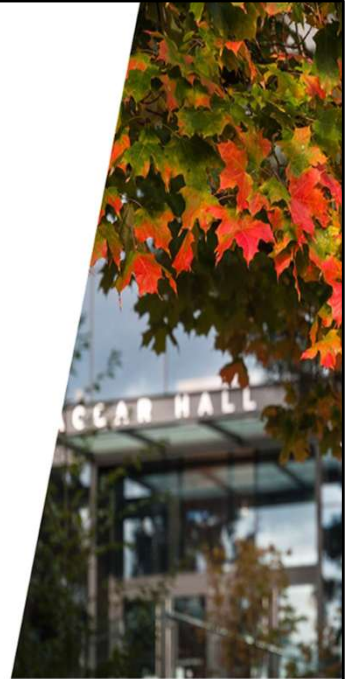


ADVANCED ETL AND TEXTUAL ANALYSIS

Data Analytics for Professional Accountants
(ACCTG 522)

Class 4 | MPAcc Class of 2026

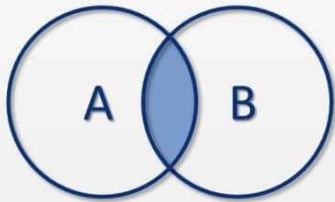


Welcome!

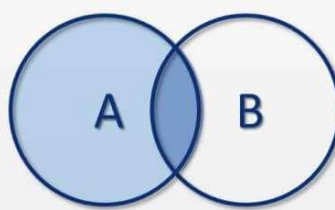
Class Agenda

- **(Extended) Review: Joins Case**
 - Mini-Lab Alteryx Joins
- Case Overview: Enron Emails Case
- Lab: RegEx parsing of email header & subject/body
 - Tools (RegEx, DateTime)

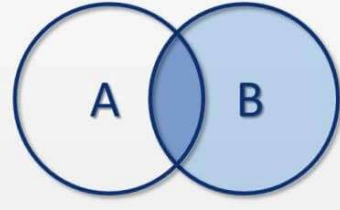
INNER JOIN



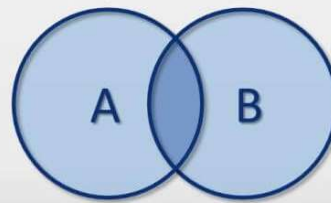
OUTER JOIN



LEFT



RIGHT



FULL

Tableau Relationships

⊖ JELinItems+ (Analytics_mindset_case_studies_ET...

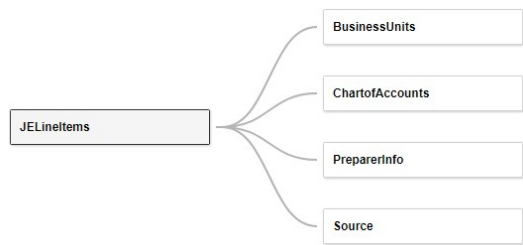
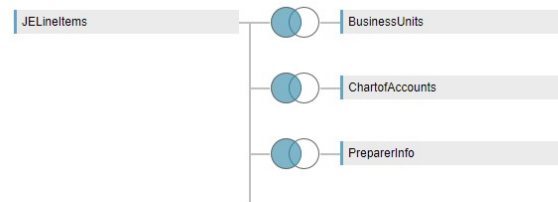
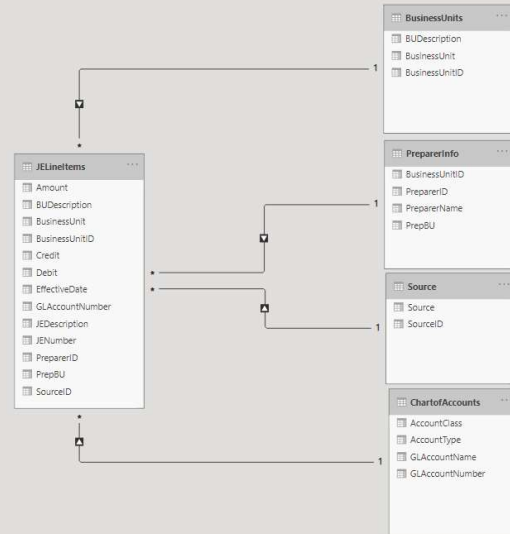


