Recommendations for the Seattle Police Department

Arnav Garg, Michelle Lee, Teresa Rexin,
Thant Zin Oo

Our Team

Michelle Lee

Teresa Rexin

Arnav Garg

Thant Zin Oo

EDA

Analyze race

-> Geolocation data (Heat map) -> Visualize dangerous neighborhoods - Michelle

Analyze # of fatal (in relation to age, rank, gender, race) - Teresa
Are cases handled by more senior officers (rank or years of experience) more likely to result in a fatality? - Teresa (normalize by number of each officer in dataset)

Sentiment analysis of summaries to see if biased (pick max score of each sentence. Average sentiment of summary))

When officers were reported, if overturned/not disciplined when recommended

-> Who starts violence in crime, affected by police

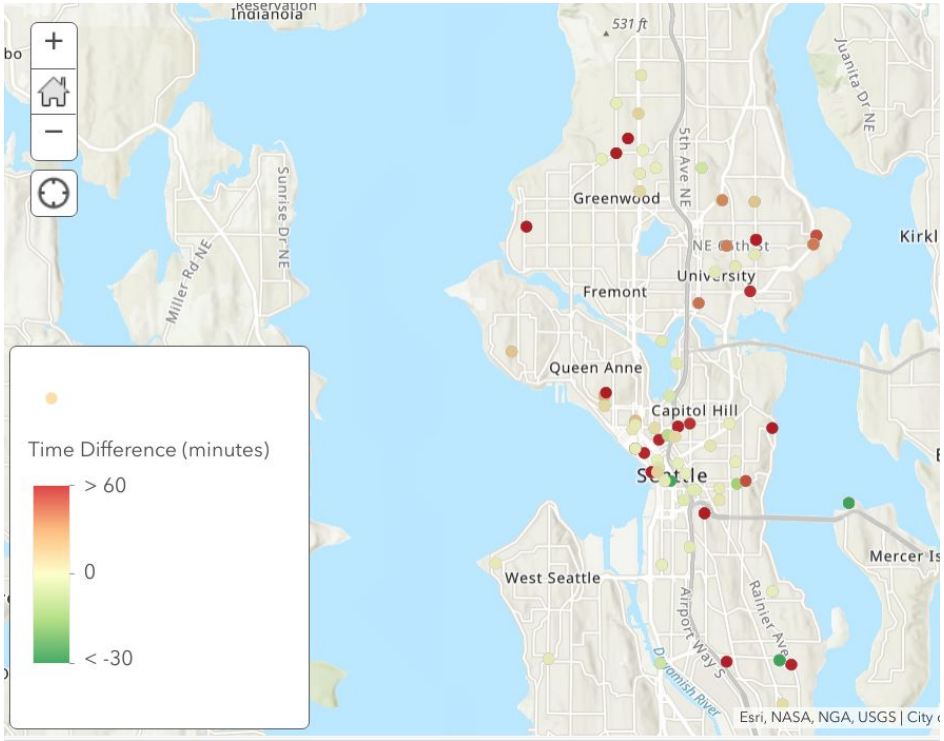
-> if certain race of officers target certain race of subjects [subject to some bias]

How time of day factors into crime frequency - Andy

Reported Time Difference

Used Azure's NER

Red dots indicate a greater time difference between reported time by an officer and the actual time an incident occurred

