




VOICE OF THE POLICYGENIUS CUSTOMER

A stylized lightbulb icon with a yellow glow, a black base, and a small red detail on the base.

April 26, 2023

Lexin Lu
Michelle Van
Godwin Anguzu
Deekshita Saikia

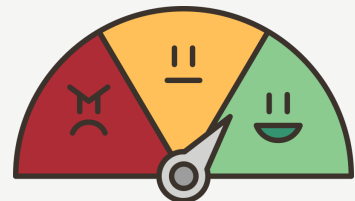
BACKGROUND

Surveys sent out to customers periodically to collect feedback

→ **Customers who withdrew**

Customer Satisfaction (CSAT) Survey

How would you rate your overall satisfaction with the insurance application process?



1: Extremely Unsatisfied



5: Extremely Satisfied

BACKGROUND

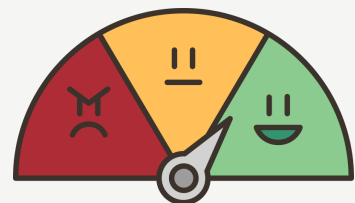
Surveys sent out to customers periodically to collect feedback

→ **Customers who withdrew**

→ **Customers who applied successfully**

Customer Satisfaction (CSAT) Survey

How would you rate your overall satisfaction with the insurance application process?



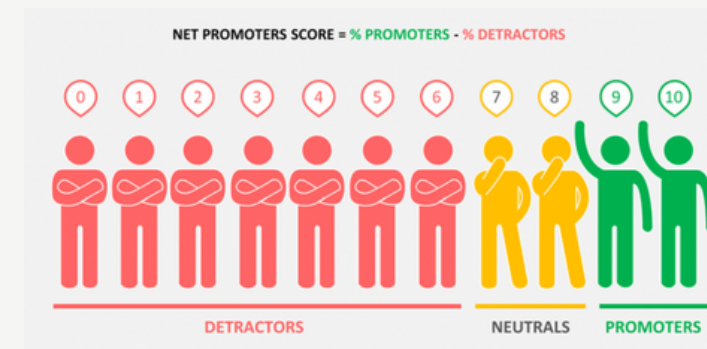
1: Extremely Unsatisfied



5: Extremely Satisfied

Net Promoter Score® (NPS) Survey

How likely would you be to recommend our company to a friend?"



- 0-6: Detractors
- 7-8: Passive
- 9-10: Promoters

CURRENT PROCESS

Analyzing customer feedback manually to extract topics

Comment	Overcontact	High Quote	Slow Process
Stop emailing me after I asked over and over again.	<input checked="" type="checkbox"/>		
The price was high.		<input checked="" type="checkbox"/>	
Costs were too much for my budget		<input checked="" type="checkbox"/>	
I have been waiting since May 2021 !			<input checked="" type="checkbox"/>
Harassing calls, texts, and emails.	<input checked="" type="checkbox"/>		
Long horrible experience & still didn't offer a policy.			<input checked="" type="checkbox"/>

Problems: • Challenging • Labor-Intensive • Time-Consuming

EXECUTIVE SUMMARY



GOAL

1

Built text analysis models to identify
key topics and explore historical trends

EXECUTIVE SUMMARY



GOAL

1

Built text analysis models to identify key topics and explore historical trends

2

Developed an automated computational topic extraction pipeline

EXECUTIVE SUMMARY



GOAL

1

Built text analysis models to identify key topics and explore historical trends

2

Developed an automated computational topic extraction pipeline

3

Designed an analytics dashboard

EXECUTIVE SUMMARY



GOAL

1

Built text analysis models to identify key topics and explore historical trends

2

Developed an automated computational topic extraction pipeline

3

Designed an analytics dashboard



TIMELINE

REQUIREMENTS
AND DATA
GATHERING

1

OCTOBER
2022

OCTOBER
2023

2

DATA CLEANING

TEXT PRE-
PROCESSING

3

NOVEMBER
2023

FEBRUARY
2023

4

COMPARISON OF
MODELING
APPROCHES

BERTOPIC
MODELING

5

MARCH
2023

MARCH
2023

6

AGGREGATE
RESULTS

PIPELINE
DEVELOPMENT

7

APRIL
2023

APRIL
2023

8

DASHBOARD
DESIGN

METHODS



Topic Modelling

Unsupervised and automated ML technique to identify semantic structures and conduct thematic analysis

Performs clustering on textual data in depth

Identifies key issues and topics present in a large corpus



Available Methods

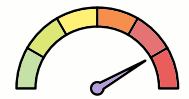
**Latent Dirichlet Analysis
(LDA)**

**Non-negative Matrix
Factorization (NNMF)**

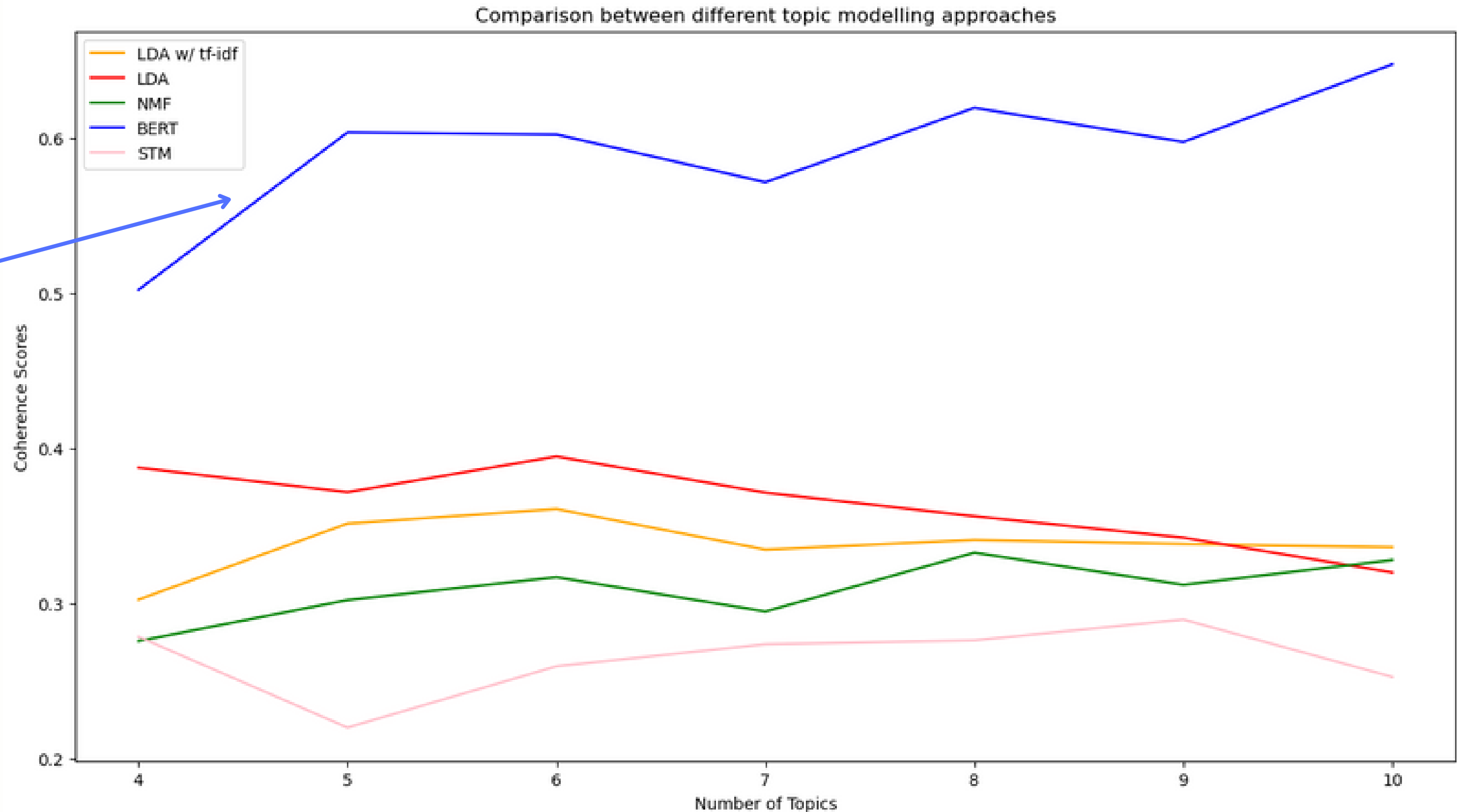
**Structured Topic Modeling
(STM)**

BERTopic

COMPARISON OF METHODS



COHERENCE SCORE : DEGREE OF SEMANTIC SIMILARITY AMONG WORDS IN A TOPIC



BERTopic

METHODS



Topic Modelling

Unsupervised and automated ML technique to identify semantic structures and conduct thematic analysis