Tableau Joins

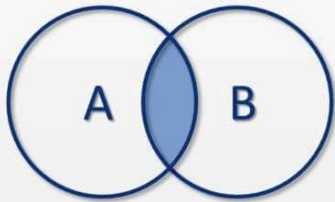
JELineItems is made of 5 tables. ⓘ



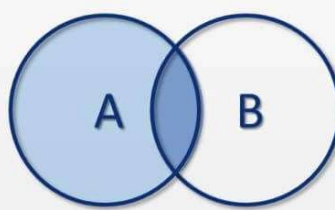
Power BI Data Relationships



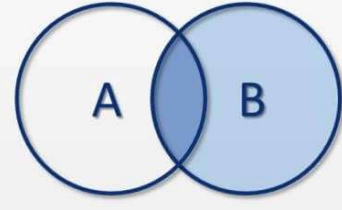
INNER JOIN



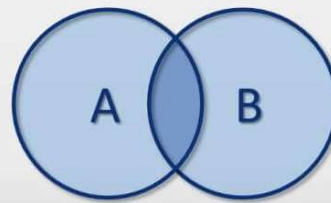
OUTER JOIN



LEFT

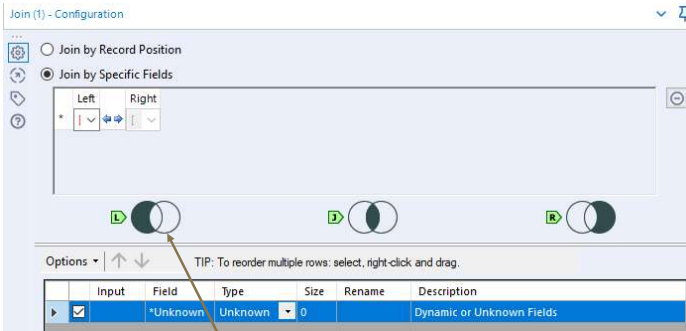


RIGHT

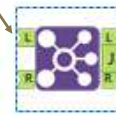


FULL

Alteryx Join Tool:



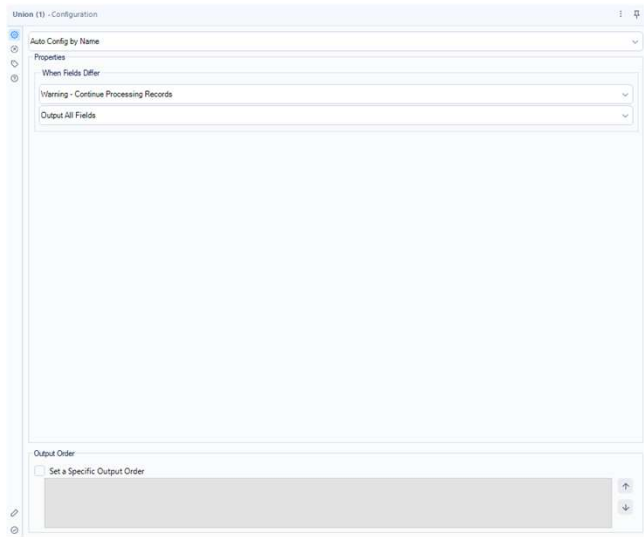
"L" input node



"L" output node

Alteryx explains what is kept for each node, e.g. "L" output

Alteryx Union Tool:

