Tippie College of Business

Data Mining Presentation & Writing Guidance

Carl Follmer

Associate Director, Frank Center

Items for discussion

Presentation best practices

- Audience expectations

- Storyline and team dynamics

- Slide design and data visualization

Written report best practices for:

- Executive summary

- Problem description

- Conclusion

Presentation: Audience Expectations

Only 8 minutes to present = high-level messaging

Who is your intended audience according to the assignment?

Audience is customer, not instructor

- What is the difference between an instructor's needs and a customer's needs?

- What does the customer care about that is different from an instructor?

Presentation: Audience expectations

Customer isn't overly interested in models, code, what you did with the data, or how long you worked on the project

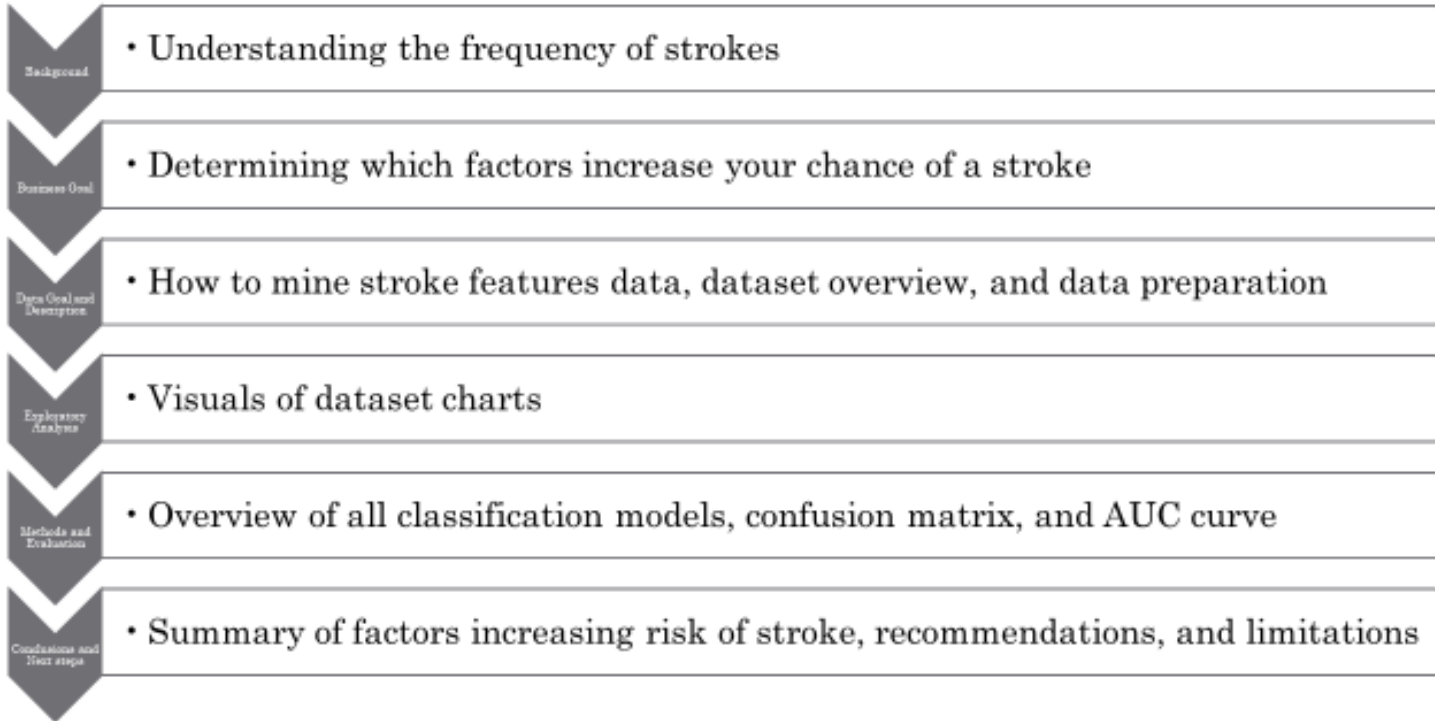Emphasis on results, not process (no more than 30 seconds on data)

- Most of the focus on recommendations/findings

Mid-level technicality appropriate for client

Presentation: Storyline as guide

What elements are needed for an effective introduction?

