Pet Product Subcategorization by Review Analysis

Namiko Nakashima

Table of Contents

1. Problem Statement
1.1. Problem

E-commerce companies set up categories for their products
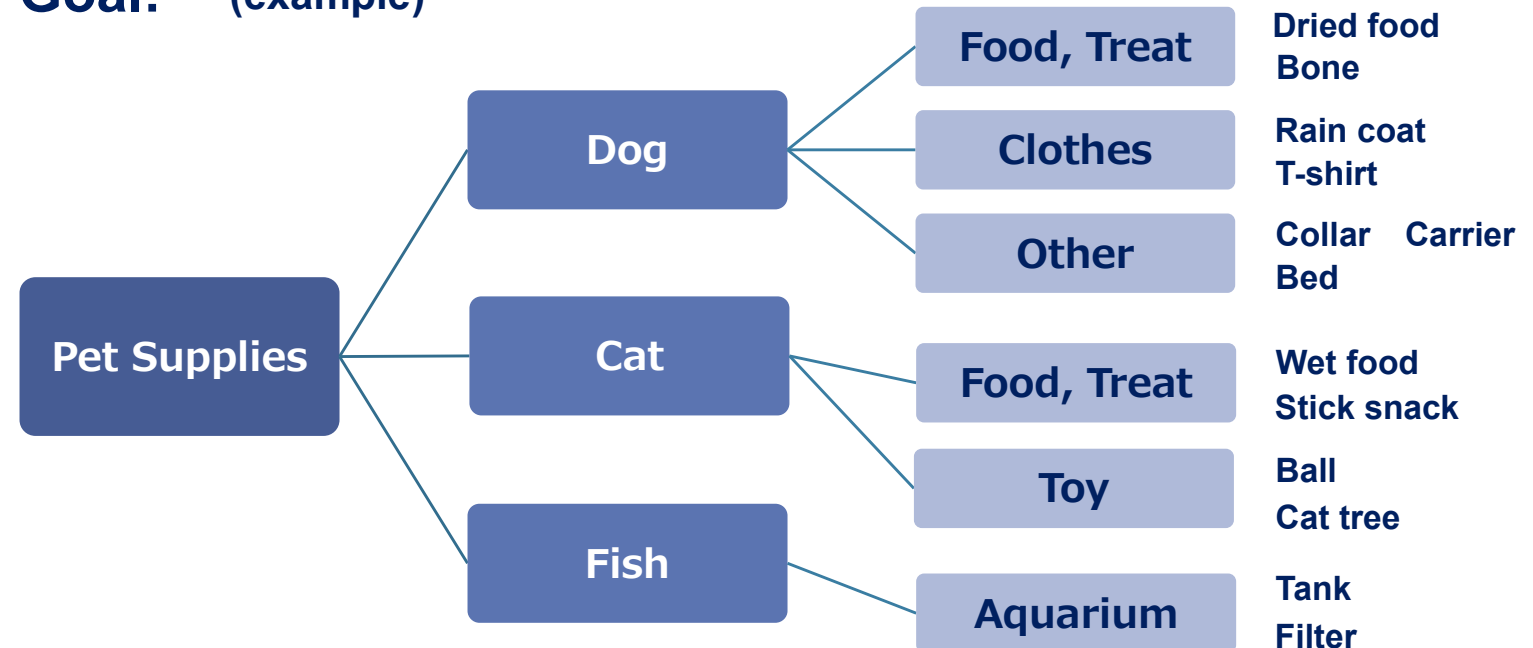
(e.g. Clothing, Beauty, Books, Pet Supplies, etc.)

More and more products in a category → subcategorize them

Now:

Pet Supplies

Dried food, Litter box, Bed, Filter, Ball, Bone, Collar, Carrier, Treat, Rain coat, Tank, T-shirt, Cat tree, Wet food

Goal: (example)

Pet Supplies
- Dog
  - Food, Treat — Dried food, Bone
  - Clothes — Rain coat, T-shirt
  - Other — Collar, Carrier, Bed
- Cat
  - Food, Treat — Wet food, Stick snack
  - Toy — Ball, Cat tree
- Fish
  - Aquarium — Tank, Filter

Goal: Subcategorize products by reviews

1.2. Client

1. E-commerce companies

- Improve analysis of trends and customer needs to a specific field

- Increase customer satisfaction by easy access to products

2. Manufacturer

- Extract a specific kind of products to analyze their competitions

2. Data Preprocessing
2.1. Data source

From [AWS](#)

Amazon Pet Supplies reviews (gz file)

https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Pet_Products_v1_00.tsv.gz

- **From 1995 to 2015**
- **In U.S.A.**
- **1,705,229 reviews with product ID and title**
- **128,995 kinds of pet products**

	marketplace	product_id	product_title	product_category	review_body	review_date
0	US	510387886	(8-Pack) EZwhelp Belly Band/Wrap	Pet Products	Best belly bands on the market! These are a g...	2015-08-31
1	US	912374672	Warren Eckstein's Hugs & Kisses Vitamin Minera...	Pet Products	My dogs love Hugs and Kisses. However, the la...	2015-08-31
2	US	902215727	Tyson's True Chews Premium Jerky - 12 ounce Ch...	Pet Products	I have been purchasing these for a long time. ...	2015-08-31

2.2. Data Wrangling 1

Extract necessary rows and columns:

1. **Extract necessary columns**
2. **Extract data collected in 2014 or 2015**
3. **Drop missing values and duplicate data**

Collect suitable reviews:

4. **Remove short and long reviews**
5. **Extract 10 reviews per product**

■ **204,030 reviews**
■ **20,403 products**

	product_id	product_title	review_body
0	510387886	(8-Pack) EZwhelp Belly Band/Wrap	Best belly bands on the market! These are a g...
1	912374672	Warren Eckstein's Hugs & Kisses Vitamin Minera...	My dogs love Hugs and Kisses. However, the la...
2	902215727	Tyson's True Chews Premium Jerky - 12 ounce Ch...	I have been purchasing these for a long time. ...
3	568880110	Soft Side Pet Crate, Navy/Tan	It is extremely well constructed, it is easy t...
4	692846826	EliteField 3-Door Folding Soft Dog Crate, Indo...	Worked really well. Very pleased with my purc...

2.2. Data Wrangling 2

Clean the tokens:

6. **Tokenization**
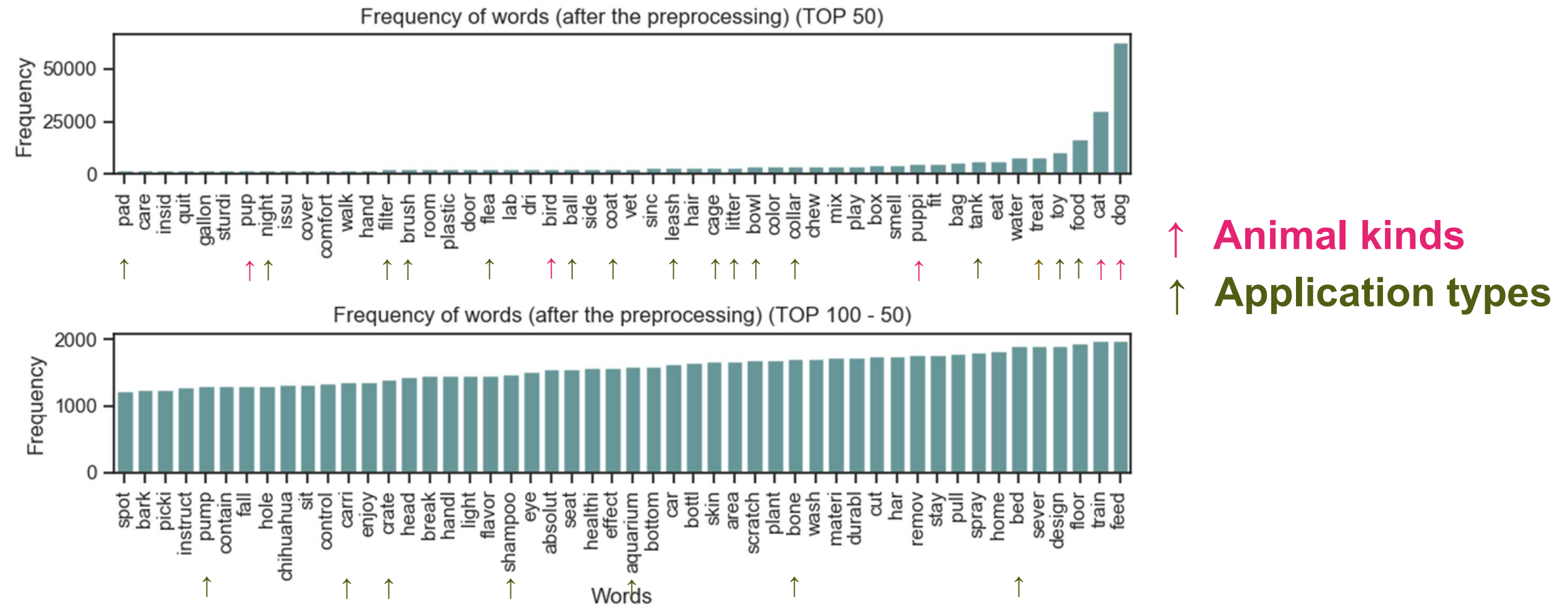7. **Remove**

- **Non-alphabet** words
- **Stop words**

8. **Stemming** and **lemmatization**
9. **Remove**

- **Non-noun** words
- Words appearing **5** times or less
- Uninformative Words in the top 200 frequent words

3. Exploratory Data Analysis
3.1. Tokens

Frequency of words (after the preprocessing) (TOP 50)

Frequency of words (after the preprocessing) (TOP 100 - 50)