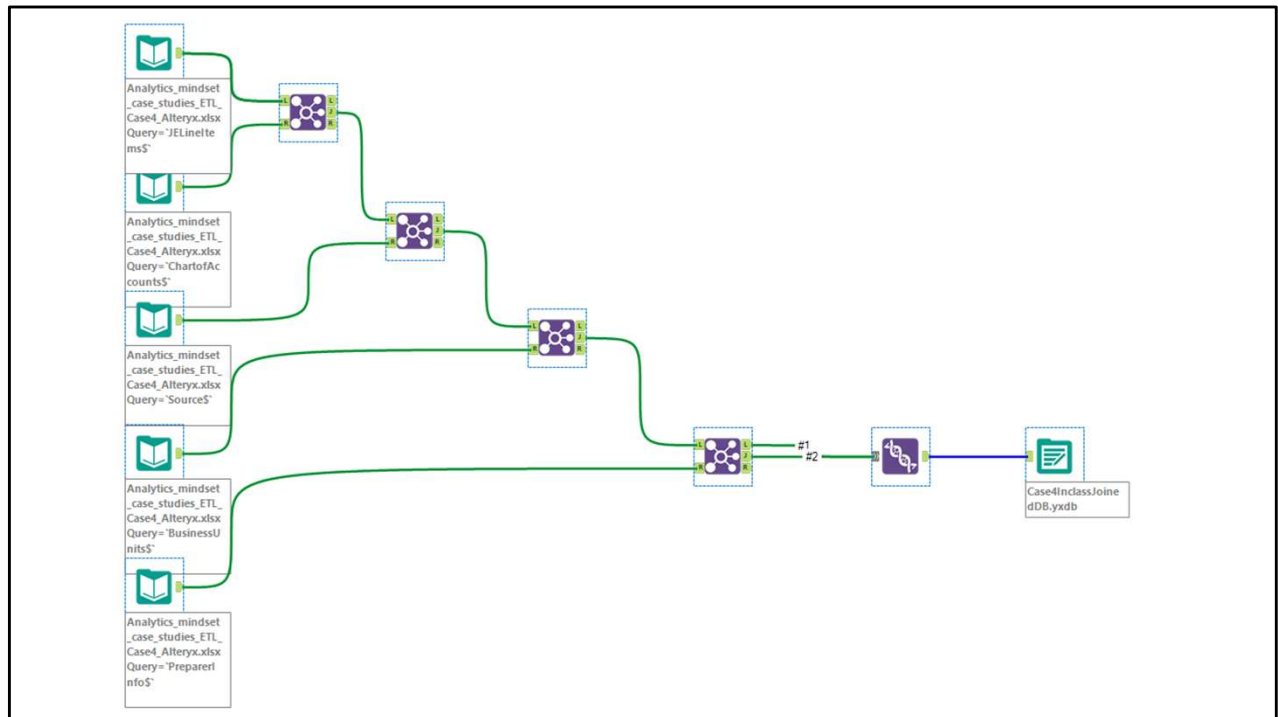


Multiple input node



Single output node





Agenda

Class Agenda

- (Extended) Review: Joins Case
 - Mini-Lab Alteryx Joins
- **Case Overview: Enron Emails Case**
- **Lab: RegEx parsing of email header & subject/body**
 - Tools (RegEx, DateTime)

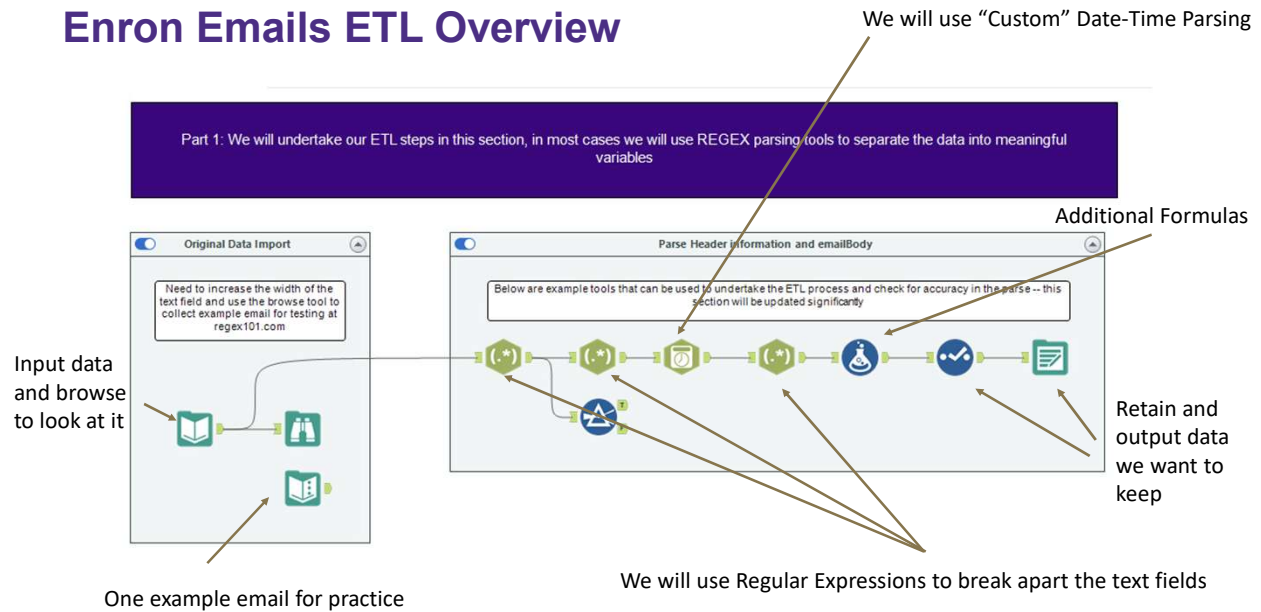
Case Overview: Enron Emails Case

Regex101

Paste the example email below into <https://regex101.com/>:

