SAILOR

SENTIMENT ANALYSIS INSIGHTS FROM LANGUAGE OF REPORTS

JASMINE BOATNER*, BENJAMIN JOHNSON*, DR. ANDREW MILLER, CHARLES WILSON*

OFFICE OF NAVAL INTELLIGENCE

*ACCENTURE FEDERAL SERVICES

Background

Climate Surveys

❑ Organization solicits feedback from employees on the workplace climate

❑ Survey consists of Likert Scale model numeric questions, as well as questions soliciting free text responses

❑ Results from survey are output as a large .pdf file

Why use AI?

❑ **Time**
- ❑ In a large organization, it is infeasible to have human readers go through all the comments
- ❑ SAILOR analyzes the text in a fraction of the time it would take a human reader

❑ **Convenience**
- ❑ Comments are stored in a meaningful manner
- ❑ We were able to build SAILOR utilizing open-source solutions
  - ❑ R (Sentimentr package)
  - ❑ UiPath

❑ **Reproducibility**
- ❑ While human readers' sentiment and emotion scores are subjective and inconsistent, using a set methodology makes our results reproducible.

How SAILOR Works

1. Text is extracted from the survey results .pdf using UiPath sequence.

2. The extracted text is converted into an excel file for analysis.

Year	Month	short_category	comment	Center	full_category
2019	May	Genesis	If you can deal with (or prefer?) rear-wheel drive, this is an amazing vehicle, and a fairly well kept secret. The 3.8 engine is wonderfully silky smooth, and very powerful. This car passes in a heartbeat, and will keep up with Charger RT hemi's on the highway (not kidding). Gas mileage is fair, and my wife achieved 30 MPG on the highway on a 5 hour trip; I generally average 24 MPG in mixed driving... not bad for a car that is every-bit a BMW or Mercedes in style and performance.	CARS	Genesis Motor LLC
2020	July	All New Fire	I like the portability of it. I take it with me when I babysit my granddaughter.	AMAZON	All-New Fire HD 8 Tablet, 8" HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta

How SAILOR Works

3. Excel file of input data and customized version of hash_nrc_emotions (Mohammad & Turney, P. D.'s 2010 emotion word list) read into R or Python.

4. Polarity is neutralized in code for common phrases such as 'yes', 'unsure', 'no comment'

5. Sentimentr assigns a sentiment score for each response based on our parameters.

```
# Do analysis (by element)

text_sentences <- sentimentr::get_sentences(text)

# Get sentiments, using the Semantic Orientation Calculator Google lexicon
sentiments <- sentiment_by(text_sentences, by = NULL,
                            averaging.function = average_weighted_mixed_sentiment,
                            polarity_dt = updated_socal_google, amplifier.weight = .9,
                            n.before = 6, n.after = 3, adversative.weight = .3)
```

How SAILOR Works

6. Extract_emotion_terms does not take modifiers into account, so we perform global substitutions on the text before analysis, i.e. "not happy" transforms into "unhappy".

7. Emotions are extracted out of the comments and restructured so each row of the data contains one emotion word.

8. We assign numeric values to the emotions for analysis and visualization purposes.

9. Final output is written to an excel file.

|	short_category	comment	Center	Year	full_category	element_id	comment_edit	ave_sentiment	sentence_id	edited_sentence	unique_id	original_sentence	Emotion	Word	Emotion as Numbers	Lemma
34043	Honda	Although	CARS	2009	Honda Mc	7111	Although�a	0.561959	1	Although�a	7111, 1	Although�ar	anticipatic	appeal	2	appeal
34044	Honda	Although	CARS	2009	Honda Mc	7111	Although�a	0.561959	1	Although�a	7111, 1	Although�ar	anticipatic	time	2	time
34045	Honda	Although	CARS	2009	Honda Mc	7111	Although�a	0.561959	1	Although�a	7111, 1	Although�ar	surprise	unique	1	unique
34046	Honda	2009 Hond	CARS	2009	Honda Mc	7112	2009 Honda	0.1955429	1	2009 Honda	7112, 1	2009 Honda A	trust	accord	3	accord
34047	Honda	I have ow	CARS	2010	Honda Mc	7113	I have owed	-0.1949834	4	unjust their a	7113, 4	Not just their	anger	unjust	-3	unjust
34048	Honda	I have ow	CARS	2010	Honda Mc	7113	I have owed	-0.1949834	6	The worst th	7113, 6	The worst thir	anger	worst	-3	bad
34049	Honda	I have ow	CARS	2010	Honda Mc	7113	I have owed	-0.1949834	10	I was forced	7113, 10	I was forced t	anger	remove	-3	remove