↑ **Animal kinds**

↑ **Application types**

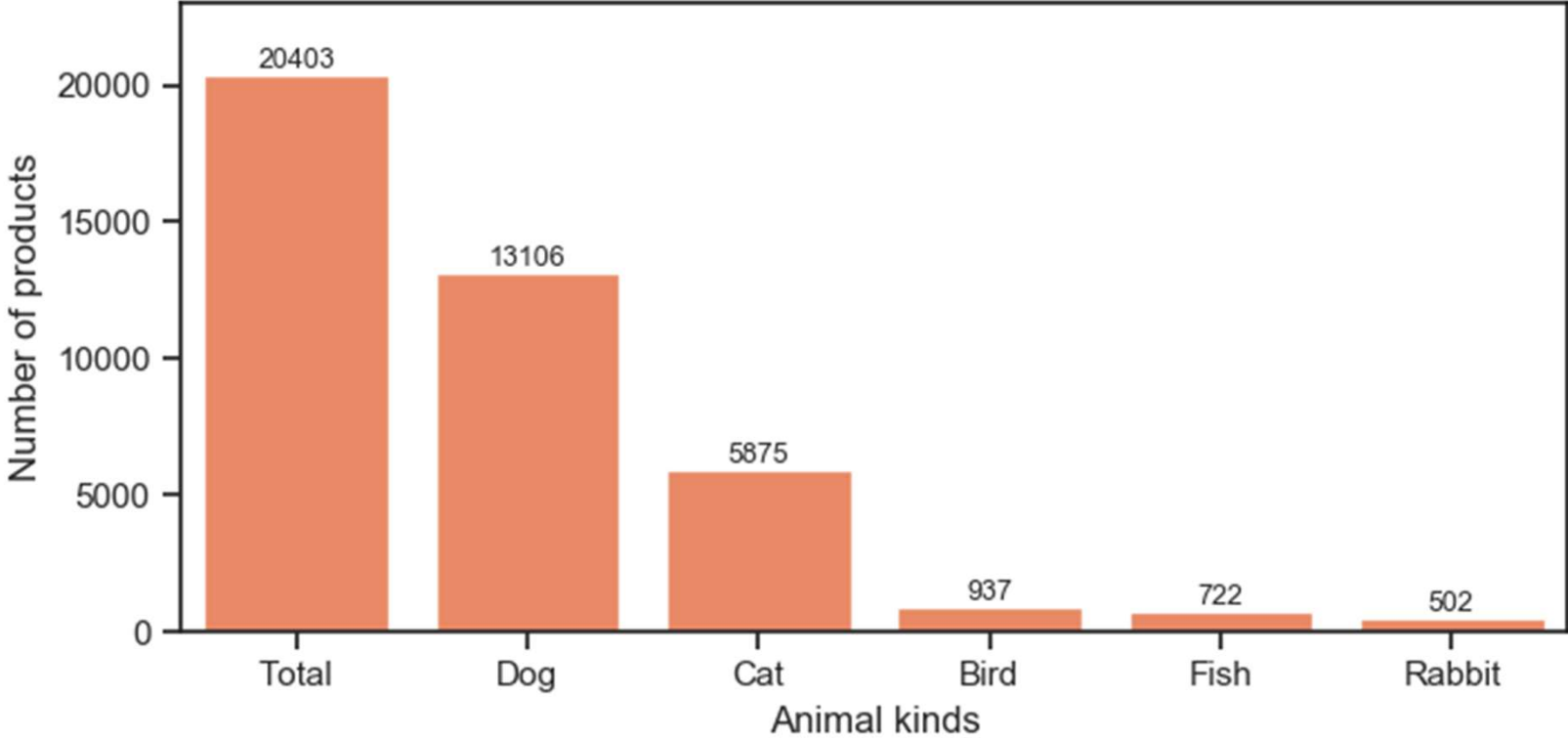**Most of frequent words were informative**

3.2. Animal Kind vs. Product Number

(e.g. Product having 'dog' as a token → Dog category)

The keywords for each category

Categories	Keywords
Dog	dog, puppi, doggi, pup
Cat	cat, kitti, kitt, kitten
Bird	bird, chick
Fish	fish
Rabbit	rabbit, bunni

The number of products in each category


Top 2: Dog, Cat

3.3. Application Type vs. Product Number

(e.g. Product having 'toy' as a token → Toy category)

The keywords for each category

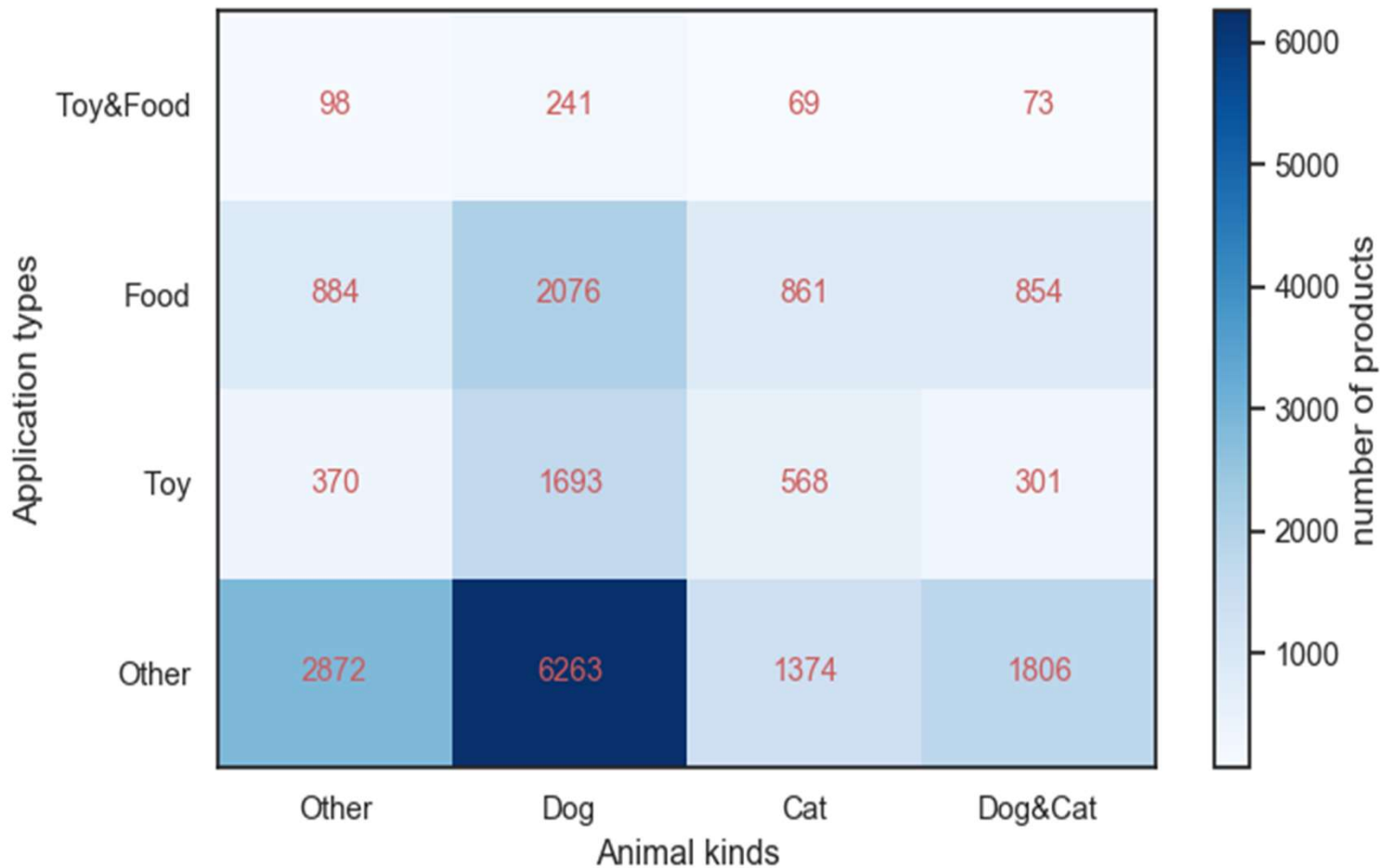| Categories	Keywords
Toy	toy, tunnel, ball, rope
Food	food, dri, wet
Treat	treat, snack, cooki
Collar & Leash	collar, leash
Clothes	shirt, coat, sweater, costum
Cage	cage, crate, carrier, kennel
Toilet	litter, pad

The number of products in each category
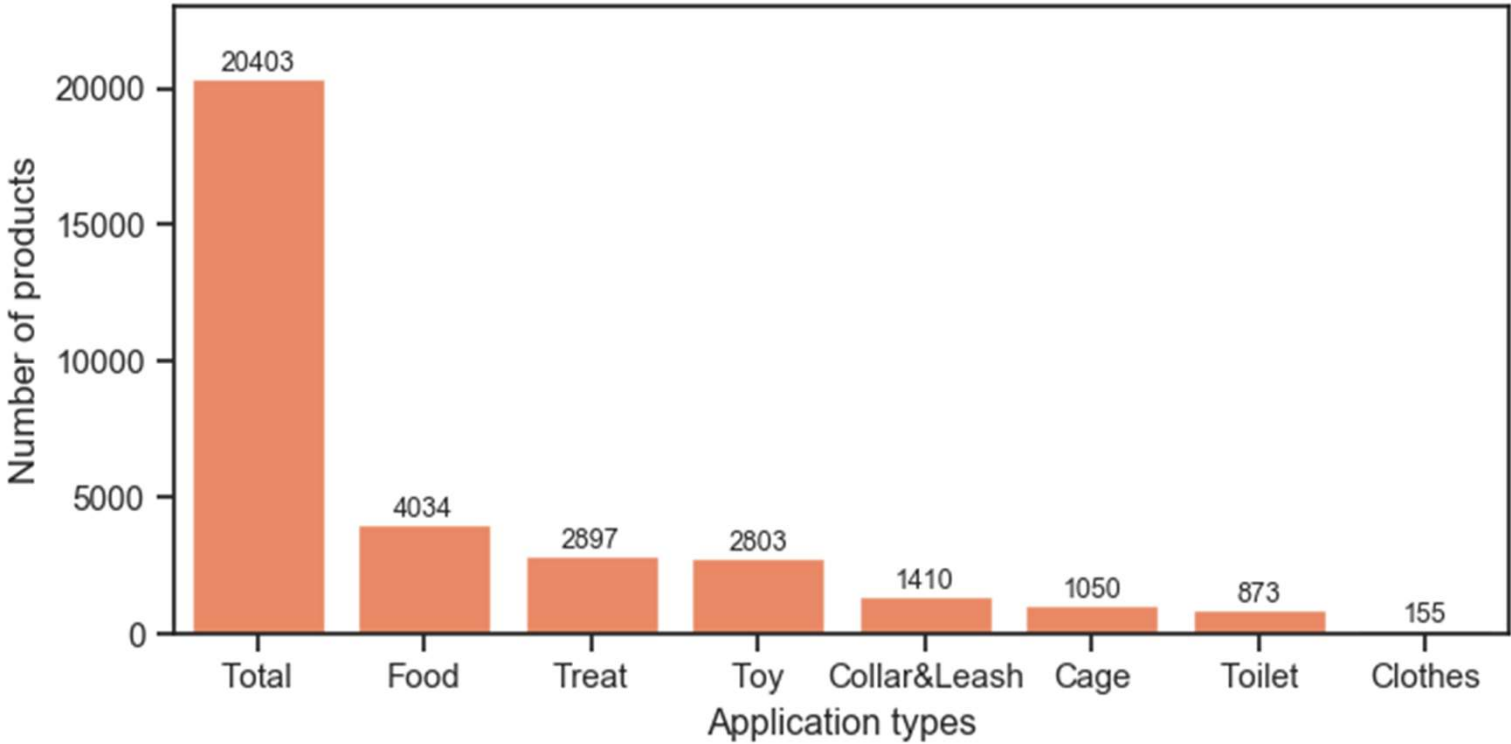
**Top 3: Food, Treat, Toy,
but relatively spread**