Performs clustering on textual data in depth

Identifies key issues and topics present in a large corpus



BERTopic

BERTopic - Density-based clustering algorithm

Semantic embeddings from pre-trained models (developed by Google) to account for a word's context

Flexible, good for noisy data

APPROACH

Prepare Data

Subsetted data to 4 segments:

- Promoters
 - Passives
 - Detractors
 - CSAT
- } NPS

APPROACH



Prepare Data

Subsetted data to 4 segments:

- Promoters
 - Passives
 - Detractors
 - CSAT
- } NPS

Model Building & Fine-tuning

- Build BERTopic models individually for the 4 segments
- Fine-tune each model to determine the number of topics by coherence score

APPROACH



Prepare Data

Subsetted data to 4 segments:

- Promoters
 - Passives
 - Detractors
 - CSAT
- } NPS

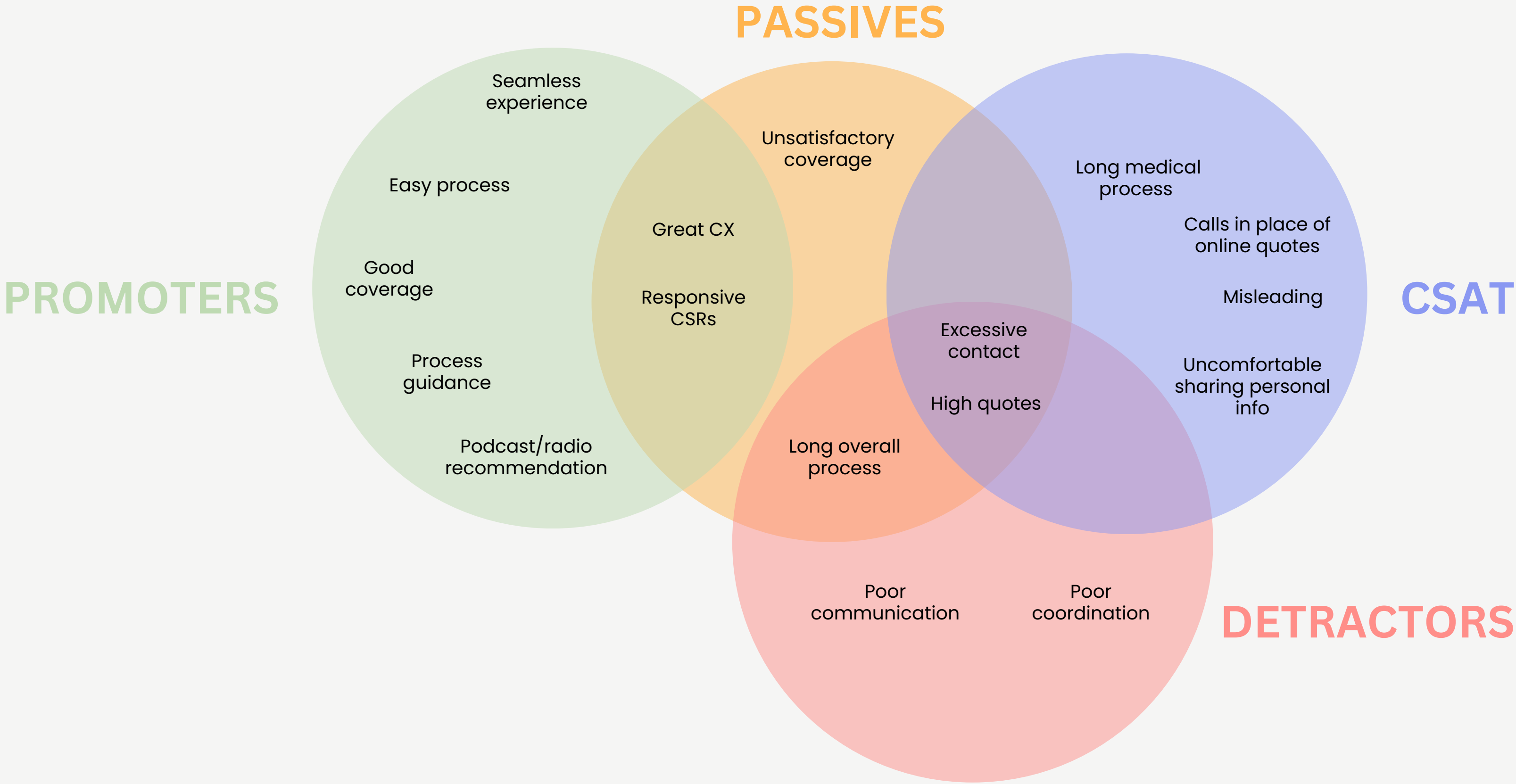
Model Building & Fine-tuning

- Build BERTopic models individually for the 4 segments
- Fine-tune each model to determine the number of topics by coherence score

Interpret Results

Employed combination of human interpretation and BERTopic results to identify key themes

