

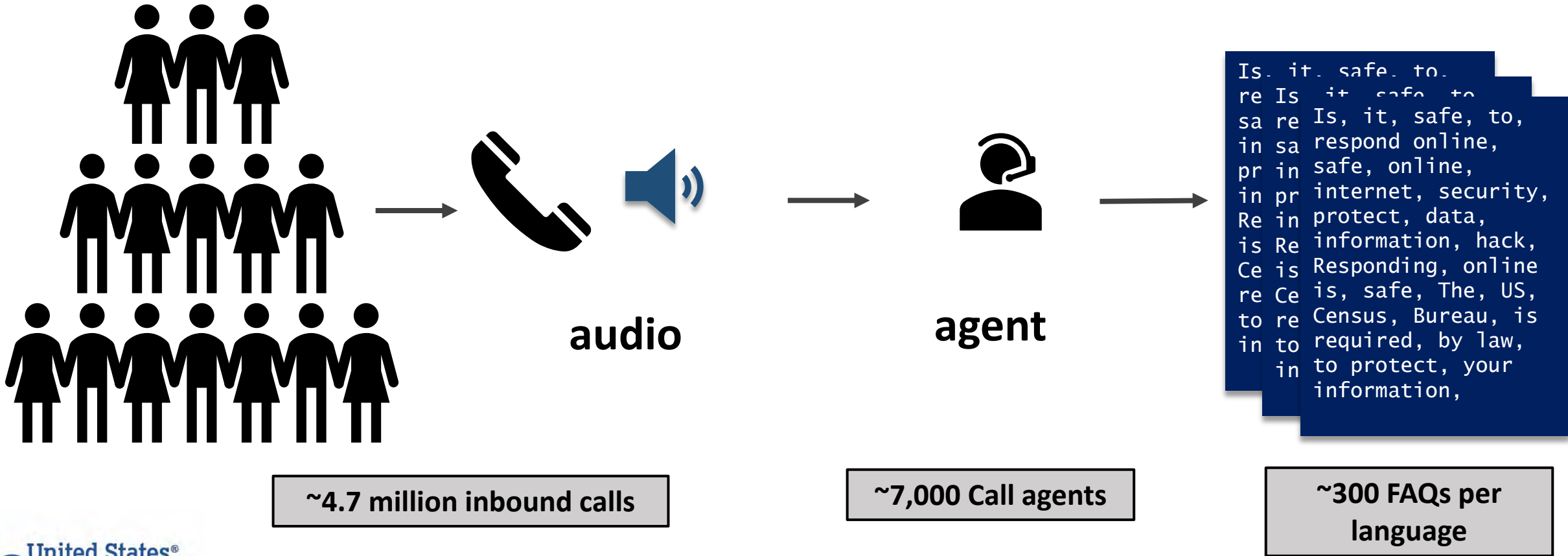
Applying machine learning language models for document linkage to evaluate FAQ Coverage

Monica Puerto*, Elizabeth Nichols⁺, Shaun Genter⁺, Kevin Zajack ,Brian Sadacca*

Accenture Federal Services* , U.S. Census Bureau⁺

FCSM October 2023

Objective: How can we use ML to help us understand if we have enough reference material to answer caller's questions in a call center?

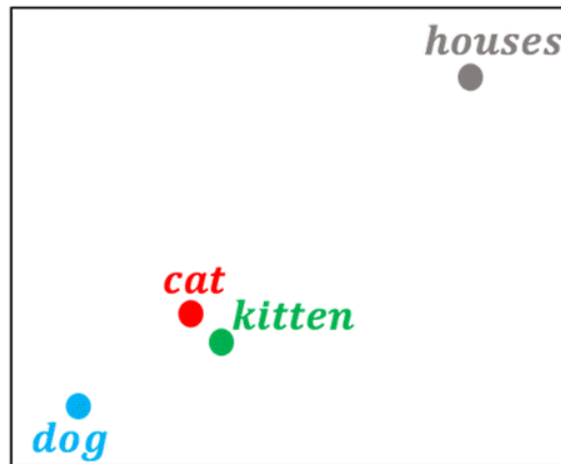


The presentation has been reviewed for disclosure avoidance and approved under CBDRB-FY24-CBSM002-035

