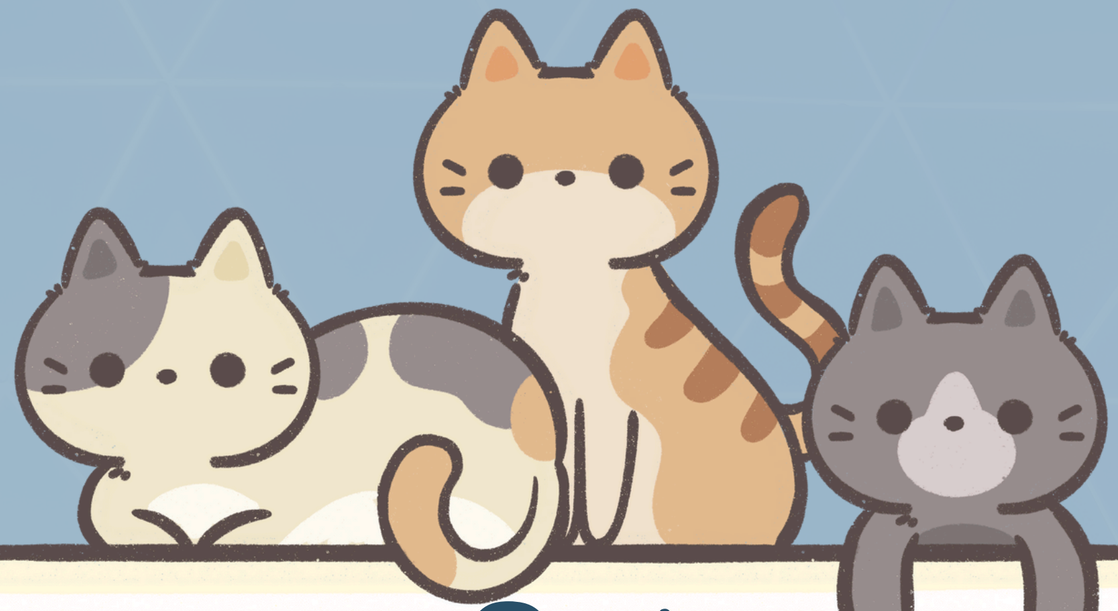


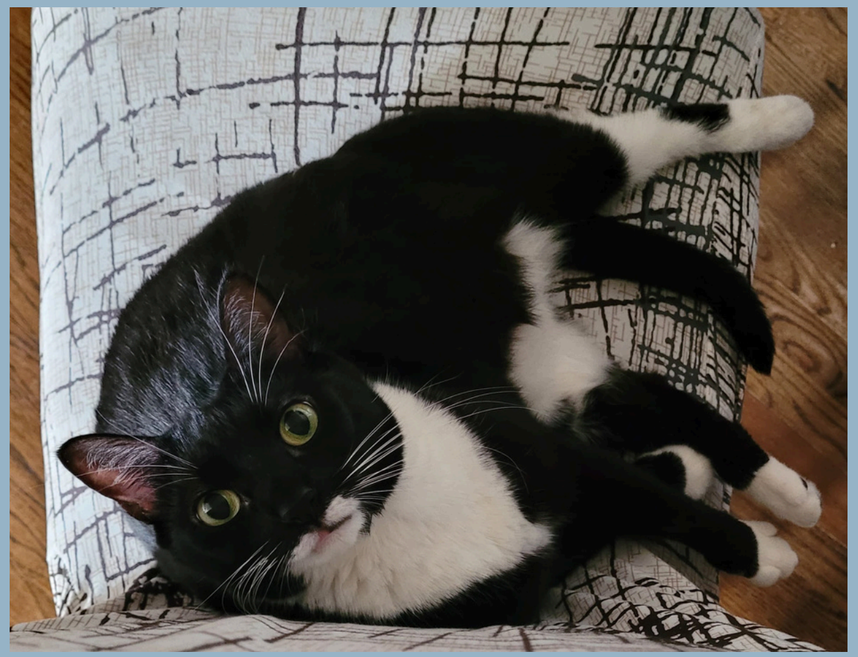
Pet Recommender System for Cats in NYC

By Sangeetha Sasikumar, Capstone Project

Research Question

Can content-based similarity methods successfully capture meaningful relationships between pets using only profile attributes?

Why conduct this research?



Data Collection & Data Cleaning

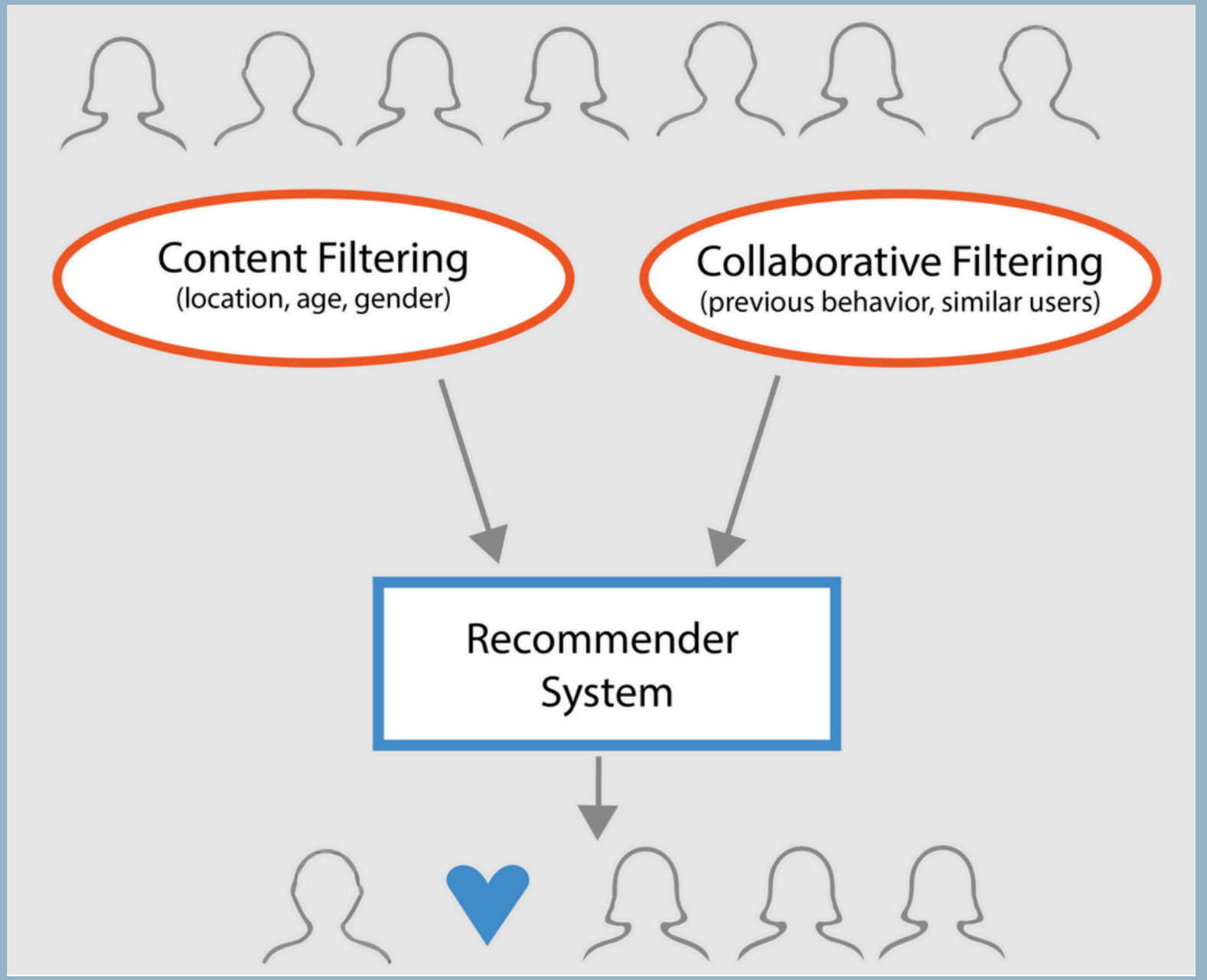
```
            val = label.parent.find_next_sibling()
            if val:
                return val.get_text(strip=True)
        return None

    def extract_gender_color(text):
        match = re.search(
            r"I\s+am\s+a\s+(\w+),\s*([a-z\s]+?)\s+[A-Z]",
            text,
            re.IGNORECASE,
        )
        if match:
            return match.group(1).strip(), match.group(2).strip()
        return None, None

    def extract_breed(text):
        match = re.search(
            r"I\s+am\s+a\s+\w+,\s*.*?\s+(.*?)(?:\.)?$",
            text,
            re.IGNORECASE,
        )
        return match.group(1).strip() if match else None
    def clean_breed_prefix(breed):
        if not breed:
            return None

        # Remove leading "and <color(s)>" phrases
        breed = re.sub(
            r"^(and\s+)?(white|black|brown|gray|grey|orange|tan|cream|buff)(\s+and\s+(white|black|brown|gray|grey|orange|tan|cream|buff))*\s+",
            "",
```