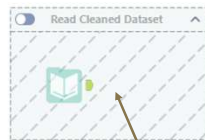
Message-ID: <25828831.1075855376669.JavaMail.evans@thyme> Date: Mon, 26 Nov 2001 07:25:16 -0800 (PST)
From: k..allen@enron.com To: jsmith@austintx.com Subject: RE: Additional properties in San Antonio Jeff, Can you
resend the info on the three properties you mailed and the one you faxed on Tuesday. I was out of the office last
week. Phillip -----Original Message----- From: "JEFF SMITH" <jsmith@austintx.com>@ENRON Sent: Tuesday,
November 20, 2001 3:43 PM To: Allen, Phillip K. Subject: Additional properties in San Antonio Phillip, I am waiting
to get info. on two more properties in San Antonio. The broker will be faxing info. on Monday. One is 74 units for
\$1,900,000, and the other is 24 units for \$550,000. Also, I have mailed you info. on three other properties in
addition to the 100 unit property that I faxed this AM. Let me know if you have any questions. Jeff Smith The Smith
Company 9400 Circle Drive Austin, TX 78736 512-394-0908 office 512-394-0913 fax 512-751-9728 mobile
jsmith@austintx.com

Enron Emails ETL Overview



Enron Emails Analysis Overview

Part 2: We will undertake our analysis steps in this section, in most cases we will be using either REGEX parsing tools or regex matches to identify the possible cybersecurity related violations.



Save dataset from part 1 and read it into Alteryx

Question 2

Question 4

Question 3

Question 5

Teams will work on each question together

Start on:

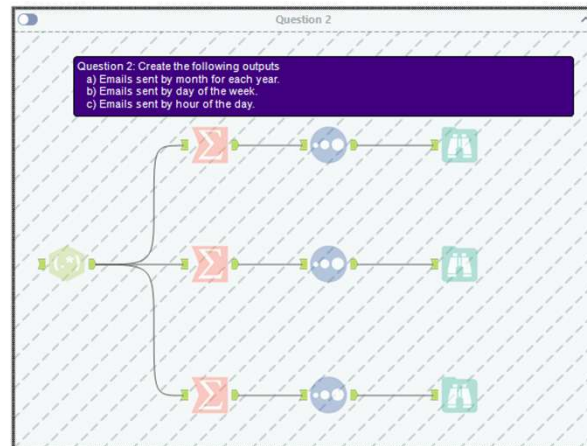
Q2 (teams 1,2,3) -> Q3

Q3 (teams 4,5) -> Q2

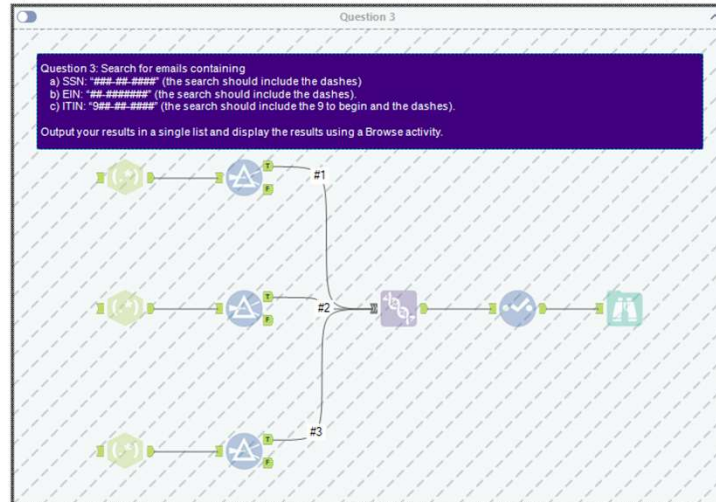
Q4 (teams 6,7) -> Q5

Q5 (teams 8,9,10) -> Q4

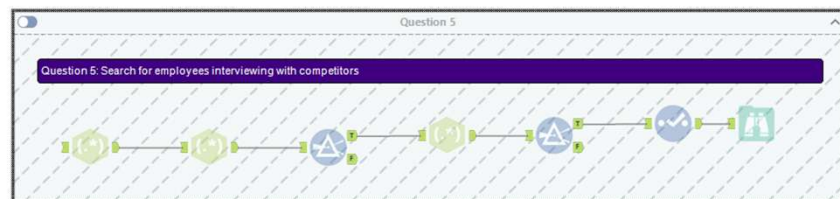
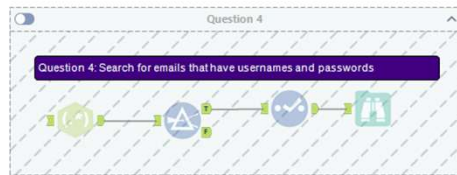
Enron Emails Analysis Hints



