

1 Introduction

Dogs, like humans, have distinct personalities. However, specific dog breeds share common traits that make selecting a breed that fits certain roles or lifestyles easier. This report uses two datasets: `breed_traits.csv` and `trait_description.csv` to discover 2 questions related to breed suitability and trait correlations.

2 Task 1: Identifying Top 10 Suitable Toy Dog Breeds

2.1 Data Filtering

To identify the top 10 toy dog breeds that best fit the given lifestyle, specific filtering rules are applied to the dataset.

I selected breeds that:

1. **Highly affectionate with their owner** (Affectionate With Family: 4–5), making them great companion dogs.
2. **Prefer familiar people and are less open to strangers** (Openness To Strangers: 1–3), suiting owners seeking a loyal, one-person dog.
3. **Adapt well to small homes and new environments** (Adaptability Level: 4–5).
4. **Have moderate energy levels** (Energy Level: 3–4), enjoying both playtime and relaxation.
5. **Playful but not overly active** (Playfulness Level: 2–4).
6. **Easy to train** (Trainability Level: 3–5), helping prevent behavior issues like separation anxiety.
7. **Comfortable being alone and don't need constant social interaction** (Good With Other Dogs: 1–3).
8. **Have a low barking tendency** (Barking Level: 1–3), ideal for quiet living environments.
9. **Have moderate mental stimulation needs** (Mental Stimulation Needs: 2–4), balancing engagement without overwhelming care.

This helps filter breeds that are perfectly **suited for small homes, preferring the companionship of a single chosen human rather than a large family**. They can be antisocial at times and may experience separation anxiety if not properly trained. While they **enjoy running around and playing**, they also **love curling up on the sofa**. These breeds are among the best top dogs.

After filtering, **22 breeds remain**. However, the dataset does not include information about the size of the dogs. By searching online, medium and large breeds are eliminated, to meet the "**small in size**" requirement. After this step, exactly **10 breeds remain**. The results after filtering are shown in the table below. The numbers 1 to 5 represent the strength of the trait, the higher the number, the stronger the trait.

Breed	Affectionate With Family	Openness To Strangers	Adaptability Level	Energy Level	Playfulness Level	Trainability Level	Good With Other Dogs	Barking Level	Mental Stimulation Needs
Lhasa Apsos	5	3	5	3	3	3	3	3	3
American Hairless Terriers	5	3	5	3	3	5	3	3	3
Maltese	5	3	4	3	3	3	3	3	3
Pekingese	5	3	4	3	4	3	3	1	3
Tibetan Terriers	5	3	4	4	3	3	3	3	4
Xoloitzcuintli	5	3	4	4	4	4	3	3	4
Lakeland Terriers	5	3	4	3	4	3	3	3	4
Swedish Vallhunds	5	3	4	4	4	4	3	3	4
Lowchen	5	3	4	3	4	4	3	3	3
Glen of Imaal Terriers	5	3	4	3	3	3	3	2	3

Figure 1: Top 10 Suitable Toy Dog Breeds

2.2 Heatmap of Key Traits

The following heatmap is used to represent the scores of various personality traits across dog breeds. The heatmap uses color intensity to show which breeds are strong in certain traits and which have lower scores. This allows for a clear comparison of how each breed performs on multiple traits.

