

2024 Data Scientist Case Study

Marius Nwobi

Presentation Outline

About me

Workflow

Key Findings (EDA)

Sentiment Analysis and LDA Model's

Areas of Satisfaction and Dissatisfaction

Recommendations

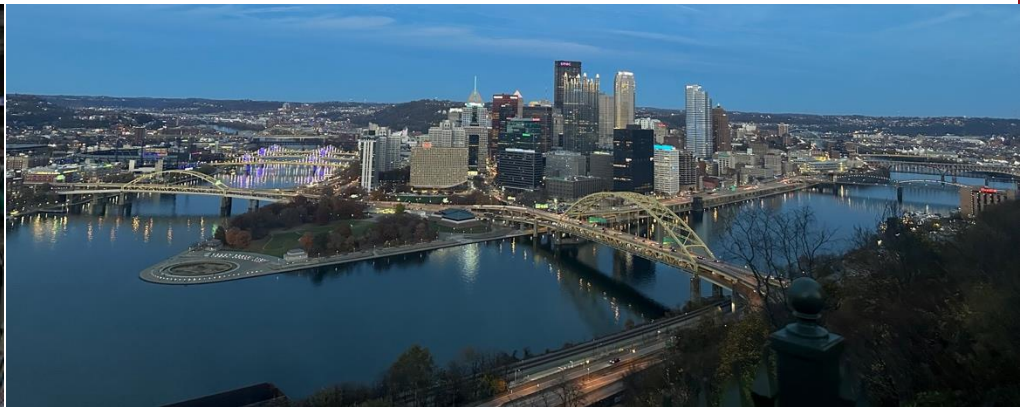
Additional Steps

Questions / Resources

About Me



Carnegie Mellon University
TARTANS



Education

- Carnegie Mellon University: Masters in Data Analytics for Science (May 25)
- Michigan State University: Bachelors in Biochemistry (Apr 24)

Work Experience

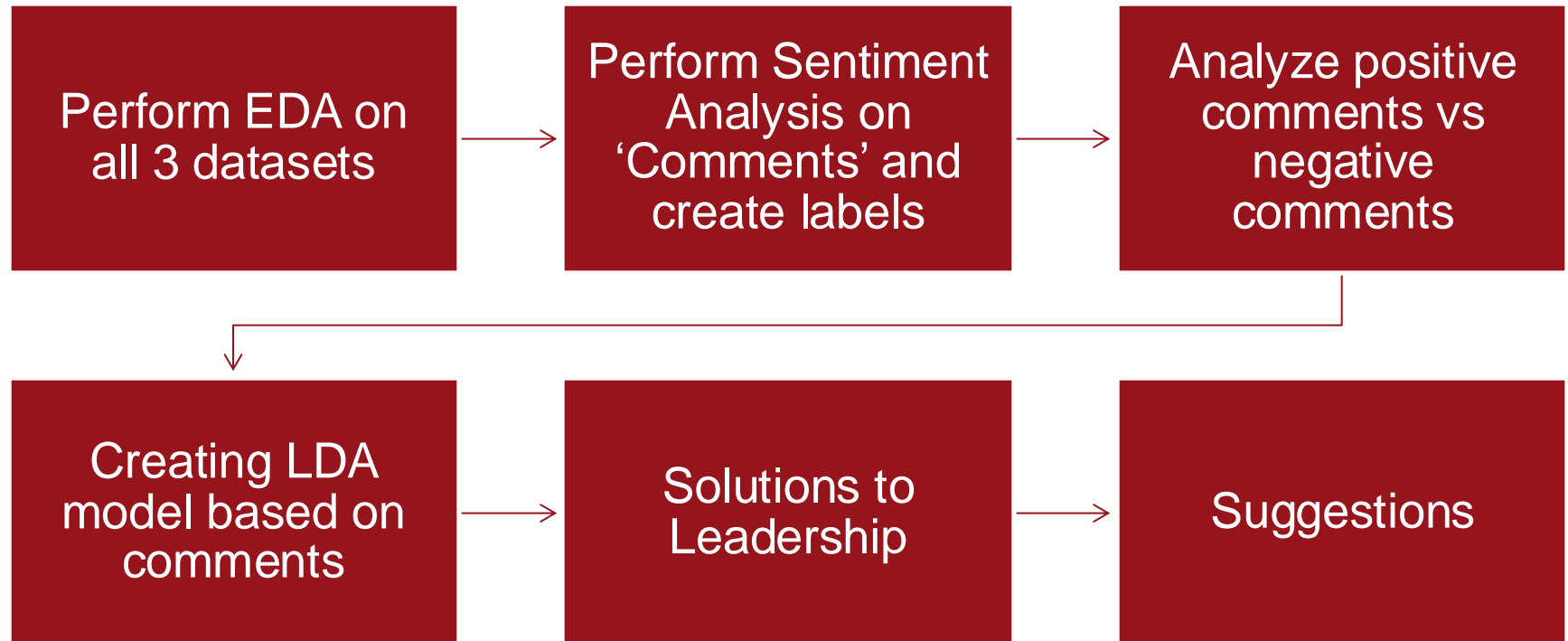
- Stellantis intern for both Data Science & Machine Learning team and Corporate Data & Analytics team (May 24)
- Undergrad research assistant in Computational Education Research Lab at MSU (Sep 22)
- Institute of Cyber-Enabled Research REU researcher at MSU (May 22)

Goal

- Continuously expand my knowledge within and around Data Science to apply these skills to real-world problems
- Become a Data Scientist and Leader who impacts and supports underrepresented communities within various domains

Things about me

Workflow



CTS Dataset (Customers who had a claim)