Enron Emails Analysis Hints



Enron Emails Analysis Hints

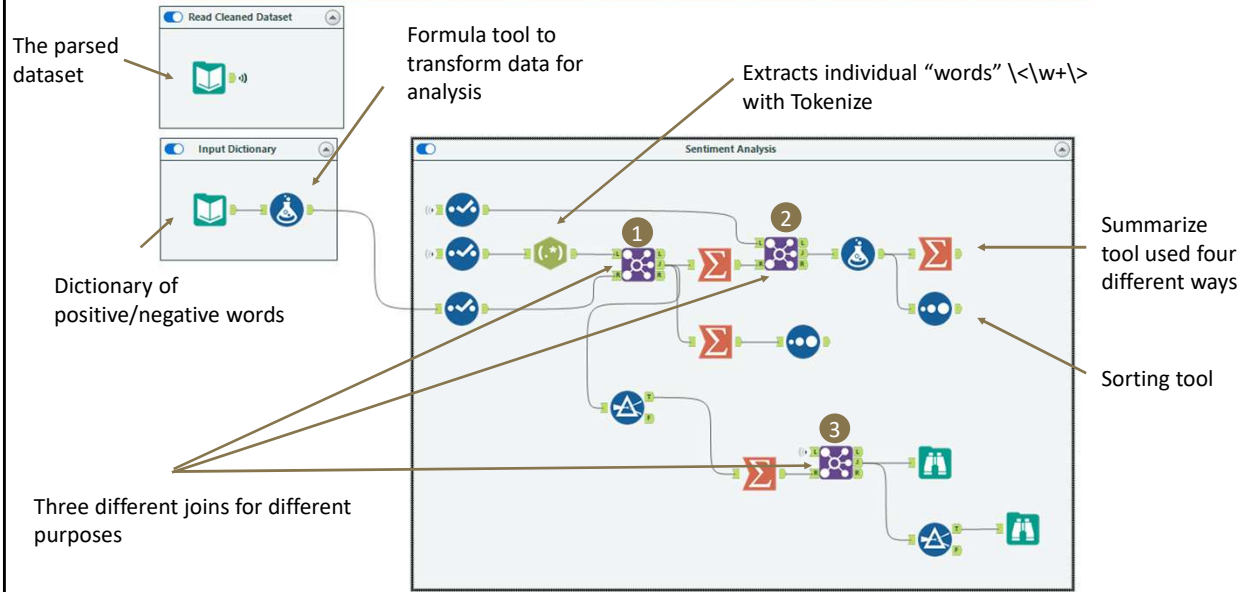


Extension: Enron Email Case

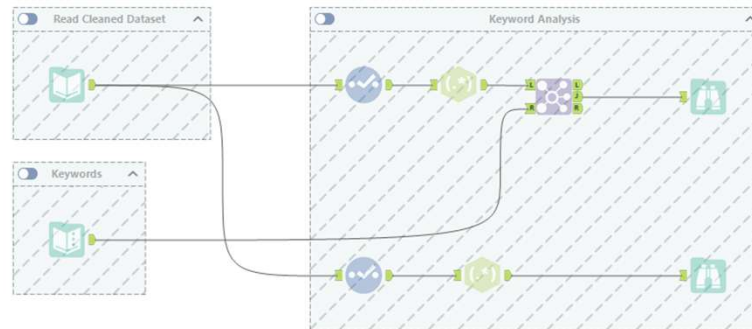
Extracting data from unstructured text, like emails, requires more advanced ETL tools. We can use regular expressions to:

- **Undertake open-ended exploration for suspicious email and negative sentiment/tone over time.**

Part 3: Examine email tone or sentiment over time and identify key words to search for suspicious or target emails. In this example, we will calculate negative sentiment as negative words / total words, and look for email relating to SEC investigations.



Part 4: Keyword searches, undertake Forensic approach to analysis, find suspicious email conversations, potential ethical concerns or discussions of SEC investigations. As a bonus, see if you can find the poem! Hint: One Fish, Two Fish, Yellow Fish, Enron



Conclusion and look-ahead

We have started using advanced tools to perform analyses on unstructured (email) data.

We will continue this on Wednesday with keyword and sentiment analyses.

Thank you!