Visualizations

- ❑ We visualize the final output using Tableau
  - ❑ Tableau allows us to control who has access to what data
  - ❑ Tableau allows the end user to interact with their data
  - ❑ Comments where no emotions are detected, such as 'no comment', 'n/a', are automatically filtered out
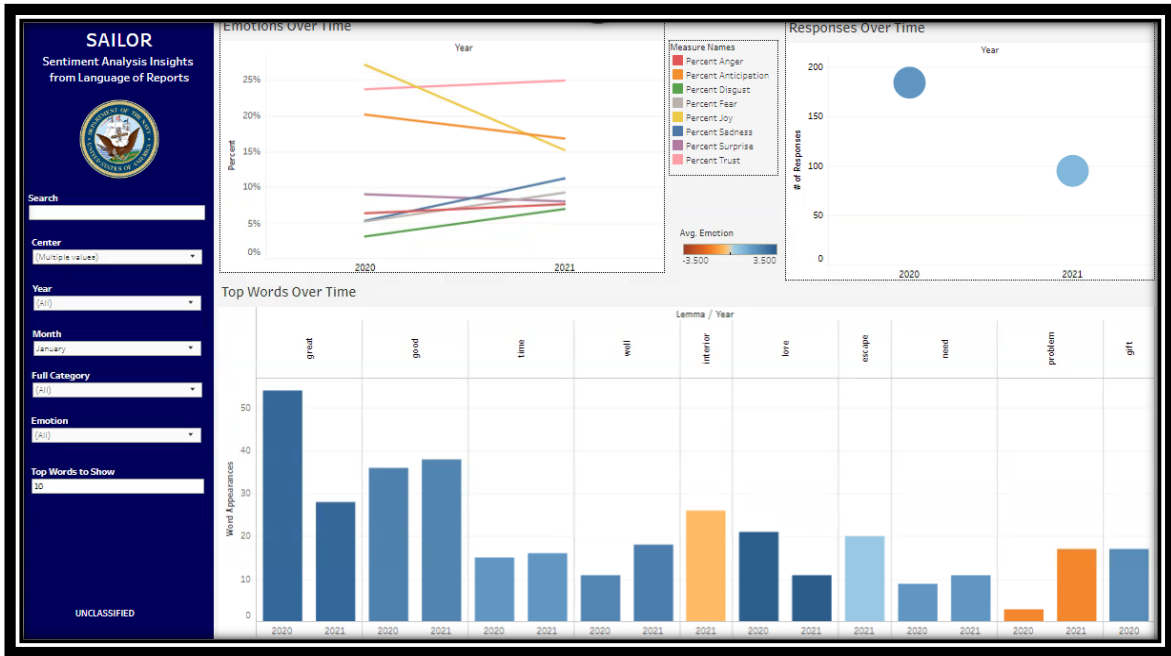
Analysis