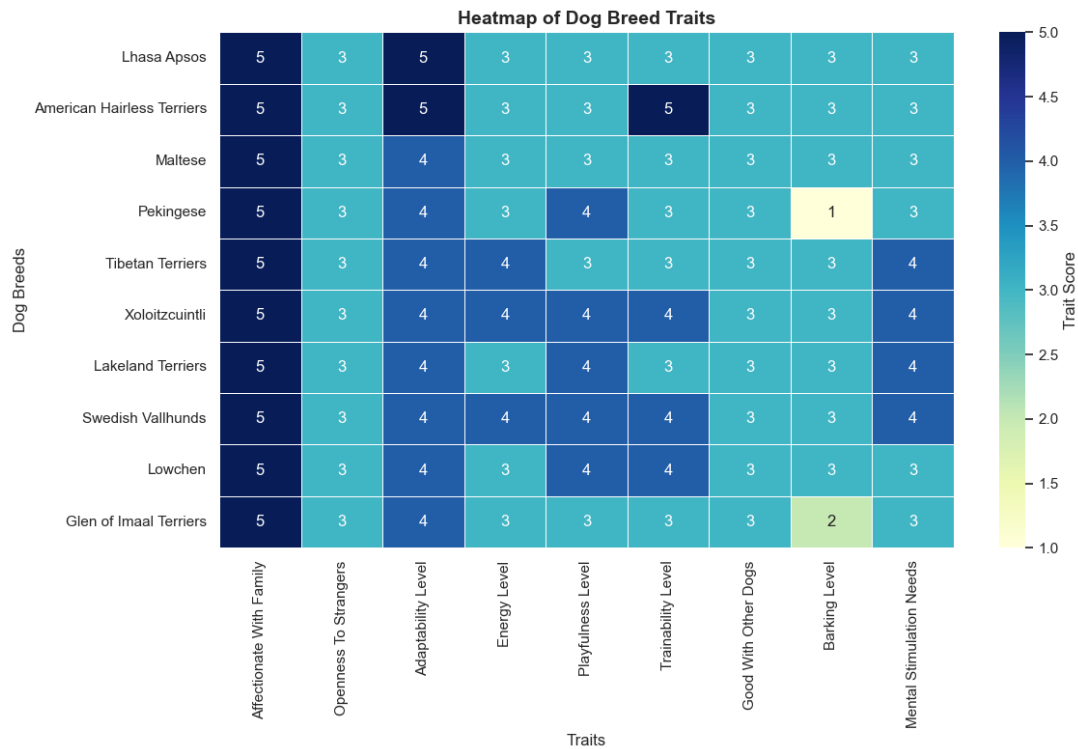


Figure 2: Heatmap of Key Traits for Top 10 Toy Dog Breeds

From the plot, we can see that all 10 breeds have the same scores in the traits "Affectionate With Family," "Openness To Strangers," and "Good With Other Dogs." This indicates that the breeds suitable as toy dogs highly overlap in these three characteristics.

To further explore the other traits of the top 10 breeds, radar charts will be used in the next part.

2.3 Radar Chart Visualization

Radar charts are used to visualize the other 6 traits (excluding the 3 traits that were the same across breeds) that were used to filter the suitable breeds.