3.4. Animal Kind & Application Type

The number of products in each category

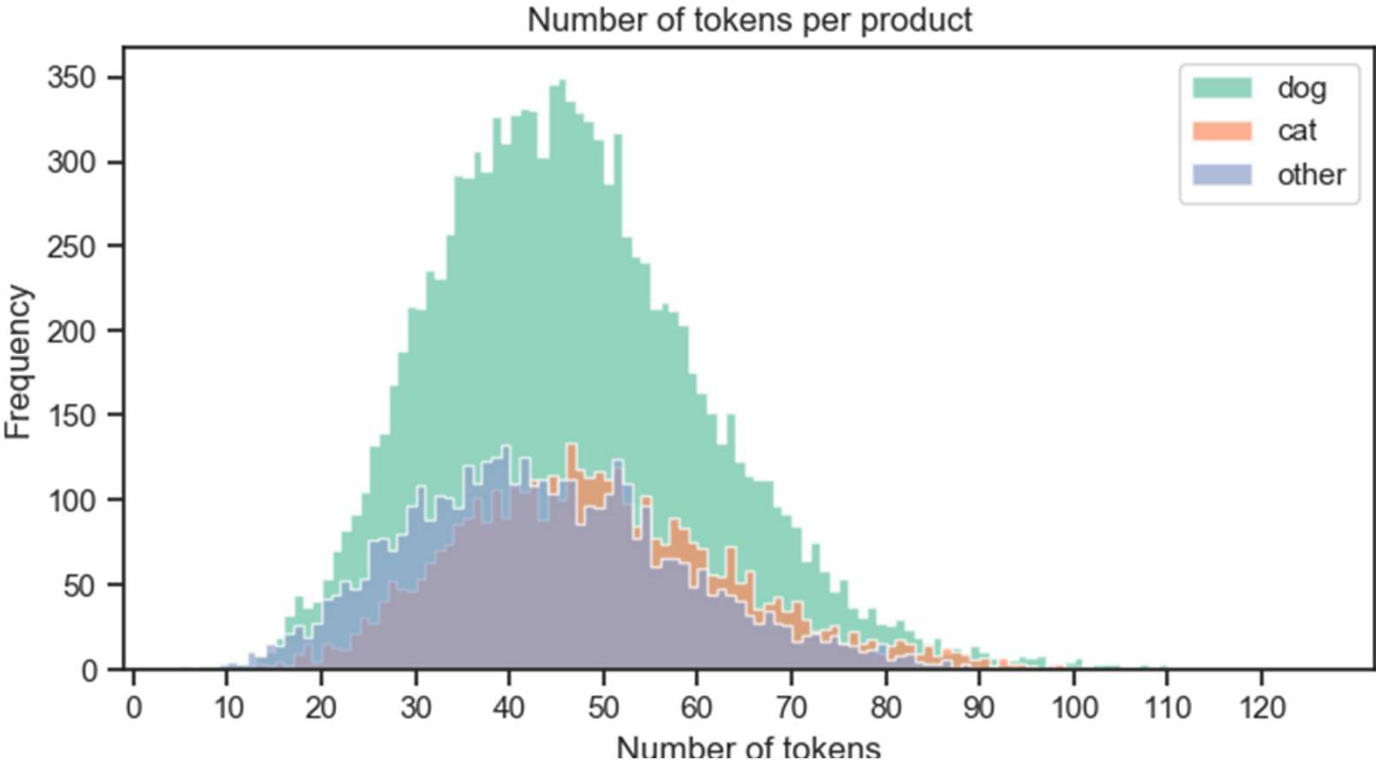■ **Top 2: Food and Toy for dogs**

■ **Many products were in**
  • **Other for dogs**
  • **Other for the other animals**

■ **Multi-features products**
  (Toy & Food, Dog & Cat)

3.5. Dog, Cat, Other datasets

Divided into 3 datasets: Dog, Cat, Other

(e.g. Product having 'dog' as a token → Dog category)

Dataset	Total products	Total tokens (unique)
Dog	11,916	6,682
Cat	4,099	6,001
Other	4,388	5,914

Number of tokens per product

Dataset	Min	25 %	50 %	75 %	Max
Dog	7	36	45	55	126
Cat	9	38	47	58	120
Other	5	32	42	52	98

Used them for clustering separately

4. Clustering

Goal Image

Animal Category:

Big Category:
Less than 10 big categories each

Small Category:
Less than 50 small categories each

Pet Supplies

Dog

Cat
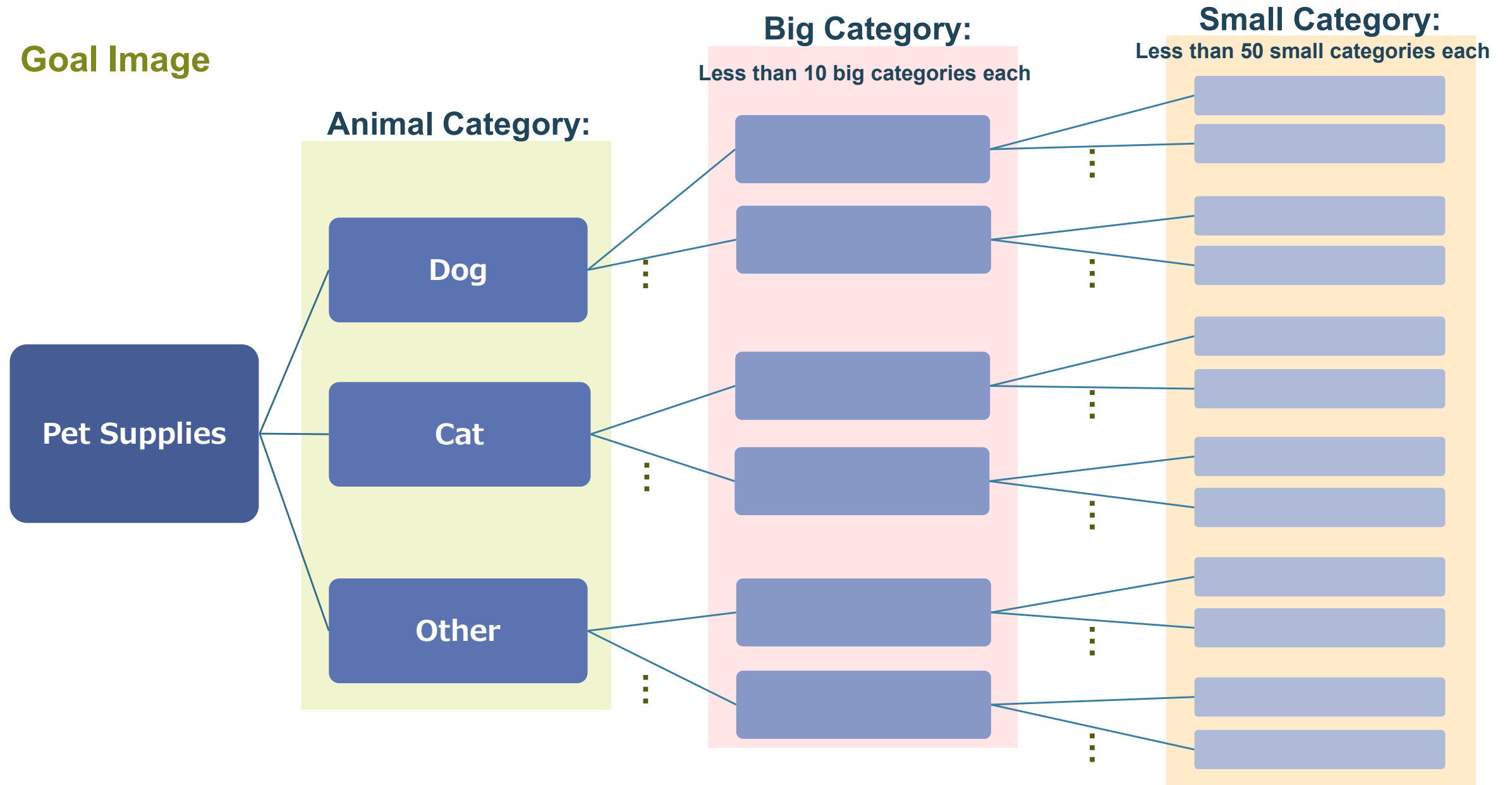
Other

4.1. Vectorization

Conditions:

- **Count vectorizer**

Category	N grams	Min_df	Max_features	Feature number
Dog	1, 2	-	5,000	5,000
Cat	1, 2	6	-	4,665
Other	1, 2	6	-	5,156

e.g. Dried food, litter box   **To control the number of tokens**

- **Normalized**

To standardize the difference of the number of tokens among products

4.2. Hierarchical Clustering

Hierarchical Agglomerative Clustering:

- **Unsupervised Learning**
- **Segment data into some groups**
- **Without any pre-labels**
- **Nested clusters (dendrogram)**
- **Time complexity**

Category	Total products
Dog	11,916
Cat	4,099
Other	4,388

Acceptable!