Data Source: Sean Casey Animal Rescue's cat profiles & my Survey

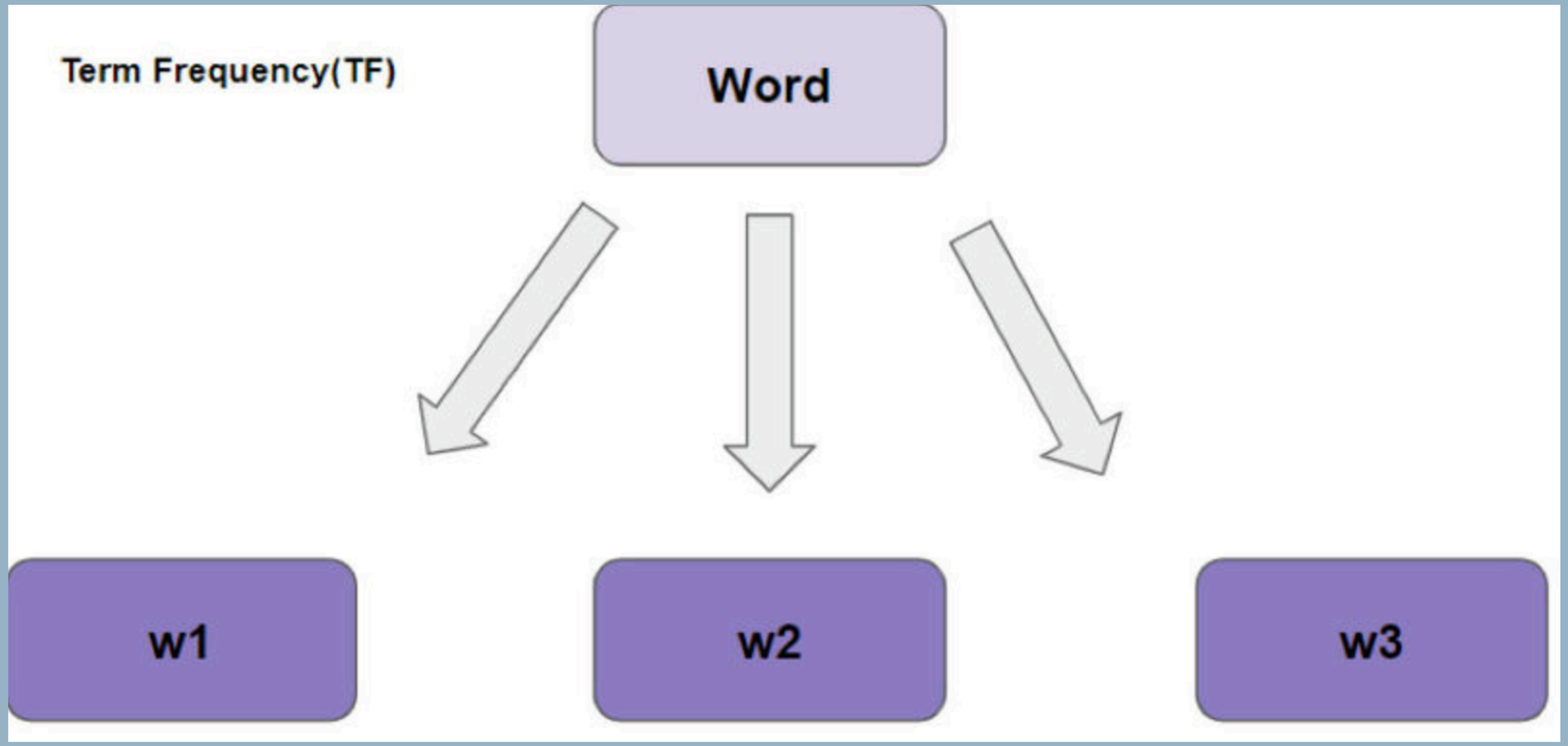
Short haired

Medium haired

Long haired


Content Filtering
(location, age, gender)
Collaborative Filtering
(previous behavior, similar users)
Recommender System

TfidfVectorizer



Content based filtering

Harley

female, white Domestic Shorthair mix.

1 year and 1 month old.

Milkyway

female, white and black Domestic Shorthair mix.