EXTRACTED TOPICS



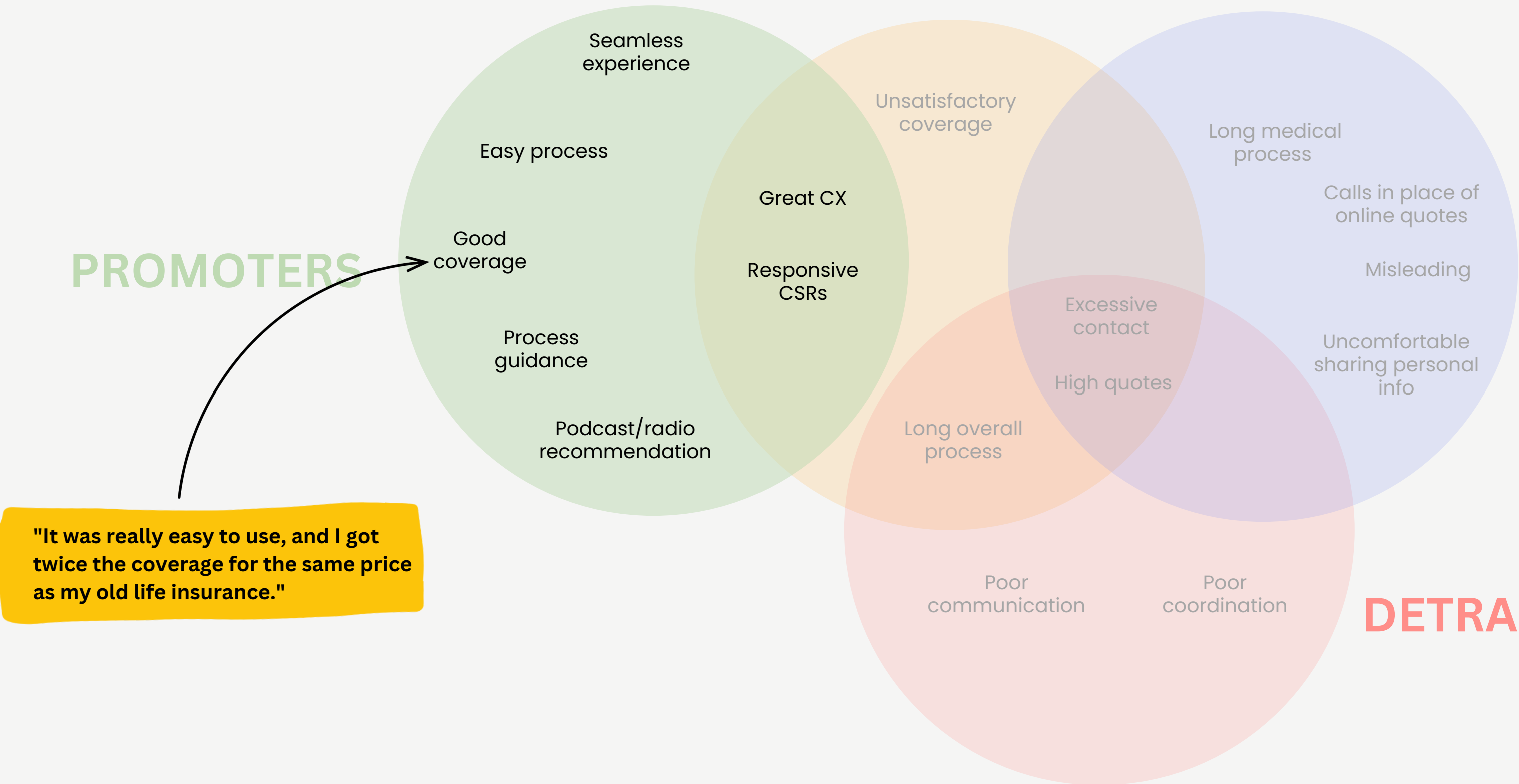
EXTRACTED TOPICS

PASSIVES

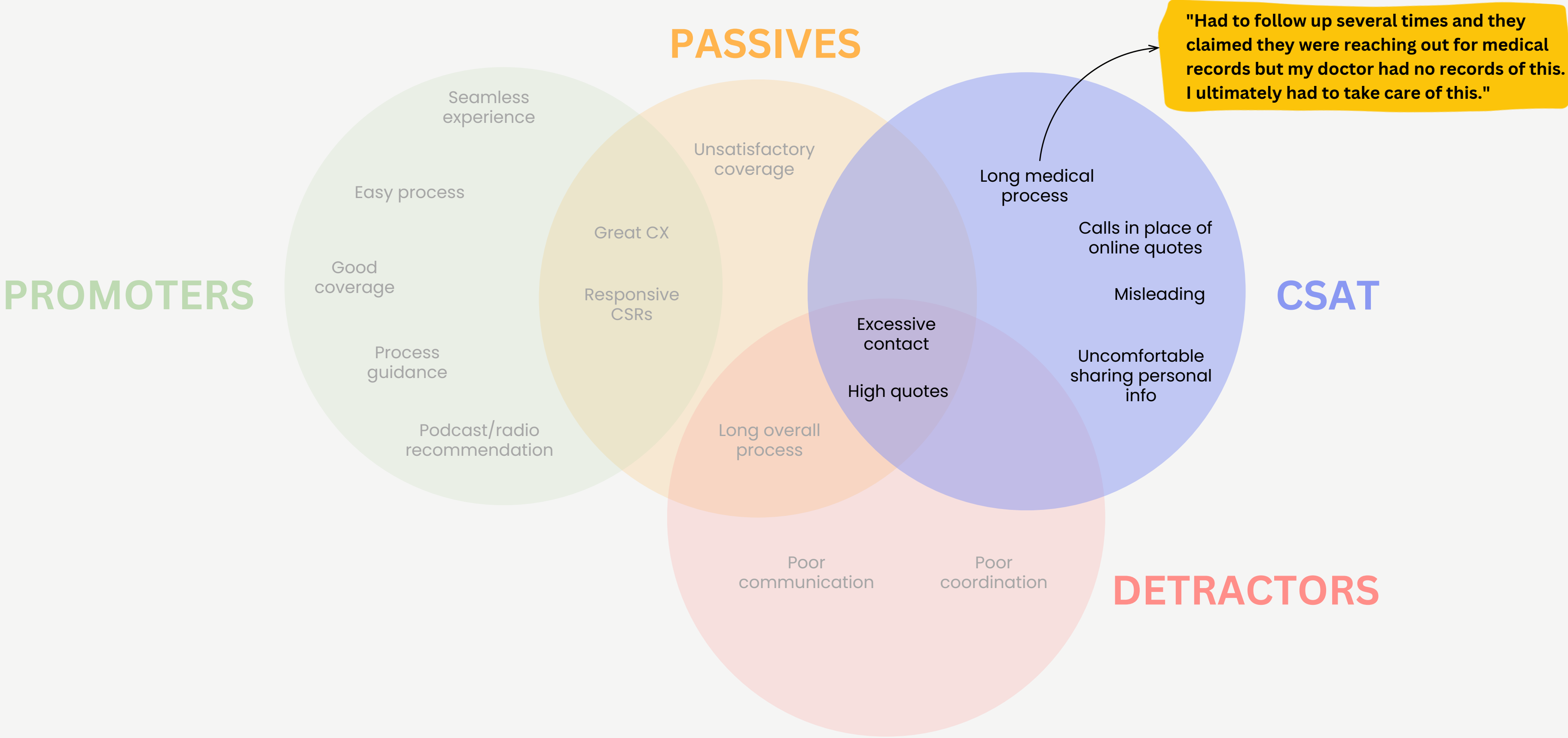
PROMOTERS

CSAT

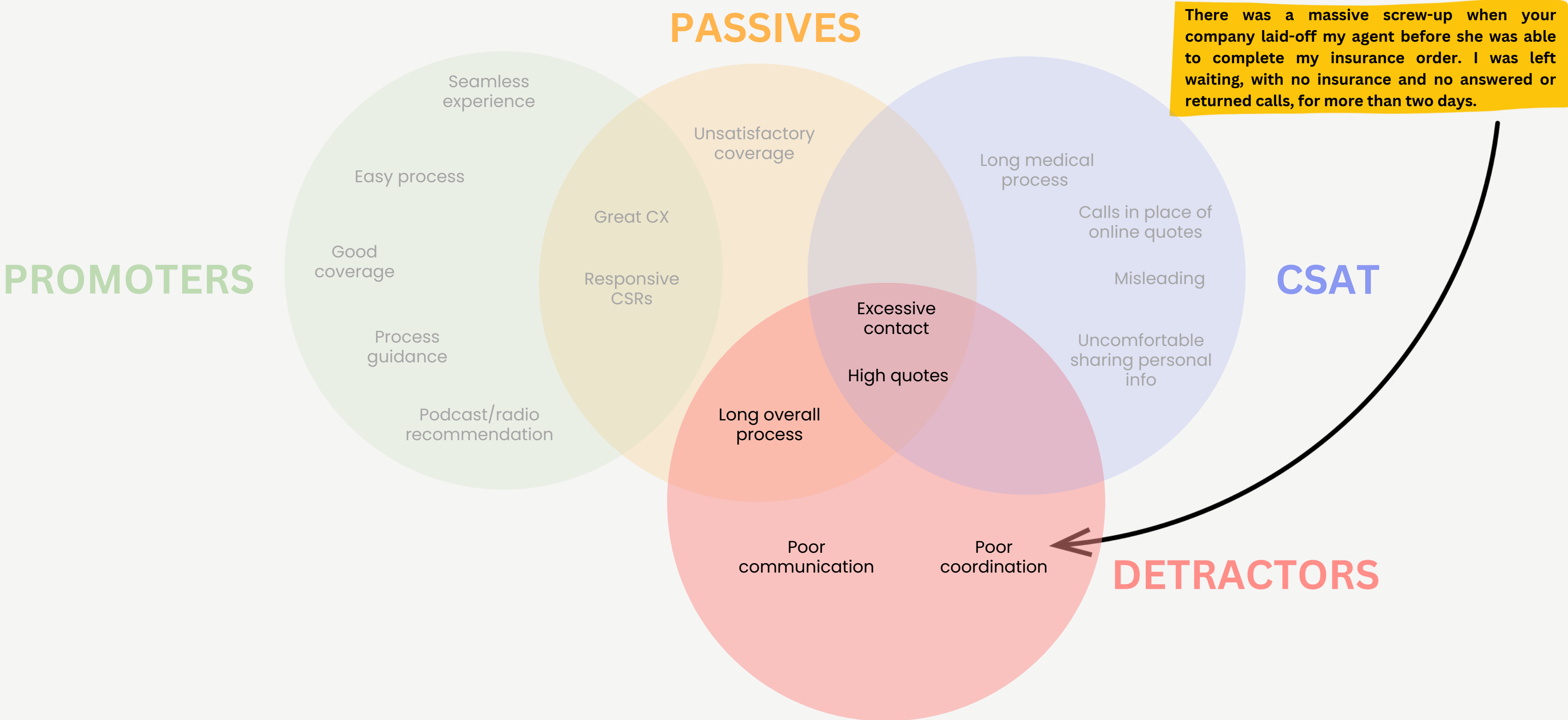
DETRACTORS



EXTRACTED TOPICS



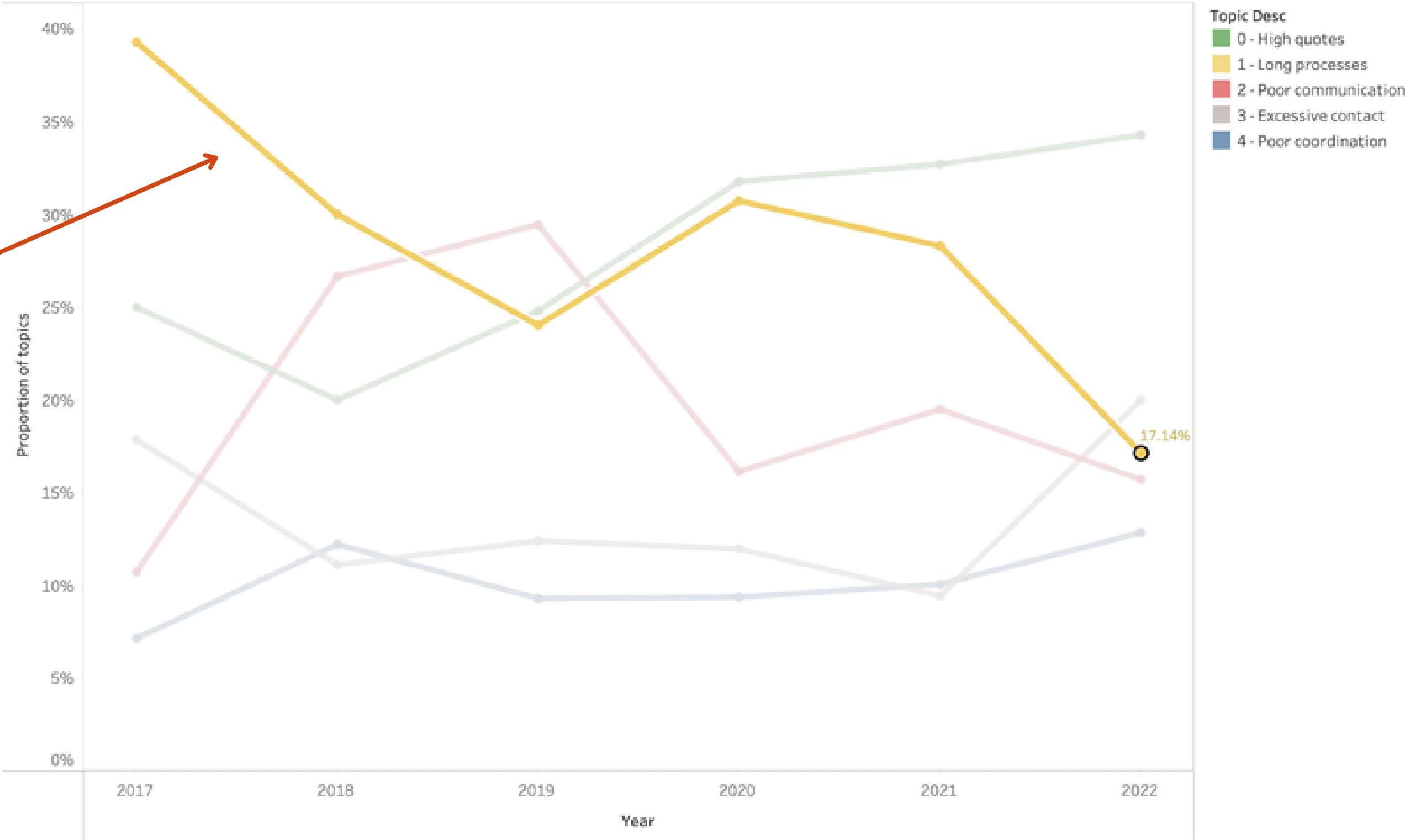
EXTRACTED TOPICS



TRENDS OF TOPIC PROPORTIONS- DETRACTORS

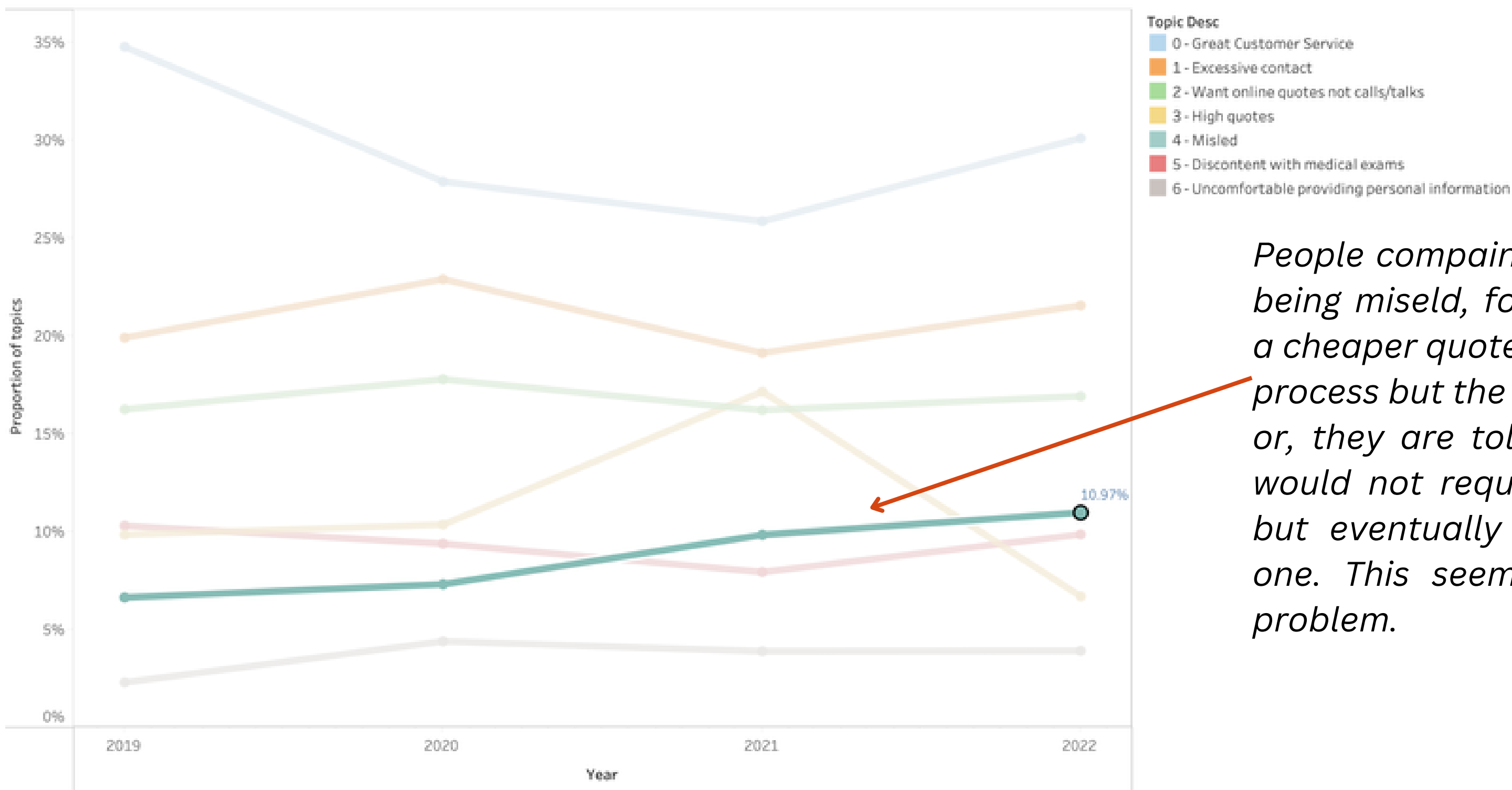
Downward trend confirmed by Pg, updates to the application process were rolled out since 2019

Yearly Trends



TRENDS OF TOPIC PROPORTIONS- CSAT

Yearly Trends



People complain that they are often being misled, for example, they see a cheaper quote when they start the process but the final quote is higher, or, they are told initially that they would not require a medical exam but eventually have to undertake one. This seems to be a growing problem.