Conditions:

- **Texts → Cosine similarity (affinity)**
- **Different size clusters → Average linkage (linkage)**

4.3. Clustering steps

Steps:

1. Calculate Silhouette scores from k = 5 to 20
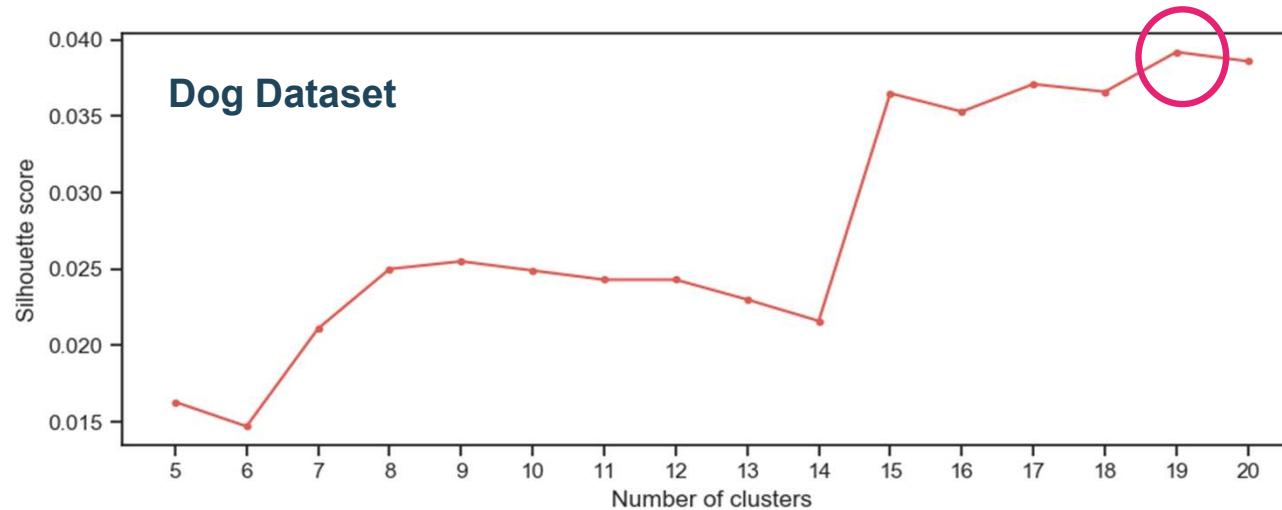
Dog Dataset

2. Clustering with the max. k

3. Decide big categories and category labels from:

- ✓ **Silhouette plot**
- ✓ **T-SNE with labels**
- ✓ **Frequent words**
- ✓ **Browsing product titles**

4. Repeat 1 to 3 to the big clusters to get small categories (k = 15 to 50)

Data
e.g. Dog dataset

Other

Big categories

Other

Small categories

4.4. Dog Dataset (Big Category 1)
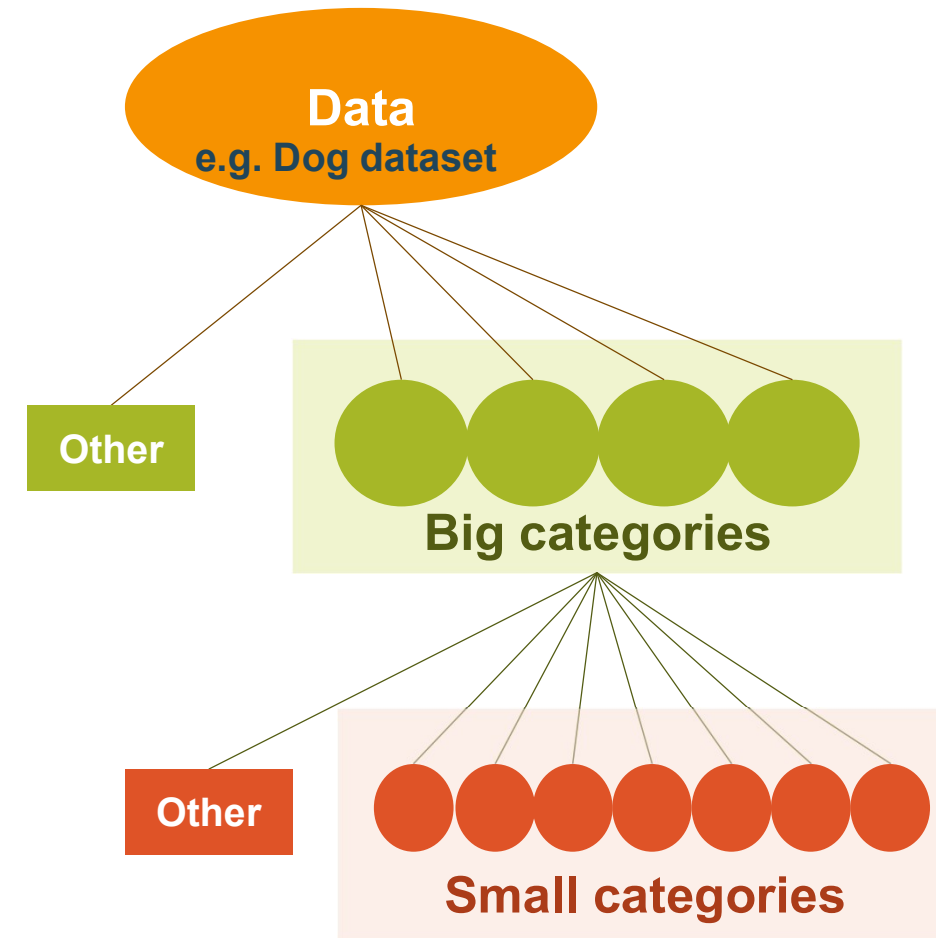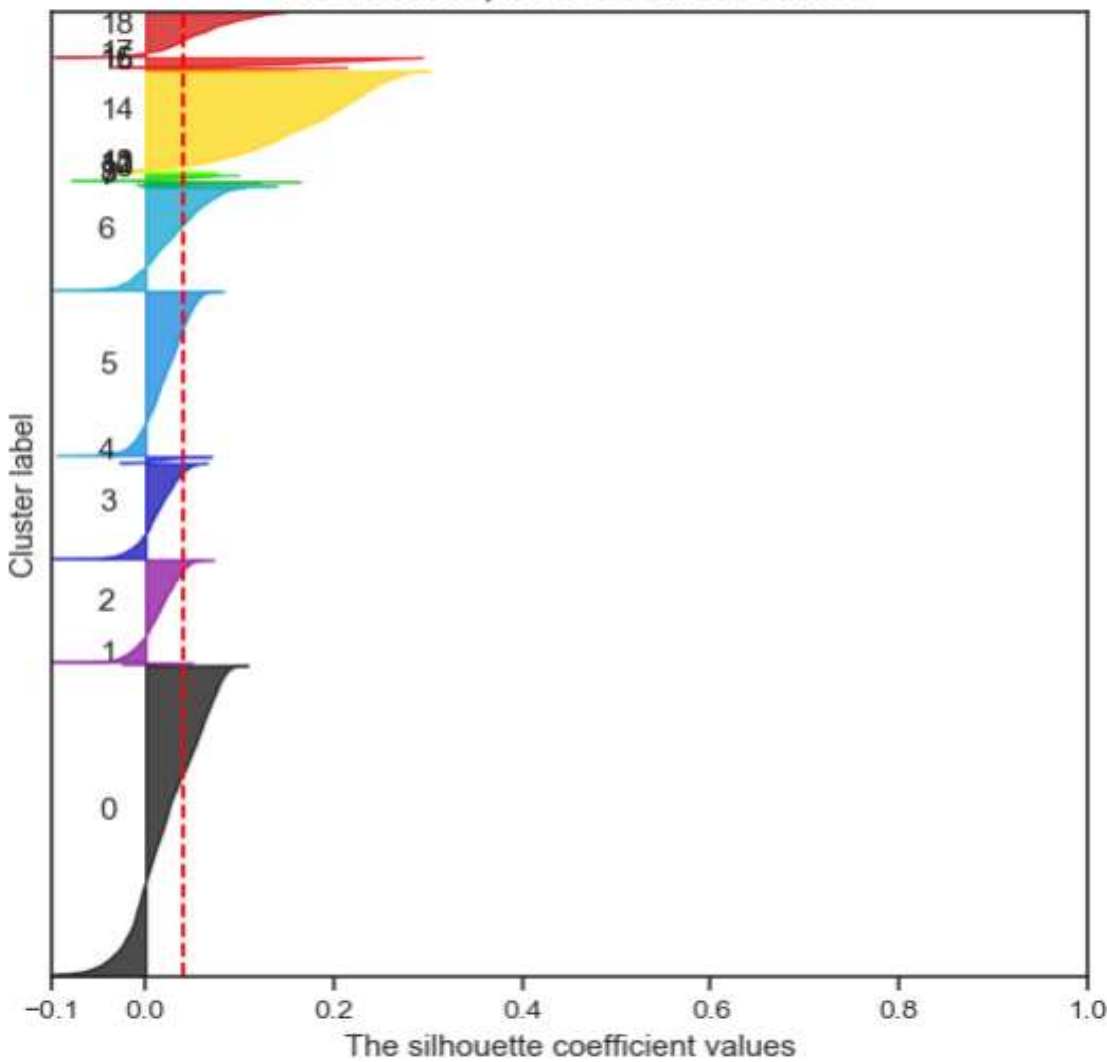
k = 19

✓ **Silhouette plot**

The silhouette plot for the various clusters

✓ **Frequent words**

cluster number	frequent words	number of products
0	(food, treat, eat, bag, mix, chew, vet, sinc, lab, pup)	3896
1	(area, air, devic, track, home, cat, tagg, call, batteri, let)	22
2	(bed, pad, fit, crate, wash, cat, door, cover, room, comfort)	1285
3	(fit, mix, materi, strap, car, comfort, seat, stay, chihuahua, ador)	1200
4	(storm, mix, thunder, calm, firework, anxieti, effect, vet, thunderstorm, rescu)	76
5	(collar, leash, walk, fit, train, color, pull, comfort, har, neck)	2065
6	(smell, coat, skin, bath, shampoo, hair, dri, flea, vet, spray)	1293
7	(camera, video, gift, watch, app, monitor, night, fun, view, connect)	10
8	(beauti, pass, gift, photo, belov, person, urn, collar, metal, display)	27
9	(kit, powder, aid, bleed, emerg, hand, home, satisfi, prepar, vet)	6
10	(cabl, ground, stake, tie, hook, rust, line, pull, mix, camp)	70
11	(daughter, gift, rock, instruct, name, person, fit, lab, grandson, miniatur)	6
12	(boston, terrier, run, bulldog, hang, sturdi, owner, tree, mom, color)	9
13	(pup, eat, sinc, compar, someth, seller, slip, odor, avail, hound)	9
14	(toy, play, chew, squeaker, durabl, chewer, ball, lab, mix, fun)	1242
15	(tank, beard, gallon, suppos, dragon, lamp, light, veil, heat, bulb)	7
16	(water, pool, hole, drain, plastic, wast, summer, empti, step, sturdi)	18
17	(eye, stain, face, tear, remov, vet, angel, head, poodl, fit)	116
18	(hair, brush, groom, cut, cat, coat, clipper, tool, comb, fur)	559