Creating embeddings for topic modeling / clustering

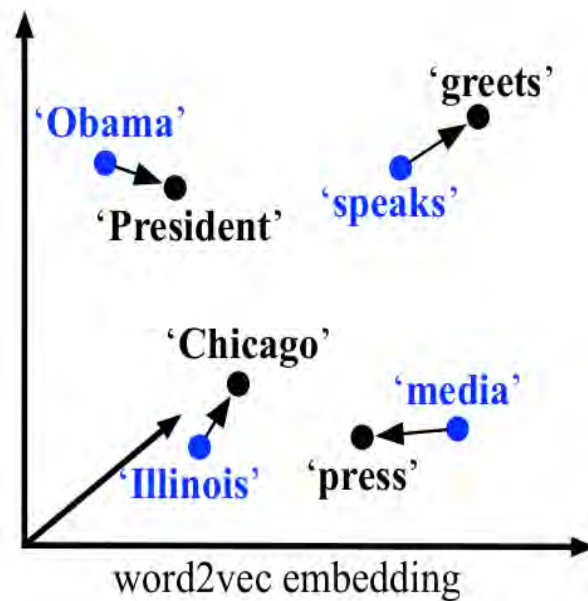
<i>cat</i>	0.6	0.9	0.1	0.4	-0.7	-0.3	-0.2
<i>kitten</i>	0.5	0.8	-0.1	0.2	-0.6	-0.5	-0.1
<i>dog</i>	0.7	-0.1	0.4	0.3	-0.4	-0.1	-0.3
<i>houses</i>	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8