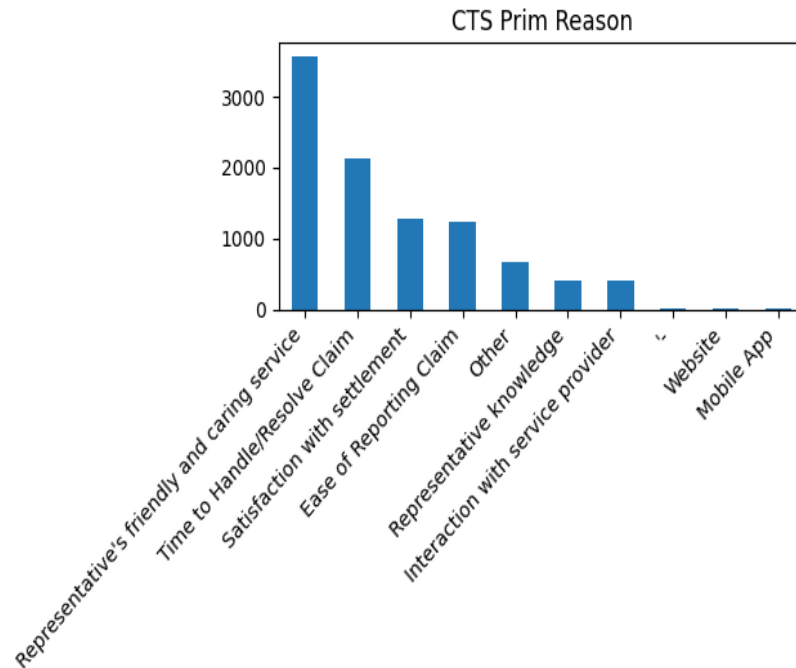
EDA Key Findings

A little over 1/3 of the Primary Reasons was 'Rep's friendly and caring service'

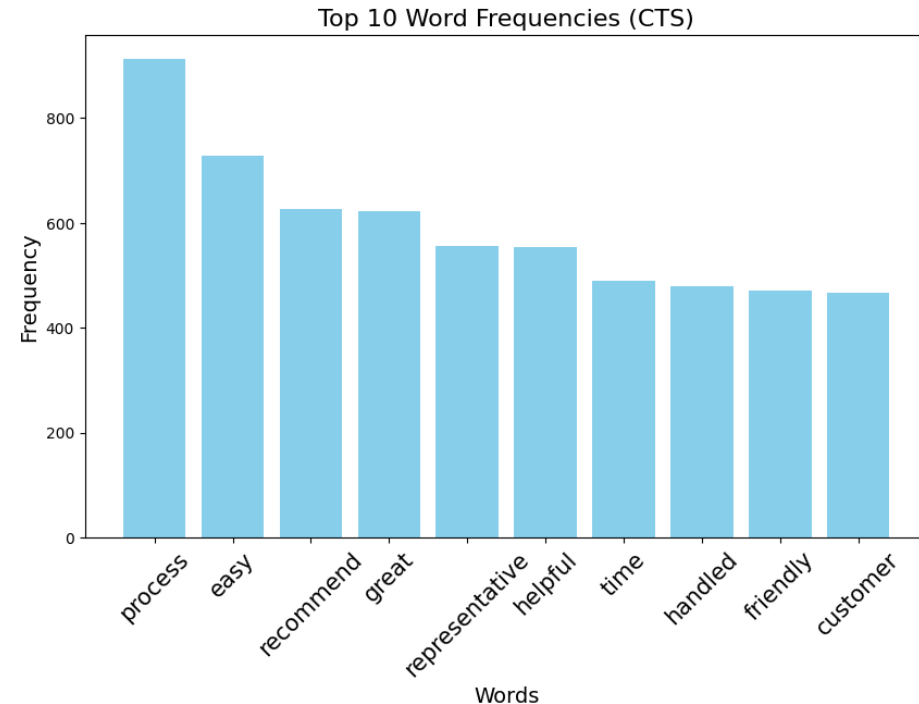
72% of the comments were filed out

Majority of the words were directed towards the process and rep's

Primary Reason Distribution



Word Frequency Plot on Comments



SCS Dataset (Customers who had a policy change)

