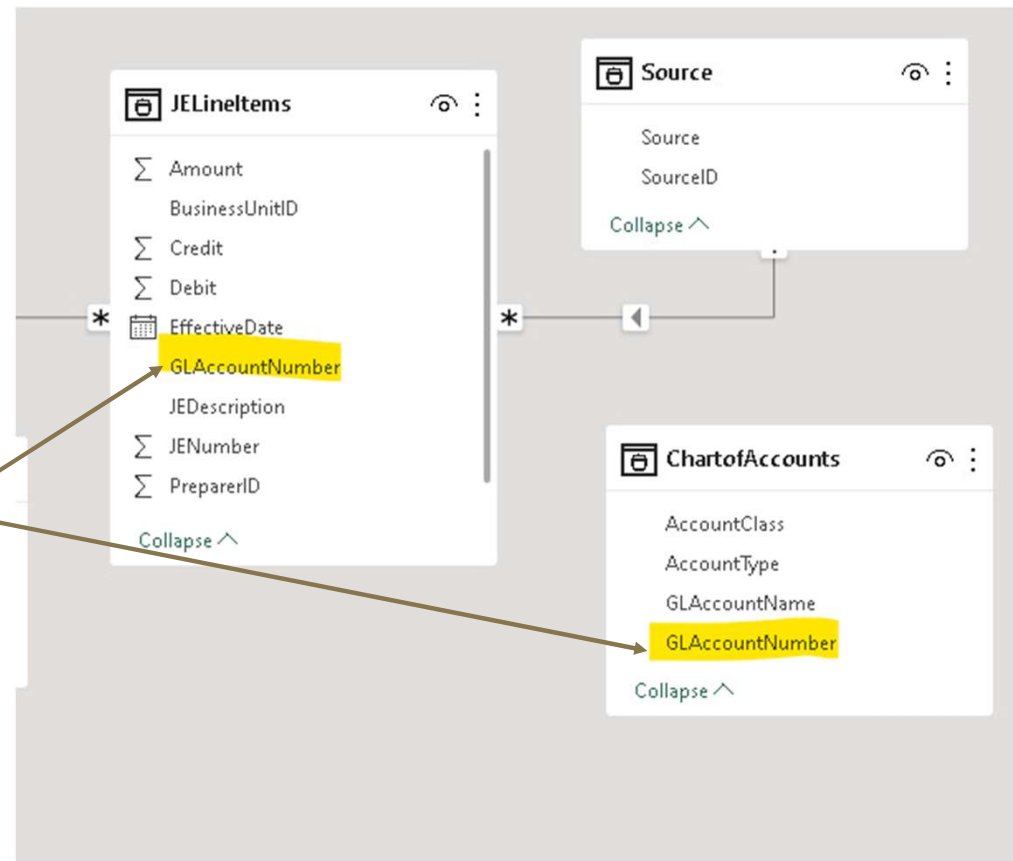
Text, left aligned

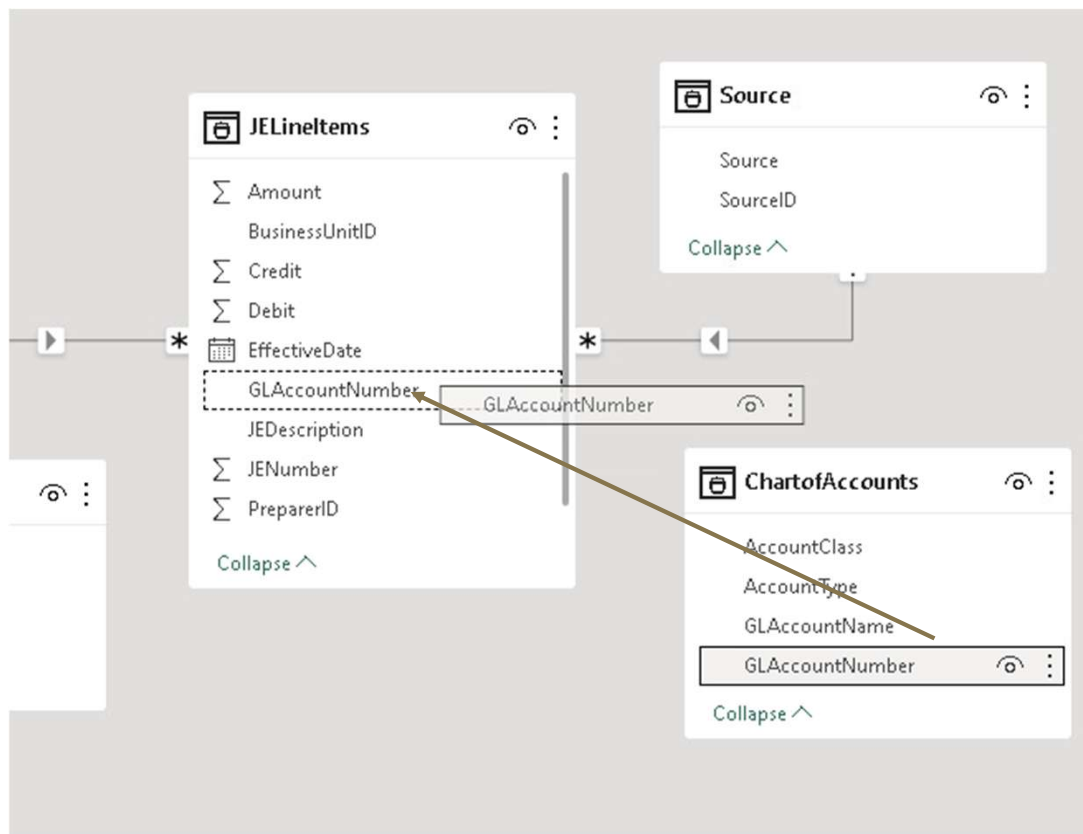
Numbers, italicized, right aligned

JENumber	GLAccountNumber	Bus
2033520	11845	
	11845	
	11845	
2033552	11845	
2033556	11845	
2033559	11845	
2033560	11845	
2033561	11845	

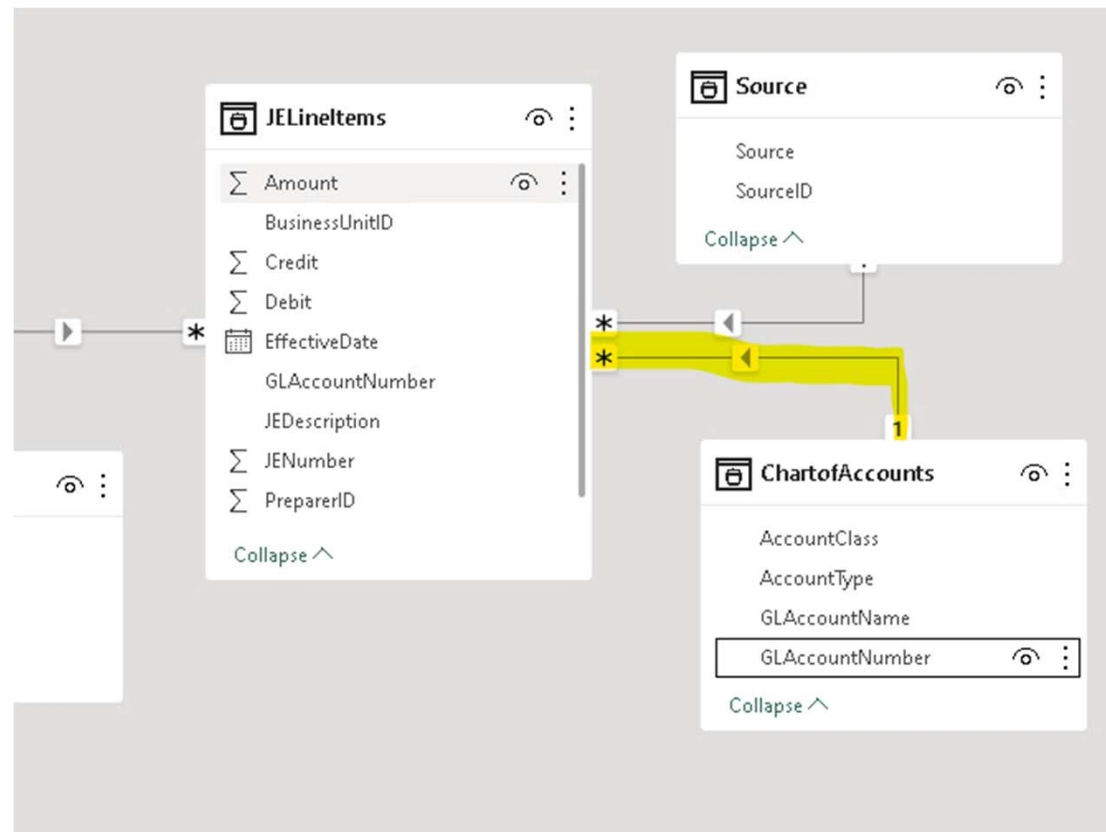