✓ **Browsing the titles** (example)

product_id		product_title	label_19
5309	442792109	Oster Narrow Blade For Finisher Trimmer (Model 59) CL-76913566Oster Narrow Blade For Finisher Tr...	18
7310	615356020	Safari Shed Magic De-Shedding ToolSafari Shed Magic De-Shedding ToolSafari Shed Magic De-Sheddin...	18
11373	955301771	Dematting Tool for Dogs - The Best Dog Grooming Comb for Undercoat Removal – Professional ...	18
1032	82031098	FURemover Mitt and Brush Set, 2-sided Rubber-like Lint Brush for Couch and Clothes, Cat and Dog ...	18
2281	185416403	Millers Forge Pet Hair Cutting Scissors, 7-1/2-Inch, StraightMillers Forge Pet Hair Cutting Scis...	18

4.4. Dog Dataset (Big Category 2)

✓ **T-SNE**

t-SNE of dog products colored by the labels

0. Food, Treat, Treatment

14.Toy

6. cleaning,
17,18. body care

4

7

15

2. Bed, Crate, Gate

8   10

3. Clothes

5. Collar, Leash
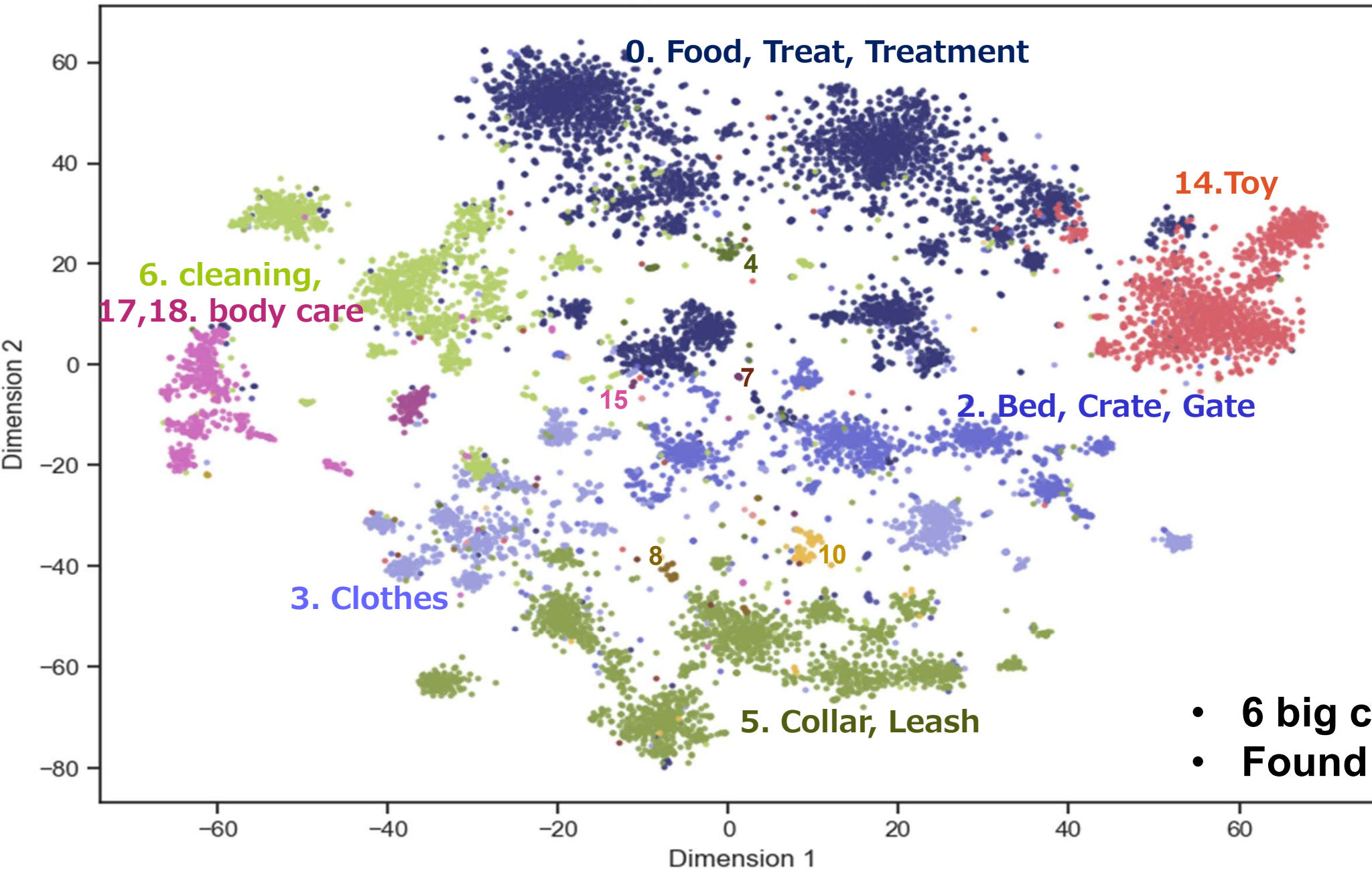
"Other"	
Cluster No.	Cluster label
4	calming
7	monitoring
8	memorial
10	tie out
15	aquarium
1, 9, 11, 12, 13, 16	other

- **6 big categories**
- **Found outliers (cluster_15)**
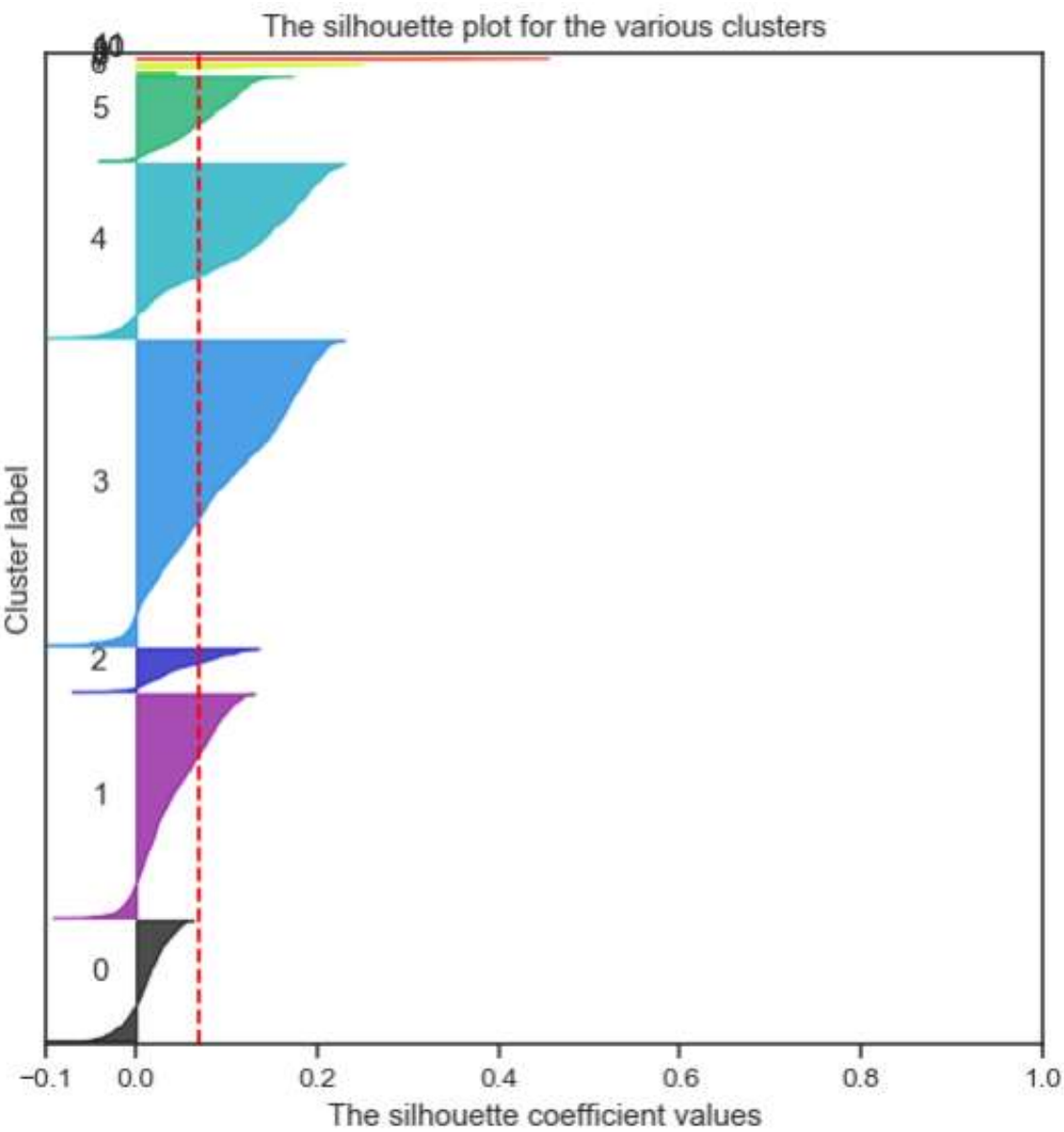
4.4. Dog Dataset (Small Category)