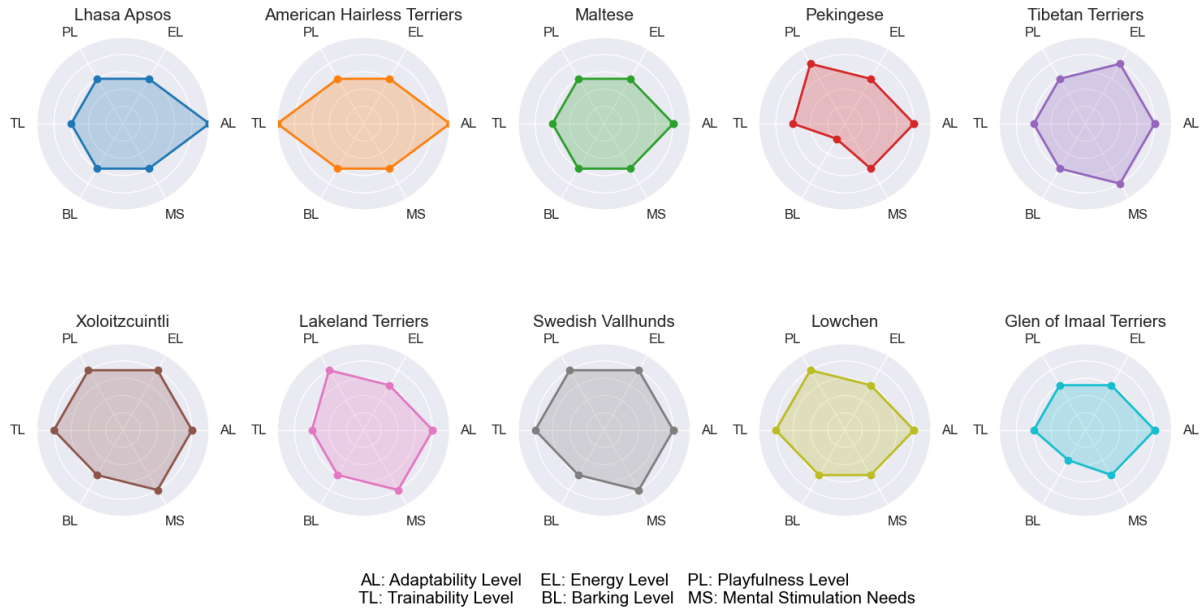


Figure 3: Radar Chart of Selected Traits for Top 10 Toy Dog Breeds

This type of plot is ideal for comparing the differences between breeds and highlighting their strengths and weaknesses. Additionally, radar charts are very useful for selecting a specific breed.

If someone values the adaptability of a dog most, they may choose the American Hairless Terrier, as it performs best in the trait "Adaptability Level" according to the radar plot. If someone prefers a quiet dog, particularly for small spaces, they could opt for the Pekingese due to its lowest barking level compared to the other 9 breeds.

3 Task 2: Correlation Analysis of Coat Characteristics and Other Traits

3.1 Preprocessing

Since some of the indicators are not numerical (ranging from 1-5), but text-based, I apply the following preprocessing methods to prepare the data for correlation coefficient calculations:

- The "Coat Type" variable is transformed into a quantified grooming difficulty scale to represent the care requirements. Each coat type is mapped to a numerical value reflecting the level of difficulty involved in grooming.

- The "Coat Length" variable is transformed into a scale of 1, 3, and 5, where higher values indicate longer coat lengths.
- The "Coat Grooming Frequency" variable originally had values of 1, 2, and 3 to represent frequency. To standardize this into a 1-5 scale, we remapped the original values accordingly.

Here are the mappings used for conversion:

- Coat Type Mapping: 'Hairless' → 1, 'Smooth' → 2, 'Double' → 2, 'Wavy' → 3, 'Curly' → 3, 'Rough' → 4, 'Silky' → 4, 'Wiry' → 5, 'Corded' → 5
- Coat Length Mapping: 'Short' → 1, 'Medium' → 3, 'Long' → 5
- Coat Grooming Frequency Remapping: 1 → 1, 2 → 3, 3 → 5

3.2 Correlation Analysis

The variables of the dataset are ranked categories instead of continuous numerical values. Spearman's rank correlation is more suitable for such data because it measures the strength and direction of monotonic relationships between variables, regardless of whether the data is normally distributed. On the other hand, Pearson's correlation assumes a linear relationship and requires continuous, normally distributed data.