1 return to model view

Confirm
GLAccountNumber is text
in both (neither include
the summation sign)





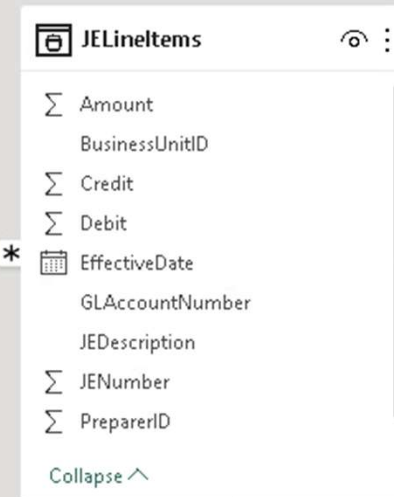
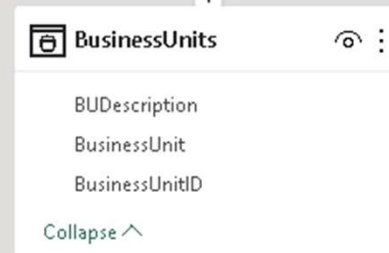
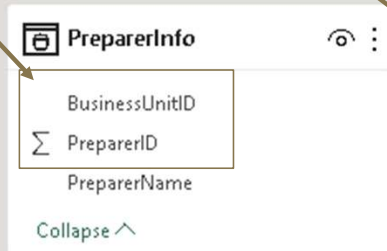
Drag GLAccountNumber from ChartOfAccounts to JELineItems and match it with GLAccountNumber