Dimensionality reduction of word embeddings from 7D to 2D



document 1

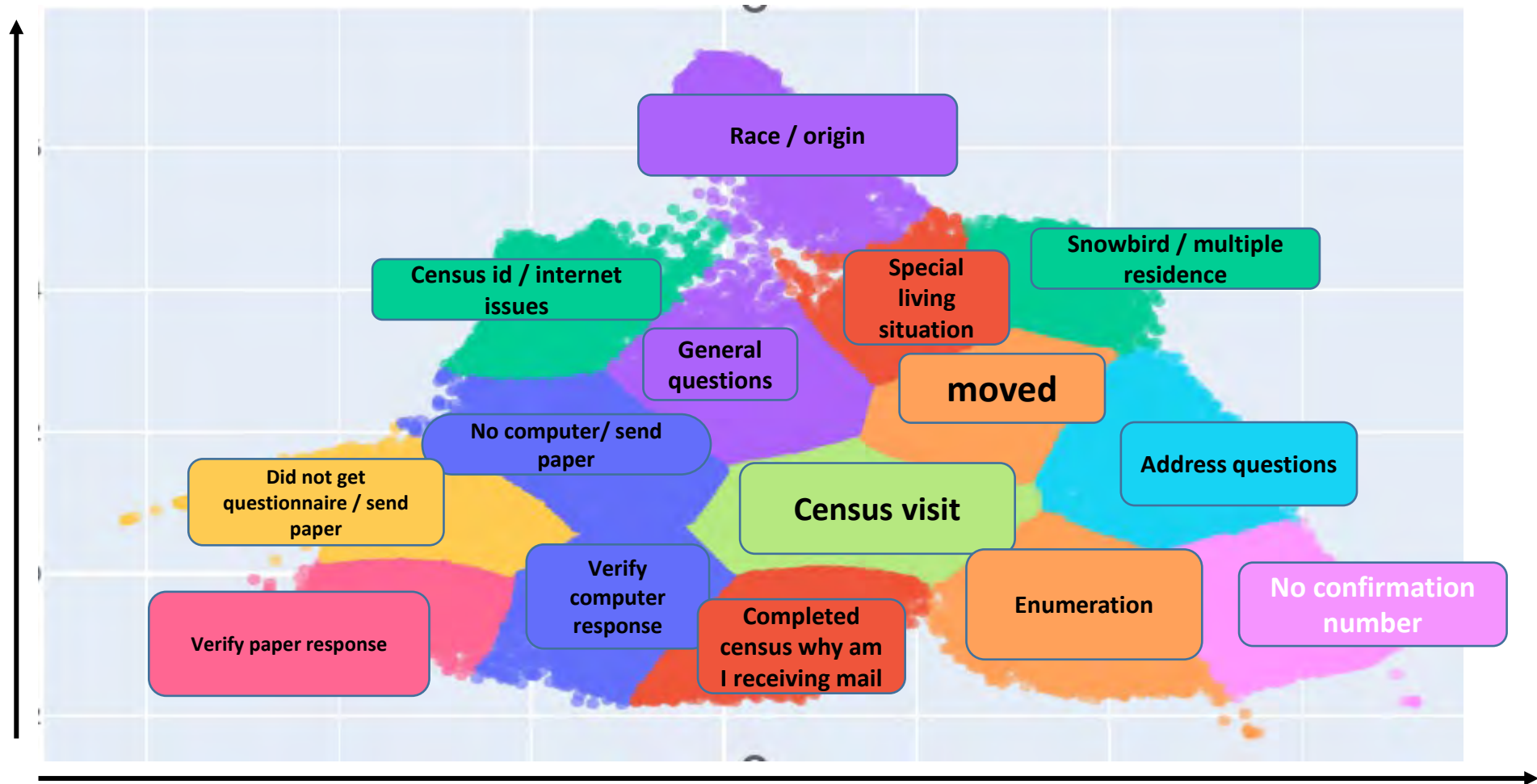
Obama
speaks
to
the
media
in
Illinois



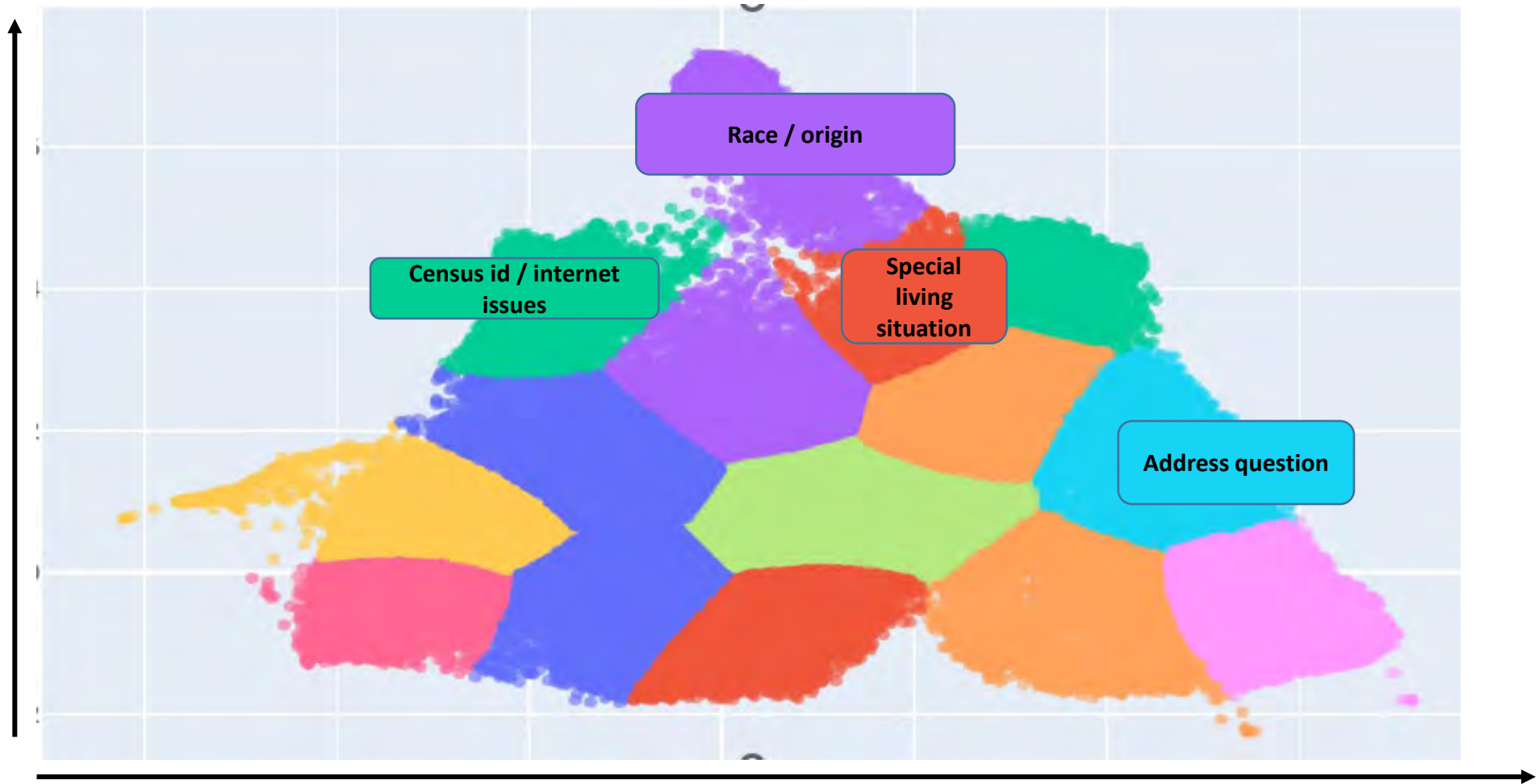
document 2

The
President
greet
the
press
in
Chicago

65K caller transcripts clustered into a 15 Topic Model



65K caller transcripts clustered into a 15 Topic Model



Top Words of sample of topics from the caller

Race / Origin Topic

- Mexican
- Hispanic
 - Race
- Spanish
- American

Census ID/ Internet issues Call Topic

- id
- Tried
- complete
- census
- invalid

Address Questions Call Topic

- passed
- away
- deceased
 - Died
- vacant

Special Living Situation Call Topic

- nursing
- facility
- assisted
- home
- living

Understanding FAQ Coverage via means of FAQ Categories

why doesn't my census id work?

how do i answer the hispanic origin question?

I live in assisted living how do i respond?

where do i find my census id??

how do i answer the race question?

i live on a military base how do i respond?

can you provide me with a new different census id?

i am _____ but i don't see that on the race question what do i do?

i live in an rv, boat, or hotel etc how do i respond?

what do i do if i can't find my census id?

what if i know this person's race but not their origin?

what are group quarters?

Understanding FAQ Coverage via means of FAQ Categories