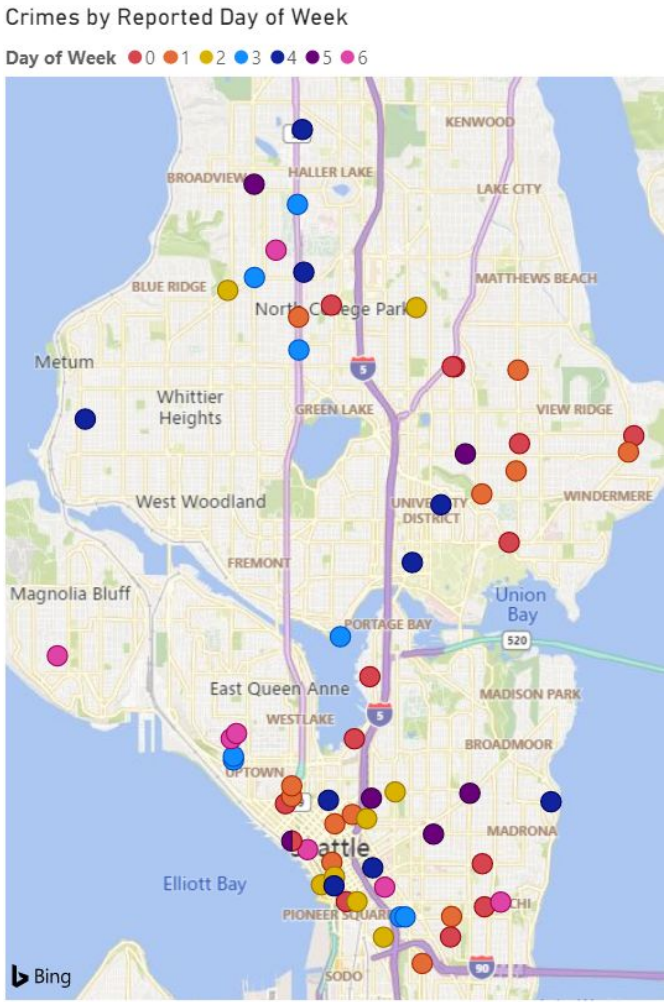


Day of Week

Points colored by day of the week. 0 is Sunday, 1 is Monday, etc.

Regions tend to have Officer Involved Shootings (OIS) on particular days of the week.

The East and Southeast has many incidents on Sunday and Monday.

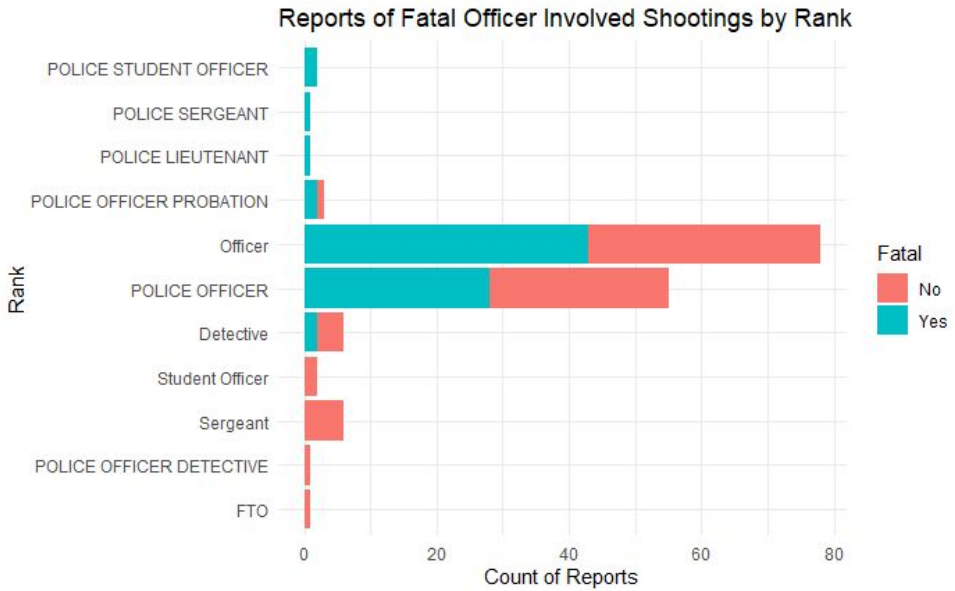


Sentiment Analysis

Used Azure's Sentiment Analysis

- Sentiment based on wording not tone
  - Ex: "The officer observed a small group of individuals in the intersection with two subjects firing handguns."
  - Negative 0.98
- Overall, most summaries were had a negative sentiment

