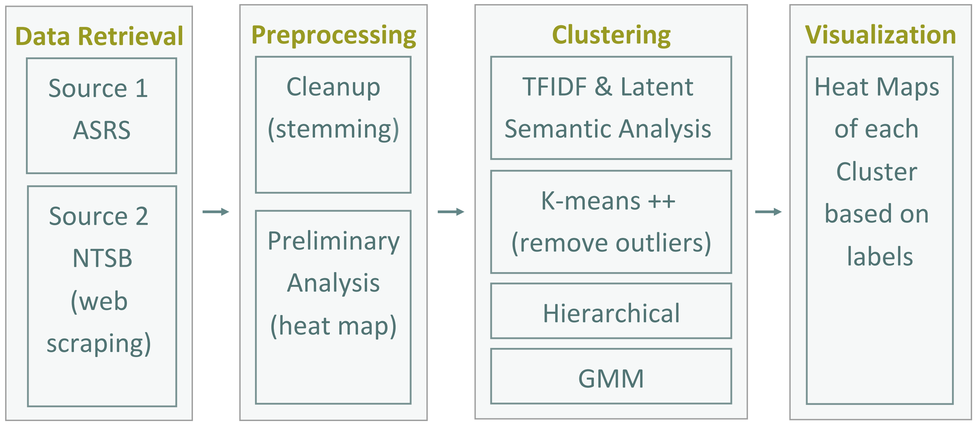
Aircraft Collision Avoidance System

Lingyi Gu
Wei Wei
Haoyu(Jerry) Wu

Objective

In the US, the VFR(Visual Flight Rules) allows pilots almost unlimited freedom to fly anywhere without filing a flight plan. In 2016, there were 179 reported near midair collisions, which led to critical consequences. Our main objectives are to **generate insight from the narratives of collision reports** that help researchers reinforce their collision avoidance strategy to alleviate the situation.



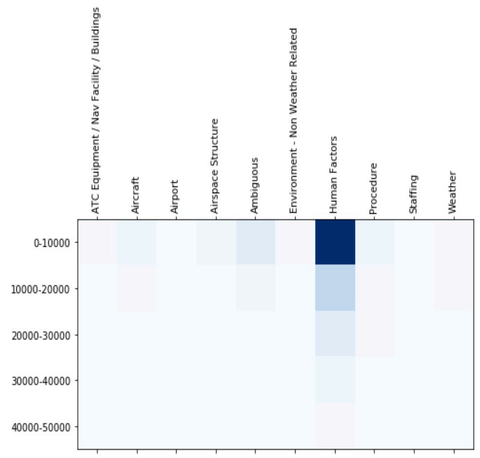
Data Retrieval

NASA's Aviation Safety Reporting System: We downloaded the collision data from the system and mainly utilized the **narratives** and **summaries** of each report, along with the **altitude** and the **relevant distance between aircrafts**.

The U.S. National Transportation Safety Board: We also scraped informative reports on aviation incidents, which serve as a useful aid to supplement our understanding of the data sourced from the ASRS database.

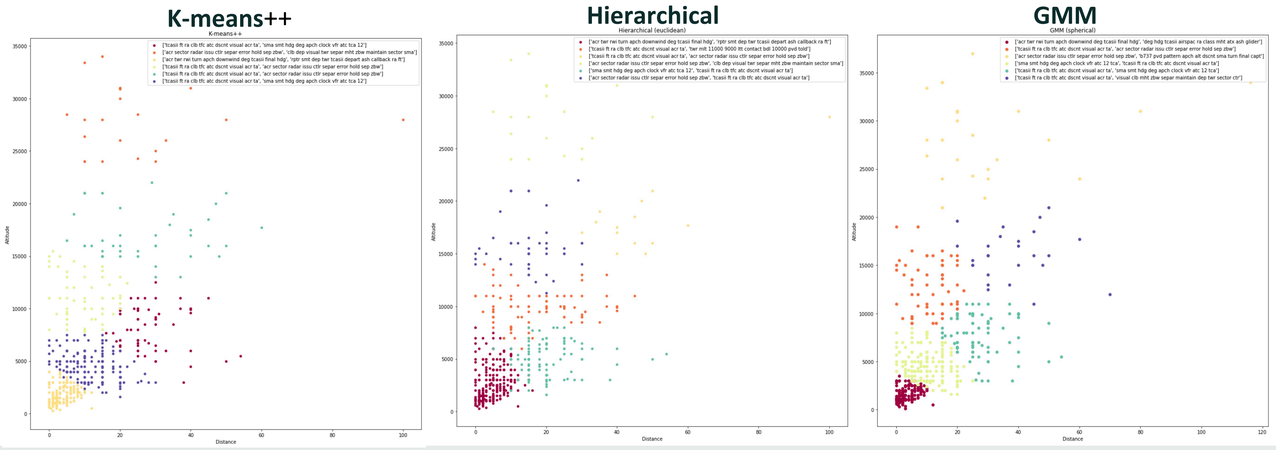
Preprocessing

1. **Stemming**: Eliminate noises in the text.
2. **Stop words**: remove common English words. (E.g. We, are)
3. **Preliminary analysis**: Before analyzing the narratives, we want to learn what factors contribute to collisions the most. From the graph on the right, we can see that **human error** is the most prominent reason for midair collisions. However, we still need to analyze the narratives to see if the finding corresponds to this heat map.



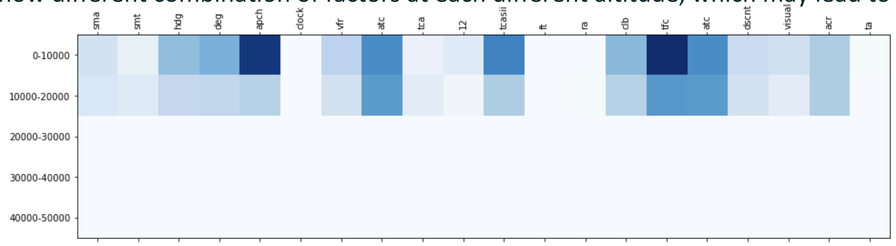
Clustering

After performing latent semantic analysis and singular value decomposition, we clustered the incidents by three metrics. We confirmed that human error is the most common reason for collision. Since K-means ++ has the **highest Silhouette Score**, we chose it for further analysis.



Heat Map of Clustering Results

Clustering gave us an intuition of how different combination of factors at each different altitude, which may lead to a collision. By plotting a heat map for each cluster, we can observe the number of occurrence of these words and determining the important ones.



Conclusion

- The lower the altitude, collision is **more** likely to happen because aircrafts are taking off or landing.
- Among low altitude collision (<20,000) during climbing and descending, the parameters which may cause a collision are: **downwind, tower, true airspeed, aircraft turning, heading mode, traffic control, visual** and collision frequently happens among **single engine piston**.
- Among high latitude collision (20,000 ~ 50,000), collisions likely happen during sector (a portion of an itinerary) or climb, and associates more with **the arrangement of the traffic control center and airport**. **Visual** and **radar** may also play an important role here.

Limitation & Future Work

- Outlier detection improvement.
- Text vectorization improvement. (E.g. remove numbers)
- Analyze based on phase rather than single word

Reference

[1] https://asrs.arc.nasa.gov/
[2] https://www.ntsb.gov/investigations/AccidentReports/Pages/aviation.aspx
[3] https://github.com/mcrovella/CS506-Computational-Tools-for-Data-Science