DASHBOARD

Demo

LIMITATIONS



Sampling Bias

All customers may not be equally likely to receive a survey

LIMITATIONS



Sampling Bias

All customers may not be equally likely to receive a survey



Non-response Bias

Customers responding to surveys were different from those who didn't

LIMITATIONS



Sampling Bias

All customers may not be equally likely to receive a survey



Non-response Bias

Customers responding to surveys were different from those who didn't



Limited Data

Only ~13000 records in NPS and ~3000 in CSAT with feedback

LIMITATIONS



Sampling Bias

All customers may not be equally likely to receive a survey



Non-response Bias

Customers responding to surveys were different from those who didn't



Limited Data

Only ~13000 records in NPS and ~3000 in CSAT with feedback



Effect of Noise

Difficult to isolate the effect of noise in trends

CONCLUSION

What do we have?

- An automated pipeline to process raw text and extract key topics and trends
- Positive outcome: Our findings correlate with those of Pg



CONCLUSION

What do we have?

- An automated pipeline to process raw text and extract key topics and trends
- Positive outcome: Our findings correlate with those of Pg

Where do we go from here?

- Deliver an end-to-end production ready visualization tool using our pipeline
- Possible future work around applying this pipeline to ingest data from social media



That's a wrap!



Questions?



Special Thanks

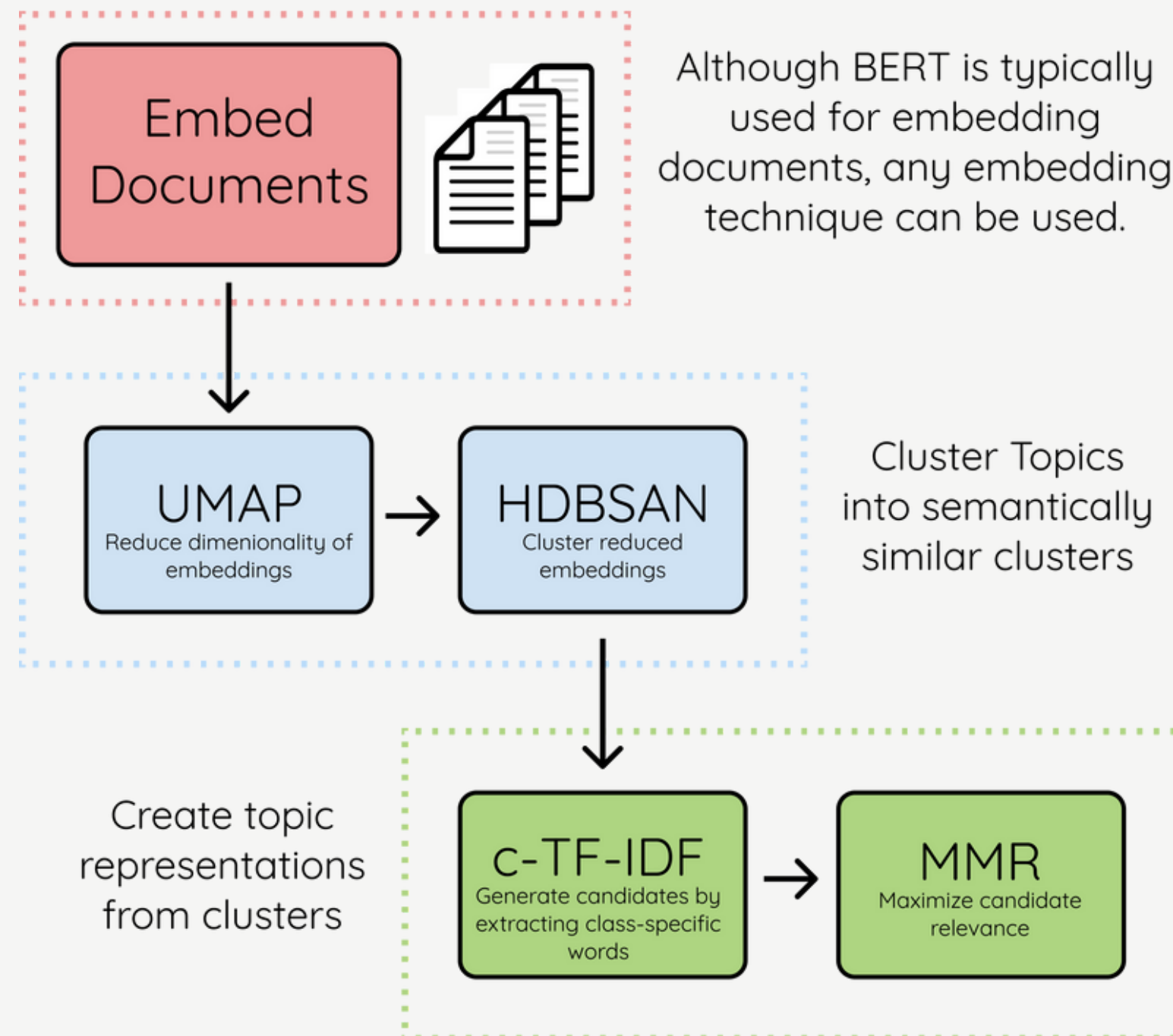
Brenna Hayes, Policygenius
Emily Nightingale, Policygenius
Dustin Tucker, Policygenius