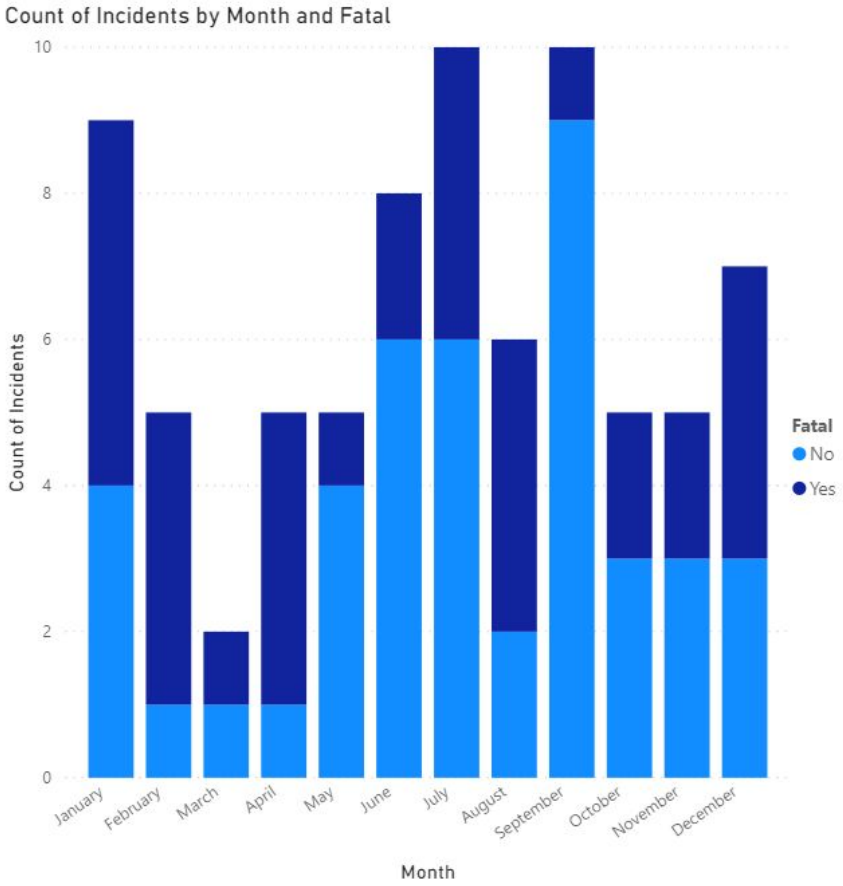
Fatal Officer Involved Shootings by Rank



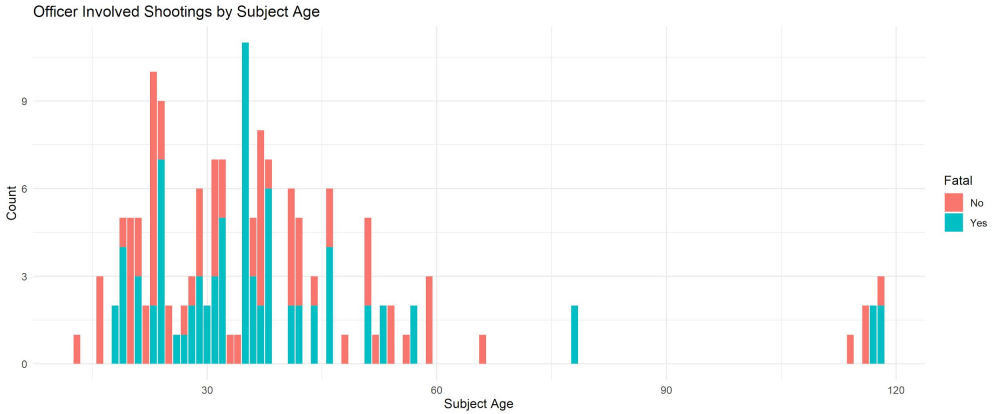
Officers and Police Officers have the largest volumes of fatal Officer Involved Shootings (OISs).