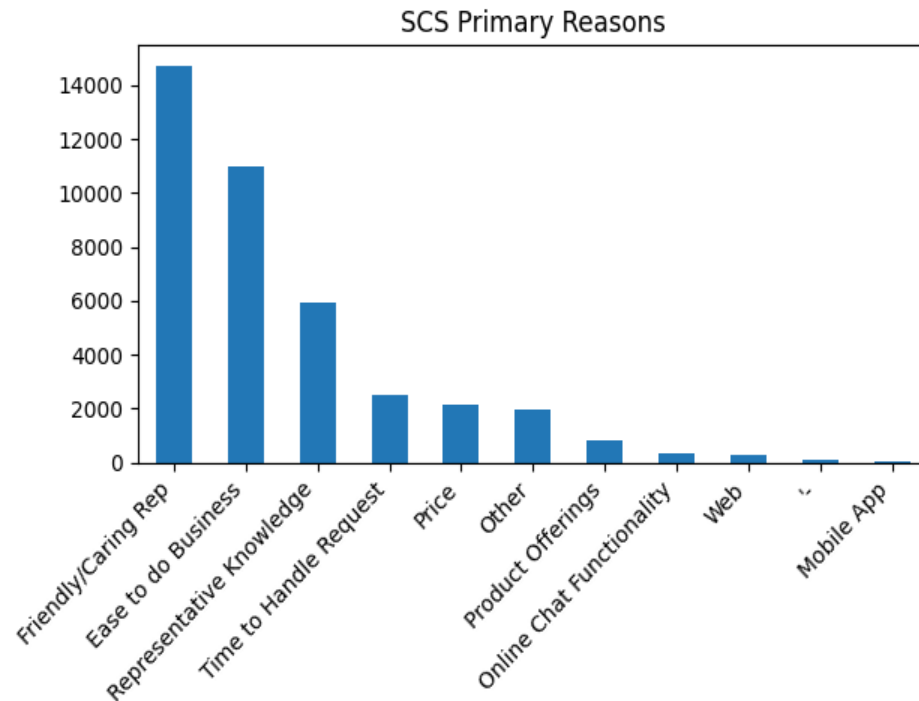
EDA Key Findings

27% of the surveys reported 'Friendly/Caring Rep' as a primary reason

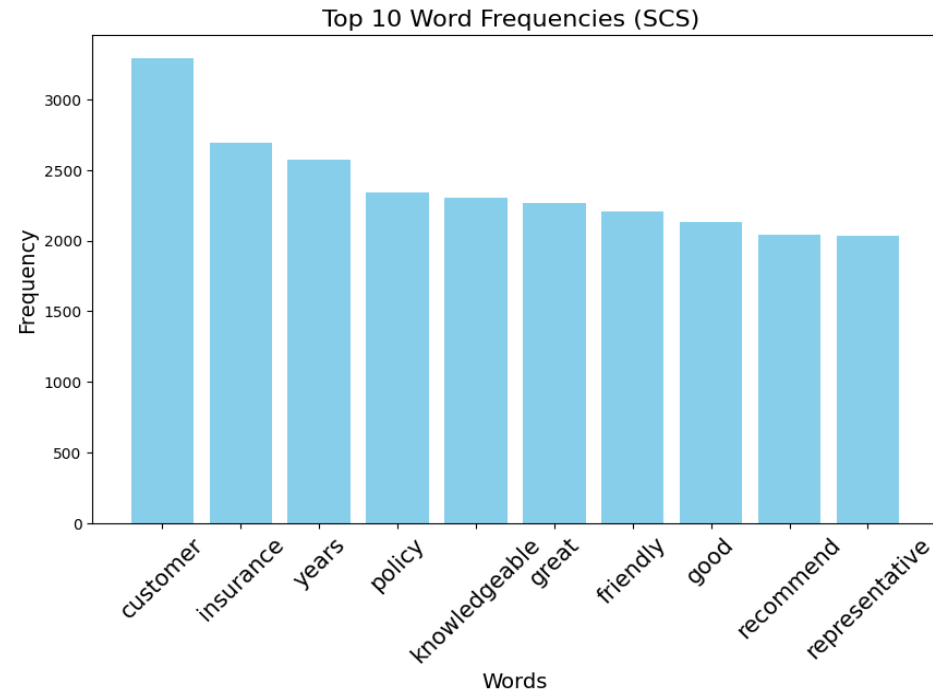
47% of the comments were filed out

Majority of the words were directed towards the policy and and reps

Primary Reason Distribution



Word Frequency Plot on Comments



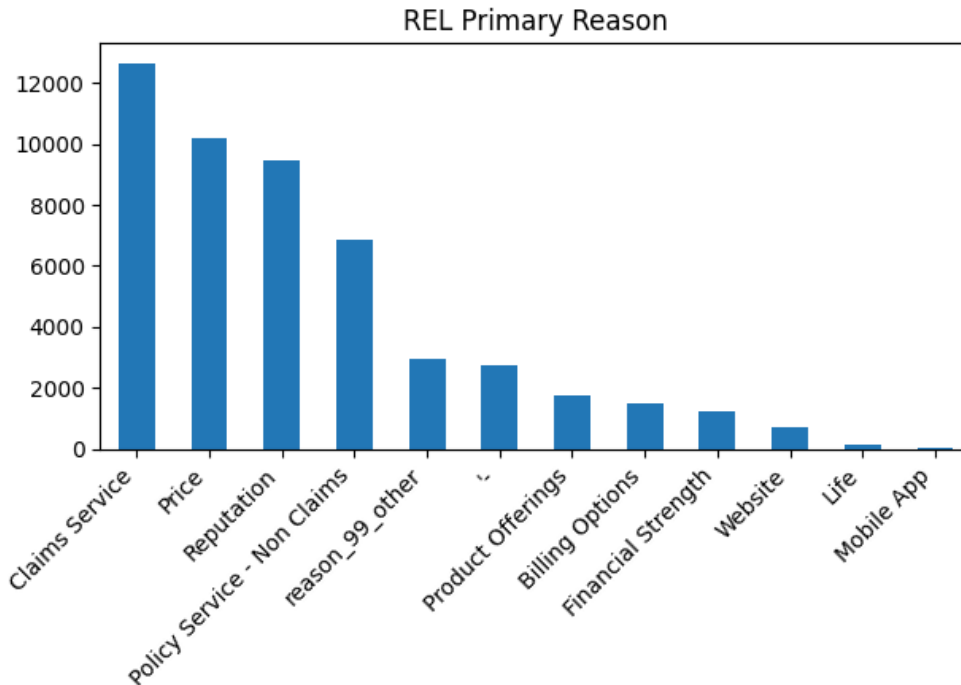
EDA Key Findings

25% of the surveys reported 'Claims Service' as a primary reason