Census ID FAQ Category

why doesn't my census id work?

where do i find my census id??

can you provide me with a new different census id?

what do i do if i can't find my census id?

Race / Origin FAQ Category

how do i answer the hispanic origin question?

how do i answer the race question?

i am _____ but i don't see that on the race question what do i do?

what if i know this person's race but not their origin?

Special Living Situation FAQ Category

I live in assisted living how do i respond?

i live on a military base how do i respond?

i live in an rv, boat, or hotel etc how do i respond?

what are group quarters?

Which FAQ / FAQ Category maps the best to this Topic?

Census ID FAQ Category

why doesn't my census id work?

where do i find my census id??

can you provide me with a new different census id?

what do i do if i can't find my census id?

Race / Origin FAQ Category

how do i answer the hispanic origin question?

how do i answer the race question?

i am _____ but i don't see that on the race question what do i do?

what if i know this person's race but not their origin?

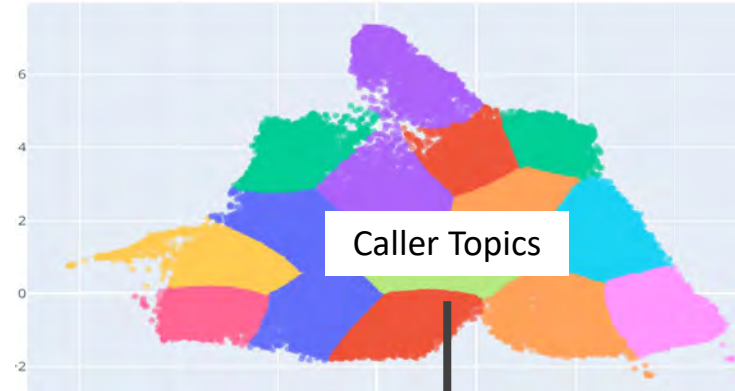
Special Living Situation FAQ Category

i live in assisted living how do i respond?

i live on a military base how do i respond?

i live in an rv, boat, or hotel etc how do i respond?

what are group quarters?