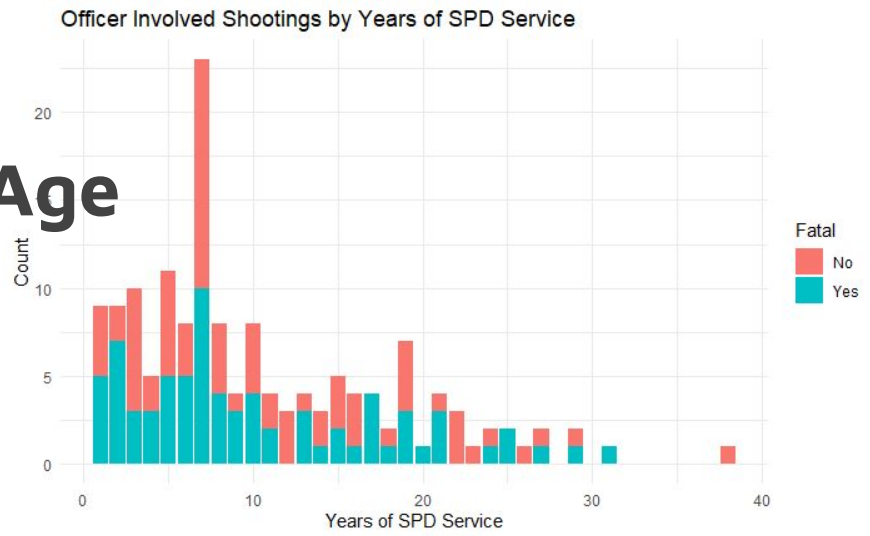
Fatal OIS by Month

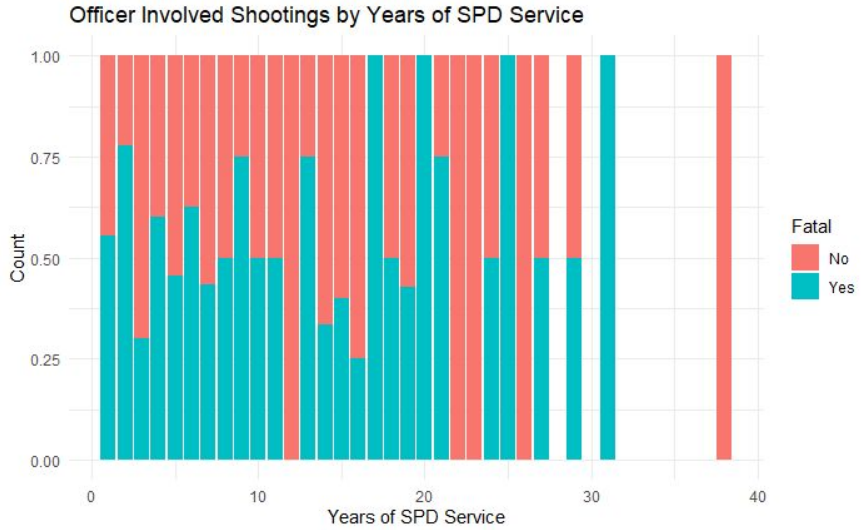
Over the 14 years represented in the dataset, there tends to be more incidents in the summer (June-July) and holiday months (Dec-Jan).