Overview

Background	• Understanding the frequency of strokes
Business Goal	• Determining which factors increase your chance of a stroke
Data Goal and Description	• How to mine stroke features data, dataset overview, and data preparation
Exploratory Analysis	• Visuals of dataset charts
Methods and Evaluation	• Overview of all classification models, confusion matrix, and AUC curve
Conclusions and Next steps	• Summary of factors increasing risk of stroke, recommendations, and limitations

What is this slide missing that a customer would want?

Presentation: Storyline as guide

For the body consider why information is presented in a certain order
→ E.g. presenting a project before discussing the cost

Contextual transitions help in demonstrating intentionality

→ Do NOT say "Now I'm gonna hand it off to Brett."

→ <u>Better</u>: "Now that Miranda has discussed the benefits of our proposal, I will talk about the commitment it will take from us to implement this plan"

Presentation: Slide design for retention

You are the presentation, the PPT is a visual resource

Short phrases (4-9 words) – otherwise 0-5% info retention
- Use "fade" or "appear" animation to keep audience's attention

No copying/pasting from proposal/report draft
- Why not?

Slide as notecards for you and resource for audience

Presentation: Slide design for retention

DATA DICTIONARY

Features we used in our models

Offensive
- ☐ *FGPCT_O* (field goal %)
- ☐ *TURNPCT* (turnover %)
- ☐ *TORD* (steal rate)
- ☐ *ORB* (offensive rebound %)
- ☐ *FTR* (free throw %)
- ☐ *2P_O* (2 point shooting %)
- ☐ *3P_O* (3 point shooting %)
- ☐ *ADJOE* (adj. Offensive efficiency)
- ☐ *WAB* (wins above bubble)

Defensive
- ☐ *ADJDE* (adj. Defensive efficiency
- ☐ *DRB* (defensive rebound %)
- ☐ *FTRD* (defensive free throw %)
- ☐ *2P_D* (defensive 2 pt shooting %)
- ☐ *3P_D* (defensive 3 pt shooting %)
- ☐ *FGPCT_D* (defensive field goal %)

Teams/Identifiers
- ☐ *TEAM* (Name of school)
- ☐ *YEAR*
- ☐ *TEAMYEAR***

Rank/Standing
- ☐ *PowerRating* (overall team rating)
- ☐ *SEED* (tournament seeding)
- ☐ *Tournament?* (if team made tournament)
- ☐ *power_conf* (if team is in power conf.)
- ☐ *Elite8** (whether or not team made elite 8)
- ☐ *Games* (# games played)
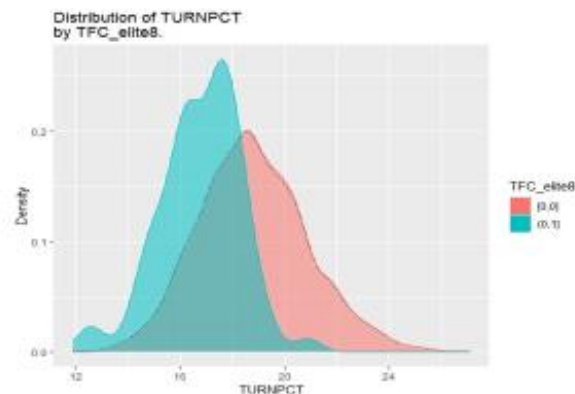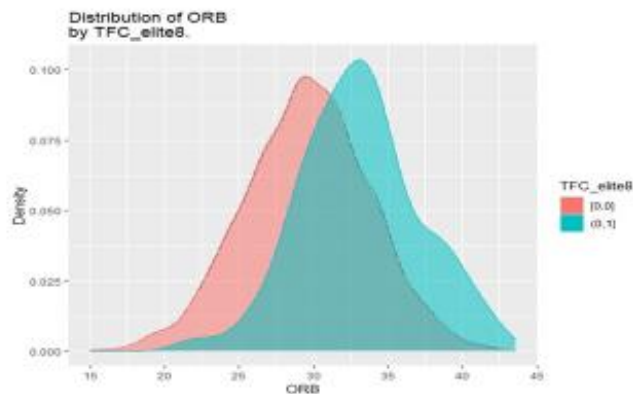- ☐ *Wins* (# of team wins)

*Target Variable
**Identifier

What is the takeaway from this slide?
How could this slide be presented effectively?