14 weeks old

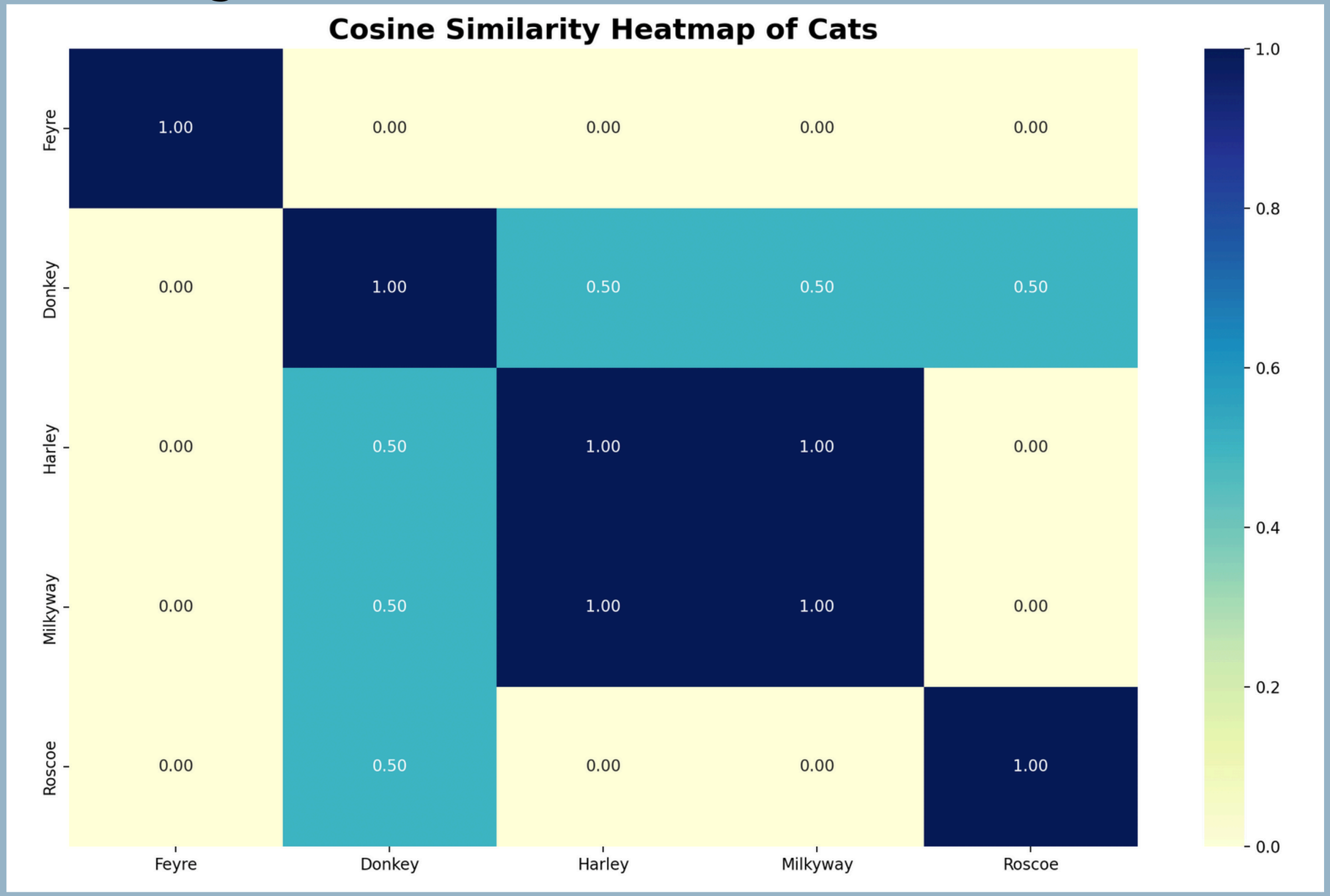
Elain

female, white and brown Domestic Shorthair mix.