Fatal OIS by Subject Age







Text Analytics and Modeling

Named entity recognition

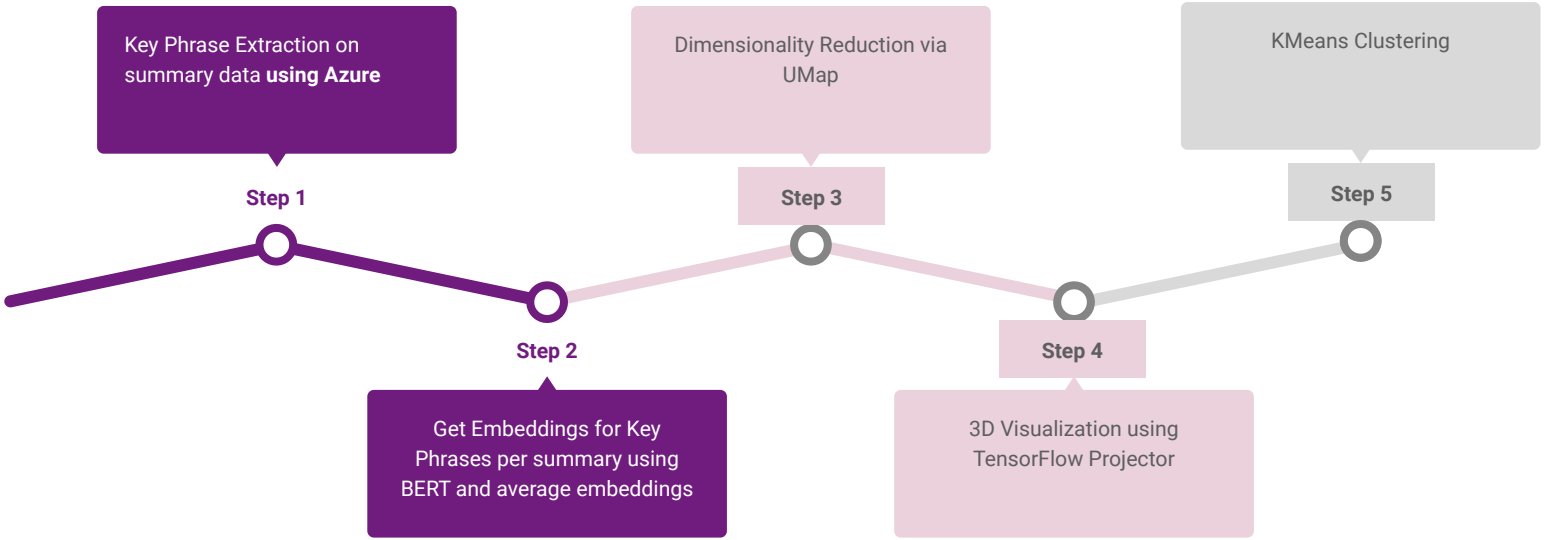
Reported Time Difference (Average time difference for when crime occurred and crime reported)

Average time crime occurred by weapon

More fatalities if longer response time?

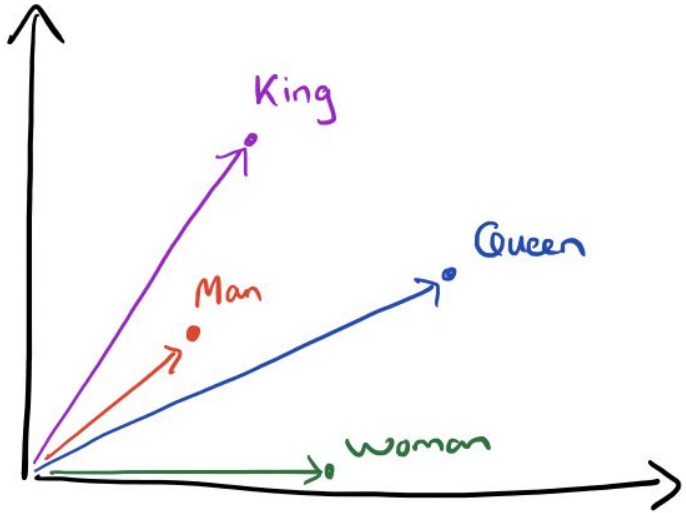
Different locations have different response times

Modeling


Key Phrase Extraction on summary data **using Azure**
Step 1
Step 2
Get Embeddings for Key Phrases per summary using BERT and average embeddings
Dimensionality Reduction via UMap
Step 3
Step 4
3D Visualization using TensorFlow Projector
KMeans Clustering
Step 5