Slide Deck: Making effective visuals to maximize readability

Distribution of ORB by TFC_elite8.

Distribution of TURNPCT by TFC_elite8.

GRAB THOSE REBOUNDS

Grabbing over 33% of offensive rebound opportunities gives you a great chance to win.

THE TURNOVER BATTLE

Keeping your turnover rate under 18% of possessions keeps you playing in the tournament

How does this slide enhance readability?

Slide Deck: Making effective visuals to maximize readability

Include a message title that establishes the visual's takeaway message

- Ineffective: Profits

- Effective: Net income down 13% in 2020 Q4

If impractical/too unorthodox, use short phrases in remaining space on slide to establish takeaways
  - Avoid complete sentences unless very short
  - If you must include a longer sentence, it <u>must be read aloud to audience</u>

Eliminate unnecessary elements/chartjunk

Eliminating unnecessary elements to maximize readability*
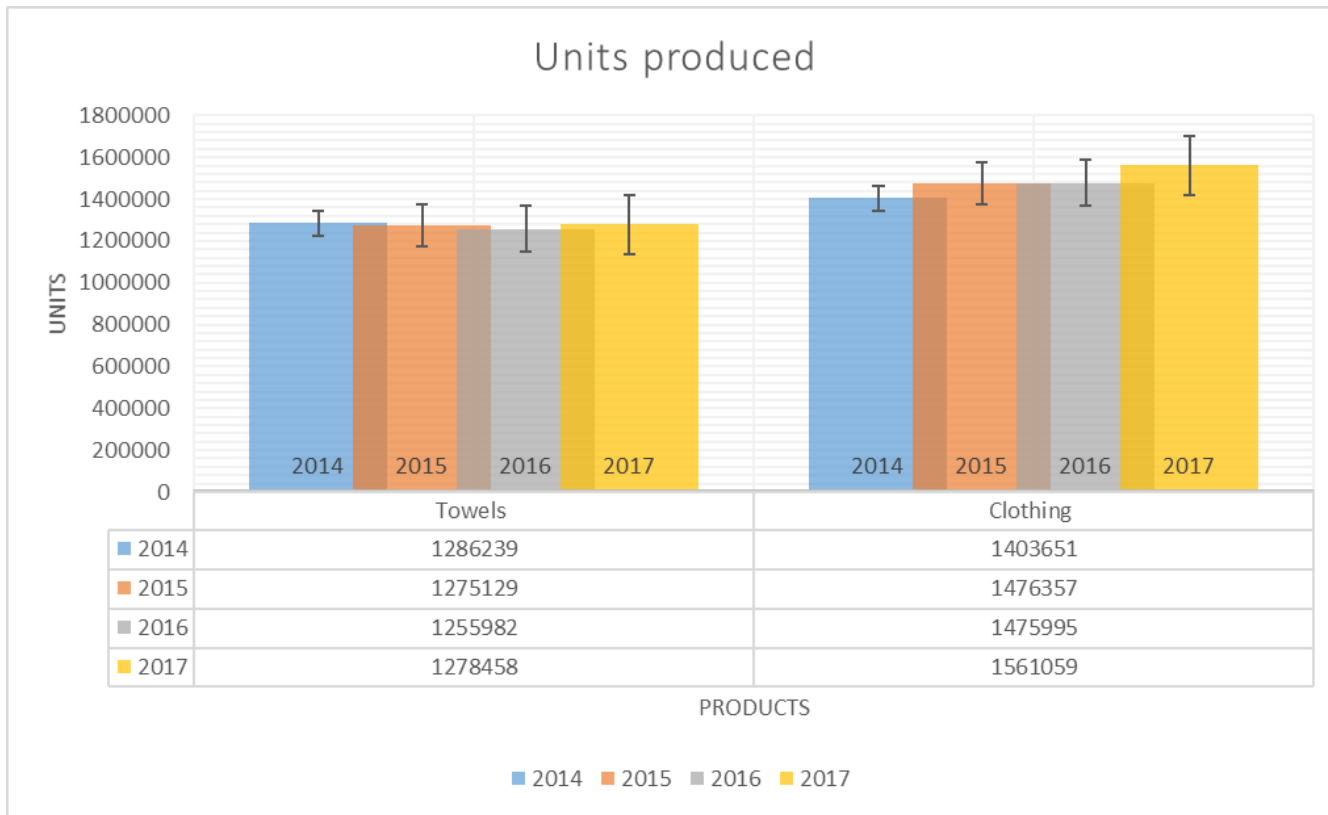
Zeros on y-axis