New Relationship
Arrow appears in the
data model

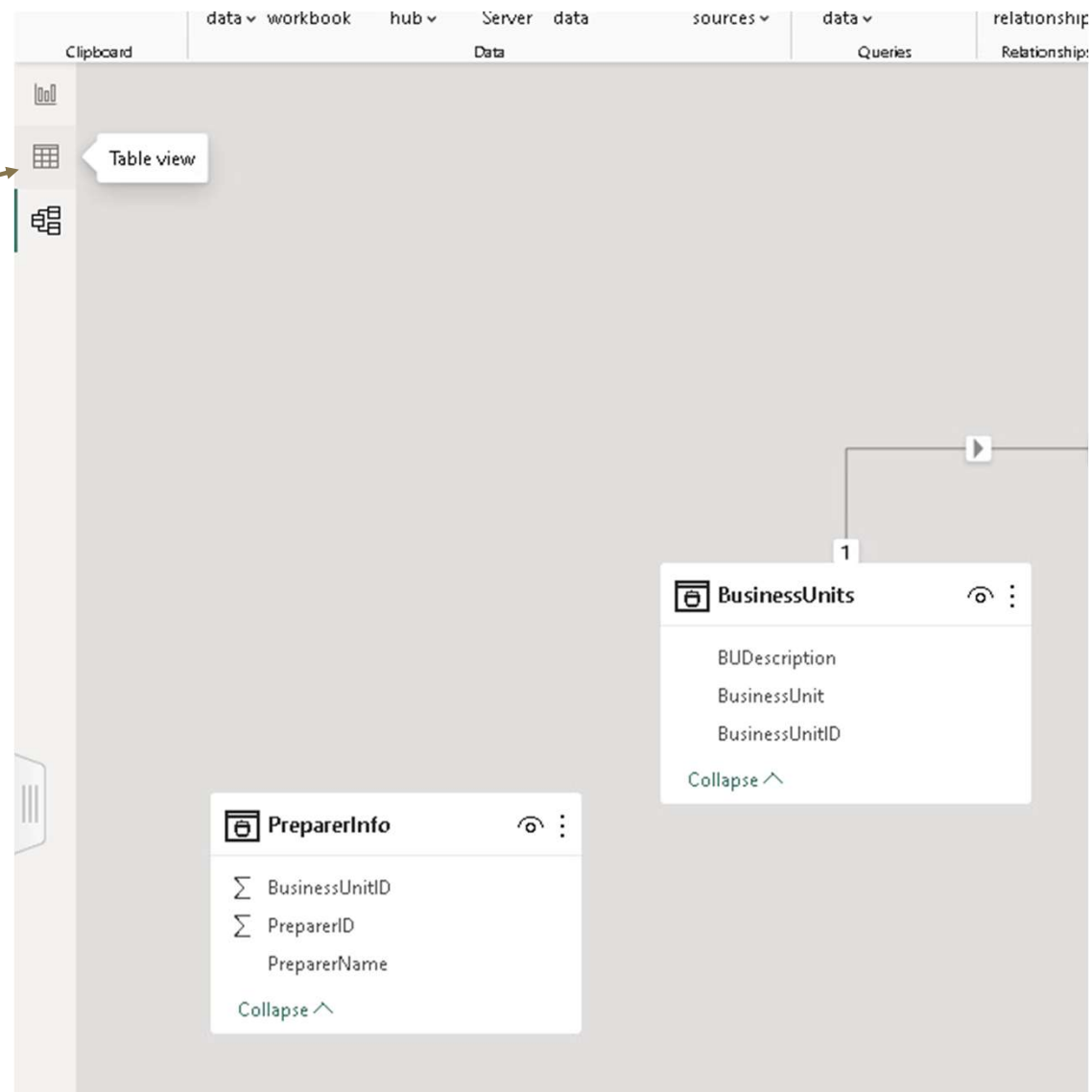
Like before we want:
BusinessUnitID AND PreparerID

1. Delete the current
connection



Let's make it possible
to connect
PreparerInfo directly
(in case we just want
preparer names)

2. Return to Table View



3. Confirm you are in the JELinItems dataset: Arrow pointed down, contents of dataset visible
Which is the most likely default given we just edited GLAccountNumber which is likely selected.