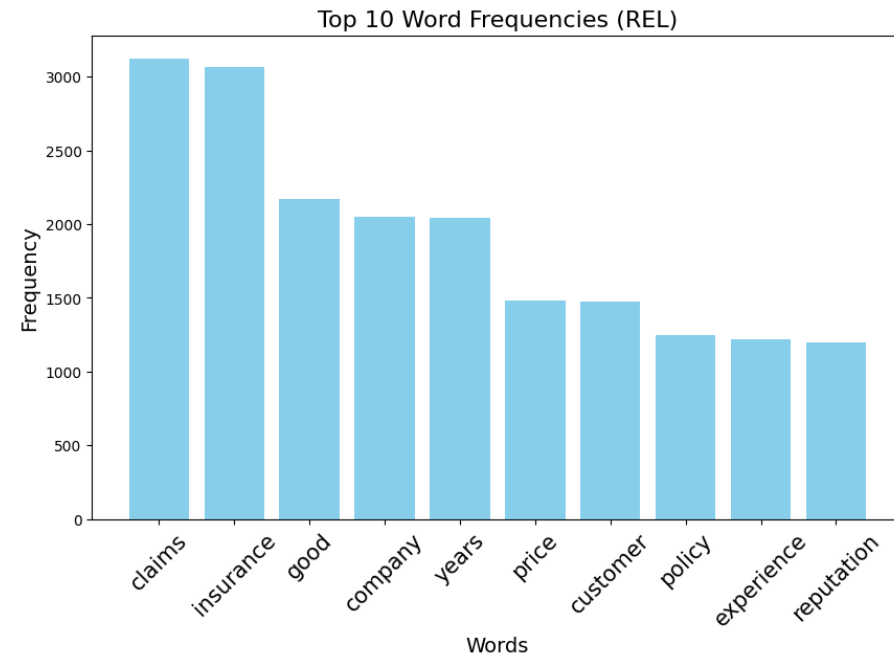
57% of the comments were filed out

Majority of the words were directed towards the policy and company

Primary Reason Distribution

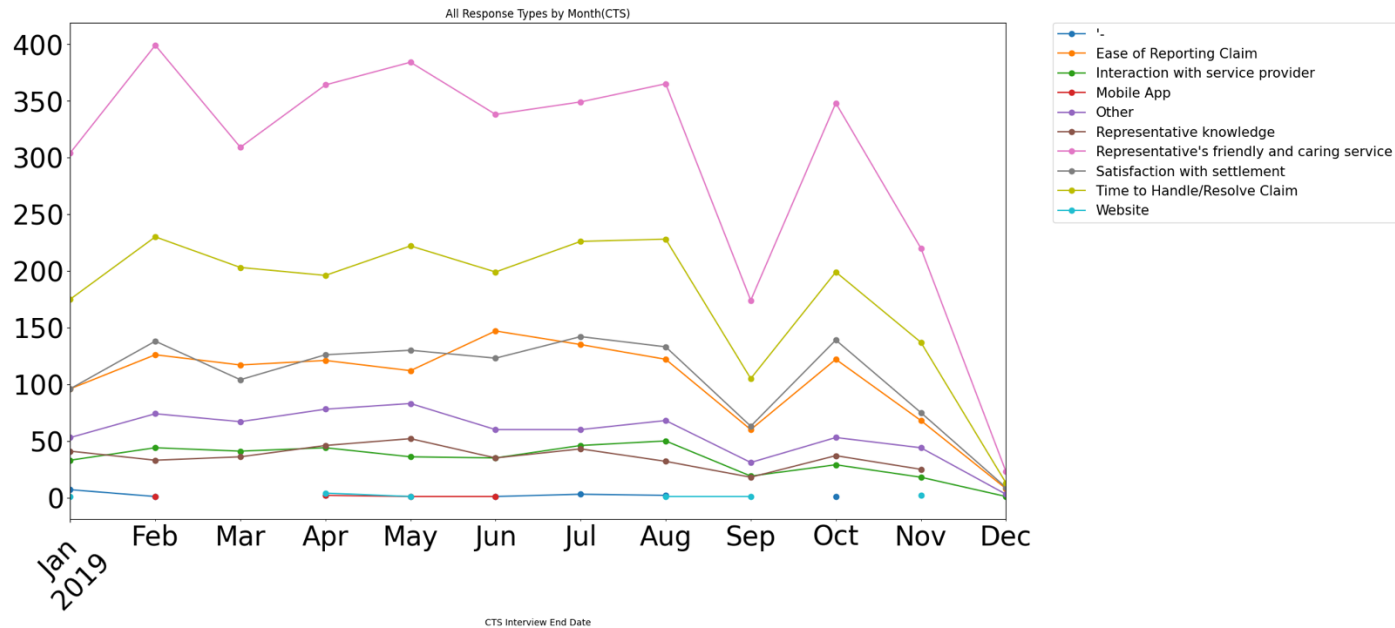


Word Frequency Plot on Comments



CTS Dataset (Customers who had a claim)

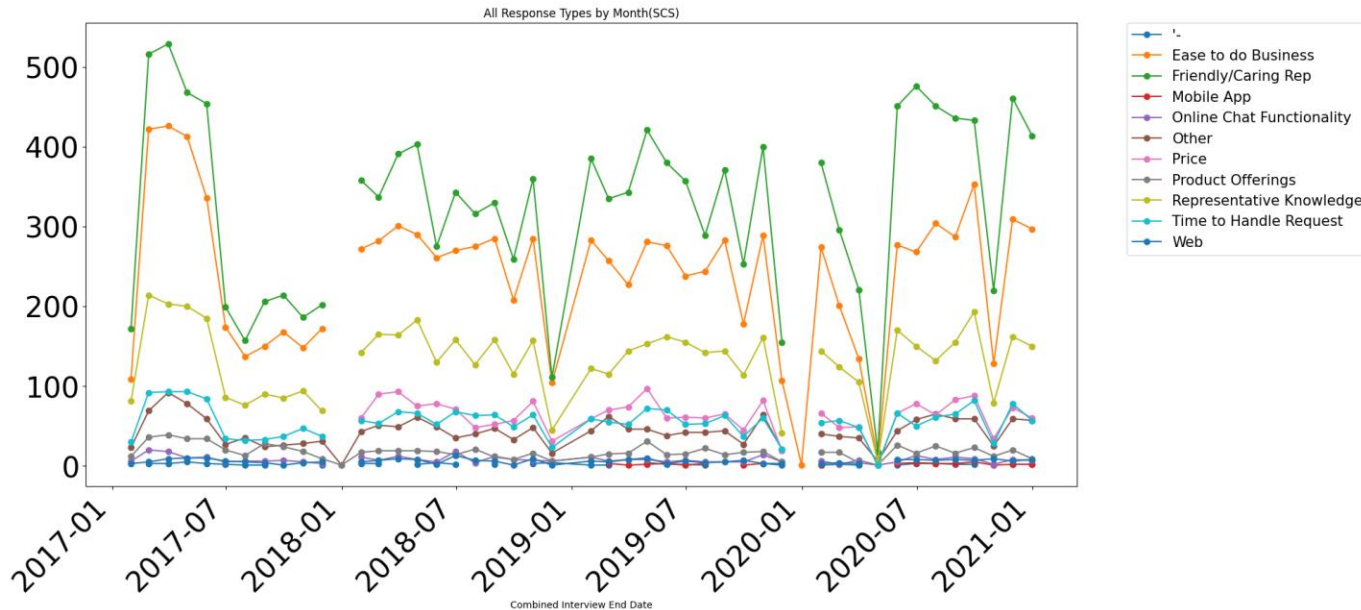
Survey Responses Over Time



Severe decrease
in reports being
filed in
September

SCS Dataset (Customers who had a policy change)