k = 34
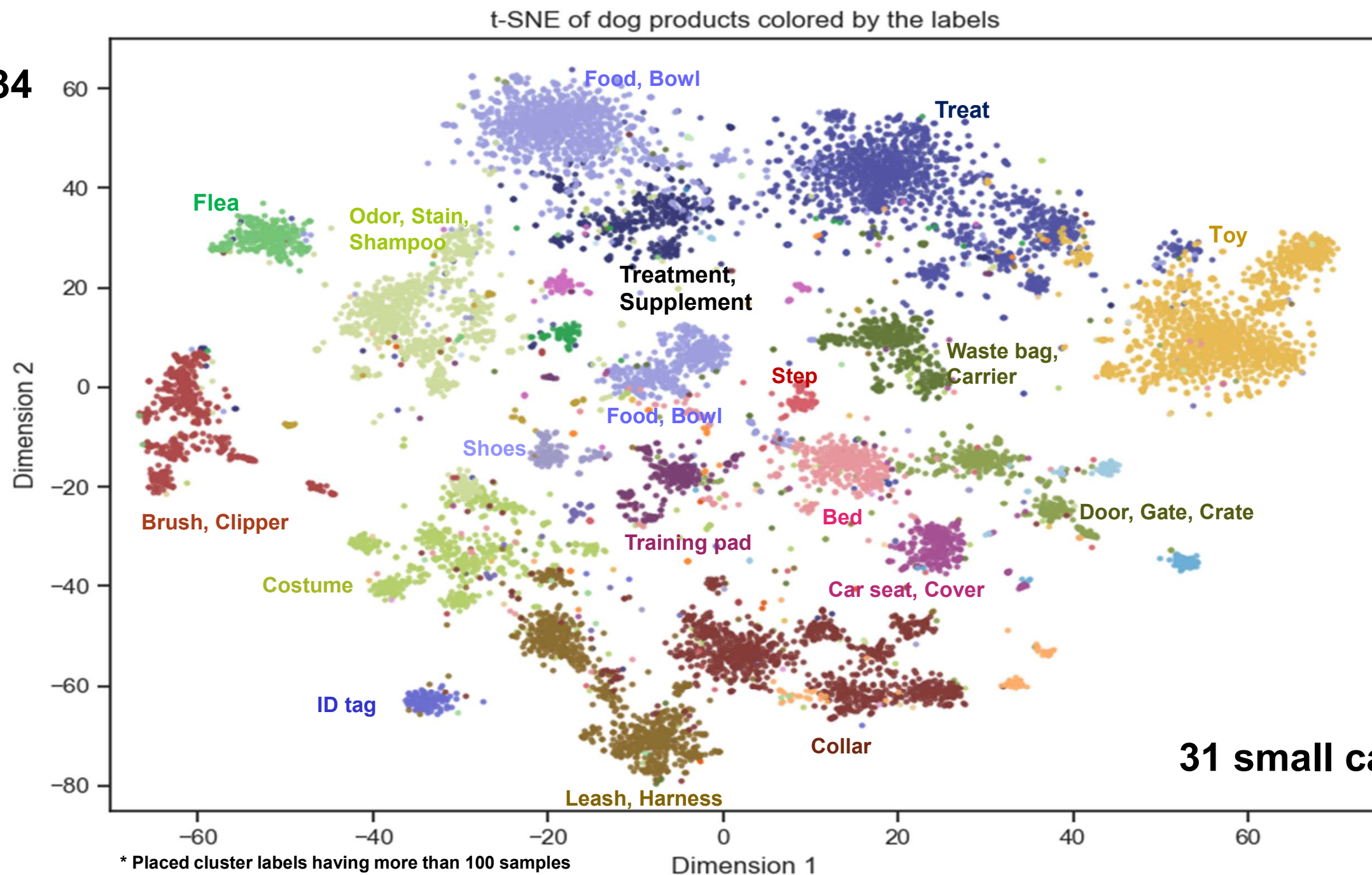
t-SNE of dog products colored by the labels

31 small categories

* Placed cluster labels having more than 100 samples

4.5. Cat Dataset (Big Category 1)

k = 12

✓ **Silhouette plot**

The silhouette plot for the various clusters

✓ **Frequent words**

| cluster number	frequent words	number of products
0	(dog, fit, door, side, room, comfort, insid, bed, carrier, vet)	518
1	(play, toy, scratch, box, sturdi, post, fun, catnip, room, jump)	956
2	(collar, color, fit, bell, neck, dog, design, stay, comfort, tag)	185
3	(food, eat, dri, dog, feed, treat, water, flavor, vet, bowl)	1304
4	(litter, box, smell, odor, scoop, clump, bag, floor, spray, room)	744
5	(dog, hair, flea, brush, fur, groom, remov, comb, effect, coat)	363
6	(chair, dog, tool, train, sound, furnitur, spray, choic, petsmart, custom)	4
7	(import, jus, ask, bring, pleas, letter, manufactur, wife, sell, vitamin)	1
8	(tank, aquarium, gallon, water, color, plant, quit, glass, home, con)	17
9	(turtl, antibiot, beardi, sugar, energi, filler, past, feed, exot, betterthi)	1
10	(bubbl, chase, form, bottl, blow, wait, soap, make, liquid, catnip)	5
11	(paper, book, glad, enjoy, kid, learn, copi, son, color, daughter)	1

✓ **Browsing the titles**   (example)

	product_id	product_title	label_12
1823	437636310	Kitty Solutions Cardboard Cat Scratcher (2 Pack)Kitty Solutions Cardboard Cat Scratcher (2 Pack)...	1
1606	383921799	Two 7 x 4 1/2 inch Cat Scratch Guards with Pins ---- THE ORIGINAL CAT SCRATCH GUARD FURNITURE PR...	1
1171	283977862	CLOSE OUT SALE Premium Cat Scratcher Toy with Catnip and Mouse Combo Set - Toy Mouse and Catnip ...	1
3901	953653364	Go Cat Teaser Cat Catcher Wand Cat ToyGo Cat Teaser Cat Catcher Wand Cat ToyGo Cat Teaser Cat Ca...	1
2738	680408952	Whirly Gig Spinning Bat Toss and Fetch Cat Toy by PetstagesWhirly Gig Spinning Bat Toss and Fetc...	1
3004	742358789	The Bird Catcher Pro Pets Can Play Interactive Cat Toy Super Wand Fishing Pole Teaser with Two F...	1
270	66486636	Electronic Remote Control Mouse Toy for Trick/Playing with Cat GrayElectronic Remote Control Mou...	1

4.5. Cat Dataset (Big Category 2)

✓ **T-SNE**

t-SNE of cat products colored by the labels
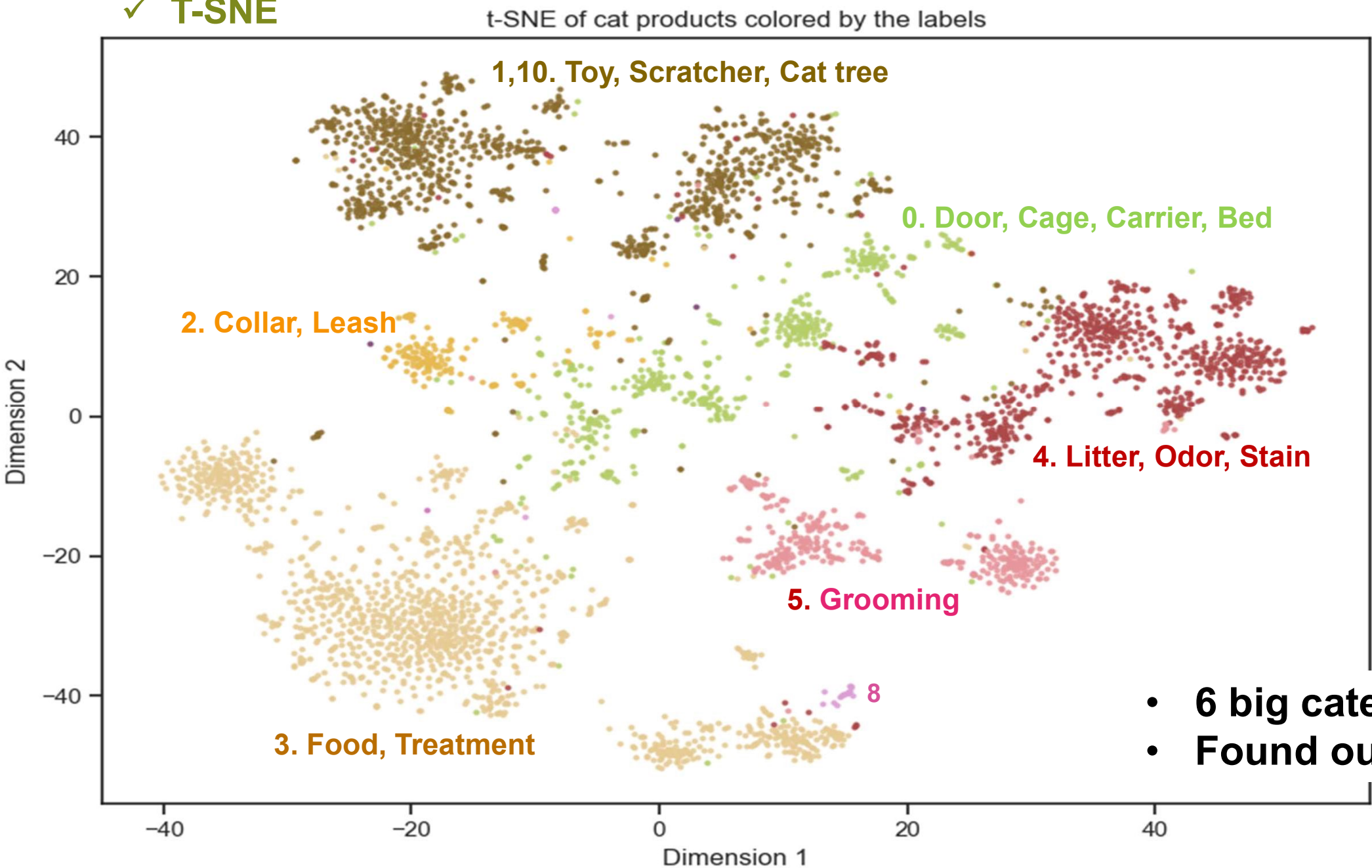
"Other"	
Cluster No.	Cluster label
8	aquarium
6, 7, 9, 11	other