Call Topic #4

- nursing facility
- assisted home living

Most Similar FAQ Category:

Special Living Situation FAQ Category

Using cosine similarity between embeddings to measure semantic similarity

Special Living Situation FAQ Category

I live in assisted living how do i respond?

.66

i live on a military base how do i respond?

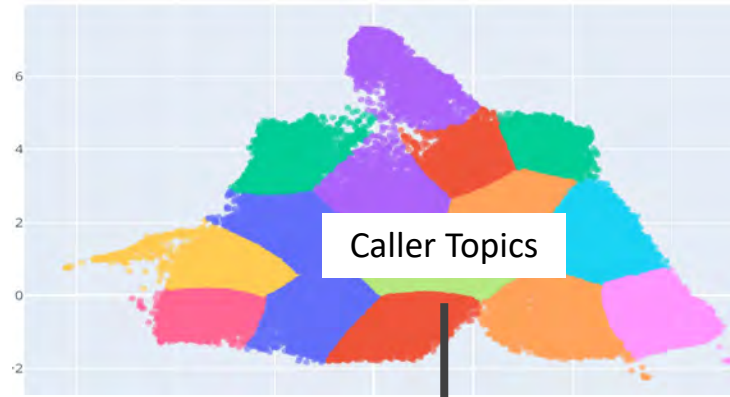
.62

i live in an rv, boat, or hotel etc how do i respond?

.60

what are group quarters?

.55



- Call Topic #4
- nursing
 - facility
 - assisted
 - home
 - living

What is the mean cosine for all the FAQ Categories to this topic?

Census ID FAQ Category

why doesn't my census work?

Avg Cosine for this group: .10

what do i do if i can't find my census id?

Race / Origin FAQ Category

how do i answer the hispanic origin question?

Avg Cosine for this group: .20

what do i know this person's race but not their origin?

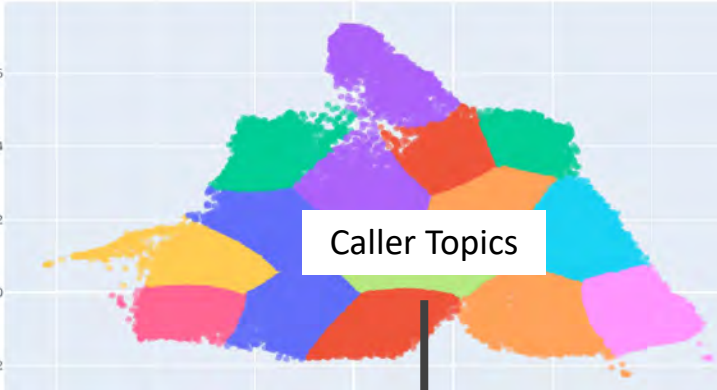
Special Living Situation FAQ Category

I live in assisted living how do i respond?

Avg Cosine for this group: .60

what are group quarters?

AVG COSINE OF ALL FAQ CATEGORIES TO TOPIC #4 : .30



Call Topic #4

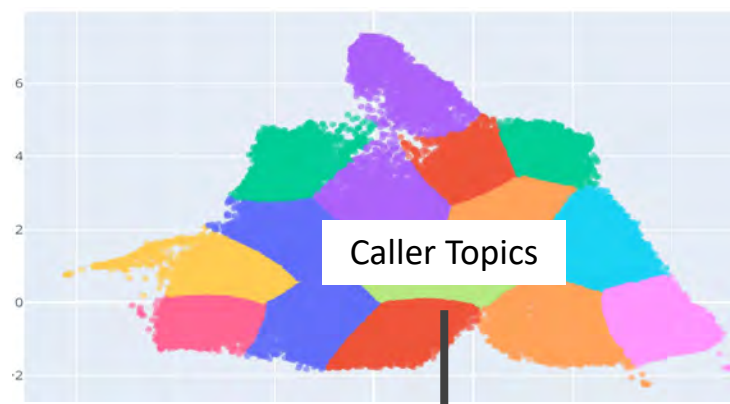
- nursing
- facility
- assisted
- home
- living

Most Similar FAQ Category:

Special Living Situation FAQ Category



What is the mean difference of all the FAQs if we drop the FAQ group of special living FAQs?