7 months old

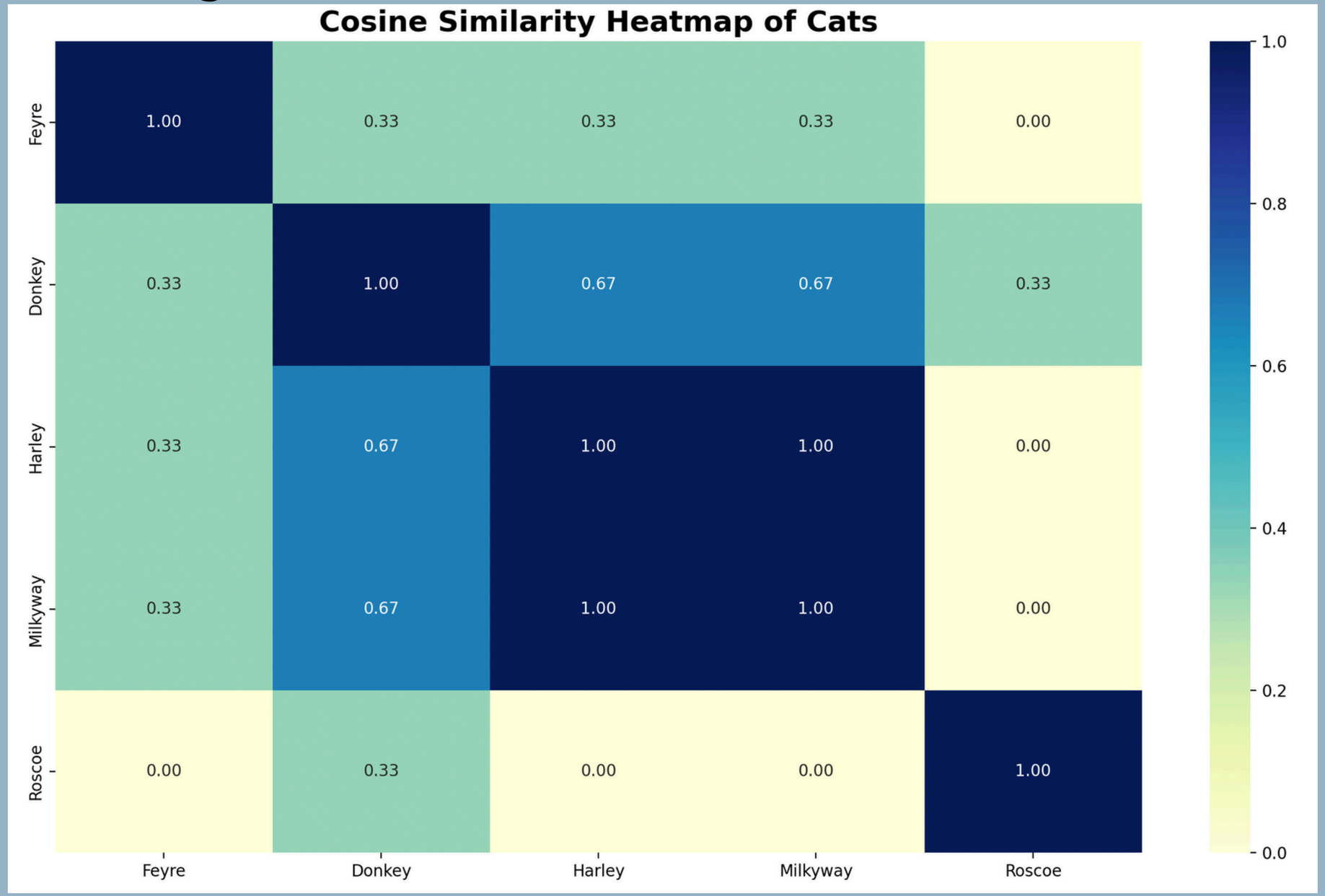
Data Modeling

Breed,
color



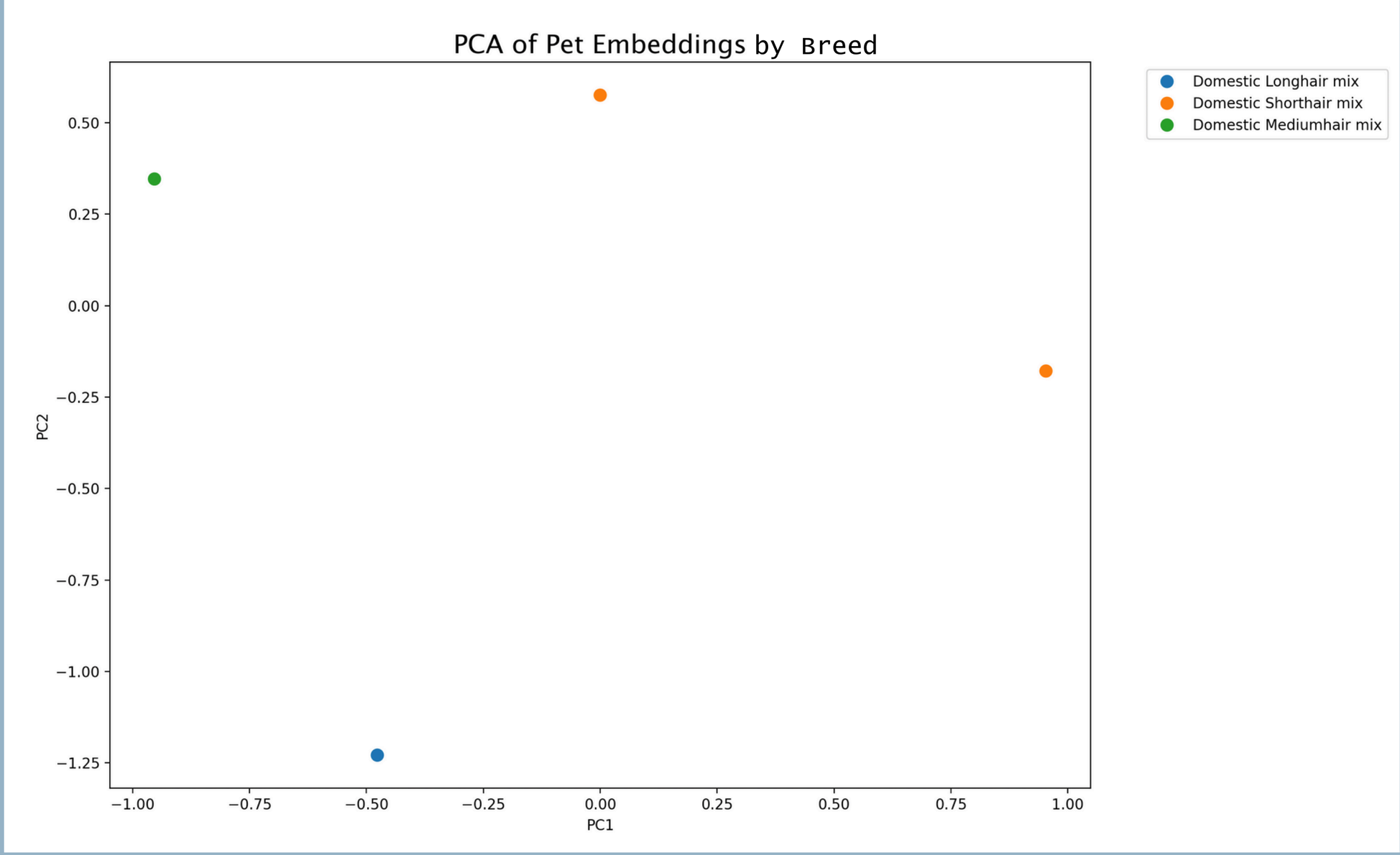
Data Modeling

Breed,
color,
gender



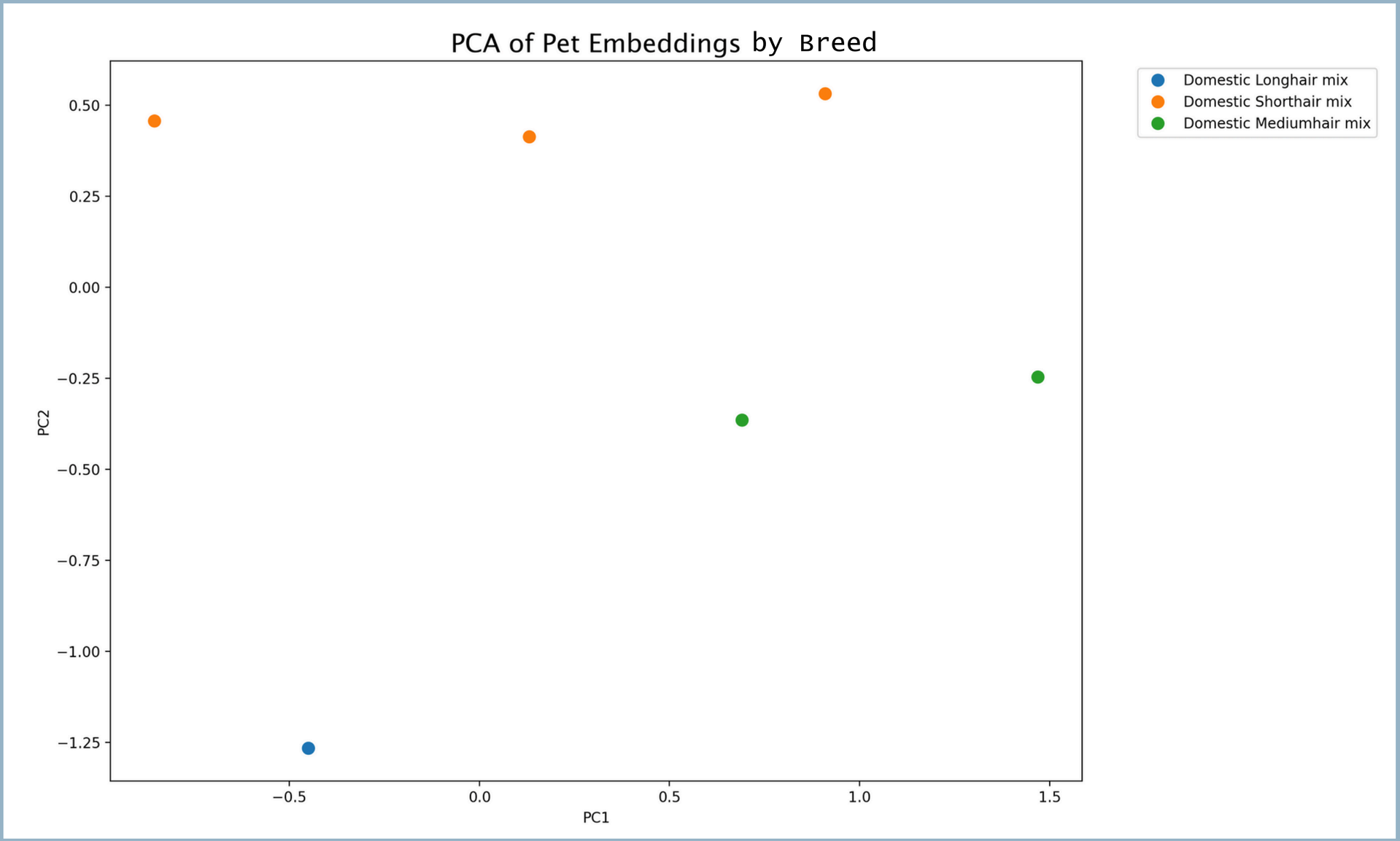