Data

Search

- > BusinessUnits
- > ChartofAccounts
- ✓ JELinItems
 - Σ Amount
 - BusinessUnitID
 - Σ Credit
 - Σ Debit
 - > EffectiveDate
 - GLAccountNumber
 - JEDescription
 - Σ JENumber
 - Σ PreparerID
 - SourceID
- > PreparerInfo
- > Source

4. Confirm or navigate to Table Tools

File Home Help **Table tools** Column tools

Name JELinItems

Structure

Mark as date table
Calendars

Manage relationships
Relationships

New measure
Calculations

Quick measure
Calculations

New column

New table

5. Create a new column

JENumber	GLAccountNumber	BusinessUnitID	PreparerID	SourceID	EffectiveDate	JEDescription	Debit	Credit	Amount
2033520	11845	1	101	1002	Saturday, June 20, 2020	TELECOM GTA JUN 2016		-36.9	-36.9
2033545	11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-6.37	-6.37
2033550	11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-45	-45
2033552	11845	1	101	1002	Monday, June 22, 2020	FEDEX		-55	-55
2033556	11845	1	101	1002	Monday, June 22, 2020	SA647 CARD JUN		-100	-100
2033559	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9
2033560	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9
2033561	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9
2033562	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9
2033563	11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9
2033565	11845	1	101	1002	Monday, June 22, 2020	EMILY R.ANDERSON P		-223	-223
2033570	11845	1	101	1002	Monday, June 22, 2020	EMILY R.ANDERSON P	-268.39		-268.39
2033575	11845	1	101	1002	Monday, June 22, 2020	STATE NATURAL GAS	-579.15		-579.15
2033576	11845	1	101	1002	Monday, June 22, 2020	COMPEAT INC	-695		-695

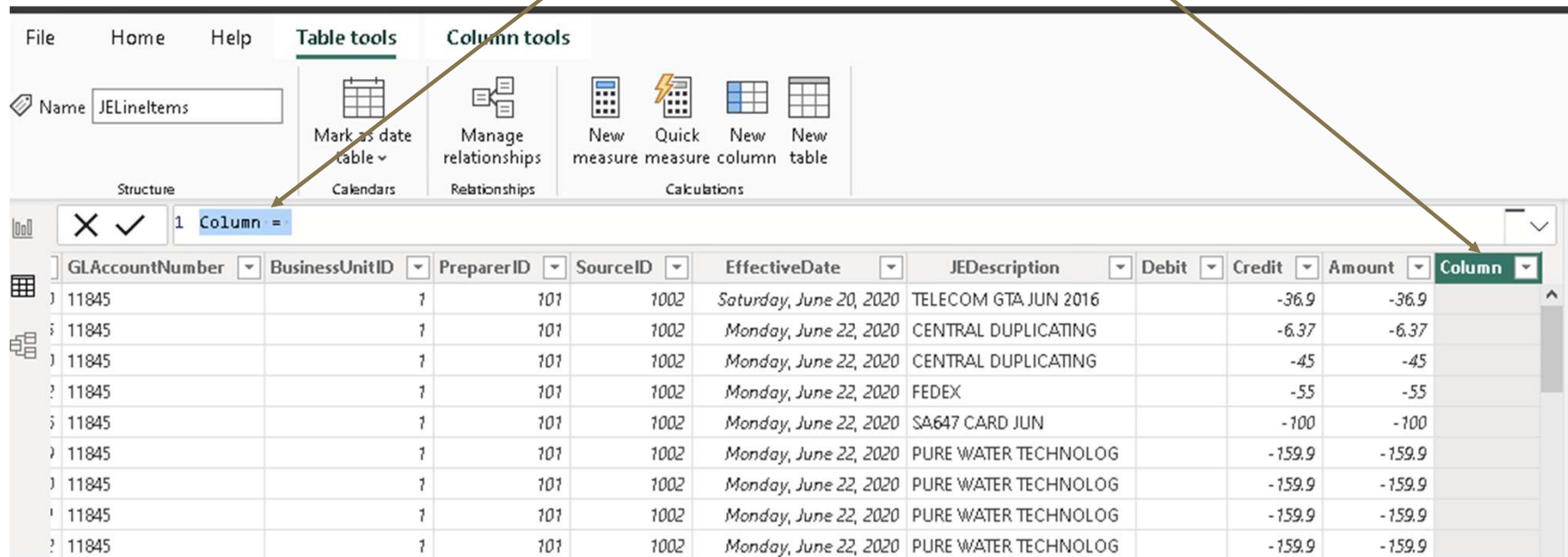
Data

Search

- BusinessUnits
- ChartofAccounts
- JELinItems
 - Amount
 - BusinessUnitID
 - Credit
 - Debit
 - EffectiveDate
 - GLAccountNumber**
 - JEDescription
 - JENumber