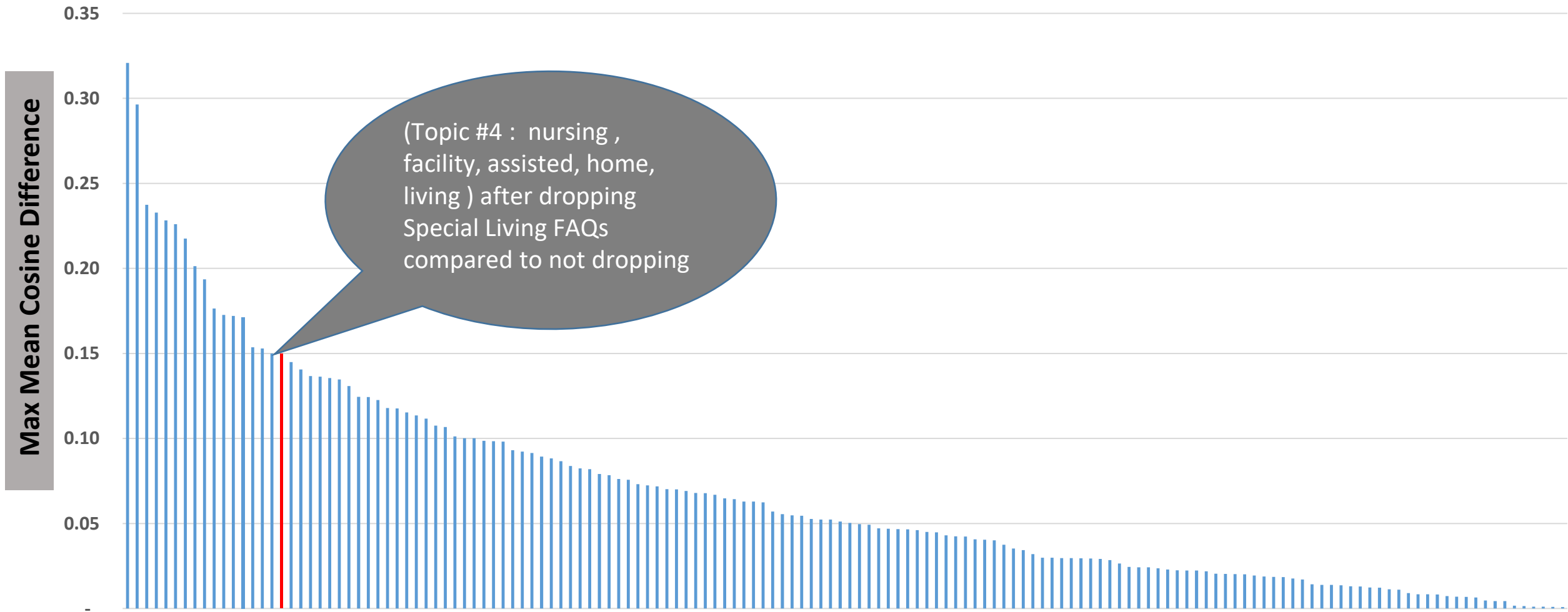
- Call Topic #4**
- nursing
 - facility
 - assisted
 - home
 - living