Data Evaluation

Breed, color



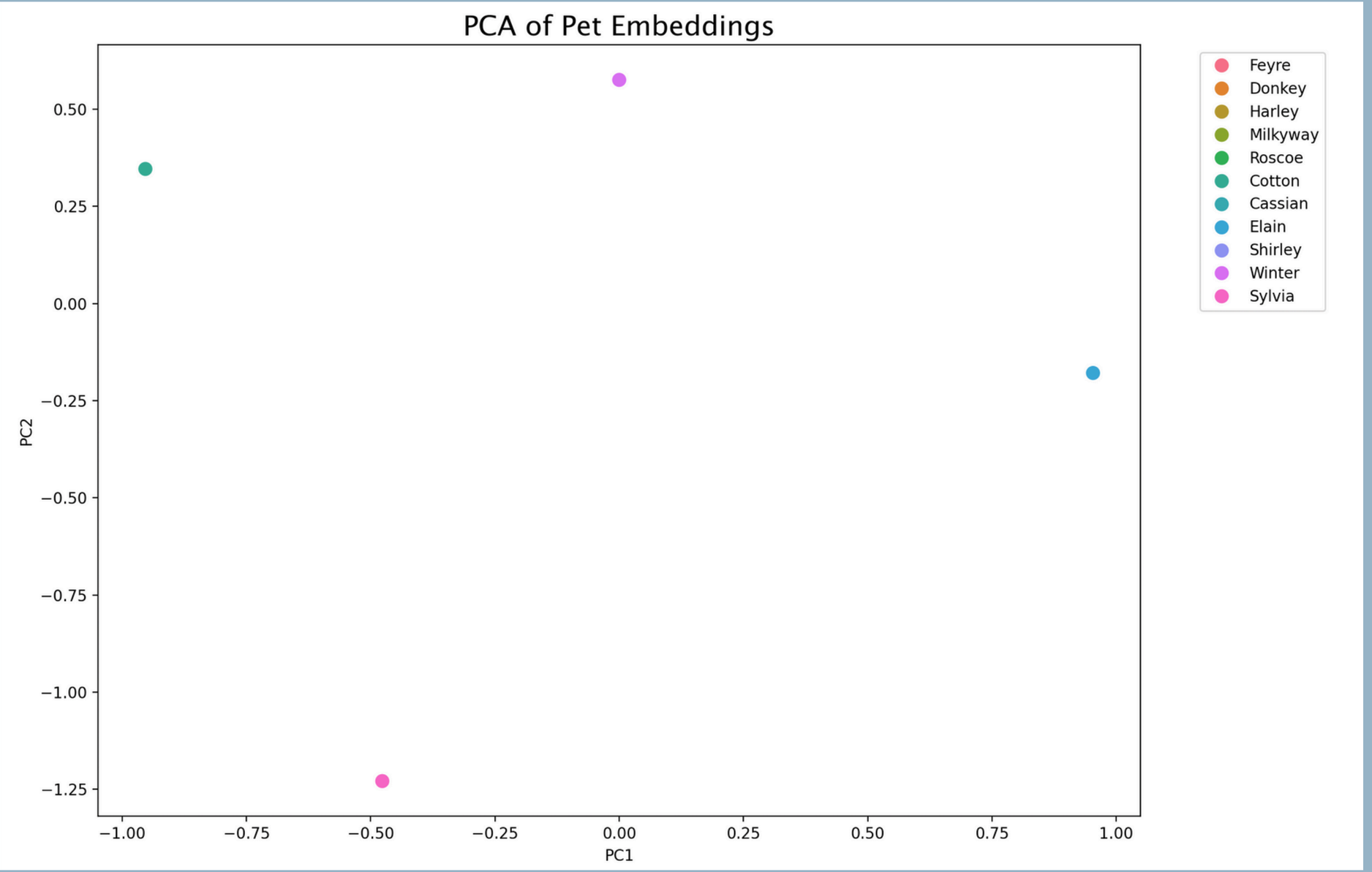
Data Evaluation

Breed,
color,
gender



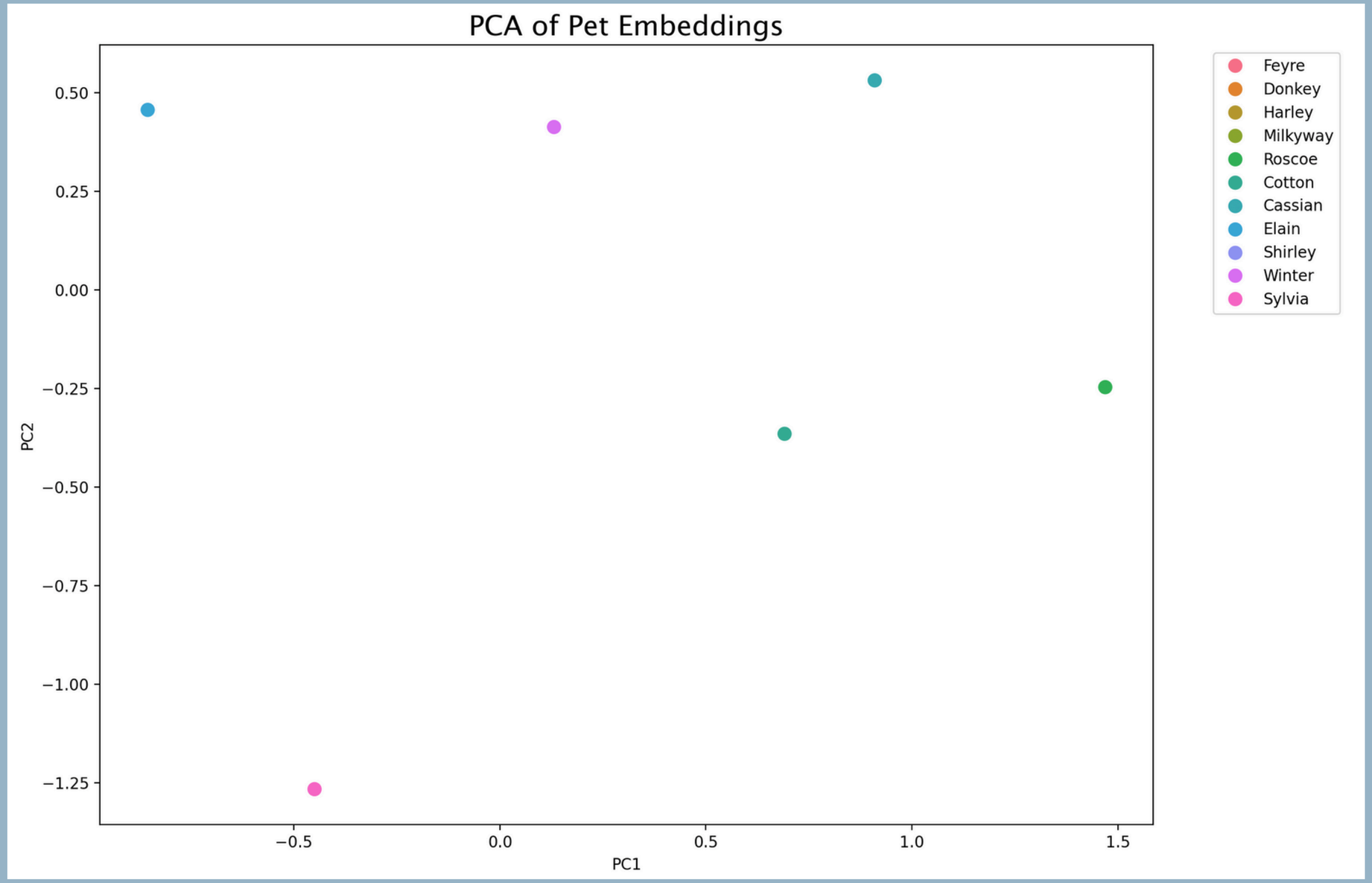
Data Evaluation