Embedding

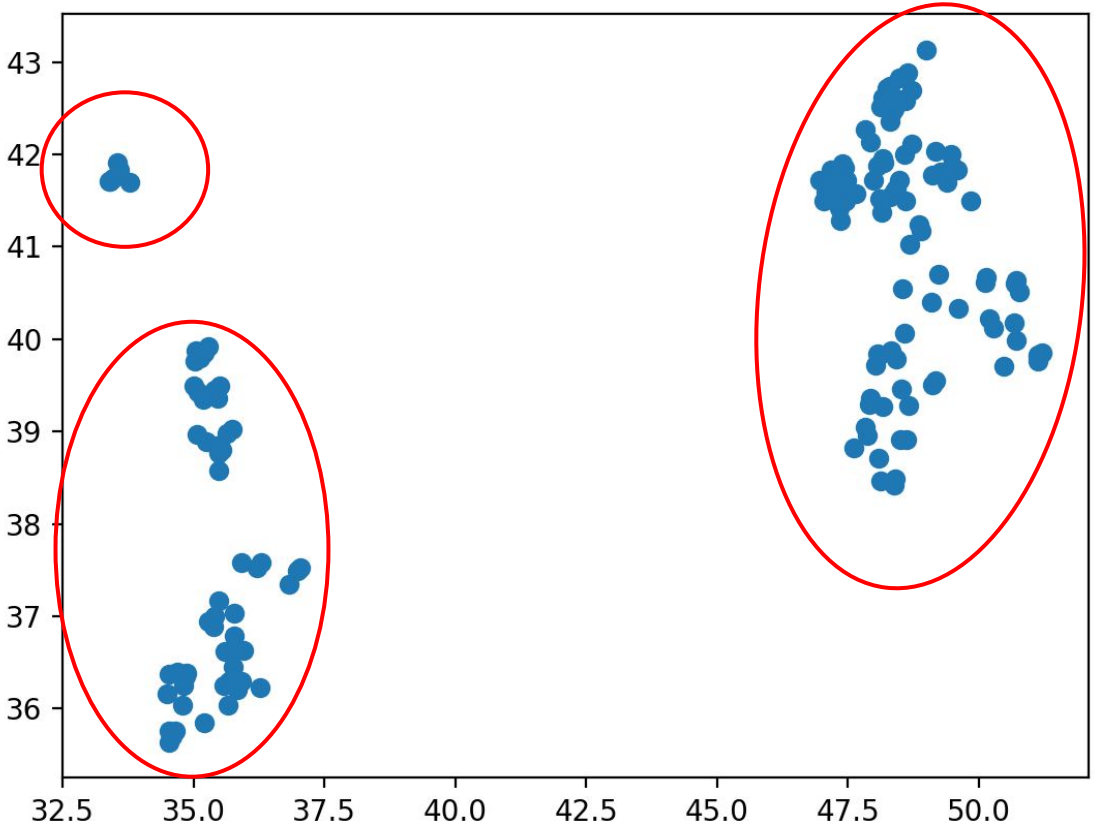
A way of mapping text data to a numeric vector