Greg Herschlag, Duke
Jason Byers, Duke
Ryan Huang, Duke

Appendix

BERTOPIC

BERTopic

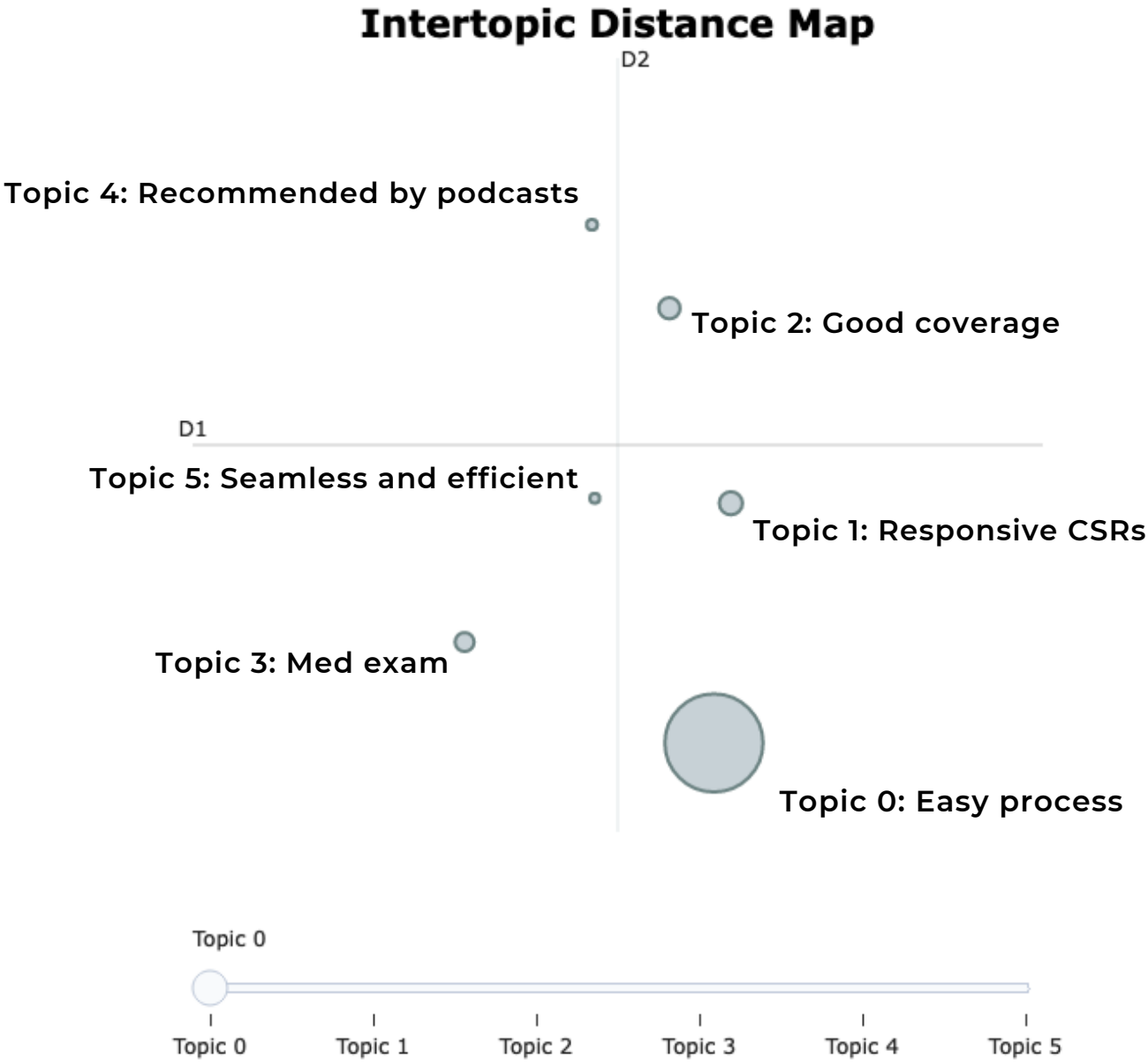


Features:

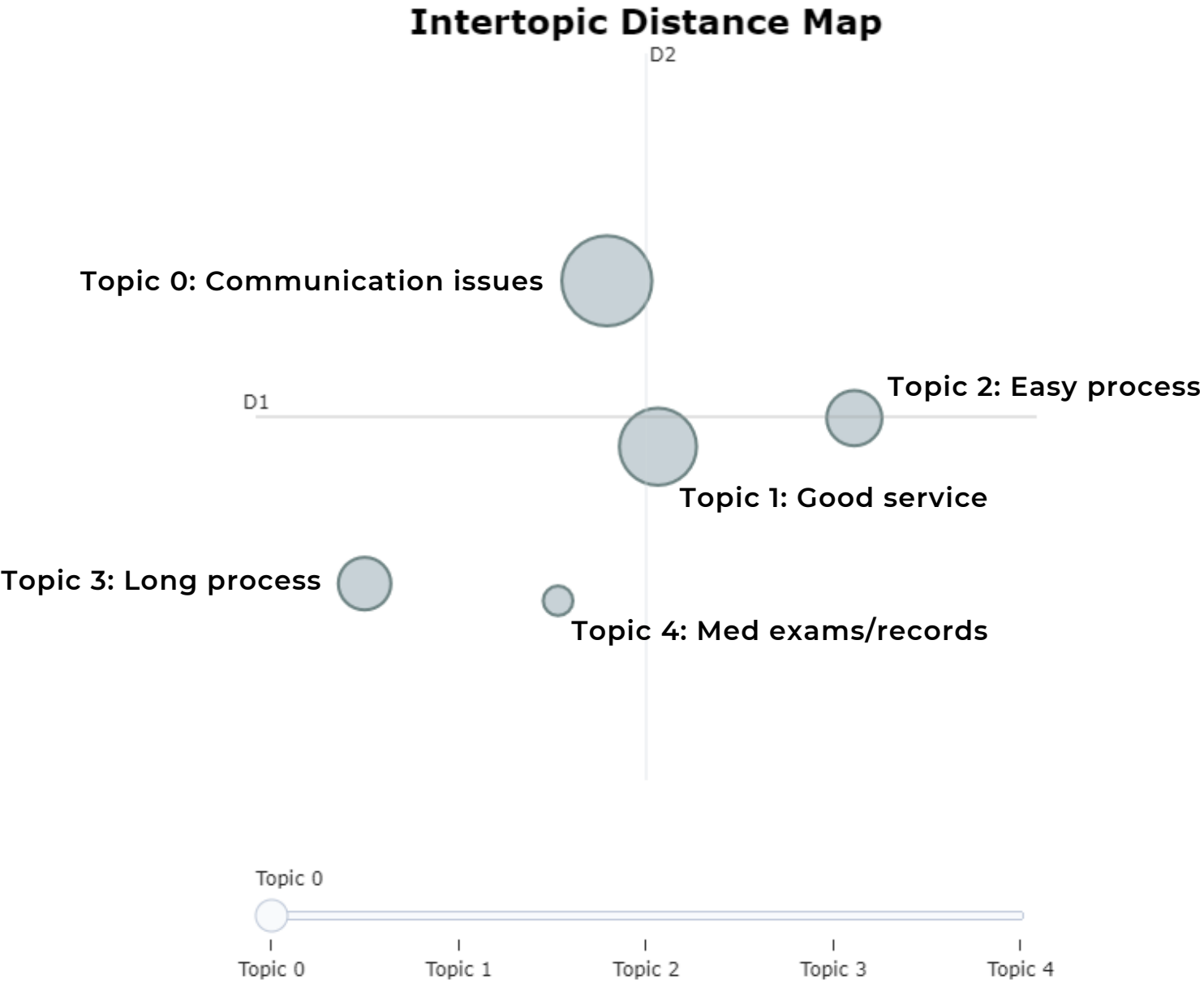
- Density-based clustering algorithm
- More flexible & interpretable
- Semantic embeddings to account for a word's context
- Better at handling short texts
- Good for noisy datasets