Table below graph

Error bars

Years x3

Grid marks

*The audience determines what elements are unnecessary, but redundancies can always be eliminated

Optimized graph w/ same data set

Clothing produced up 11% 2014-17

Units in Millions

Legend: Towels, Clothing

Notes on <u>presenting</u> visuals

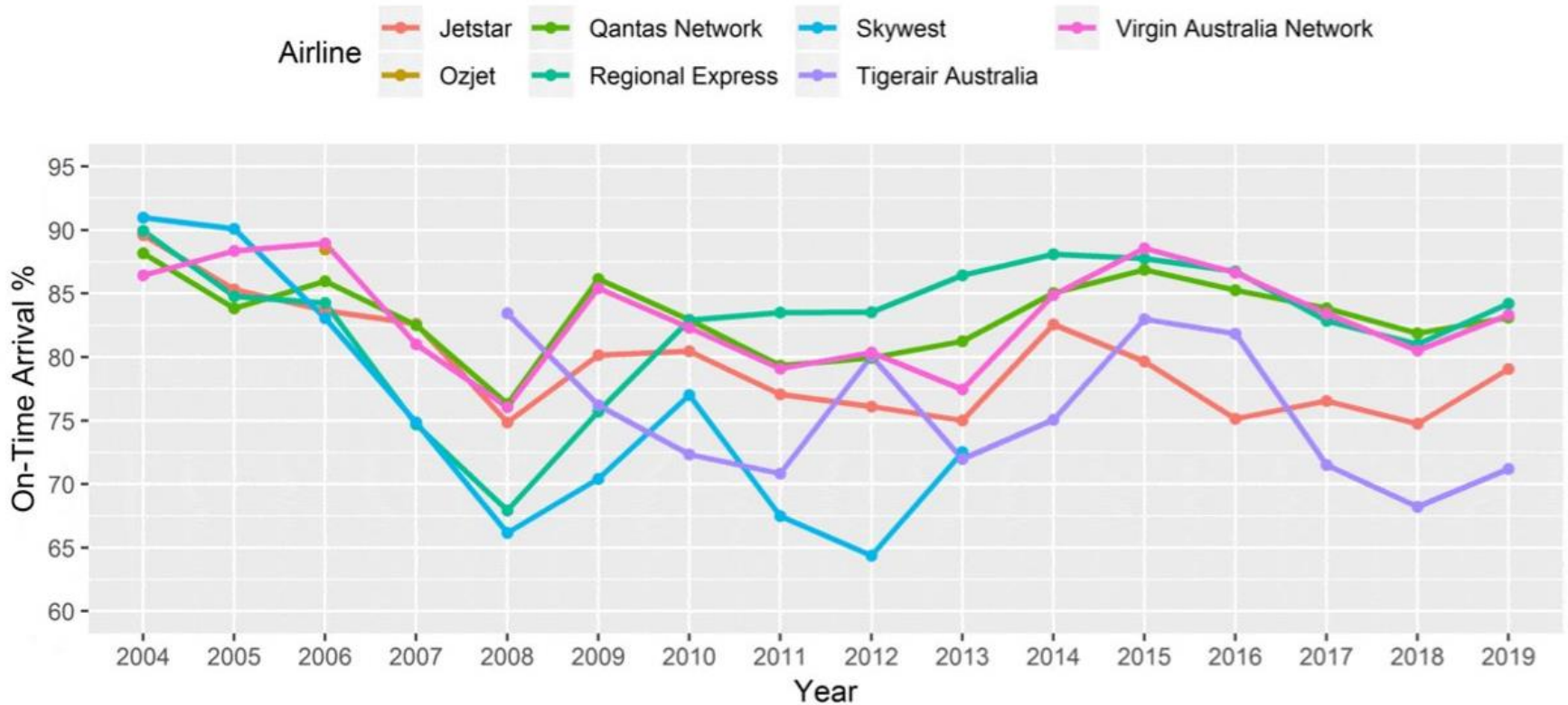Transition: "Now that we understand (principle), let's move on to (example)"