6. Confirm new column and ...

... 7. Edit the column formula here



File Home Help **Table tools** **Column tools**

Name: JELineItems

Structure: **Column =**

GLAccountNumber	BusinessUnitID	PreparerID	SourceID	EffectiveDate	JEDescription	Debit	Credit	Amount	Column
11845	1	101	1002	Saturday, June 20, 2020	TELECOM GTA JUN 2016		-36.9	-36.9	
11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-6.37	-6.37	
11845	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-45	-45	
11845	1	101	1002	Monday, June 22, 2020	FEDEX		-55	-55	
11845	1	101	1002	Monday, June 22, 2020	SA647 CARD JUN		-100	-100	
11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	
11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	
11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	
11845	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	

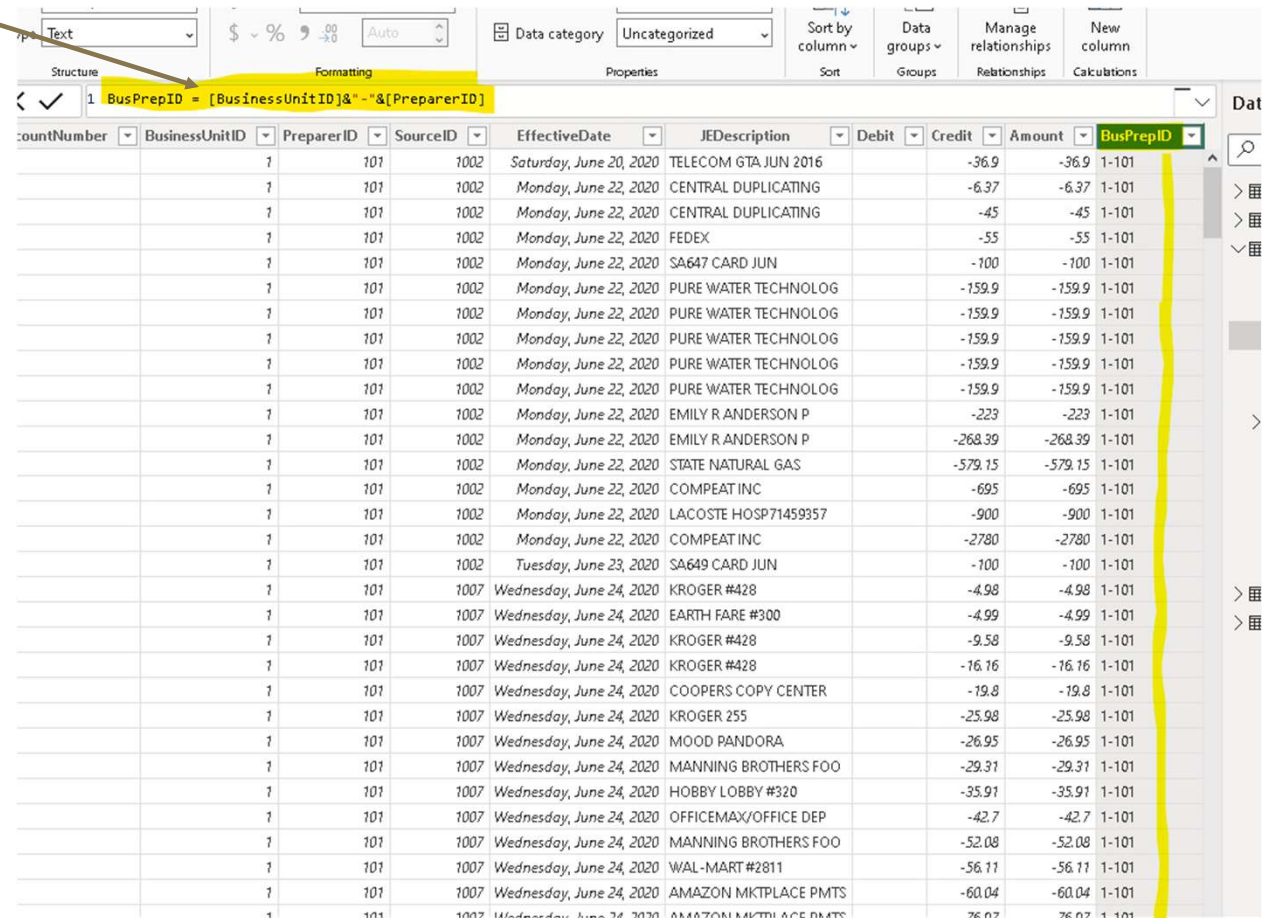
Our formula goal is to concatenate BusinessUnitID and PreparerID to create a unique key i.e., 1-101, 2-101 etc.

Our formula goal is to concatenate BusinessUnitID and PreparerID to create a unique key i.e., 1-101, 2-101 etc.

