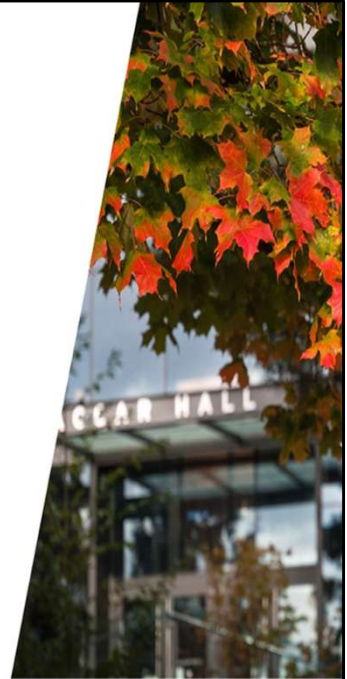


ADVANCED ETL AND TEXTUAL ANALYSIS 2

Data Analytics for Professional Accountants
(ACCTG 522)

Class 5 | MPAcc Class of 2026



Welcome!

Class Agenda

- **Case Review: Enron Emails Case**
- Case Deliverable Overview
- Lab Tool and workflow overviews:
 - DateTime parsing functions
 - Sentiment Analysis
 - Phrase Analysis
 - Individual Forensic Workflow
- Lab Forensic Analysis

Case Overview: Enron Emails Case

Parsing Text using Regex

Our test email:

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

Parsing Text using Regex

Our test email:

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

Unique Key

Unstructured data

Key identifier variables

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

Parsing Text using Regex

Our test email:

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

Unique Key

Unstructured data

Key identifier variables

What patterns can we use to extract meaning from the Subject/Body of the email (just this email)?

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

What patterns can we use to extract meaning from the Subject/Body of the email (just this email)?

Pattern meaning



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

Another example email

Subject: Rick: As I am sure you have read by now, Jeff Skilling has chosen to leave Enron. Jeff is resigning for personal reasons and his decision is voluntary. I regret his decision, but I accept and understand it. I have worked closely with Jeff for more than 15 years, including 11 here at Enron, and have had few, if any, professional relationships that I value more. I am pleased to say that he has agreed to enter into a consulting arrangement with the company to advise me and the Board of Directors. I want to assure you that Jeff's departure will in no way affect the relationship that Enron has with Microsoft and more specifically MSN. I have been briefed on the MSNIA project, by my team, and they assure me that we are on track to meet your Broadband Internet Access roll-out plans in early October. I look forward to getting more personally involved with Microsoft to help insure that our ongoing relationship and partnership continues to mature and flourish. I will contact you in the near future to discuss with you this transition. In the meantime, if you have any questions or concerns please do not hesitate to contact my office. Ken

Welcome!

Class Agenda

- Case Review: Enron Emails Case
- **Case Deliverable Overview**
- Lab Tool and workflow overviews:
 - DateTime parsing functions
 - Sentiment Analysis
 - Phrase Analysis
 - Individual Forensic Workflow
- Lab Forensic Analysis



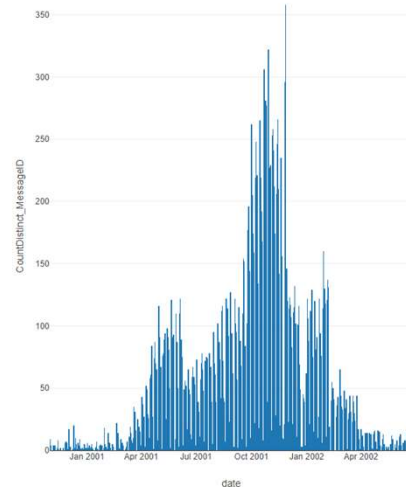
**VERYTHING'S
BIGGER IN**



TEXAS

Enron Key Timeline

- 1999–2000 → Enron at peak: rapid growth, accounting complexity.
- **Aug 2001** → **Skilling resigns, Watkins memo.**
- **Oct 2001** → **Massive loss and equity reduction disclosed.**
- **Oct–Nov 2001** → **SEC investigation begins.**
- **Nov 2001** → **Restatement and Dynegy merger collapse.**
- **Dec 2001** → **Bankruptcy (at time biggest ever).**
- 2002 → Andersen indicted and collapses.
- 2002–2006 → Enron Task Force prosecutes key executives.



Case Deliverable Question

Conduct a forensic review of internal communications from the Enron email corpus. What evidence, if any, points to failures of governance, accountability, or integrity within the organization? Submit a memo and attach supporting exhibits from your analysis.

