# USAA x Data Mine
# Spring 2024
# Arizona State University and Purdue University

Sreeram Kondapalli, Melissa Lin, Alex Elledge, Leslee Still, Nikki La, Dan Gibson, Chris Park, Jasmine Sharron, Giovanni Aratico, Saif Elsaady

## Introduction

About USAA: Founded in 1922, USAA stands as a premier provider of insurance, banking, and financial solutions to more than 13 million members of the United States military community, including active service members, veterans, and their families. Dedicated to enhancing the financial well-being of its members, USAA offers a blend of competitive products, outstanding customer service, and reliable advice with the goal of becoming the preferred choice for the military and their relatives.

## Objectives

Analyze Linguistic Patterns in customer calls to find main themes.

Create Word Clouds for top 20 words in high and low friction calls.

Identify Frequent Words in high vs. low friction interactions.

Apply Topic Modeling to group interactions into clusters.

Use N-grams (unigrams and bigrams) for deeper context.

Determine Cluster Number with Silhouette and Elbow methods.

## Data Source

The data is sourced from call data and transcripts provided by USAA.

## Methodology

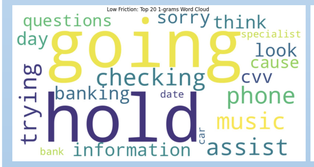
Phase 1: To form a word cloud of the top 20 most frequently used words, we had to first clean the data by filtering out the variety of stop words. Then we used n-grams (both unigrams and bigrams) to be able to contextualize certain phrases and words that may be insightful when evaluating call performance. Two word clouds were propagated, one for high friction score (top 30%) and one for low friction score (bottom 70%).

Phase 3: In topic modeling, we initially divide the data into segments with high friction scores (top 30%) and low friction scores (bottom 70%). Subsequently, we employ Silhouette and Elbow methods to ascertain the optimal number of clusters (k). Following the visualization process, we eliminate stop words and implement additional tuning techniques to enhance the utility of the results.

## Results and Discussion

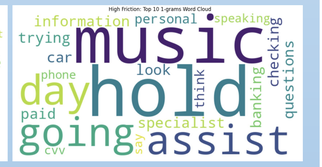
1-Gram

Word Cloud (Low Friction)




- The word cloud shows common words from customer feedback, highlighting "hold," "music," and "sorry," which suggest wait times and apologies are common. Words like "going," "checking," and "assist" show that the service is active in helping customers. Overall, the findings hint at a focus on prompt and caring customer service, without negative feedback.

Word Cloud (High Friction)




- The word cloud reveals keywords from difficult customer service calls, showing "hold," "music," and "paid" highlight issues like long waits and billing problems. "Trying," "questions," and "information" point to attempts to solve complex issues. Together, these words suggest calls often involve intricate problems needing careful handling by service reps.

2-Grams

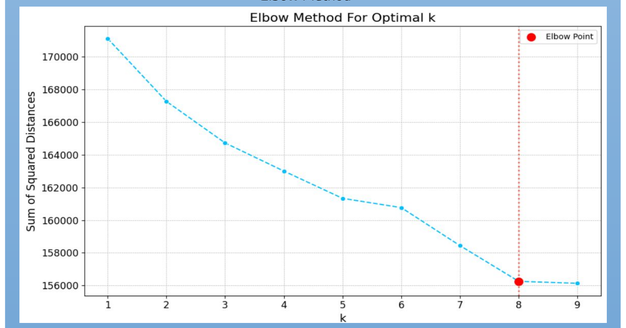
Word Cloud (Low Friction)


Phrases such as "great_day" and "pleasure_speaking" hint at happy call conclusions, showing good service. "Additional_questions" suggests customers are invited to ask for more help. Words like "uh_huh" and "brief_hold" show casual talks and quick holds, making calls more enjoyable.

Word Cloud (High Friction)


The word cloud shows top phrases from tough service calls, highlighting "brief_hold," "monitored_recorded," and "quality_assurance" which suggest recording for quality and possible waiting times, causing issues. However, "pleasure_speaking" shows staff stay positive, and "additional_questions" means calls might get complex and long.