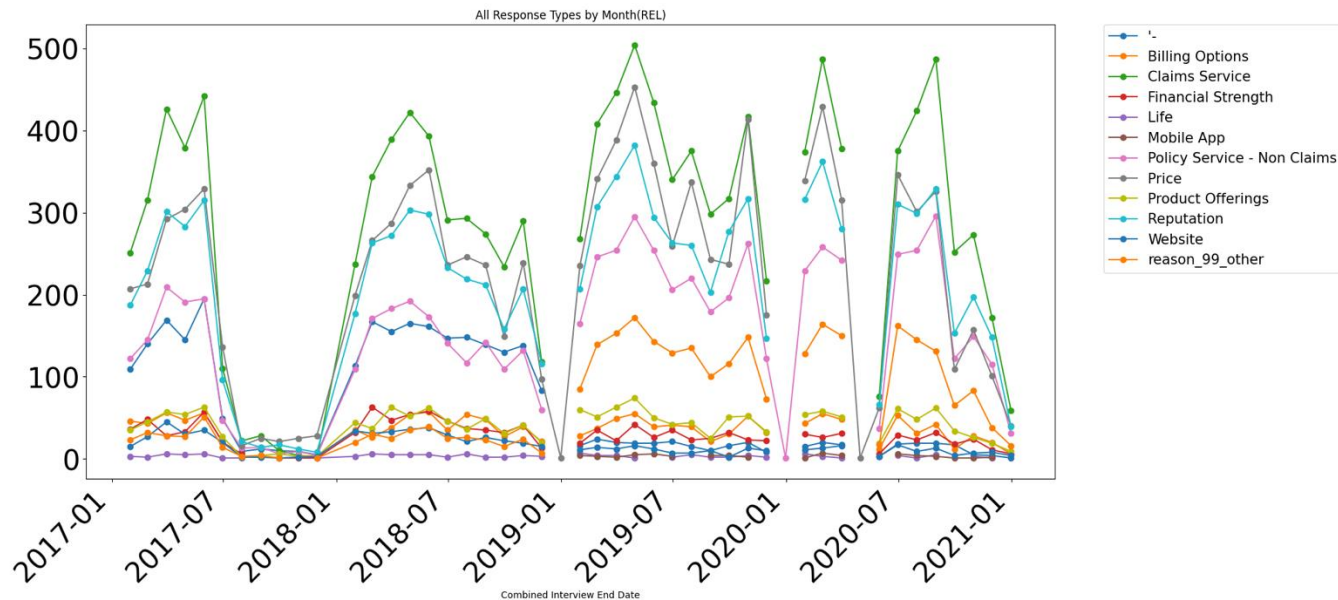
Survey Responses Over Time



Months Dec-Jan
are lows while
the peaks
change around
April-May

REL Dataset (Customers who had no policy change or claim over the year)

Survey Responses Over Time



Months May-June
are peaks while
Dec-Jan are lows

Sentiment Analysis Disclaimers

- Utilized sentiment intensity analyzer algorithm
- Used 0 as a threshold for class cut-off
- Class 0 is positive comments
- Class 1 is negative comments

Example

```
'Ms. Bradshaw has been very pleasant to work with. She has reached out to me to make sure everything is going fine.'
```

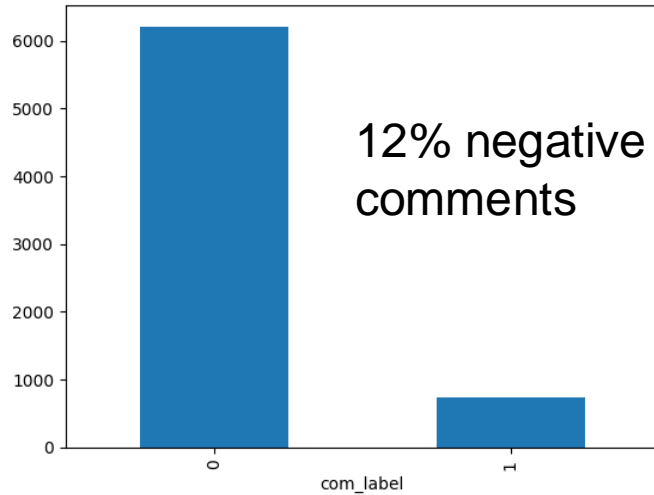
Score

```
{'neg': 0.0, 'neu': 0.664, 'pos': 0.336, 'compound': 0.796}
```

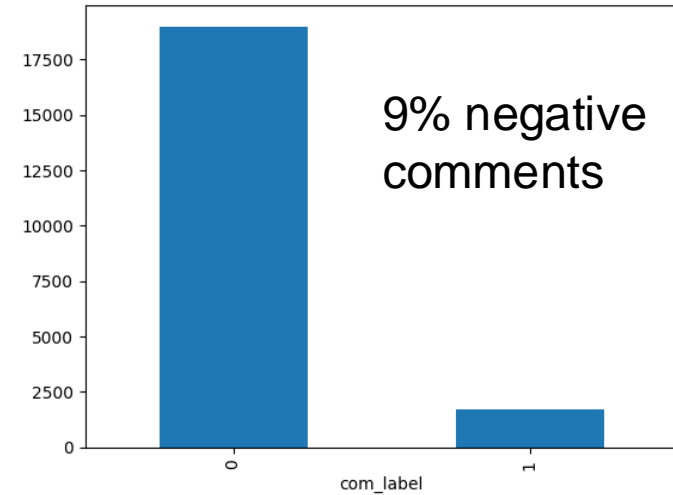
Sentiment Analysis (Class Imbalance)