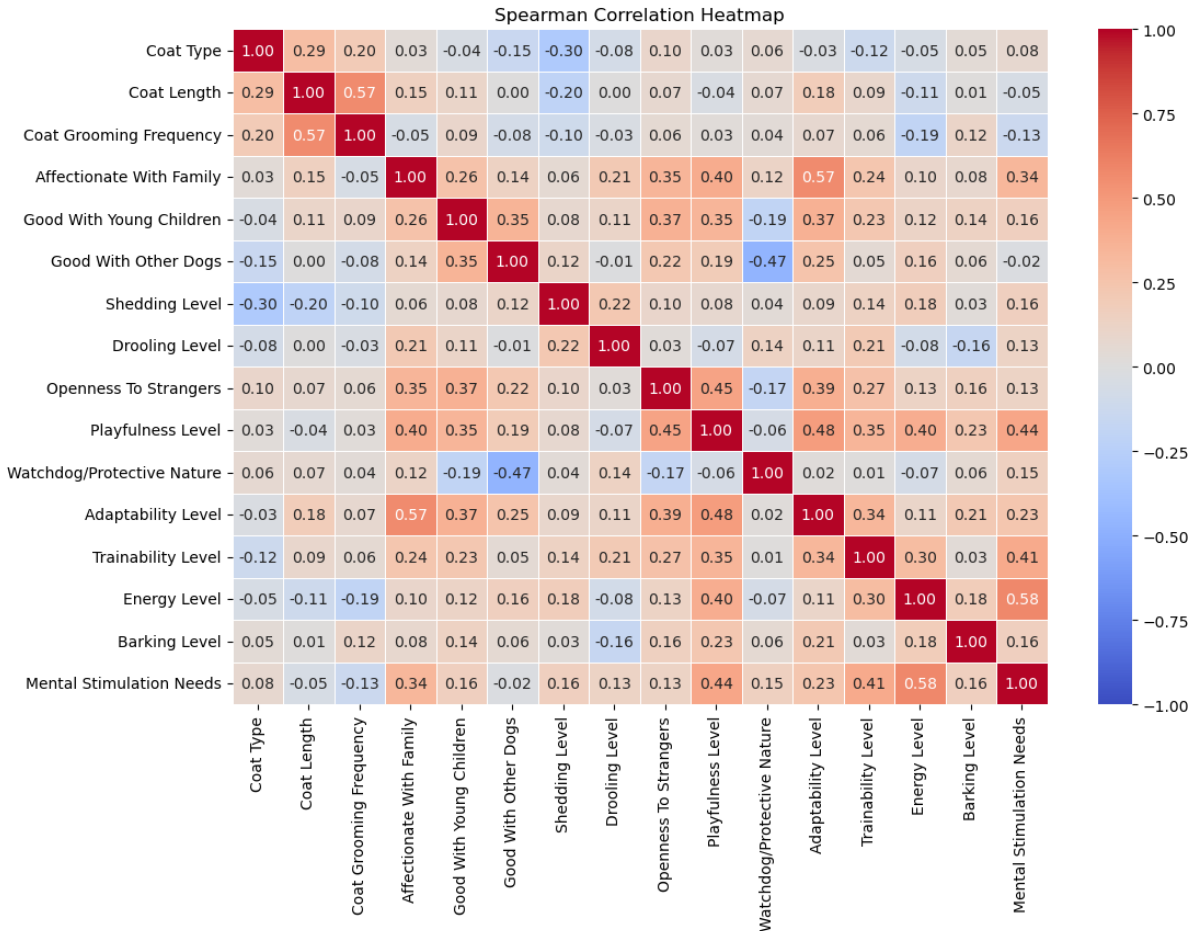


Figure 4: Correlation Heatmap between Coat Characteristics and Other Traits

After calculating the Spearman correlation coefficients, a correlation heatmap is used to visualize the results. The heatmap shows the correlations between variables, making it easy to spot strong or weak correlations and identify patterns within the dataset. The color intensity in the heatmap allows for a quick comparison of correlation values, helping to highlight important associations between the traits of different dog breeds.

The following conclusions can be drawn based on the Correlation Heatmap:

1. Coat Type

- A weak positive correlation with **Openness to Strangers** (0.099) and **Barking Level** (0.046), suggesting that dogs with more complex coats may be more vocal and open to strangers.
- A negative correlation with **Shedding Level** (-0.301), suggesting that breeds with more complex coats shed less.

2. Coat Length

- A potential positive correlation with **Adaptability Level** (0.184), suggesting that dogs with longer coats might have a slight advantage in adapting to environmental changes.

3. Coat Grooming Frequency

- A strong positive correlation with **Coat Length** (0.569), suggesting that long-haired breeds generally require more frequent grooming.
- A negative correlation with **Energy Level** (-0.195), suggesting that breeds that require more grooming may be less active or have lower energy levels.
- A positive correlation with **Mental Stimulation Needs** (0.128), suggesting that dogs that require more grooming may also need more mental stimulation.