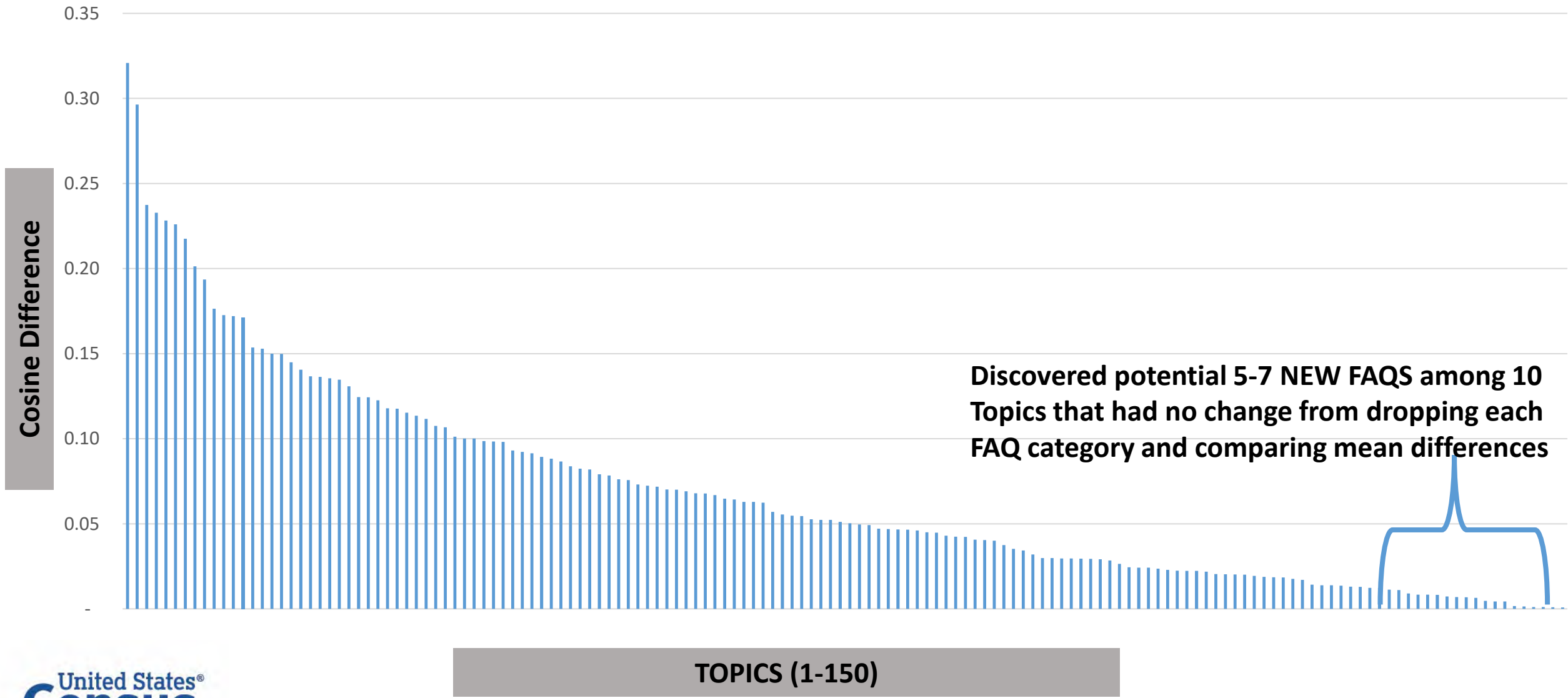
AVG COSINE OF ALL FAQ CATEGORIES TO TOPIC #4 : .15

MEAN DIFFERENCE OF FAQs: .30 - .15 = .15 Difference

150 Topics and their max mean difference after dropping each FAQ Category



150 Topics and their max mean difference after dropping each FAQ Category



The presentation has been reviewed for disclosure avoidance and approved under CBDRB-FY24-CBSM002-035

Takeaways

- We can use ML to help us understand if we have enough reference material to answer different types of conversations.
 - Gather insight from large bodies of text without having a human read through thousands of conversations after clustering them via topic modeling.
- Measuring distances between variables (embeddings) to understand the relationships and then changing the environment (dropping the FAQ category) and analyzing if there was any change from the norm to estimate coverage.
- Topic modeling at a granular level helps estimate FAQ coverage.
 - Estimating that Census had coverage of 96% of types of conversations with their 300 FAQs since after manual review of the ~10 topics that had no change, found potentially 5-7 topics that could benefit from a new FAQ (niche topics).
 - If we can anticipate the same thing for 2030 we know that what we saw in the past, we are prepared for and have a algorithmic approach to do on going evaluations during 2030 operations.

Thank you!

Contact:

monica.puerto@census.gov

Appendix

FAQ Categories and Count of FAQs under each