- **6 big categories**
- **Found outliers (cluster_8)**

4.5. Cat Dataset (Small Category)
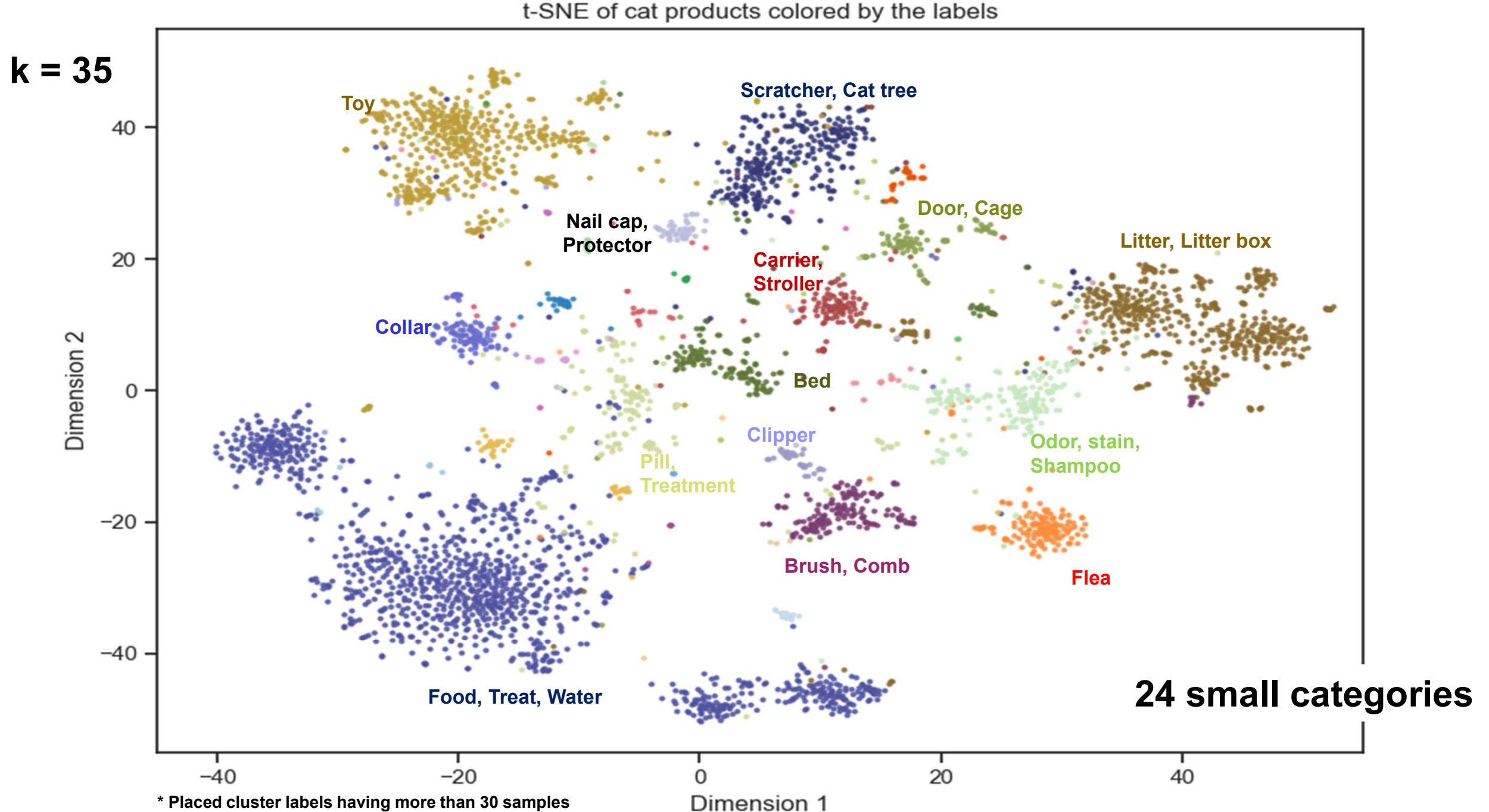
t-SNE of cat products colored by the labels

k = 35

24 small categories

* Placed cluster labels having more than 30 samples

Toy
Scratcher, Cat tree
Nail cap, Protector
Door, Cage
Litter, Litter box
Carrier, Stroller
Collar
Bed
Clipper
Odor, stain, Shampoo
Pill, Treatment
Brush, Comb
Flea
Food, Treat, Water

4.6. Other Dataset (Big Category 1)

k = 12

✓ **Silhouette plot**

The silhouette plot for the various clusters
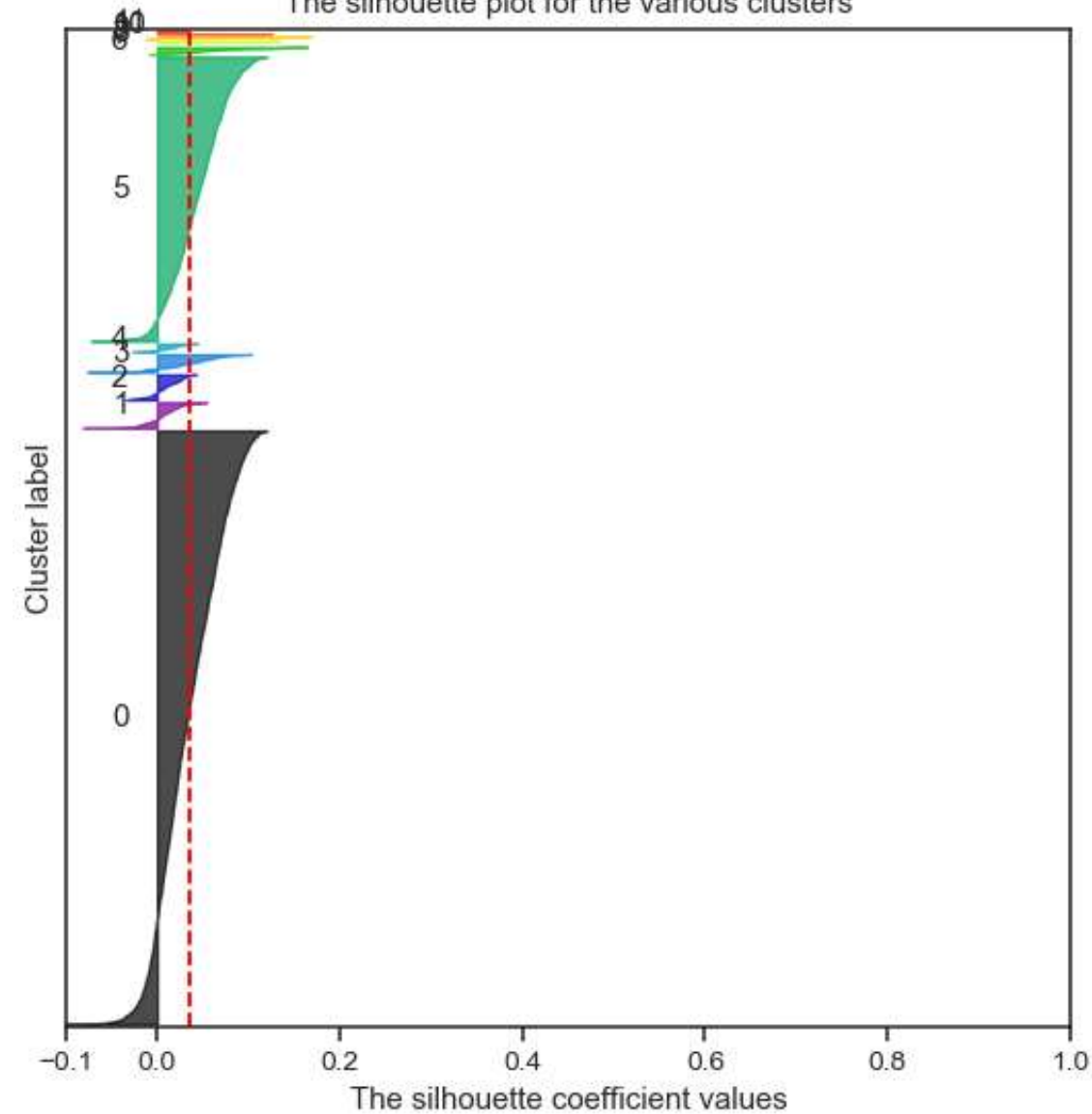
✓ **Frequent words**

cluster number	frequent words	number of products
0	(tank, water, gallon, aquarium, fish, filter, plant, color, food, pump)	2689
1	(smell, spray, hand, area, effect, bottl, water, result, sinc, vet)	118
2	(fit, comfort, car, color, side, door, head, strap, materi, clip)	116
3	(hair, groom, cut, blade, brush, clipper, fit, comb, oster, tool)	81
4	(beauti, gift, son, chain, hang, seller, tag, make, color, person)	39
5	(bird, cage, food, eat, parrot, toy, seed, rabbit, treat, guinea)	1289
6	(batteri, collar, fit, leash, box, shock, charg, train, hand, origin)	33
7	(bio, colon, alreadi, benefit, meter, provid, biolog, matrix, reus, effect)	1
8	(goat, milk, right, watch, administ, box, duti, side, dwarf, edg)	6
9	(pad, plastic, filter, origin, name, pee, absorb, leak, sound, spin)	12
10	(strip, yorki, check, prompt, vet, impress, glucos, blood, system, conveni)	2
11	(nice, bedroom, wall, leather, addit, fortun, variat, skimpi, air, craft)	2

✓ **Browsing the titles**

(example)

	product_id	product_title	label_12
4191	954892991	Exo Terra Swamp Glo Basking Spot LampExo Terra Swamp Glo Basking Spot LampExo Terra Swamp Glo Ba...	0
2904	660990044	Hydor 12V Koralia Controllable Circulation PumpHydor 12V Koralia Controllable Circulation PumpHy...	0
1389	312588126	S.T. International Aquarium Surface SkimmerS.T. International Aquarium Surface SkimmerS.T. Inter...	0
622	143449000	Lifegard Aquatics 1/2-Inch Double Threaded BulkheadLifegard Aquatics 1/2-Inch Double Threaded Bu...	0
3107	708080244	Aqueon Fern Aquarium PlantAqueon Fern Aquarium PlantAqueon Fern Aquarium PlantAqueon Fern Aquari...	0
1732	393482881	Microworms Starter Cluture. Micro Worms Live Fish FoodMicroworms Starter Cluture. Micro Worms Li...	0
1229	273345241	Lee's Airline Tubing, 25-Foot, StandardLee's Airline Tubing, 25-Foot, StandardLee's Airline Tubi...	0