8. Replace Column = with:

BusPrepID = [BusinessUnitID]&"-"&[PreparerID]

Then press Enter



countNumber	BusinessUnitID	PreparerID	SourceID	EffectiveDate	JEDescription	Debit	Credit	Amount	BusPrepID
	1	101	1002	Saturday, June 20, 2020	TELECOM GTA JUN 2016		-36.9	-36.9	1-101
	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-6.37	-6.37	1-101
	1	101	1002	Monday, June 22, 2020	CENTRAL DUPLICATING		-45	-45	1-101
	1	101	1002	Monday, June 22, 2020	FEDEX		-55	-55	1-101
	1	101	1002	Monday, June 22, 2020	SA647 CARD JUN		-100	-100	1-101
	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	1-101
	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	1-101
	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	1-101
	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	1-101
	1	101	1002	Monday, June 22, 2020	PURE WATER TECHNOLOG		-159.9	-159.9	1-101
	1	101	1002	Monday, June 22, 2020	EMILY R ANDERSON P		-223	-223	1-101
	1	101	1002	Monday, June 22, 2020	EMILY R ANDERSON P		-268.39	-268.39	1-101
	1	101	1002	Monday, June 22, 2020	STATE NATURAL GAS		-579.15	-579.15	1-101
	1	101	1002	Monday, June 22, 2020	COMPEAT INC		-695	-695	1-101
	1	101	1002	Monday, June 22, 2020	LACOSTE HOSP71459357		-900	-900	1-101
	1	101	1002	Monday, June 22, 2020	COMPEAT INC		-2780	-2780	1-101
	1	101	1002	Tuesday, June 23, 2020	SA649 CARD JUN		-100	-100	1-101
	1	101	1007	Wednesday, June 24, 2020	KROGER #428		-4.98	-4.98	1-101
	1	101	1007	Wednesday, June 24, 2020	EARTH FARE #300		-4.99	-4.99	1-101
	1	101	1007	Wednesday, June 24, 2020	KROGER #428		-9.58	-9.58	1-101
	1	101	1007	Wednesday, June 24, 2020	KROGER #428		-16.16	-16.16	1-101
	1	101	1007	Wednesday, June 24, 2020	COOPERS COPY CENTER		-19.8	-19.8	1-101
	1	101	1007	Wednesday, June 24, 2020	KROGER 255		-25.98	-25.98	1-101
	1	101	1007	Wednesday, June 24, 2020	MOOD PANDORA		-26.95	-26.95	1-101
	1	101	1007	Wednesday, June 24, 2020	MANNING BROTHERS FOO		-29.31	-29.31	1-101
	1	101	1007	Wednesday, June 24, 2020	HOBBY LOBBY #320		-35.91	-35.91	1-101
	1	101	1007	Wednesday, June 24, 2020	OFFICEMAX/OFFICE DEP		-42.7	-42.7	1-101
	1	101	1007	Wednesday, June 24, 2020	MANNING BROTHERS FOO		-52.08	-52.08	1-101
	1	101	1007	Wednesday, June 24, 2020	WAL-MART #2811		-56.11	-56.11	1-101
	1	101	1007	Wednesday, June 24, 2020	AMAZON MKTPLACE PMTS		-60.04	-60.04	1-101
	1	101	1007	Wednesday, June 24, 2020	AMAZON MKTPLACE PMTS		-76.07	-76.07	1-101

10. Still in Tables, select the PreparerInfo Dataset and create the same new column "BusPrepID"

The screenshot displays the 'Tables' view in Microsoft Dynamics 365. The main table has the following columns: Action, Debit, Credit, Amount, and BusPrepID. The 'BusPrepID' column is highlighted. The table contains various entries, including dates, descriptions, and numerical values. On the right side, the 'Data' pane is open, showing a list of datasets. The 'PreparerInfo' dataset is selected, and an arrow points from the text instruction to it.

Action	Debit	Credit	Amount	BusPrepID
1 JUN 2016		-36.9	-36.9	1-101
PLICATING		-6.37	-6.37	1-101
PLICATING		-45	-45	1-101
		-55	-55	1-101
IUN		-100	-100	1-101
TECHNOLOG		-159.9	-159.9	1-101
TECHNOLOG		-159.9	-159.9	1-101
TECHNOLOG		-159.9	-159.9	1-101
TECHNOLOG		-159.9	-159.9	1-101
TECHNOLOG		-159.9	-159.9	1-101
ERSON P		-223	-223	1-101
ERSON P		-268.39	-268.39	1-101
AL GAS		-579.15	-579.15	1-101
C		-695	-695	1-101
SP71459357		-900	-900	1-101
C		-2780	-2780	1-101
IUN		-100	-100	1-101
		-4.98	-4.98	1-101
#300		-4.99	-4.99	1-101
		-9.58	-9.58	1-101
		-16.16	-16.16	1-101
PY CENTER		-19.8	-19.8	1-101
		-25.98	-25.98	1-101
ORA		-26.95	-26.95	1-101
OTHERS FOO		-29.31	-29.31	1-101
Y #320		-35.91	-35.91	1-101
OFFICE DEP		-42.7	-42.7	1-101

The 'Data' pane on the right shows the following datasets:

