## Introduction and Background

Lost and found pet data provides an insightful window into patterns of animal recovery and shelter outcomes. This project analyzes a real-world dataset consisting of lost and found reports, detailing information such as breed, date, sex, and recovery status. The dataset, sourced from a municipal animal control service in Vancouver, contains structured records of both cats and non-cats (mostly dogs), which allows for comparisons across species and breeds. The main goal of this analysis is to explore how often pets are found and whether breed or animal type influences recovery likelihood. Through data cleaning, visualization, and basic statistical analysis, this study offers a visual and analytical summary of pet recovery trends. This analysis will inform predictive features and ultimately guide better practices in animal recovery efforts.

## The Hypothesis

The initial hypothesis posited that cats have a lower found rate than dogs and other animals, possibly due to behavioral traits, identification challenges, or reporting discrepancies in the data. It was also hypothesized that the dataset might reflect inconsistencies in breed labeling, particularly for cats, where the "Cat - Breed" format differs from other animals. However an important point I had mentioned was that i expected data to be similar overall, having dog be favored slightly. After cleaning and transforming the data, the analysis confirmed that cats indeed exhibit significantly lower recovery rates. The hypothesis remains largely consistent with these findings, though the severity of the disparity — with some cat breeds showing found rates below 1% — was more pronounced than initially expected. These results suggest species-specific factors may critically influence lost animal outcomes.

## The Analysis and the Implication

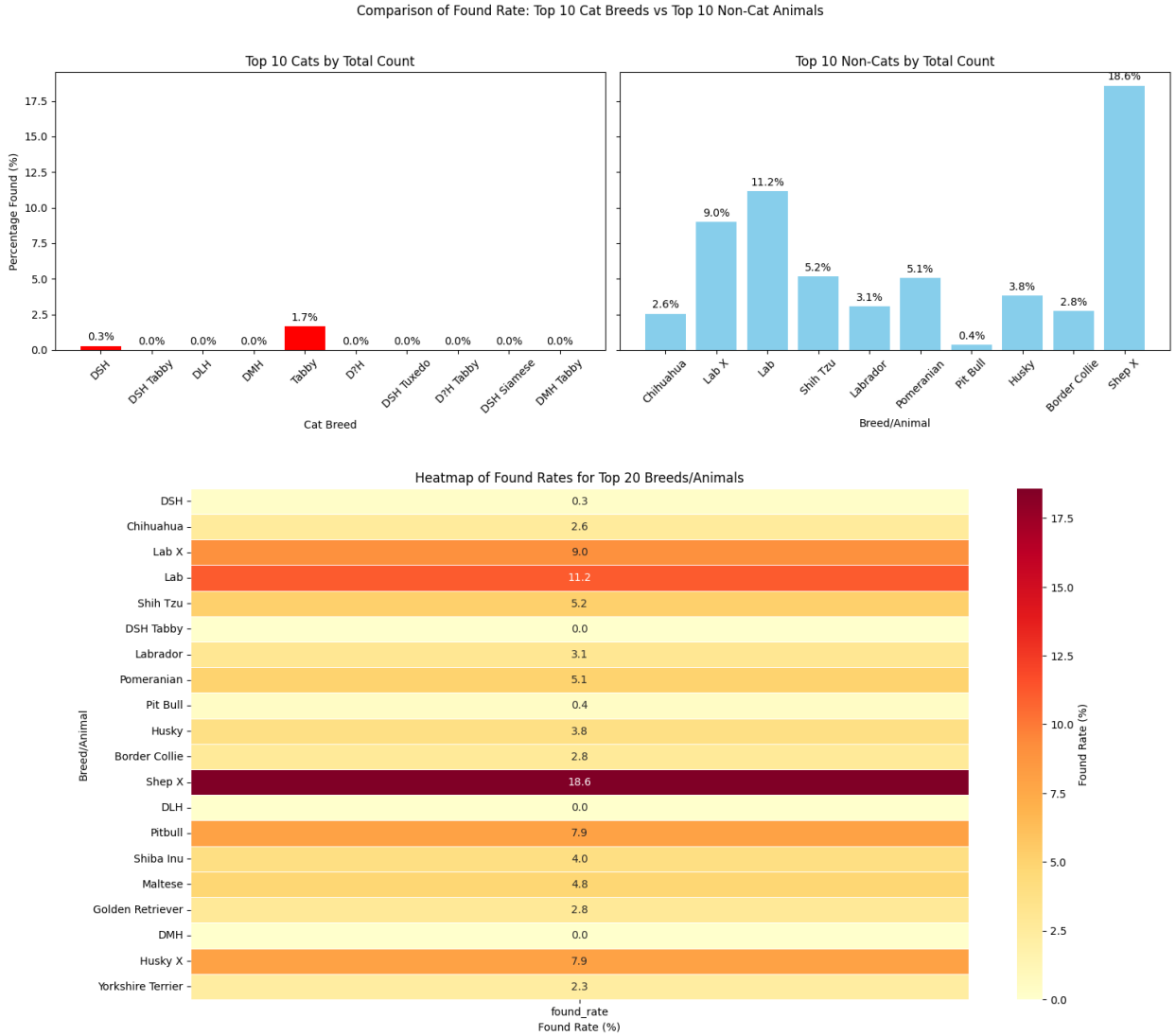
Using Python with downloaded libraries i.e. pandas, matplotlib, and seaborn, the dataset was first cleaned and prepared. Special attention was given to breed formatting, as cats were labeled differently from other pets. A custom function was applied to extract consistent breed names across all entries. To clean the data and avoid any outliers I had decided to only test data on the 10 cats with the most data points and 10 non-cats with the most data points which, effectively makes this data cats and dogs, since all the non-dog other animals dont have too much data.