Class Distribution

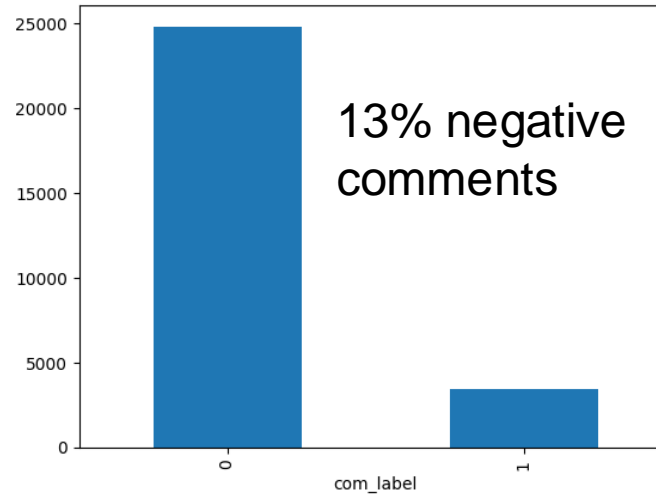
CTS Class Distribution



SCS Class Distribution



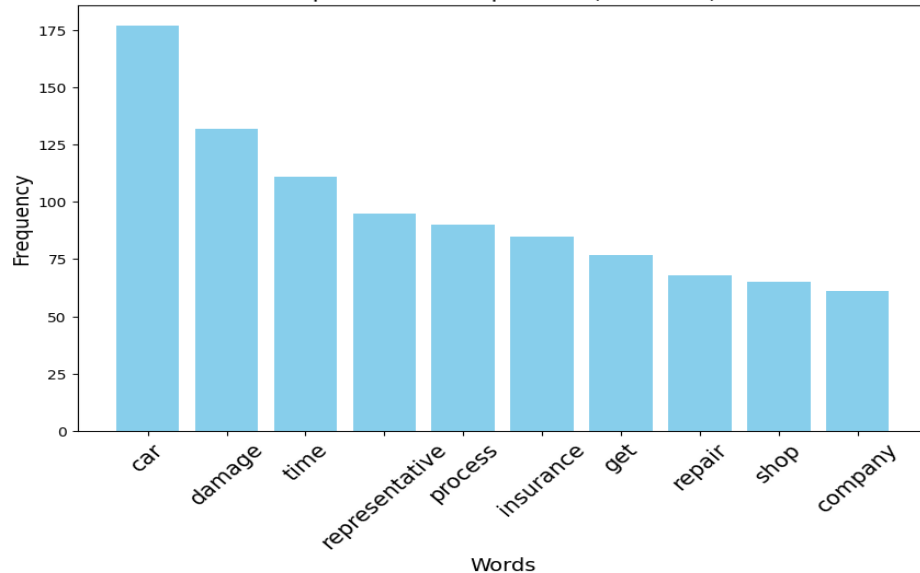
REL Class Distribution



Key Findings from LDA and Sentiment Analysis (BAD comments)

Word Frequency Plot on Comments

Top 10 Word Frequencies(CTS BAD)



Keywords are
tailored towards
reps and car
insurance

LDA results

```
Model Perplexity: 475.841570061303 , Model loglikely: -41077.96110729336
Topic: 0
    time, representative, stressful, claims, stress, process, easy, car, took, years
Topic: 1
    car, damage, repair, shop, time, rental, called, said, phone, body
Topic: 2
    car, insurance, loss, total, damage, time, bad, feel, pay, representative
Topic: 3
    damage, settlement, covered, insurance, contractor, check, work, adjuster, rep,
    hassle
Topic: 4
    process, problems, problem, rep, know, hit, communication, resolved, really,
    collision
```

Possible topics

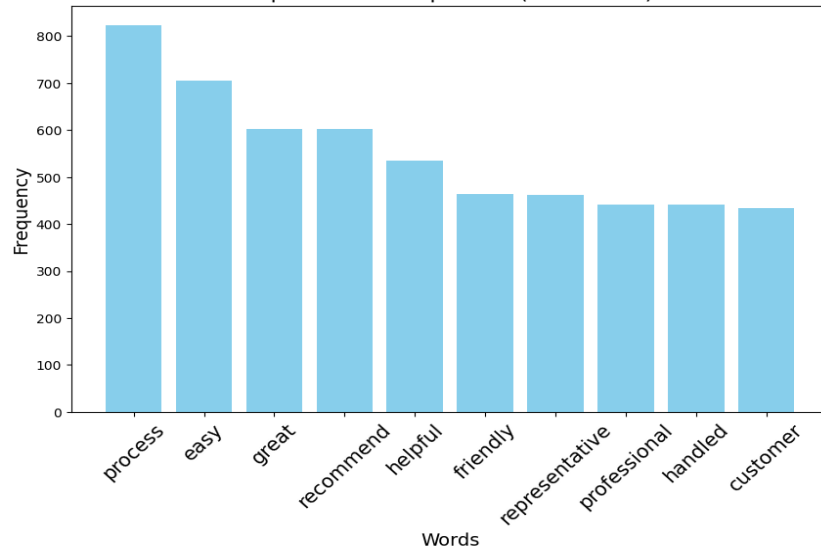
Confusion about what the insurance is covering on car

How stressful or hassling it is to contact a representative

Key Findings from LDA and Sentiment Analysis (GOOD comments)

Word Frequency Plot on Comments

Top 10 Word Frequencies(CTS GOOD)



Words are tailored towards the representative heavily

LDA results

Model Perplexity: 311.5572896581311 , Model loglikely: -184230.1813176791

Topic: 0

easy, handled, quickly, process, settlement, fair, quick, timely, fast, work

Topic: 1

recommend, excellent, insurance, company, best, customer, kept, ease, informed, years

Topic: 2

process, took, went, time, representative, make, shop, feel, care, like

Topic: 3

customer, good, car, years, great, satisfied, experience, shop, rental, rep

Topic: 4

helpful, friendly, professional, knowledgeable, great, prompt, representative, quick, efficient, response

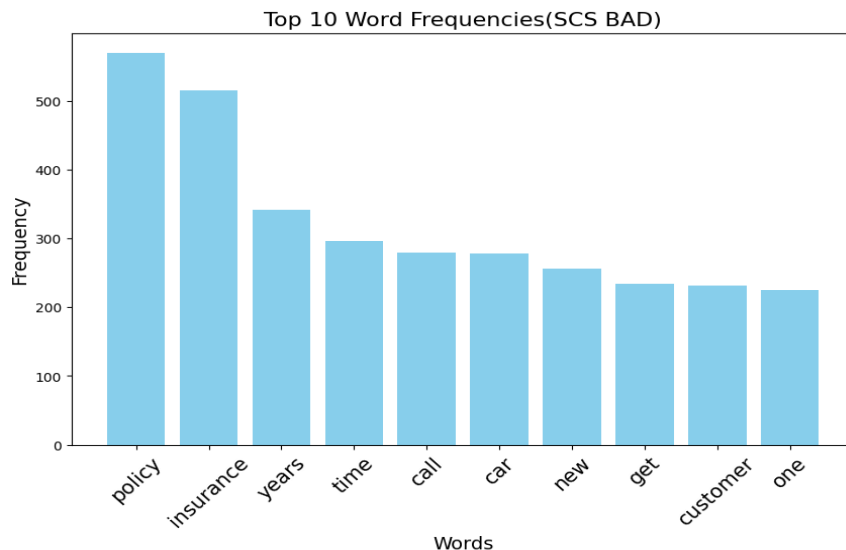
Possible topics

Good traits about the representatives such as knowledgeable, friendly, caring, etc.