- BusinessUnits
- ChartofAccounts
- JELinItems
 - Amount
 - BusinessUnitID
 - BusPrepID
 - Credit
 - Debit
 - EffectiveDate
 - GLAccountNumber
 - JEDescription
 - JENumber
 - PreparerID
 - SourceID
- PreparerInfo
- Source

File Home Help **Table tools**

Name **PreparerInfo**

Structure

Mark as date table

Manage relationships

New measure Quick measure New column New table

Calculations

PreparerName	PreparerID	BusinessUnitID
Carolyn Slater Food Operations Manager	101	1
Alexander Knox Assistant Manager	101	2
Trevor Roberts Housekeeping Supervisor	101	4
Gavin McGrath Dining Commons Manager	101	3
Rodney Wilkins Maintenance Supervisor	101	6
Elizabeth Rees Food Operations Manager	102	1
Brian Gibson Assistant Manager	102	2
POS	102	3
David Brown Accounting Manager	103	2
GUESTSYS	104	2
Victoria Simpson Accounting Manager	105	2
Yvonne Chapman Catering Manager	106	2

Name: BusPrepID
 Data type: Text
 Format: Text
 Summarization: Don't summarize
 Data category: Uncategorized

1 BusPrepID = [BusinessUnitID]&"-"&[PreparerID]

PreparerName	PreparerID	BusinessUnitID	BusPrepID
Carolyn Slater Food Operations Manager	101	1	1-101
Alexander Knox Assistant Manager	101	2	2-101
Trevor Roberts Housekeeping Supervisor	101	4	4-101
Gavin McGrath Dining Commons Manager	101	3	3-101
Rodney Wilkins Maintenance Supervisor	101	6	6-101
Elizabeth Rees Food Operations Manager	102	1	1-102
Brian Gibson Assistant Manager	102	2	2-102
POS	102	3	3-102
David Brown Accounting Manager	103	2	2-103
GUESTSYS	104	2	2-104
Victoria Simpson Accounting Manager	105	2	2-105
Yvonne Chapman Catering Manager	106	2	2-106

Update formula

- Note that this column, BusPrepID, is now:
1. In JELineItems, and
 2. PreparerInfo

Data

Search

- > BusinessUnits
- > ChartofAccounts
- > JELineItems
 - Σ Amount
 - BusinessUnitID
 - BusPrepID
 - Σ Credit
 - Σ Debit
 - > EffectiveDate
 - GLAccountNumber
 - JEDescription
 - Σ JENumber
 - Σ PreparerID
 - SourceID
 - > PreparerInfo
 - Σ BusinessUnitID
 - BusPrepID
 - Σ PreparerID
 - PreparerName
 - > Source

Name: BusPrepID Format: Text

Data type: Text

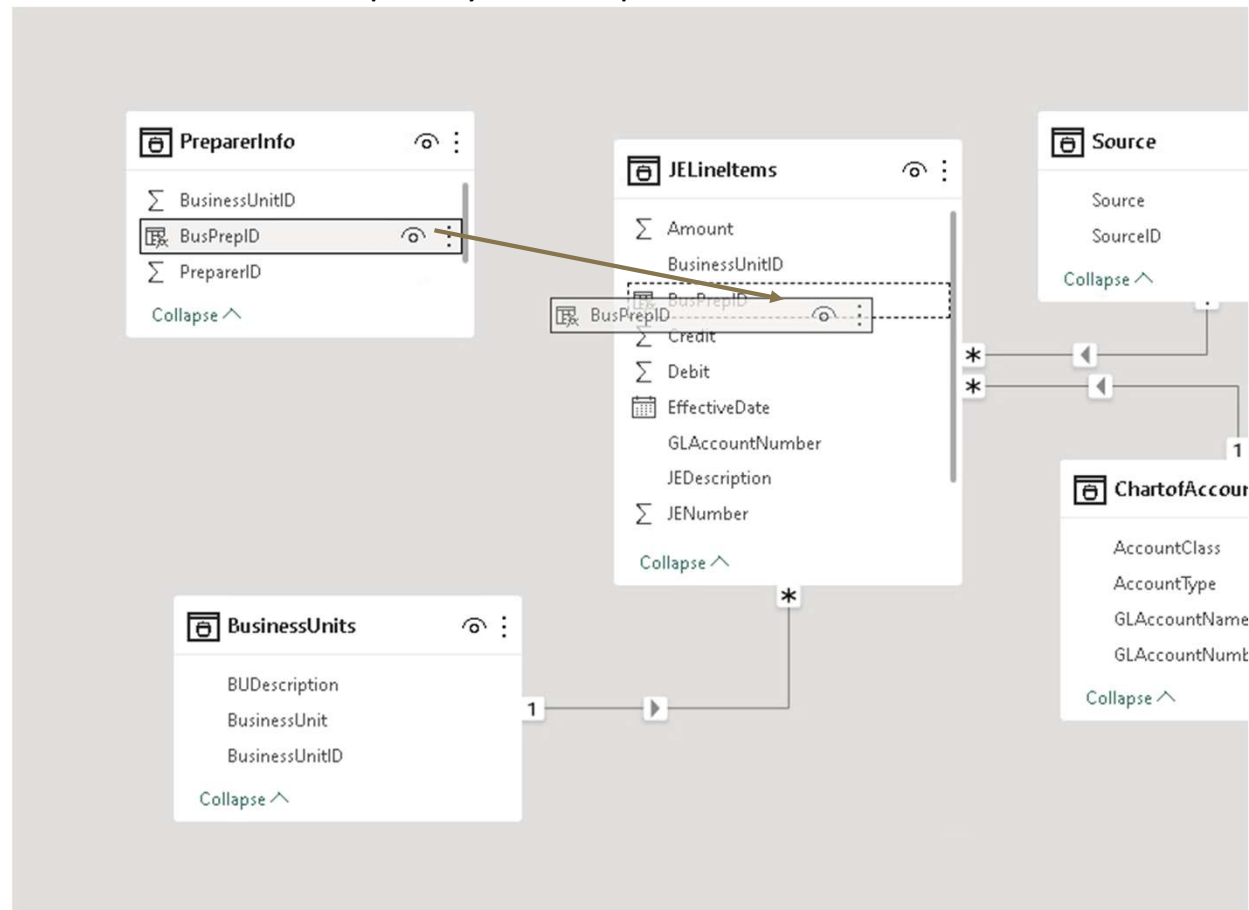
Structure: 1 BusPrepID = [BusinessUnitID]&"-"&