Breed, color

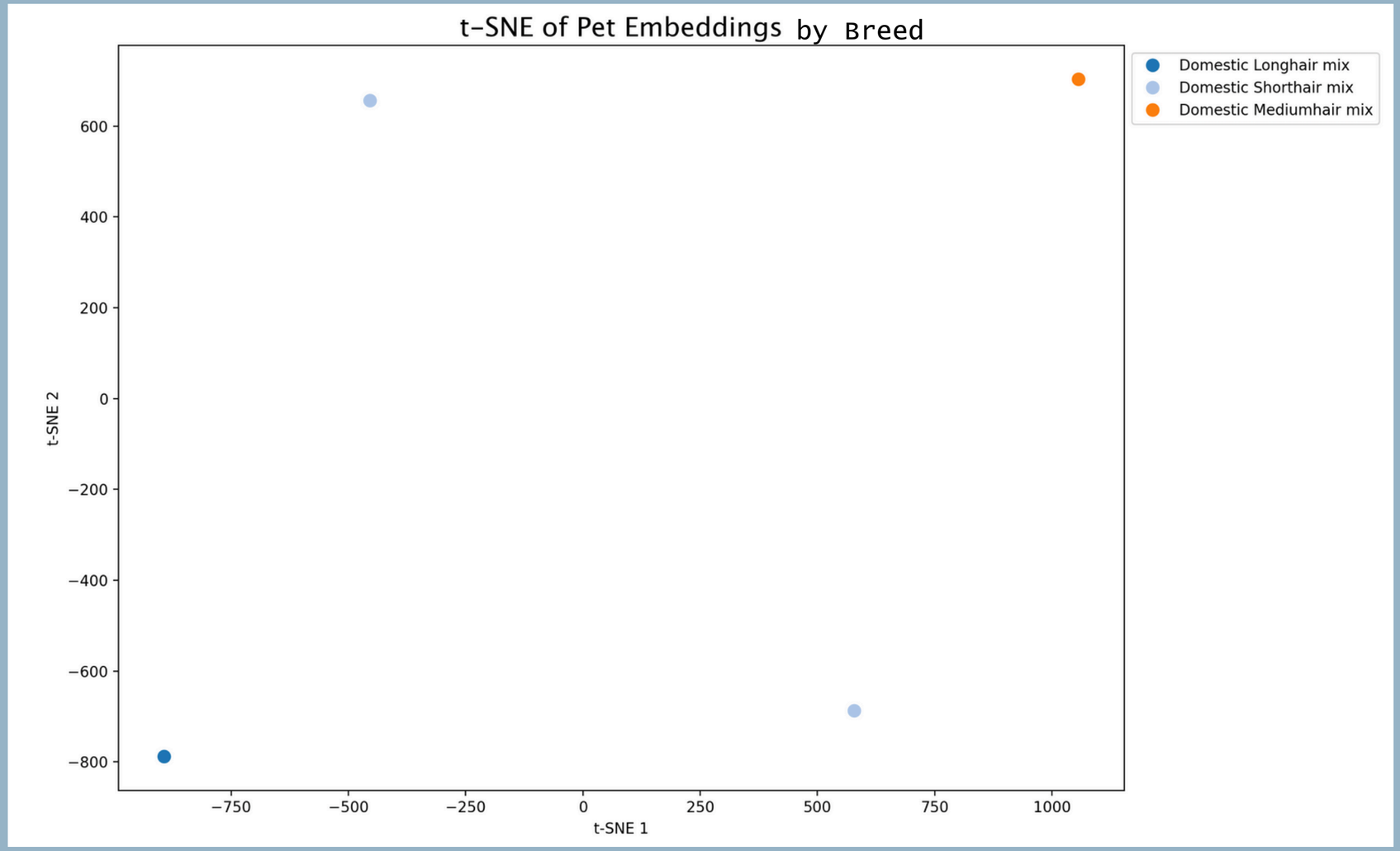


Data Evaluation

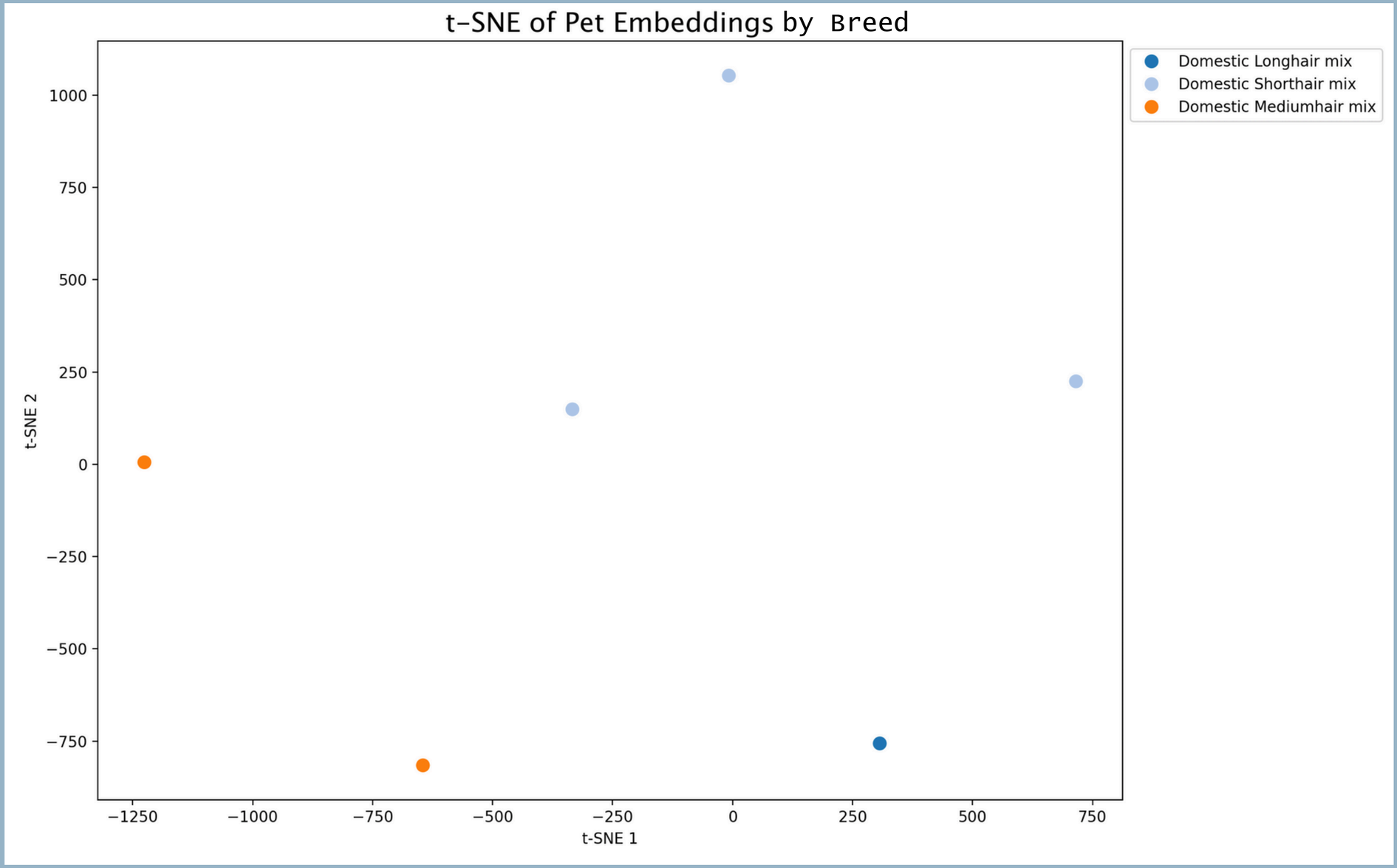
Breed,
color,
gender



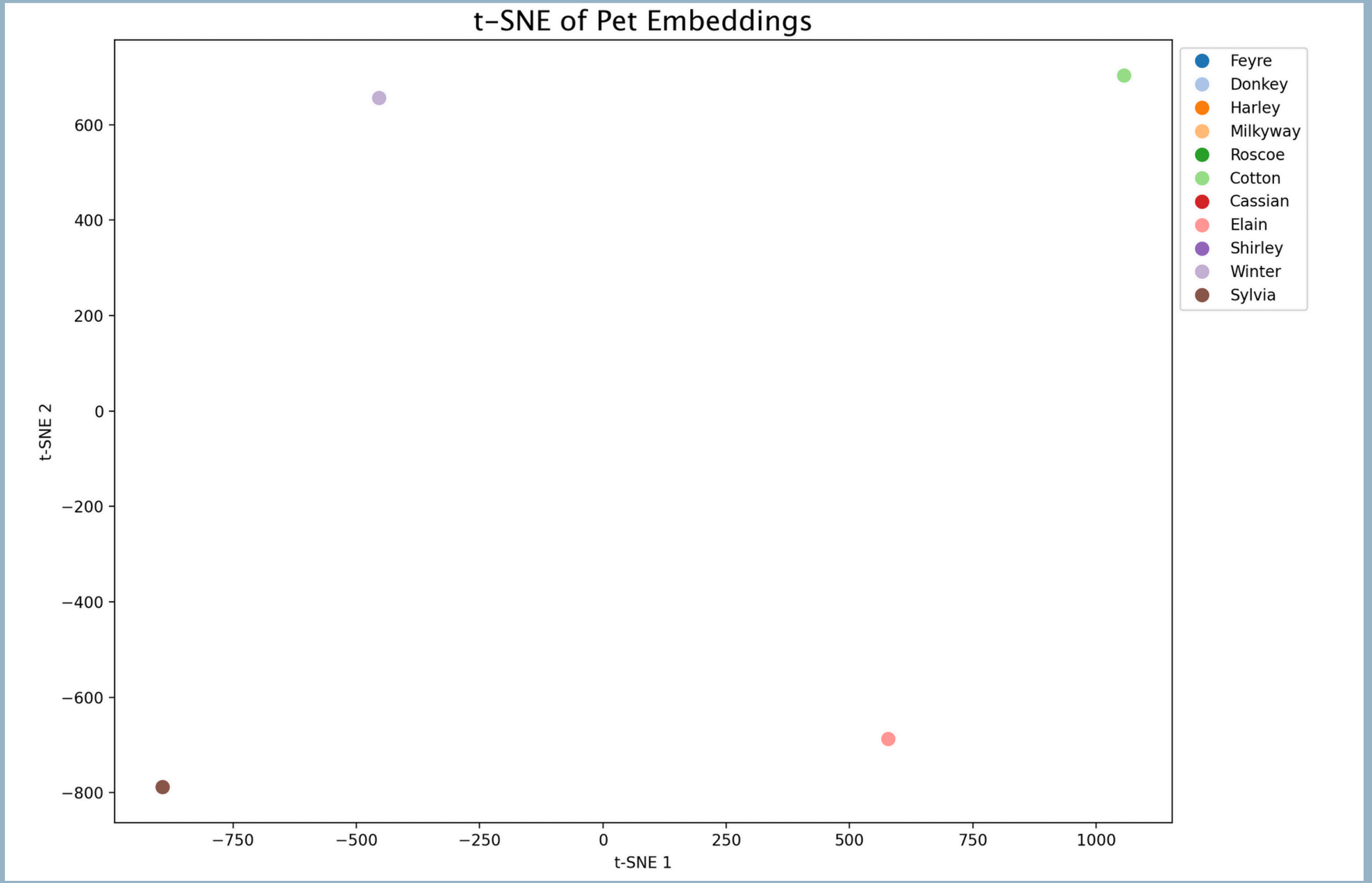
Breed,
color


t-SNE of Pet Embeddings by Breed
Domestic Longhair mix
Domestic Shorthair mix
Domestic Mediumhair mix
t-SNE 2
600
400
200
0
−200
−400
−600
−800
−750
−500
−250
0
250
500
750
1000
t-SNE 1