How fast and efficient the process is

Key Findings from LDA and Sentiment Analysis (BAD comments)

Word Frequency Plot on Comments



LDA results

```
Model Perplexity: 388.2618518123642 , Model loglikely: -111352.25881216701
Topic: 0
    policy, change, insurance, representative, auto, pay, years, time, information,
    payment
Topic: 1
    policy, rep, car, new, told, called, time, coverage, representative, took
Topic: 2
    reason, time, email, called, phone, policy, request, received, times, agent
Topic: 3
    price, recommend, insurance, year, rates, company, years, problem, claims,
    premium
Topic: 4
    insurance, years, customer, car, home, new, company, policy, experience, year
```

Keywords are
tailored toward
communication,
time, policy
change(auto), and
pricing

Possible topics

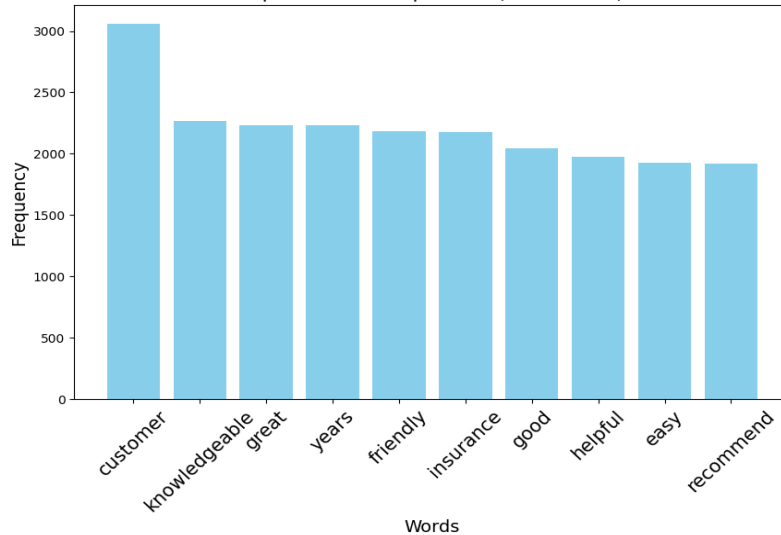
Issues with communicating with the
representative(calling, email)

Issues or confusion with policy change after x years

Key Findings from LDA and Sentiment Analysis (GOOD comments)

Word Frequency Plot on Comments

Top 10 Word Frequencies(SCS GOOD)



Words are tailored
towards the policy
and
representative,
and pricing

LDA results

Model Perplexity: 261.1408885866385 , Model loglikely: -622897.1730413557

Topic: 0

knowledgeable, friendly, helpful, easy, representative, representatives,
professional, quick, efficient, business

Topic: 1

recommend, excellent, customer, insurance, company, best, business, high,
friends, like

Topic: 2

policy, insurance, car, new, change, time, coverage, called, representative, rep

Topic: 3

great, good, customer, price, people, company, prices, rates, pricing, fair

Topic: 4

years, customer, insurance, experience, claims, company, happy, satisfied, time,
recent

Possible topics

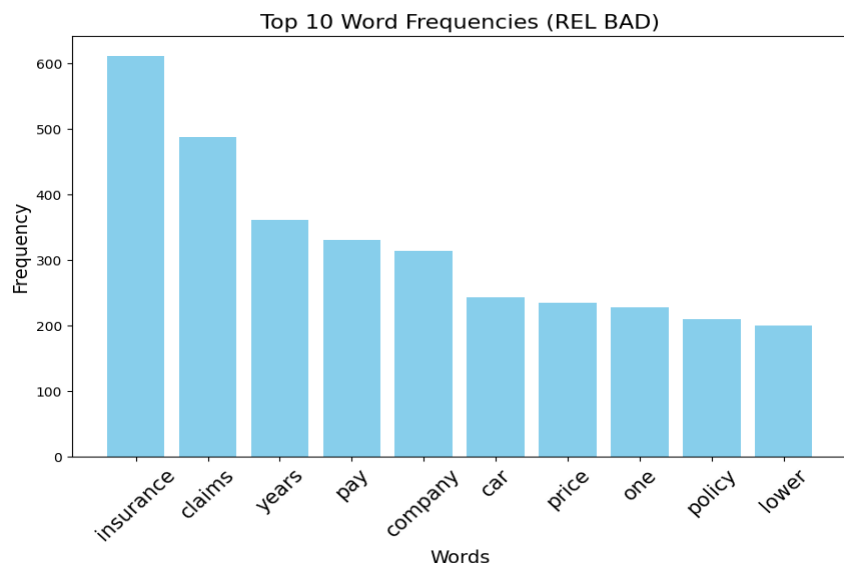
Positive comments on policy coverage

Positive comments on the representatives

Positive comments around time

Key Findings from LDA and Sentiment Analysis (BAD comments)

Word Frequency Plot on Comments



Keywords are tailored towards policy(auto/home) and price/rates

LDA results

```
Model Perplexity: 356.69981899601913 , Model loglikely: -99331.27227439127
Topic: 0
  claims, years, policy, hassle, auto, went, quickly, handled, month, customer
Topic: 1
  insurance, lower, price, rates, companies, year, reason, claims, premiums,
competitive
Topic: 2
  pay, insurance, need, company, time, low, want, like, customer, money
Topic: 3
  car, damage, problem, years, home, ago, claims, handled, paid, house
Topic: 4
  company, experience, reputation, bad, insurance, good, fault, claims, question,
difficult
```

Possible topics

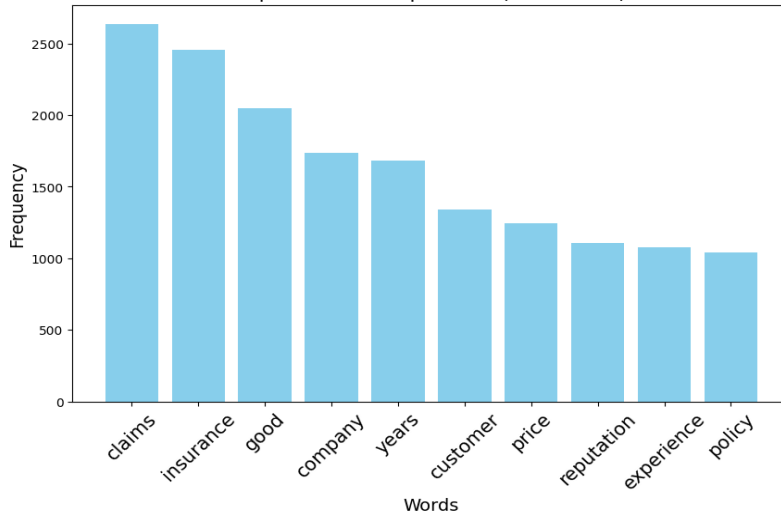
Criticism about rates or pricing

Criticism on plans/insurance

Key Findings from LDA and Sentiment Analysis (GOOD comments)