Message Title: "As you can see, over the last 10 years…"

Orient: "The blue bars on the chart represent…"

Discuss: Move into the visual's points of interest

How can we improve on this? It depends on audience needs…

On-Time Arrivals by Airlines by Year

Step 1: Add a better message title or have takeaway somewhere nearby

2008 was worst year for on-time arrivals at Australian airports

But what else?

IOWA

Direct the eye by outlining

2008 was worst year for on-time arrivals at Australian airports

Better, but what can we still fix?

2008 was worst year for on-time arrivals at Australian airports

Get rid of unnecessary elements: "Year" label, background grid
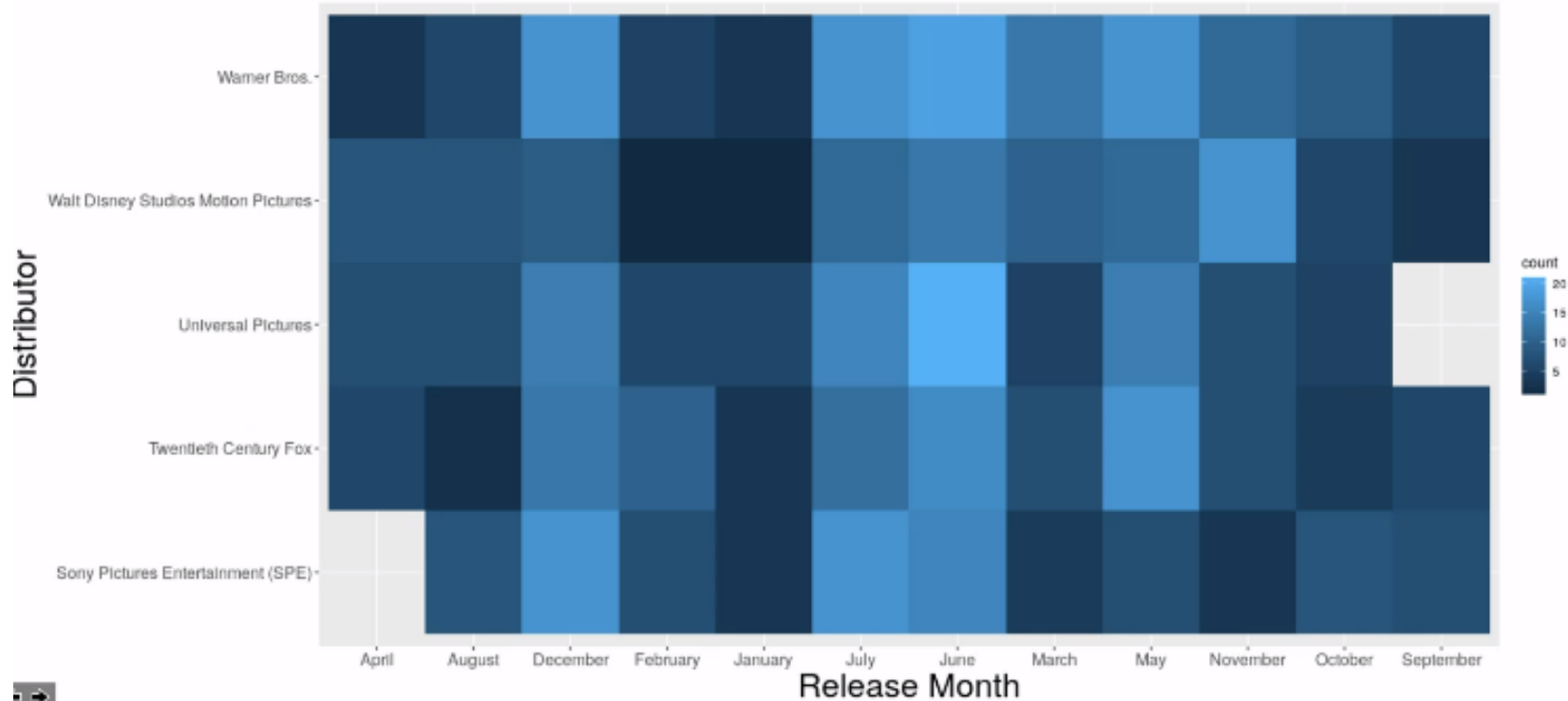
Use horizontal text on y-axis