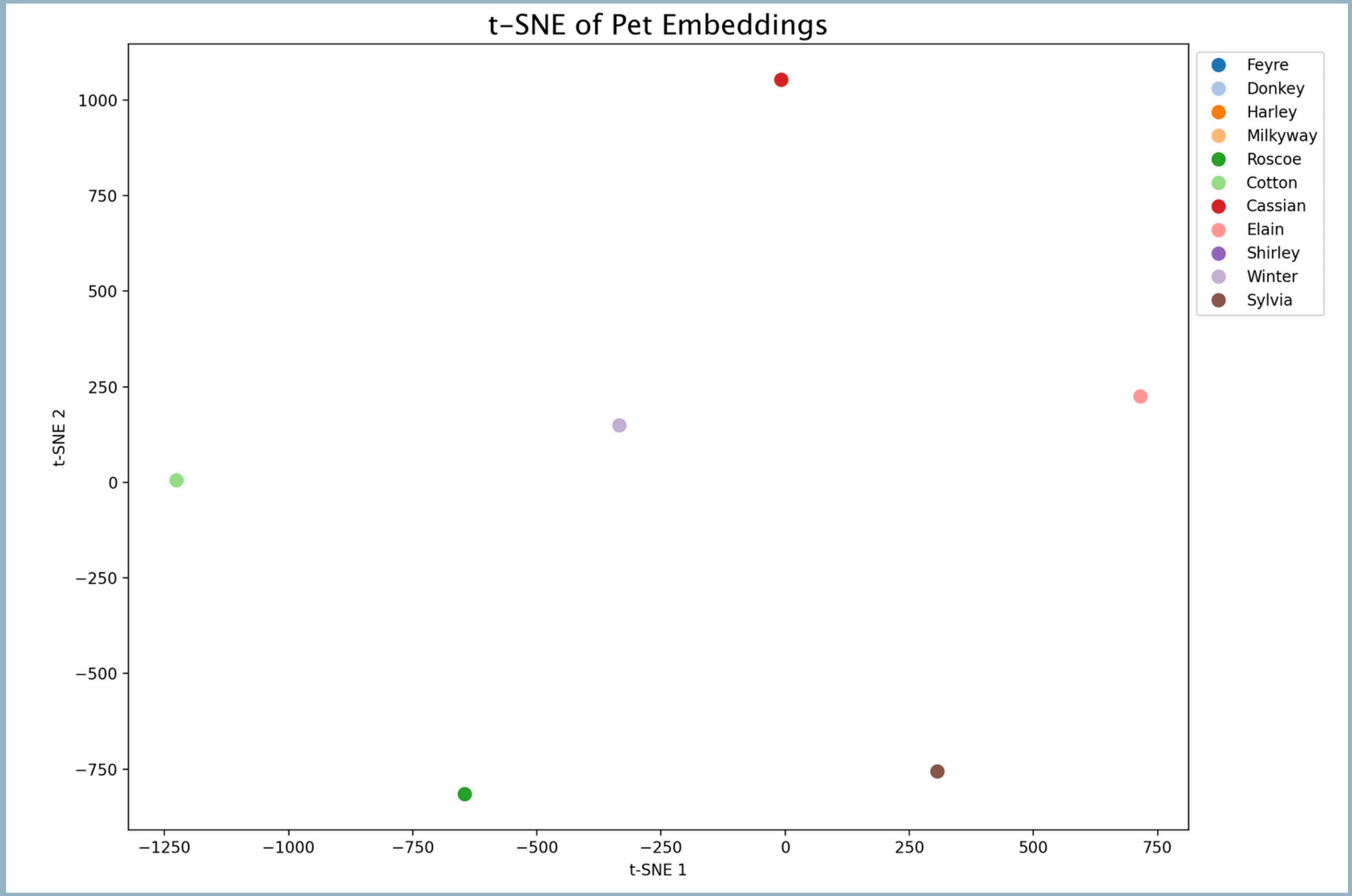
Breed,
color,
gender



Breed, color

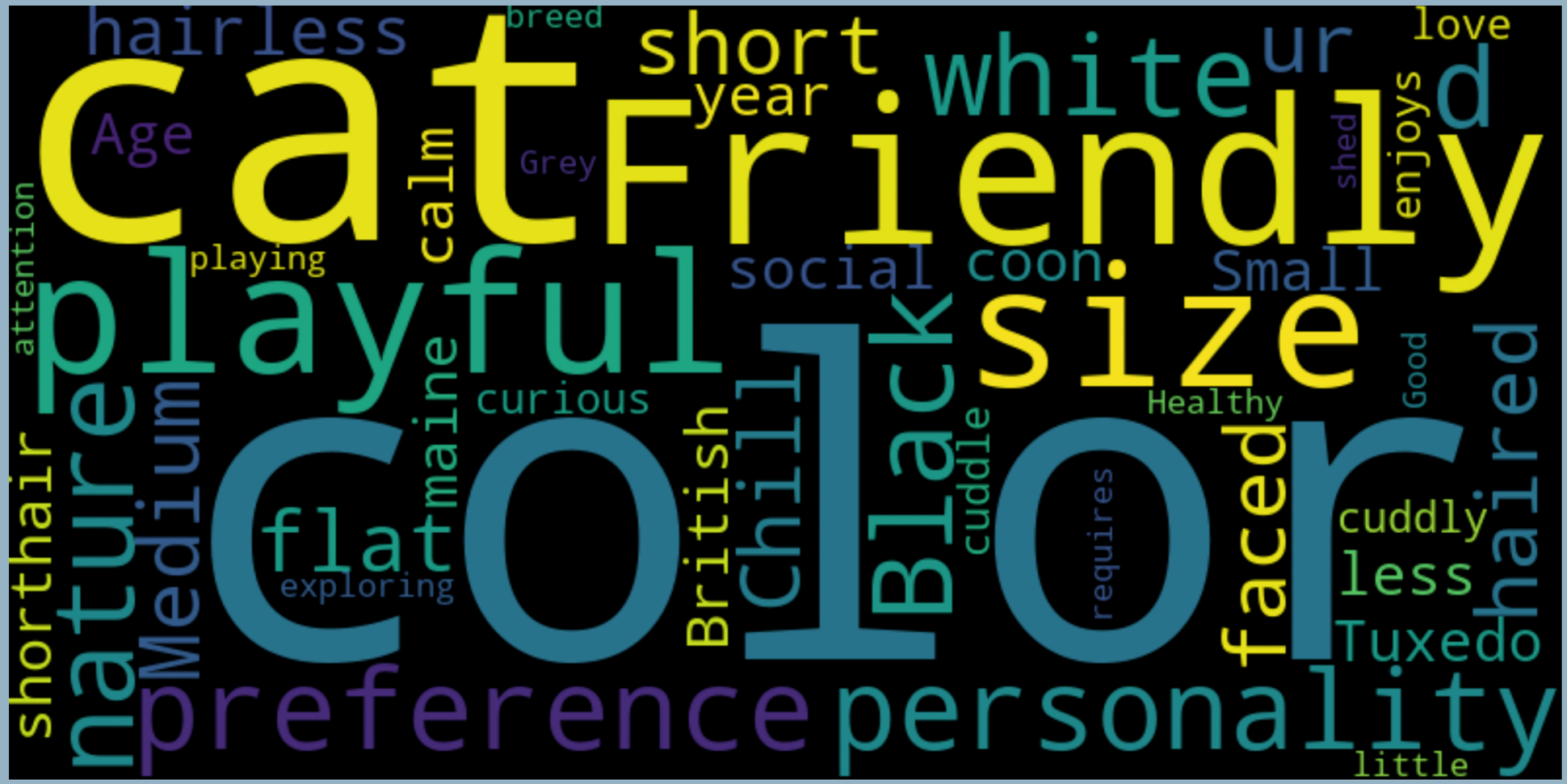

t-SNE of Pet Embeddings
t-SNE 2
t-SNE 1
600
400
200
0
−200
−400
−600
−800
−750
−500
−250
0
250
500
750
1000
Feyre
Donkey
Harley
Milkyway
Roscoe
Cotton
Cassian
Elain
Shirley
Winter
Sylvia

Breed,
color,
gender



Feature Engineering

WordCloud of Survey



Limitations/ Challenges

- Super small dataset (although it can be scaled with better data sources!) Also, I may not need a huge dataset since I am not training a model or building a production level recommender system.
- Better visualizations if I had more data/attributes
- No human/adopter data

Conclusion

In this project, a content-based pet recommendation system was developed using pet (cat) profile data. After scraping, cleaning and normalizing descriptive attributes, pets were represented as numerical embeddings and compared using cosine similarity. The findings demonstrate that attribute-based recommendations can be effective in pet adoption platforms, and provide a foundation for future enhancements incorporating richer features or user behavior data. However, with more data this will be seen more clearly!

Further Steps

- Create adopter profiles or a better survey
- Find other Pet Adopting profiles without issues, to scrape more data
- Incorporate a hybrid recommender system
- Testing if what the model recommends is “good”
- Expand to dogs as well

Thank You!