☐ With numeric scores for both sentiment and emotion, it becomes possible to *quantify* the *qualitative* free response data.
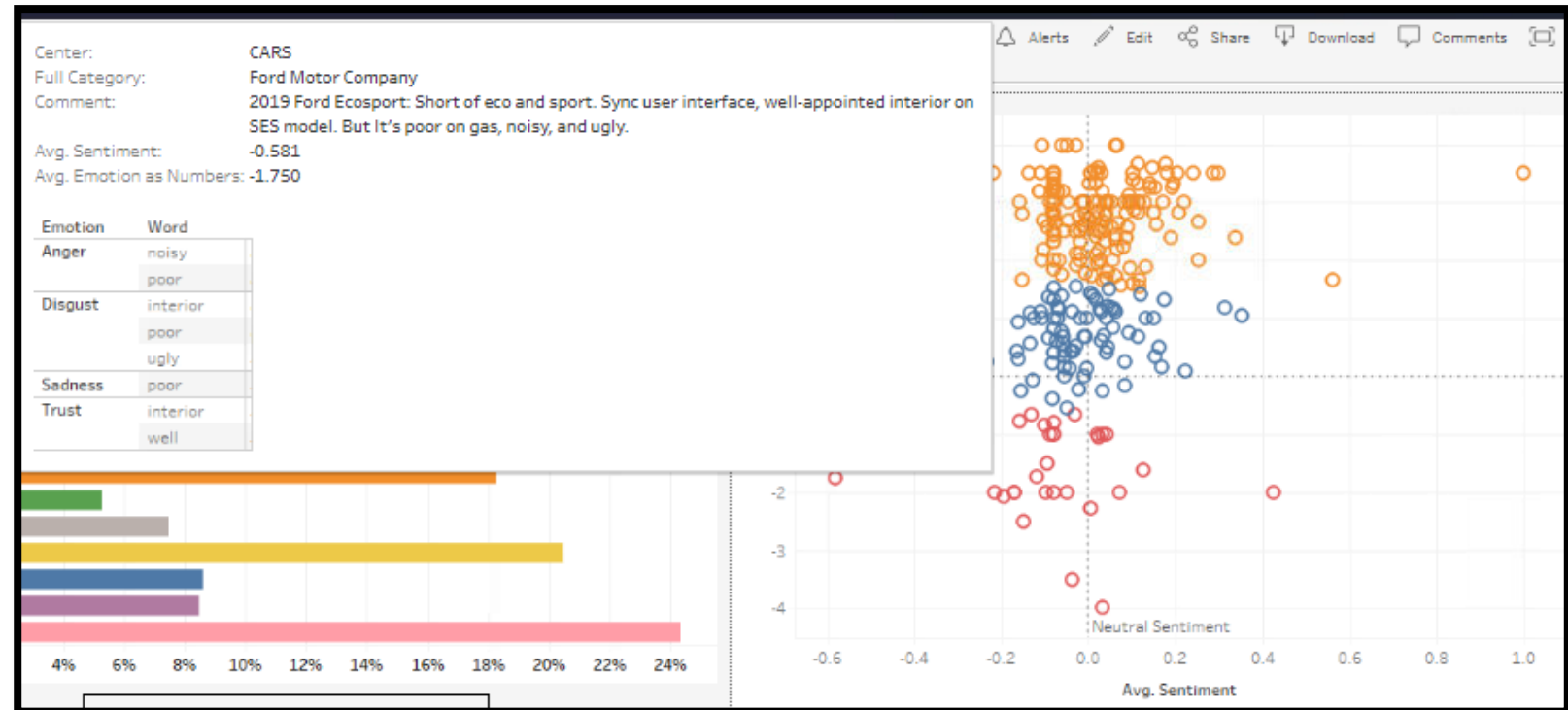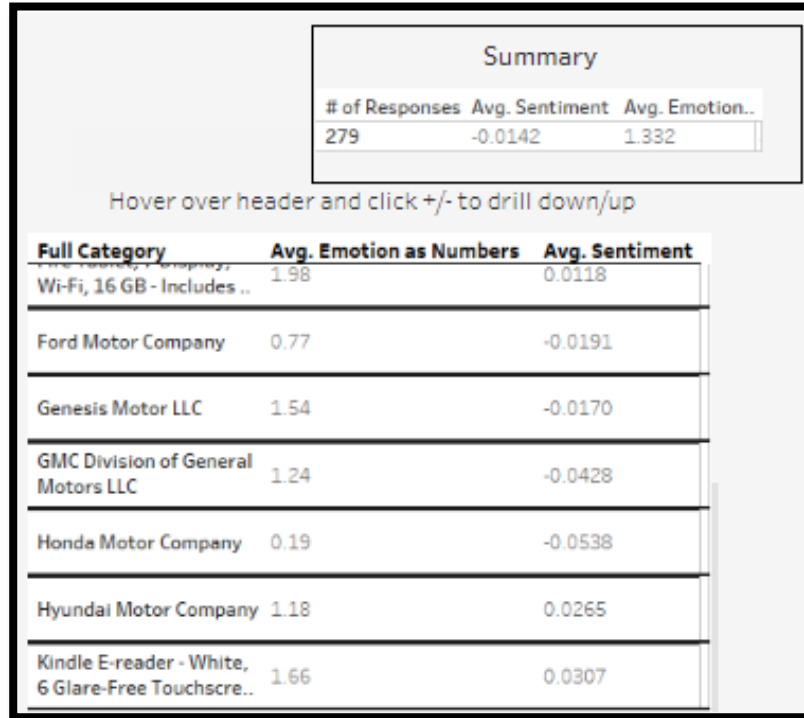
Caveats

❑ No matter how much we tweak the parameters, Sentimentr is never going to be 100% accurate

❑ More data will not make results more accurate, since this is not a machine learning model

❑ Parameters and dictionaries must be manually updated

Potential Future Updates

❑ Fully customized sentiment lexicon dictionary

❑ Markov Chains

Summary

❑ Sentiment Analysis Insights from Language of Reports (SAILOR) utilizes Sentimentr to analyze sentiment and emotion in Climate Survey responses.

❑ Assigning quantitative values to the qualitative response data allows for analysis over categories and time.

❑ Additionally, the numeric scores provide a way to rank the comments against one another. This allows a human reader the ability to focus on the most positive, or most negative, comments.

❑ Our final product is an interactive dashboard, but the data could be used for other purposes.