RESULTS

PROMOTERS

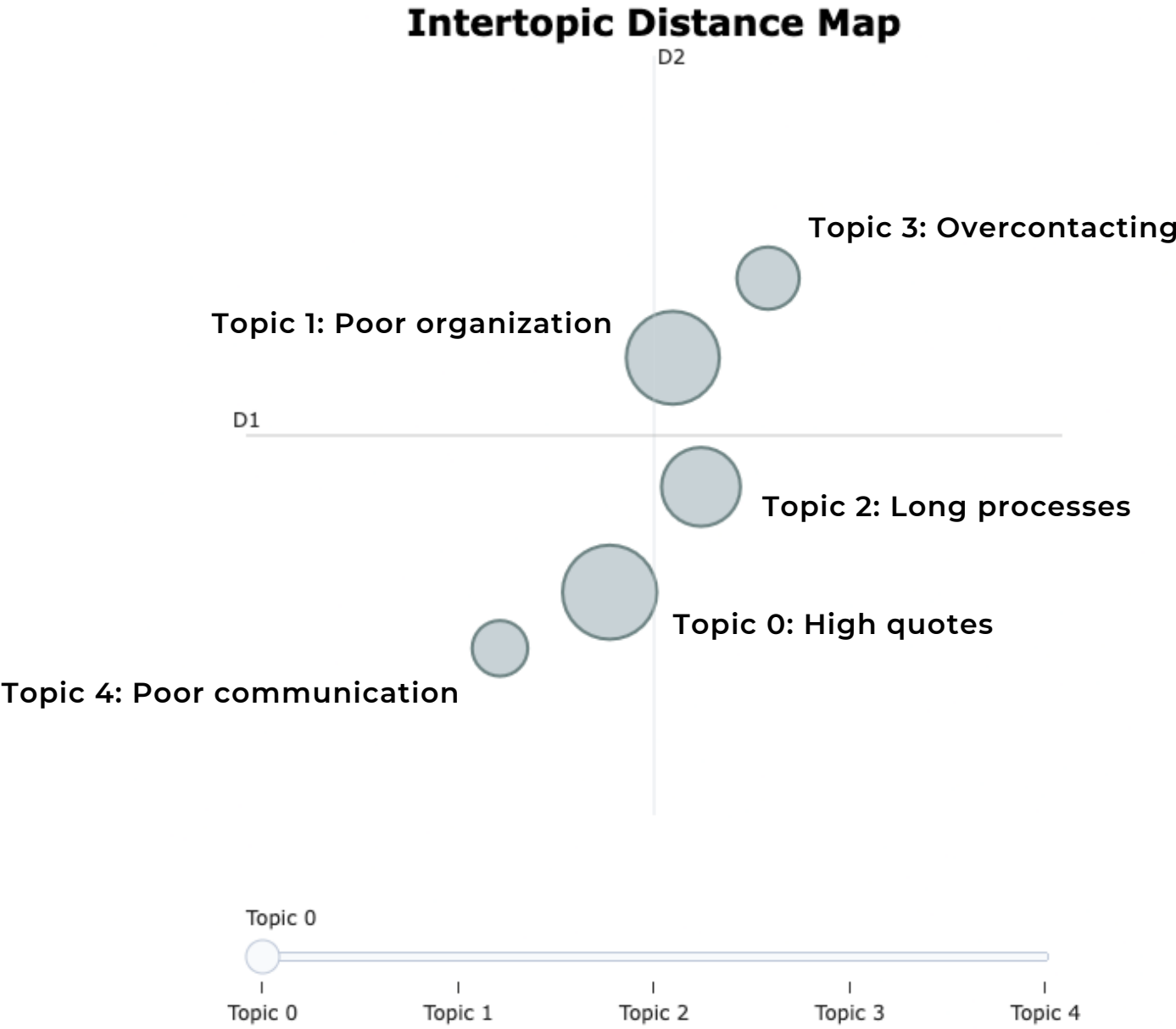


PASSIVES

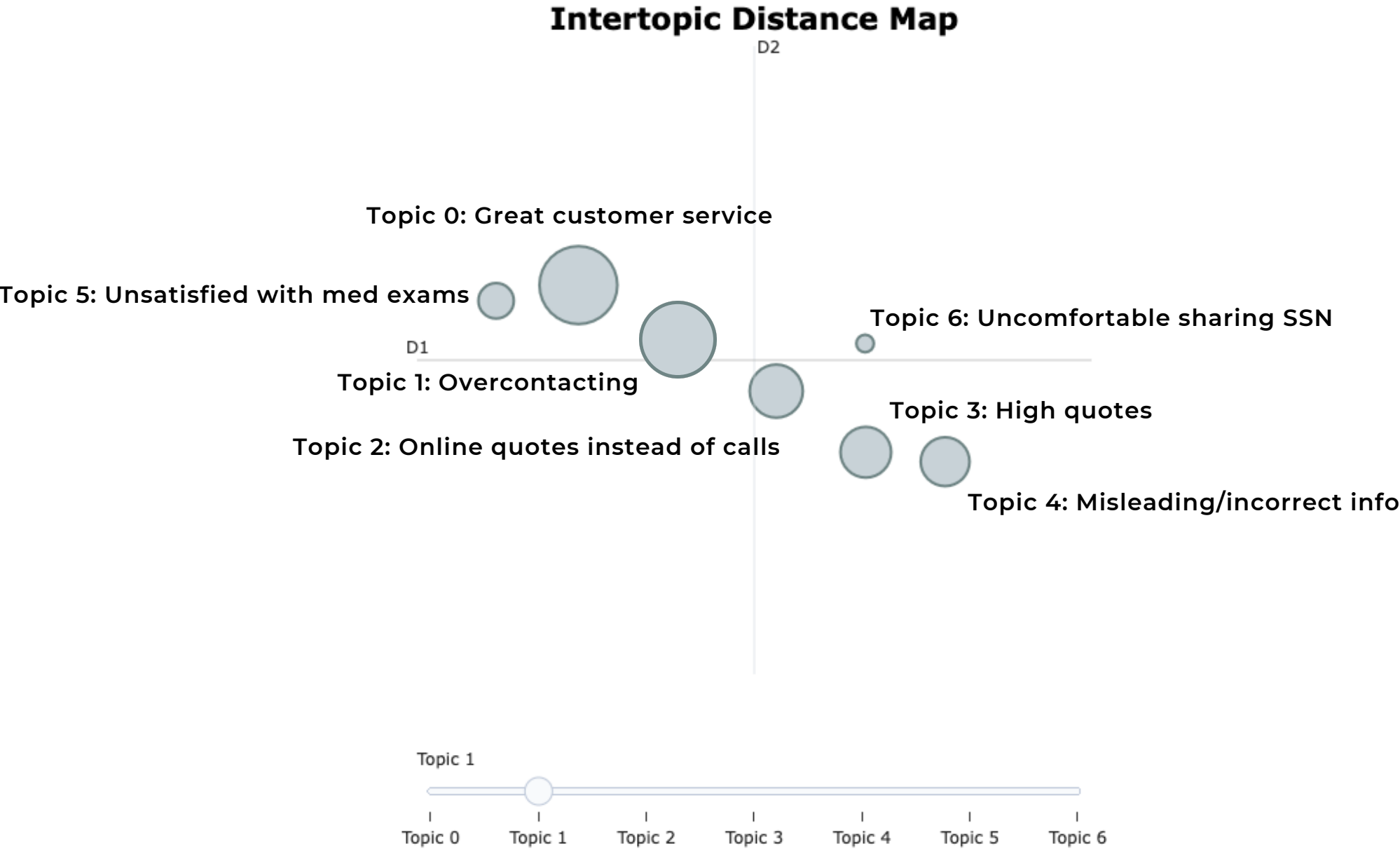


RESULTS

DETRACTORS

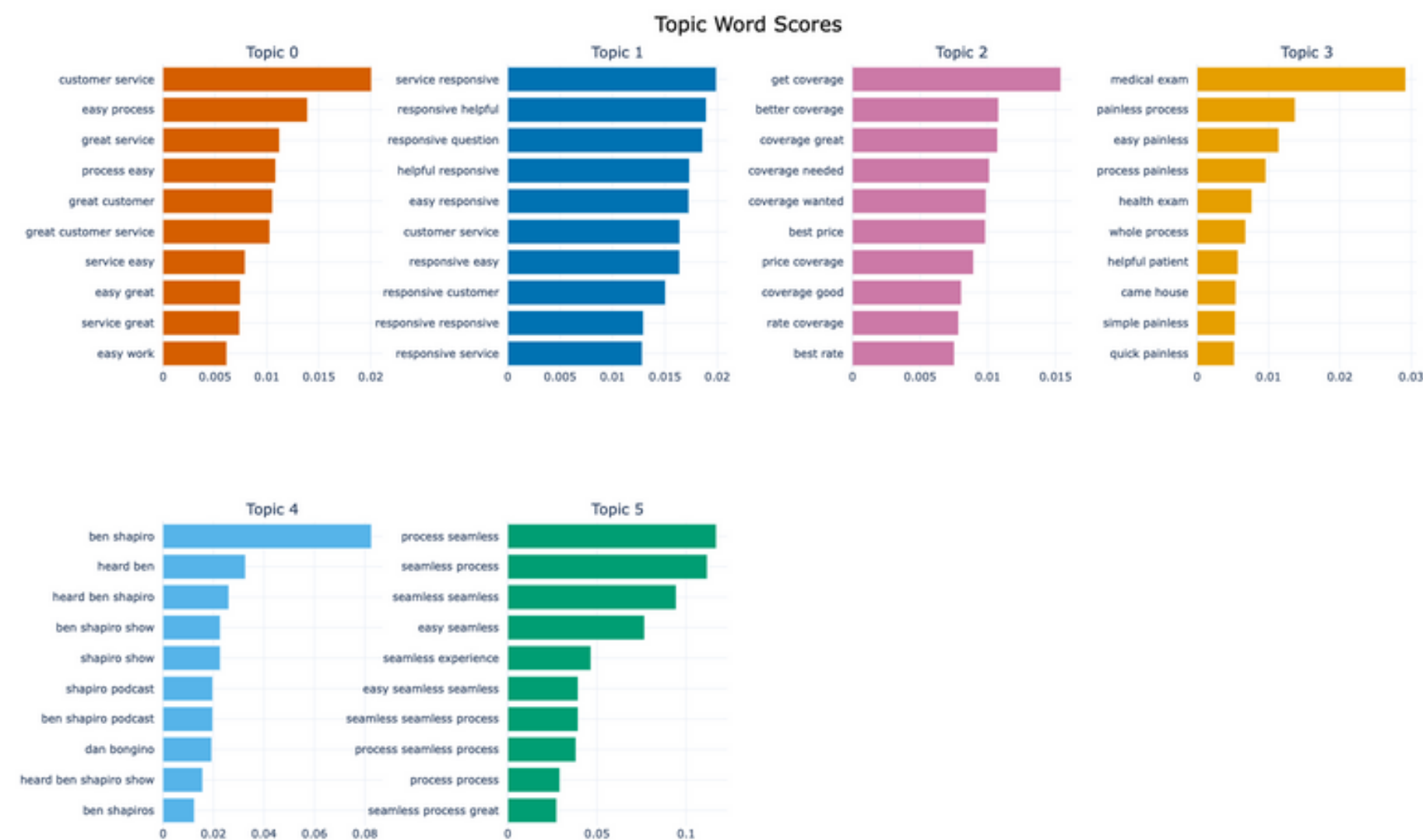