Memo Structure

To: Asher Curtis

From: [Student Name], Forensic Analyst

Subject: Summary of Evidence from Enron Email Corpus

Date: [Insert date]

Summary of Findings (1–2 paragraphs)

Brief overview of evidence discovered

Relevance to possible fraud or ethical violations

Analytical Methods Used

Regex and/or keyword and/or individual search

Sentiment analysis or topic clustering

Network mapping or anomaly detection

Evidence Highlights (with attachments)

Include 2–3 key emails or message excerpts

Visualizations: network graphs, frequency plots, etc.

Interpretation and Recommendations

Explain significance

Suggest further investigative steps (e.g., deeper audit of certain users)

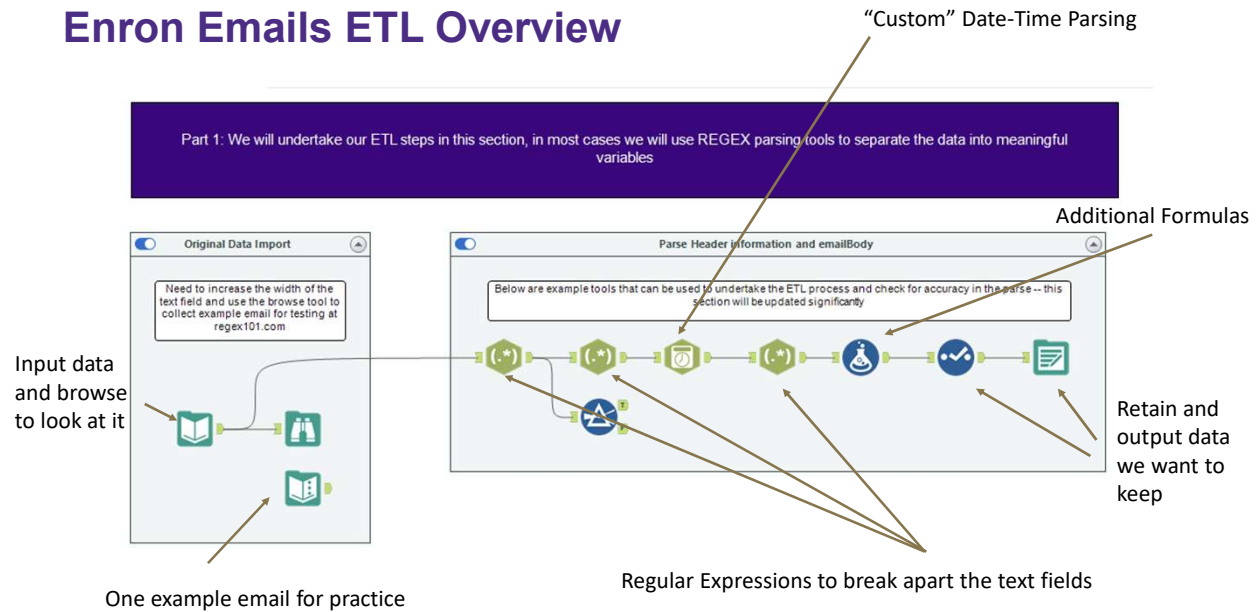
Attachments

Agenda

Class Agenda

- Case Review: Enron Emails Case
- Case Deliverable Overview
- **Lab Tool and workflow overviews:**
 - **DateTime parsing functions**
 - **Sentiment Analysis**
 - **Phrase Analysis**
 - **Individual Forensic Workflow**
- Lab Forensic Analysis

Enron Emails ETL Overview



DateTime (25) - Configuration

Select the format to convert

☐ Date/Time format to string

☒ String to Date/Time format

Select the string field to convert

emailDate

Specify the new column name

DateTime_Out

Specify your DateTime Language

English

Select the format that matches the incoming string field

MM/dd/yy h:mm:ss

dd/MM/yyyy h:mm:ss

dd/MM/yy h:mm:ss

day, dd Month, yyyy

dd-MM-yy

dd-MM-yyyy

dd-Mon-yy

dd Month, yyyy

dd-MM/yy

dd/MM/yyyy

day, Month dd, yyyy

MM-dd-yy

MM-dd-yyyy

MM/dd/yy

MM/dd/yyyy

Month dd, yyyy

Month, yyyy

yyyy-MM-dd

yyyyMMda

yyyy/MM/dd

HH:mm:ss

Custom

Specify the format of the incoming string field

dy, d Mon yyyy HH:mm:ss

Example

Sun, 2 Jan 2000 03:04:05

Output

2000-01-02 03:04:05

Note: The incoming string should match the example.