of numbers



King

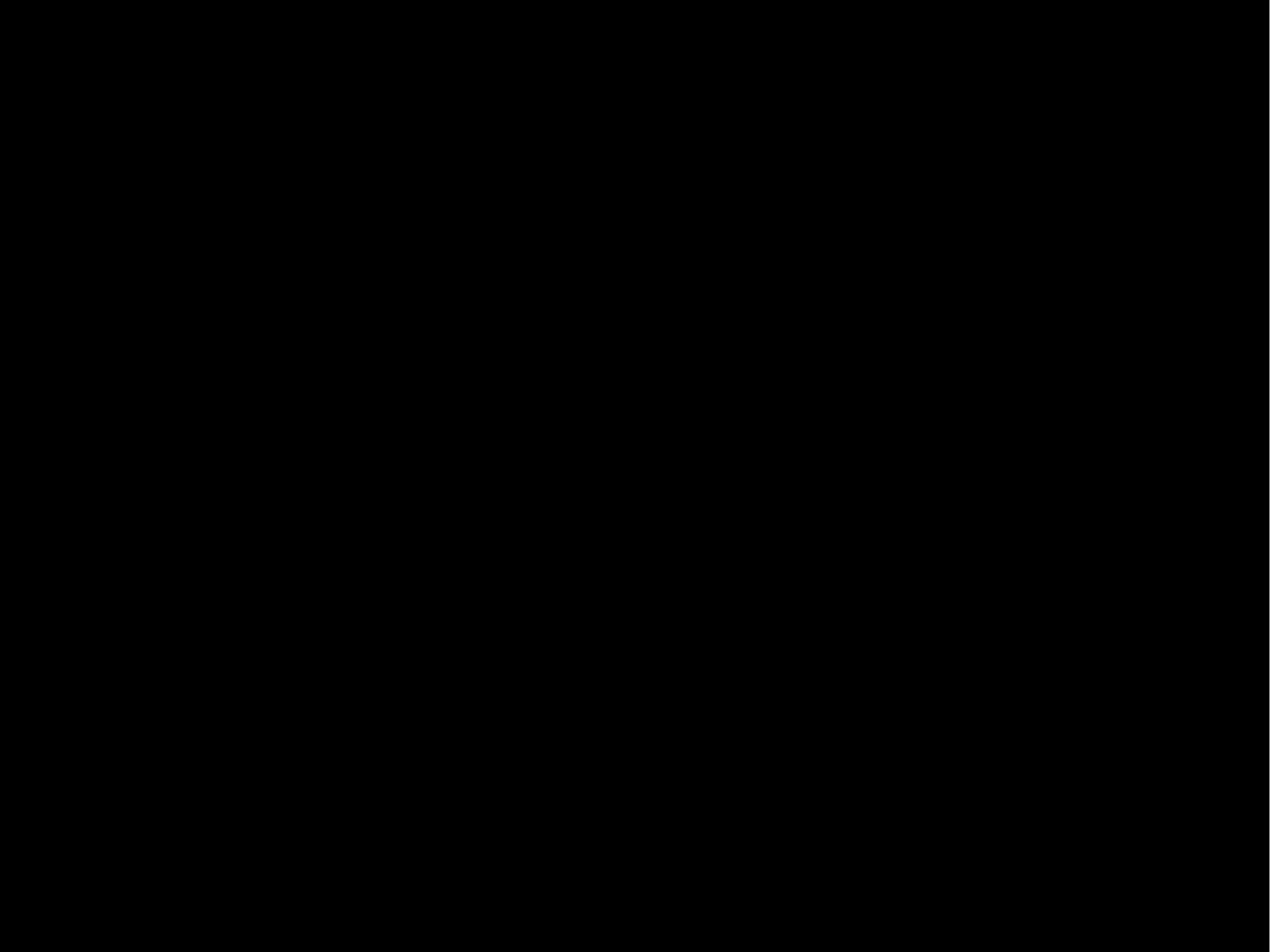
[0.23, -0.143, 0.76, 0.94, 0.002]