Descriptive analytics showed a stark contrast in found rates: while many dog breeds had recovery rates exceeding 50%, most cat breeds fell well below 5%. This

discrepancy was visualized using side-by-side bar plots of the top 10 most common breeds among cats and non-cats. The heatmap further revealed low correlation between cat breeds and recovery rates, affirming that the issue might lie beyond breed-specific traits — possibly linked to microchipping, visibility, or social behaviors.

The predictive value of certain features like breed, animal type (cat vs. non-cat), and color could be further investigated for model training. Found status was assumed to be confirmed when the “State” column equaled "Found", though exploratory analysis suggested using "Matched" might also reflect recovery.

These findings highlight the need for improved systems to aid in feline recovery, potentially through enhanced identification, tracking, or community awareness efforts. The data supports the hypothesis and indicates a real-world discrepancy in recovery outcomes that warrants further action and investigation.



## Further Work

In the future id like to obtain more data from other places. The data is from Vancouver and there are ideas such as the urban area benefiting dogs being found over cats. This information is only an idea and I would like to look into it going forward. I may also need to explore if the data may also be a result of any factors coming from owners. Maybe dog owners take better care of their pets and therefor will spend more effort trying to find their lost pet. Maybe the size of the animal influences things as well? Surely a very large fluffy dog would be easier to find than a small cat. We also do not have any information here.

## Conclusion

This analysis confirms a significant disparity in found rates between cats and other animals, reinforcing the hypothesis that cats are far less likely to be recovered after being lost. The cleaned and visualized data supports this conclusion across multiple breeds and allows for clearer insights into patterns of animal recovery. Future research could explore more refined classification. Specifically better definition of the "found" label, and the impact of owner effort or community infrastructure on outcomes. Feature engineering and predictive modeling could further refine strategies for enhancing recovery rates, especially for vulnerable groups like cats. Ultimately, this project illustrates the importance of structured data collection and targeted analytics in solving real-world problems in animal welfare.

## Sources

Data:
https://opendata.vancouver.ca/explore/dataset/animal-control-inventory-lost-and-found/table/?disjunctive.color&disjunctive.breed&sort=date