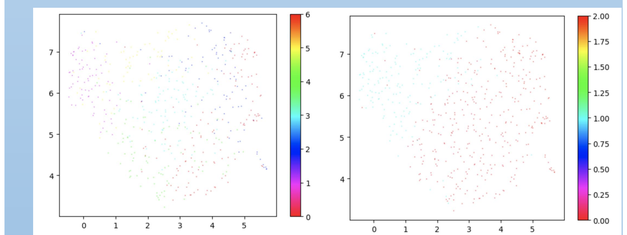
Elbow Method




## Results and Discussion (continued)

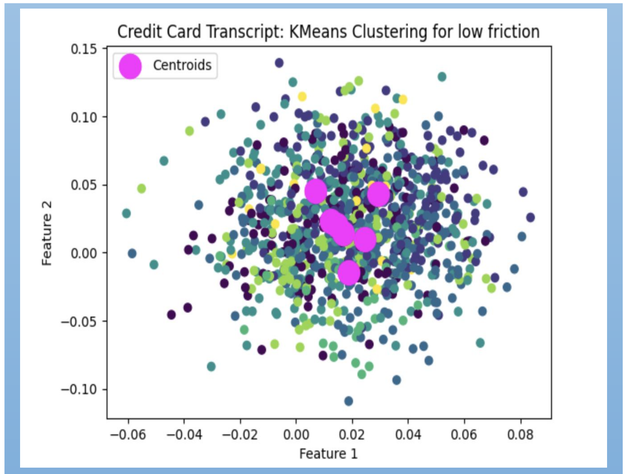
The elbow method graph illustrates the optimal number of clusters for k-means clustering. The 'elbow point' at 8 indicates that increasing the number of clusters beyond this does not significantly enhance the variance explained within the data. Therefore, 8 is the optimal cluster number, striking a balance between precision and simplicity of the model

K Clusters



This plot shows the analysis of 7 clusters.

This plot shows the analysis of 3 clusters.




This scatter plot represents KMeans clustering of low friction customer service calls based on a sample size of USAA's call center data. Each dot corresponds to a call, plotted by two key features derived from the conversation content. The color-coded clusters signify groupings of calls with similar characteristics, and the large 'Centroid' markers indicate the central point of each cluster. This visualization aids in identifying patterns within the calls, informing strategies for reducing the friction score

## Conclusion

| | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | account | alright | ma | bye | ya | laughter | thank | card |
| 1 | today | thank | am | thank | yup | sorry | help | number |
| 2 | usaa | today | thank | welcome | thank | good | appreciate | debit |
| 3 | gonna | help | today | thanks | laughter | thank | calling | phone |
| 4 | thank | good | alright | alright | alright | alright | sir | account |
| 5 | uhm | sir | help | day | account | money | today | credit |
| 6 | sir | appreciate | assist | care | good | gonna | good | pin |
| 7 | good | account | appreciate | great | hear | help | great | routing |
| 8 | please | perfect | account | sir | absolutely | account | perfect | usaa |
| 9 | check | gonna | day | usaa | problem | time | usaa | send |
| 10 | help | assist | good | calling | today | great | nope | uhm |
| 11 | correct | thanks | sorry | good | hold | work | awesome | gonna |
| 12 | sorry | uhm | usaa | rest | check | long | fine | code |
| 13 | huh | great | great | wonderful | gonna | phone | day | alright |
| 14 | transfer | hold | correct | am | card | today | assist | checking |

In our project to enhance call center performance, we meticulously analyzed transcripts from customer service interactions, focusing on those with low friction scores. Our goal was to discern the language that correlates with smooth and positive call experiences.

We utilized clustering algorithms to segment the calls based on linguistic patterns, forming distinct clusters that represent various interaction types. Within each cluster, we extracted the most common words, the language elements that signal effective communication.

By synthesizing these findings, we compiled a list of the top 20 words from all clusters combined. These words reflect affirmations, respectful acknowledgments, and efficient conversation markers which are pivotal in crafting satisfactory customer interactions.

Implementing these keywords into the call center script aims to reduce friction scores by encouraging a service language that resonates with customer satisfaction. This targeted approach is poised to not only lower friction scores but also elevate overall customer service quality.

## Acknowledgements

We extend our heartfelt gratitude to our mentors (Russell, Alex, Matthew, and Professor Samara), our teaching assistant (Dhruv), and the entire team at Data Mine for their invaluable support and guidance in orchestrating this project and assisting us throughout the semester.