“Custom” Date-Time Parsing
Specifying: dy, d Mon yyyy HH:mm:ss

****Warning:** Sometimes this tool will default back to unconfigured, losing the info in Custom. Type in the date format above exactly to have it run correctly.

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

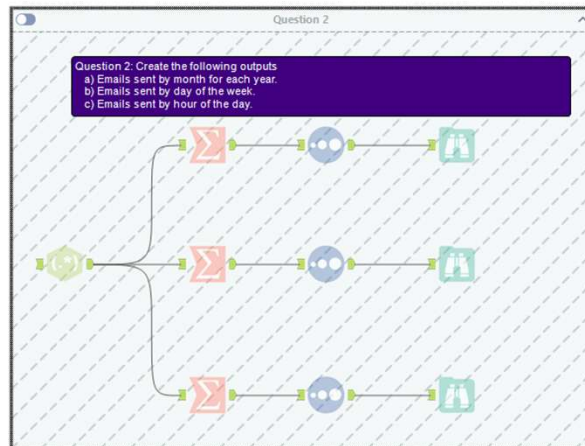
☐ Question 4

☒ Question 3

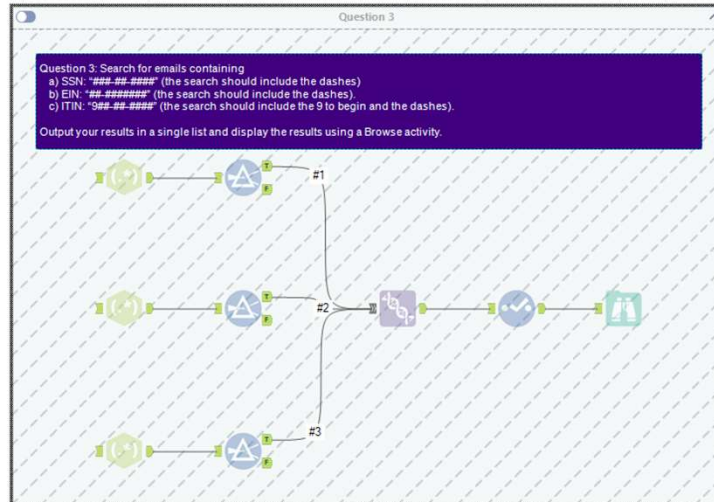
☐ Question 5

☒ Additional Examples

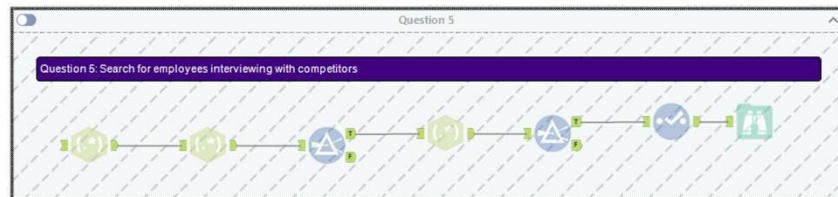
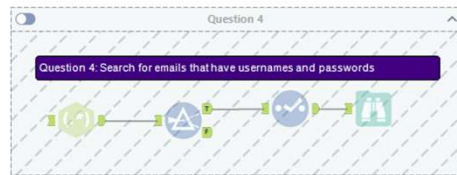
Enron Emails Analysis