2D UMap Clustering by Fatality



Data neatly divides into 2-3 clusters based on summary, indicating a simple structure within the types of crimes reported.

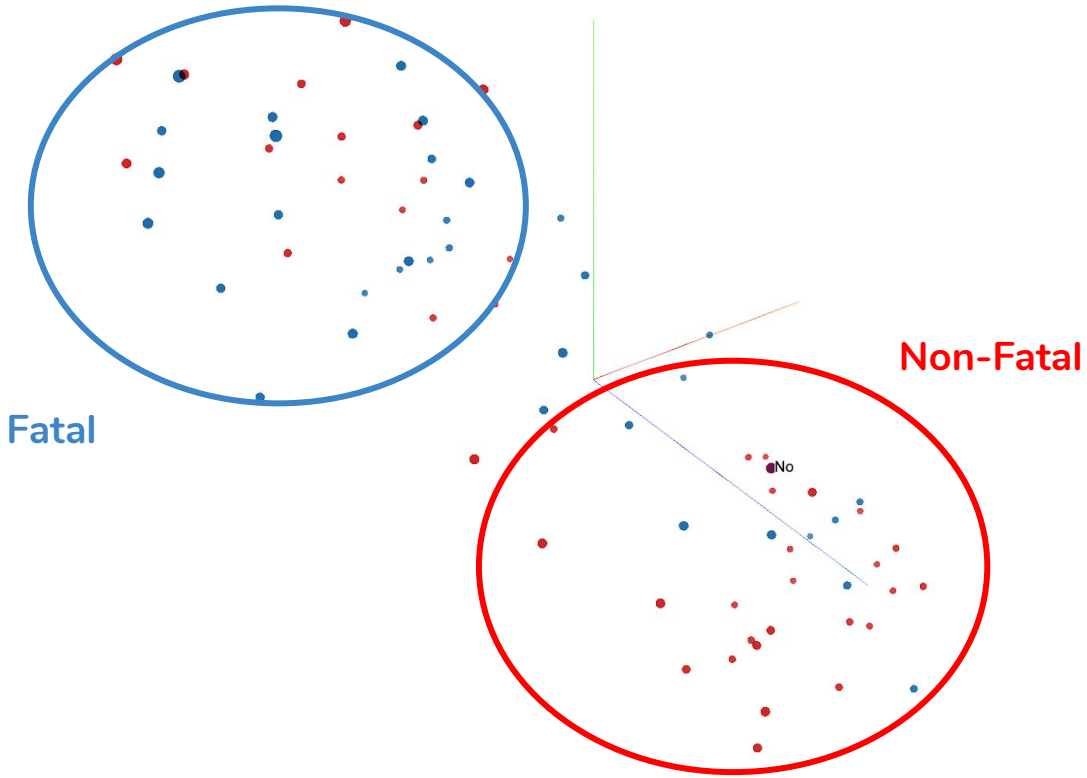
Possibly - Fatal and Not Fatal



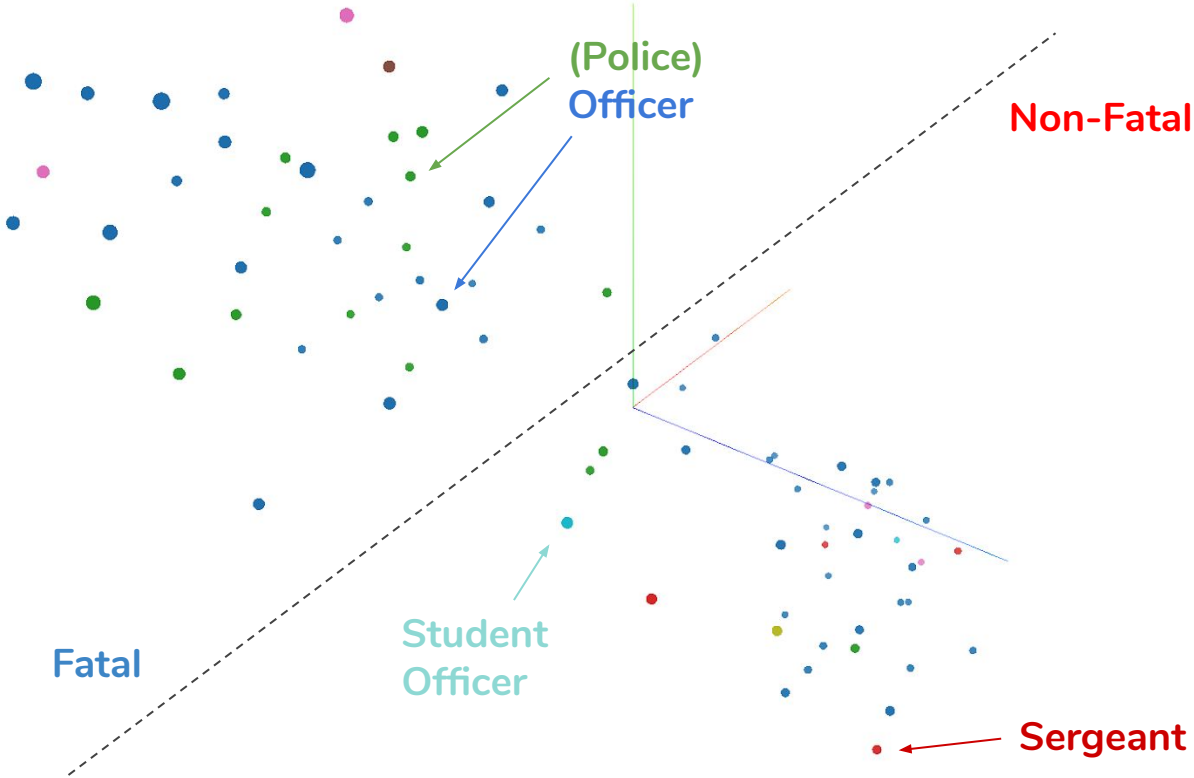
3D Visualization of Embeddings For All 77 OIS Cases

* Created using TensorFlow Projector and UMap dimensionality reduction

3D UMap Clustering by Fatality



3D UMap Clustering by Officer Rank



Final Recommendations

- Assign more officers to the regions specified by the maps
- Be aware of fatalities among different ranks and incidents by month
- Collect more data to confirm recommendations
- Validate the summaries with other data recorded and fill in missing data by leveraging NLP and ML algorithms

Any Questions?