Use blue only in key areas, gray in the rest

Ideally, limit the # of years/data points – combine all airlines?

Consider how information is organized

Written Project Report

Report: Writing Executive Summary for Busy Readers

1-2 paragraphs that outline everything the client needs to know

Mindset: assume the reader won't read anything else – what do they need to know?

Elements to include:

Main goals of the project

Outline of the report's content

Major takeaways / findings / determinations

Anticipated costs or risks (if necessary)

Exec Summary Sample

This report explains our findings on the variables that may determine whether a customer renews their contract with a telecommunications company. Using a dataset that provides both demographic and customer relationship information, we used data mining to make a model that best describes which customers will continue business with a telecommunications company. The data shows that there are three variables that most affect the customer's decision, which are total charges, monthly charges, and tenure. We recommend that telecommunications companies should increase their retention efforts for customers with a low probability of renewing their contract. In addition, they should shift their marketing spending away from customers that are already likely to continue working with the company an reallocate that spending to customers with a low renewal likelihood.

Purple = context, Gray = Methods, Blue = Findings

Exec. Summary Sample 2

Strokes are one of the leading causes of death worldwide, and multiple health risks play a role in an individual having a stroke. The dataset we found examined individuals who did and did not have a stroke in correlation with the individual's health and lifestyle. We used an individual's Gender, Age, Hypertension, Health of their heart, Marital status, Type of work, Where they live, Glucose level, BMI, and Smoking status to help predict their likelihood of having a stroke. We created a variety of models using the features above to help us predict the possibility of an individual having a stroke and which features had the most impact on someone having a stroke.

Purple = context, Gray = Methods, Blue = Findings

Report: Problem Description

1 paragraph on the background of the business problem and why it is important to the client

Separate section identifying the project's business goals and definition of success

Clear, efficient, and precise language (applies to entire document)

- Clear: conveys easily understood, audience-oriented meaning

- Efficient: only essential phrasing is used (not using extra, unnecessary words)

- Precise: avoids awkward or vague phrasing (this, these, that, etc.)

Report: Concluding comments

More than just reviewing previous content

Use positioning statements to begin paragraphs in recommendations and limitations subsections:

- <u>Description (ineffective)</u>: The data provide insight about the challenges of customer retention.

- <u>Positioning statement (effective)</u>: Customer retention can improve by [action A, B, and C].

Include: other information/data that would have been useful, unanswered questions, next steps, goodwill ending

Use of clear language

Demonstrates respect for audience

Avoid 2nd person (you/your) but some 1st person plural (we/our) is fine in the intro and the conclusion

- You/your creates confusion for the reader

"This, these, that" usage requires clarifying nouns or phrases

- Ex: This leads to success for them
- Can be confusing, esp. in technical or client-centric documents

Use of clear language

Avoiding wordiness and unnecessary phrasing

→ Excess of 1-5 letter words in a row

→ Ex: One thing to consider would be to think about using the data to make changes to boost efficiency for processes the company uses (23 words)

→ Better: This data provides a potential roadmap for more efficient processes (10 words)

Read aloud and highlight sentences that take extra time to read through or that require multiple readings

Client expectations about quality

To meet professional-level communication standards:

SPELL CHECK

Read your work out loud to find places with too many words, typos, or missing words

Check that paragraphs in conclusion begin with positioning statements that form a chain

Data Mining Project Resources

Communication consultant virtual office hours (TBA)

The Frank Center