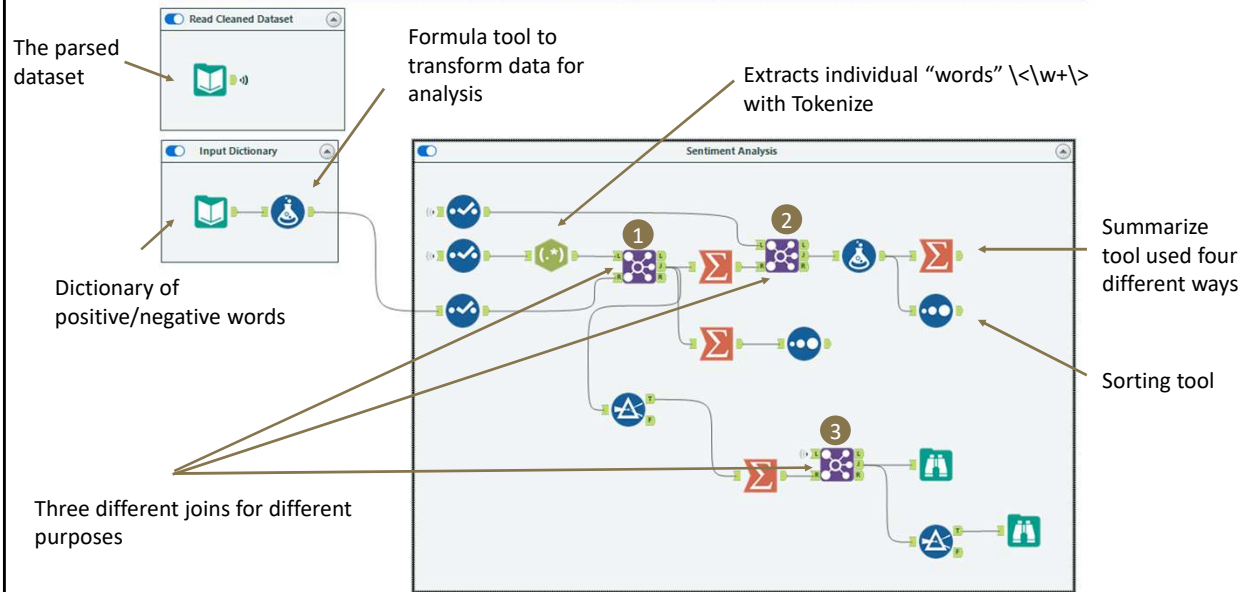
Enron Emails Analysis



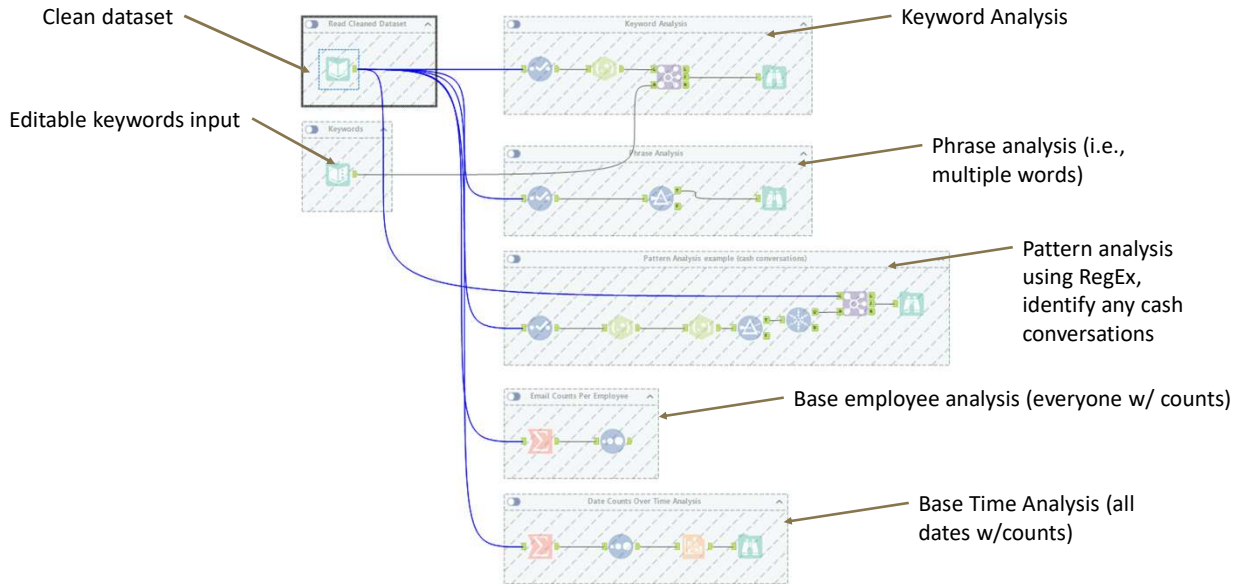
Enron Emails Analysis



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 and additional tools to help 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



Agenda

Class Agenda

- Case Review: Enron Emails Case
- Case Deliverable Overview
- Lab Tool and workflow overviews:
 - DateTime parsing functions
 - Sentiment Analysis
 - Phrase Analysis
 - Individual Forensic Workflow
- **Lab Forensic Analysis**

Conclusion and look-ahead

We have undertaken a sophisticated analysis of unstructured data.

Please submit your case deliverable by Friday.

Thank you!