CSAT

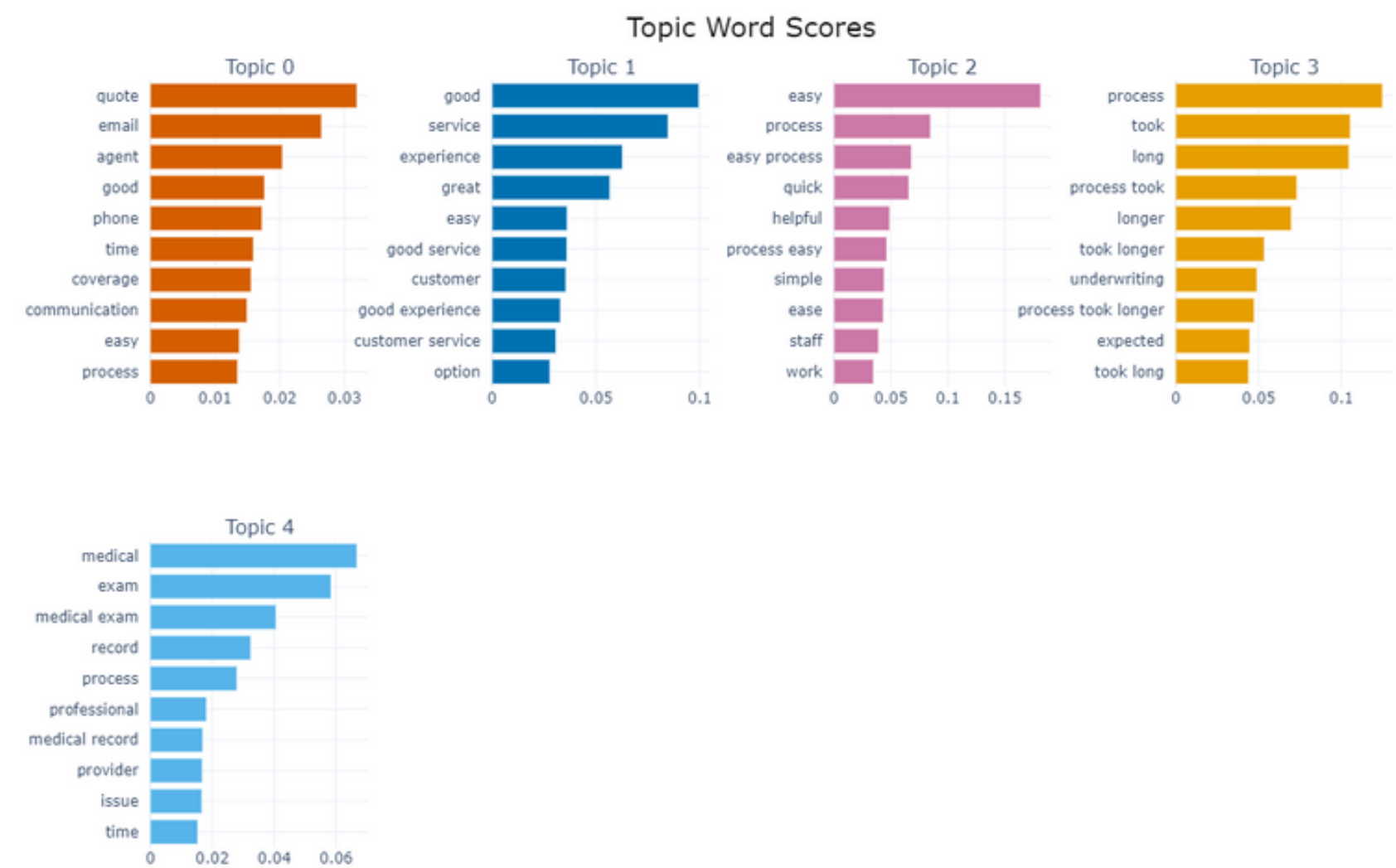


RESULTS

PROMOTERS

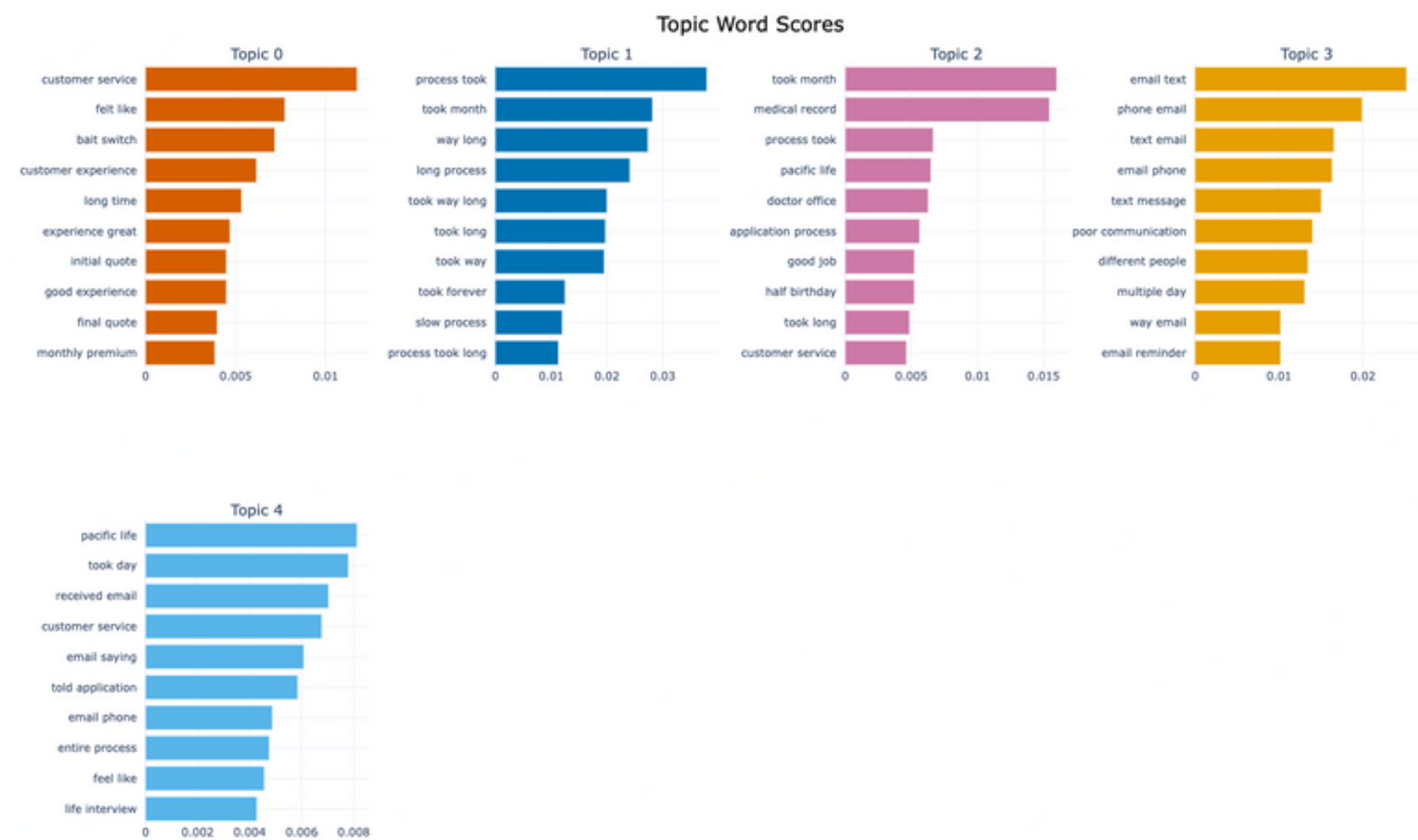


PASSIVES



RESULTS

DETRACTORS



CSAT