Word Frequency Plot on Comments

Top 10 Word Frequencies (REL GOOD)



Words are tailored towards the policy and company

LDA results

```
Model Perplexity: 305.09808066154136 , Model loglikely: -496236.33568244416
Topic: 0
  claims, years, customer, handled, excellent, helpful, quickly, time, ago, care
Topic: 1
  good, price, experience, reason, questions, customer, reasonable, competitive,
  value, best
Topic: 2
  policy, easy, insurance, like, auto, needed, home, need, good, make
Topic: 3
  reputation, coverage, response, best, answer, questions, quick, previous, fair,
  question
Topic: 4
  insurance, company, important, high, rates, companies, consumer, cost, reports,
  year
```

Possible topics

Positive feedback on quickness of process

Positive comments on pricing of products

Positive comments on the company's reputation

Areas of Satisfaction and Dissatisfaction (Based on Relationship)

CTS

Satisfied	Dissatisfied
Representative	Website
Time to resolve claim	Mobile app
	Contacting representatives
	Car/auto insurance

SCS

Satisfied	Dissatisfied
Representative	Website
Time to handle request	Mobile app
Policy	Online chat
Pricing	Confusion on policy
	Contacting representative

REL

Satisfied	Dissatisfied
Company reputation	Website
Quickness	Mobile app
Price	Policy coverage
	Rates

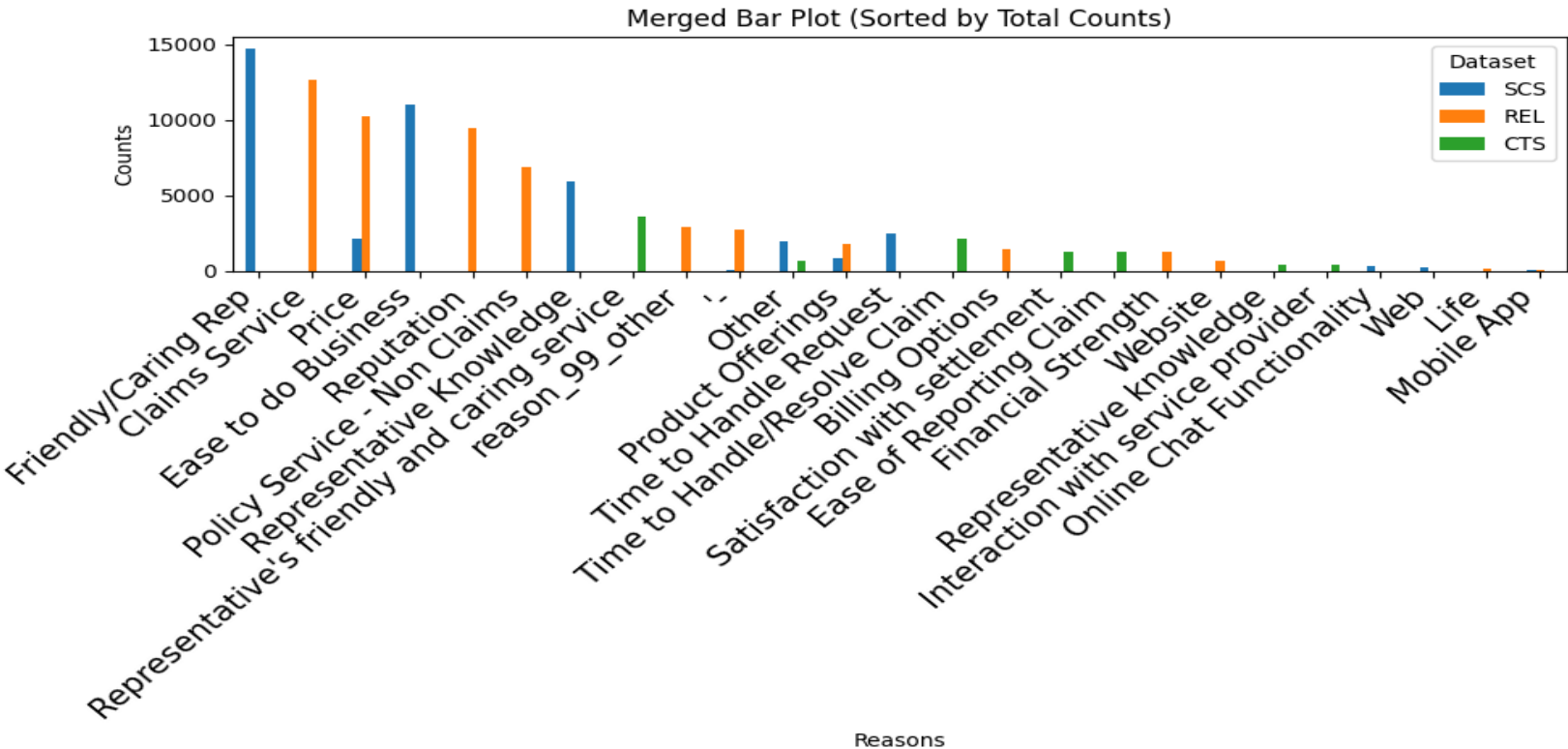
Areas of Satisfaction and Dissatisfaction (Overall)

Satisfied

- Representatives' knowledge and positive traits
- Quickness and ease of process
- Services and Reputation
- Pricing

Dissatisfied

- Website
- Mobile App
- Online Chat functionality
- Contacting representatives
- Pricing
- Coverage of policies (auto/home)



Recommendations



Improve mobile app and website



Inform customers on policy change



Improve system for contacting representatives



Constant Rates

What might be under SureStore's control to improve experiences

Additional Steps

Areas of Improvement

- Frequency plot
- LDA Model
- Vectorizers
- Look into 'Others' column
- Combine datasets for overall analysis

Questions on Dataset

- How was the data pulled
- What's the context of 'time' in this dataset
- How should I go about handling the other columns since they contain little to no data
- Are offices closed during December/January?

Questions?

Resources

- <https://hex.tech/templates/sentiment-analysis/vader-sentiment-analysis/>
- <https://scikit-learn.org/1.5/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html>