PreparerName	PreparerID
Carolyn Slater Food Operations Manager	101
Assistant Manager	101
Housekeeping Supervisor	101
Gavin McGrath Dining Commons Manager	101
Rodney Wilkins Maintenance Supervisor	101

Model view

Return to model view

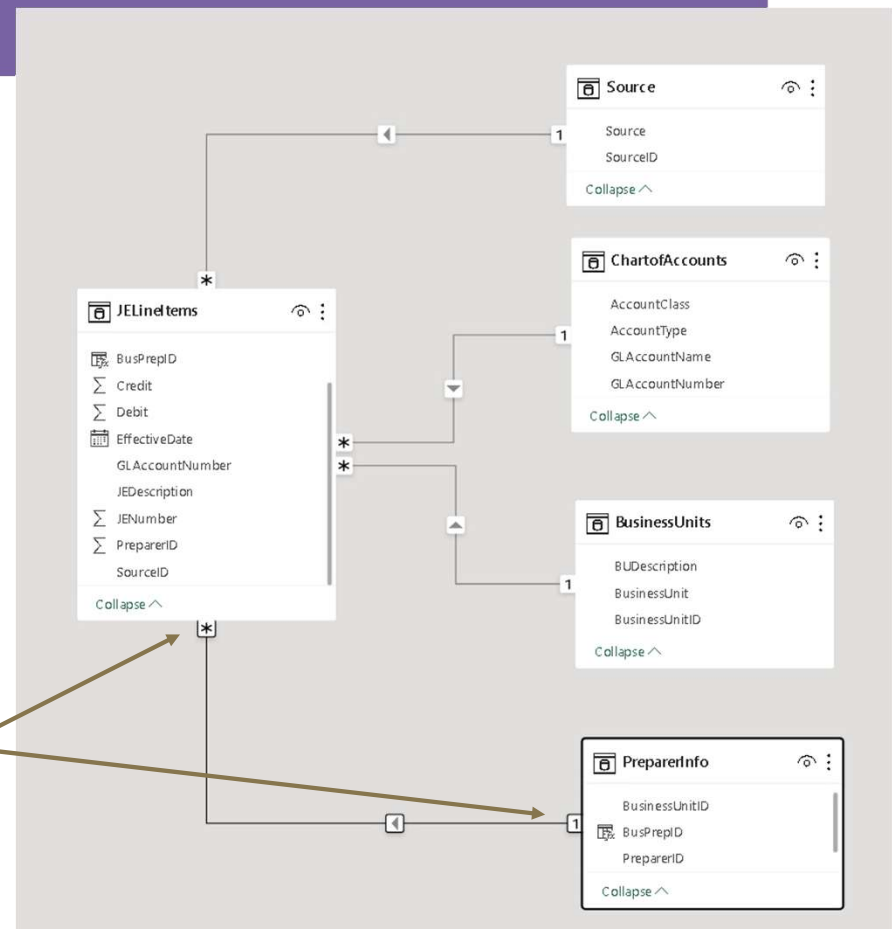
Connect PreparerInfo and JELineItems by dragging the new unique key "BusPrepID"



PowerBI Completed

Confirm the new
BusPrepID
connection and
the **data
relationship
model** is
complete.

Note it should be
one (1)
to many (*)



Conclusion and look-ahead

ETL is the first step in the analytical process – we cannot undertake data analytics without data!

ETL is especially important when the data needs to be cleaned or transformed for use in analytics.

Conclusion and look-ahead

Wednesday:

- Cases: Enron Emails Case – RegEx Heavy

Thank you and Congratulations!