4.6. Other Dataset (Big Category 2)

t-SNE of pet products colored by the labels

✓ **T-SNE**

5. Bird, Rabbit, Hamster

0,7,9. Fish, Reptile

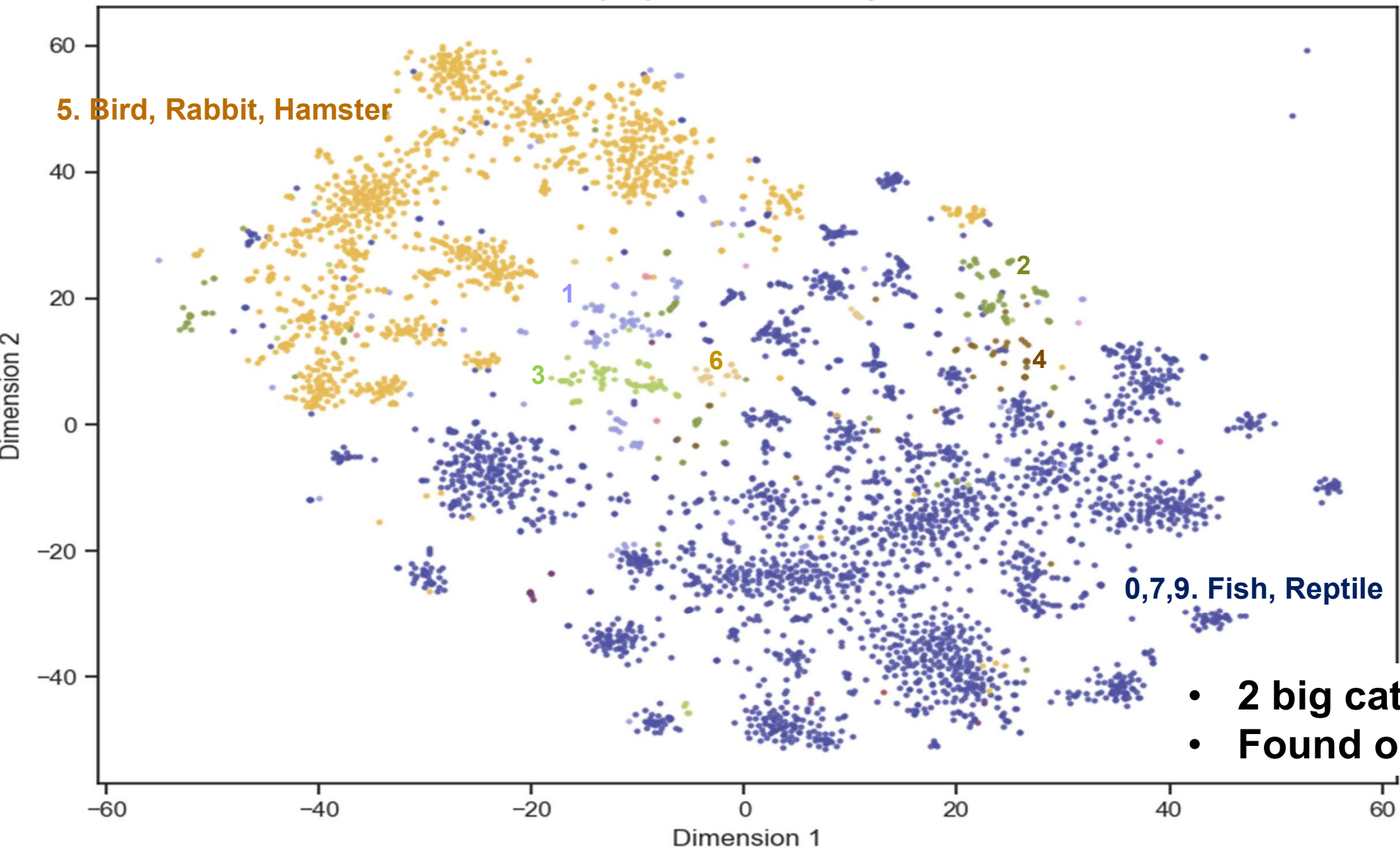| "Other"	
Cluster No.	Cluster label
1	treatment
3	brush, comb, clipper
4	memorial, tag
6	Dog training
2, 8, 10, 11	other

- 2 big categories
- Found outliers (cluster_6)

4.6. Other Dataset (Small Category)

t-SNE of pet products colored by the labels

k = 37

18 small categories

* Placed numbers of clusters having more than 30 samples
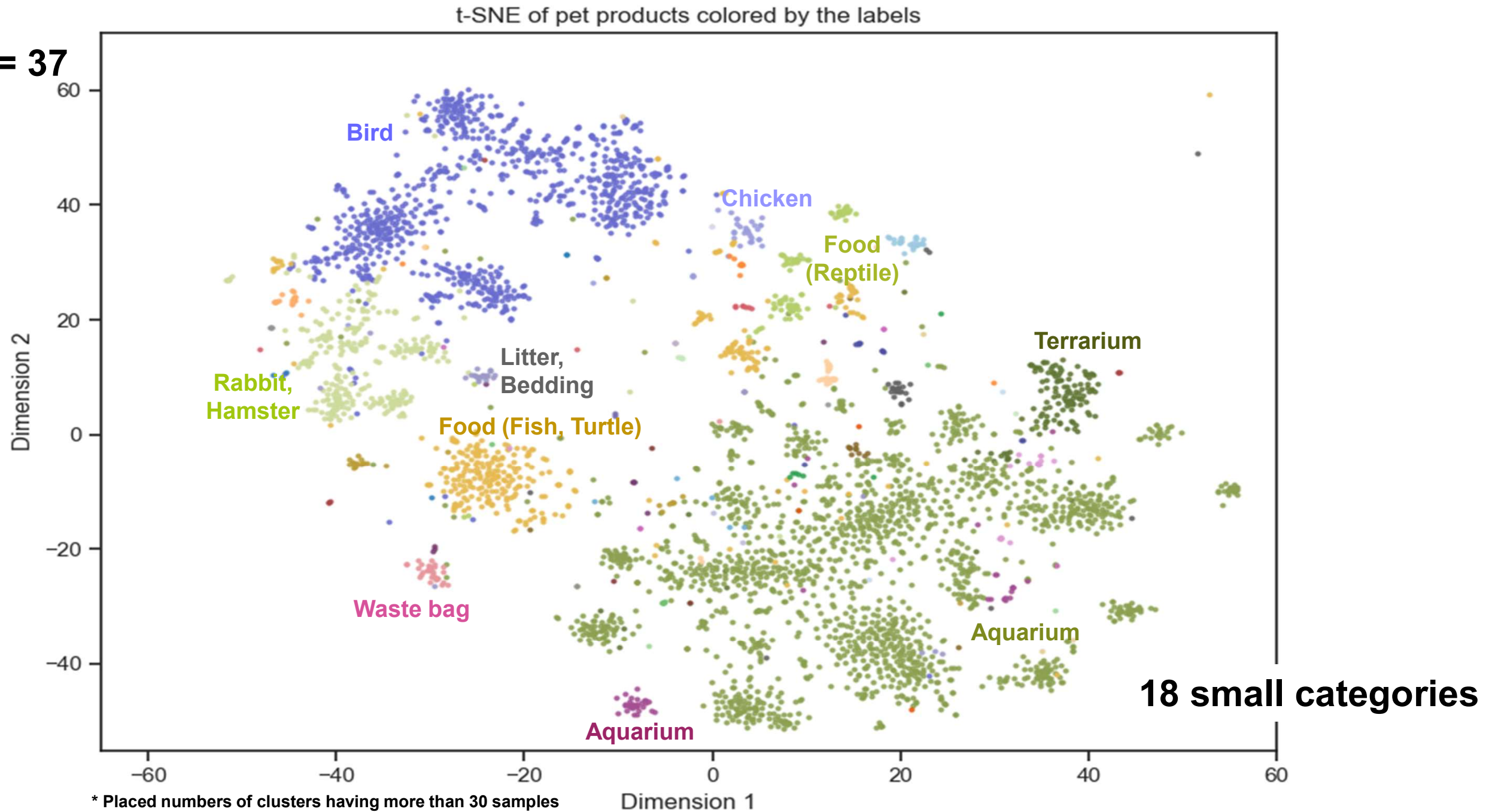
4.7. Summary Table

Each row represents one product with:
- **product id**
- **title**
- **animal category**
- **big category**
- **small categories**

End objective!

	product_id	product_title	animal	big_category	small_category
0	70064	Perfect Pet Soft Flap Cat Door with Telescoping FramePerfect Pet Soft Flap Cat Door with Telesco...	cat	door, cage, carrier, bed	door, cage
1	119780	ARK Naturals PRODUCTS for PETS 326066 4-Ounce Breath-Less Chewable Brushless Toothpaste, MiniARK...	dog	food, treat, treatment	treat
2	202371	Stella & Chewy's Freeze Dried Dog Food for Adult Dogs, Chicken Patties, 15 Ounce Bag - 2 PackSte...	dog	food, treat, treatment	food, bowl
3	291967	Premium Deshedding Brush for Dogs and Cats with Medium to Long Hair	Veterinary Approved	Rugg...	dog	body care, cleaning	brush, clipper
4	490904	Remington Coastal Pet R0206 GRN06 Rope Leash, 72-Inch, GreenRemington Coastal Pet R0206 GRN06 Ro...	dog	collar, leash	leash, harness
5	593896	Pet Food Can Covers Lids Set of 3Pet Food Can Covers Lids Set of 3Pet Food Can Covers Lids Set o...	cat	food, treatment	food, treat, water
6	674575	Scotch Pet Hair Roller 839RScotch Pet Hair Roller 839RScotch Pet Hair Roller 839RScotch Pet Hair...	other	other	brush, comb, clipper
7	690871	Petco Brooklyn 55 Gallon Metal Tank StandPetco Brooklyn 55 Gallon Metal Tank StandPetco Brooklyn...	other	fish, reptile	other
8	798322	Pet Dog Puppy Nonslip Canvas Sport Shoes Sneaker Boots Rubber Sole Size 5 Blue by MallofusaPet D...	dog	clothes	shoes
9	800175	Purina Pro Plan Focus Large Breed Formula Dry Dog FoodPurina Pro Plan Focus Large Breed Formula ...	dog	food, treat, treatment	food, bowl

* The first 10 products. 20,403 rows in the actual table.

5. Conclusion

✓ **Subcategorized 20,403 pet products in total by hierarchical agglomerative clustering**

✓ **Three-level nested categories (hierarchy)**

✓ **73 categories in total** (see slide 30)

< Further Work >

- **Fragment some larger categories:**

  E.g. Cat food and treat are in the same category now. They should be divided into different categories in some cases because the usage and purpose are different.

- **Adapt to larger scale:**

  K-means (roughly divide data) → Hierarchical clustering

Acknowledgment

This project was conducted as a part of
Data Science Career Track Course at

Thank you

Appendix

73 Categories

Animal	Big category	Small category	Count
dog	bed, crate, gate	bed	424
		door, gate, crate	475
		other	26
		step	100
		tie out	70
	body care, cleaning	brush, clipper	559
		dryer, towel	47
		ear cleaner	93
		eye care	116
		flea	275
		odor, stain, shampoo	857
	clothes	costume	672
		other	38
		shoes	133
	collar, leash	ID tag	126
		collar	1039
		leash, harness	803
	food, treat, treatment	food, bowl	1431
		oral care	78
		other	10
		treat	1494
		treatment, supplement	447
	other	calming	76
		car seat, cover	279
		memorial	27
		monitoring	10
		other	91
		stroller	74
		training	130
		training pad	264
		waste bag, carrier	436
	toy	toy	1242

Animal	Big category	Small category	Count
cat	collar, leash	collar	114
		harness, leash	28
	door, cage, carrier, bed	bed	193
		carrier, stroller	107
		door, cage	101
		other	9
		perch, shelf	26
		step	10
		tent	11
	food, treatment	cat grass	20
		food, treat, water	1220
		pill, treatment	202
		scale	6
	grooming	brush. comb	174
		clipper	37
		flea	155
		nail cap, furniture protector	49
	litter, odor, stain	litter, litter box	534
		odor, stain, shampoo	191
	other	calming	19
		memorial	19
		other	15
	toy, scratcher, cat tree	scratcher, cat tree	307
		toy	535

Animal	Big category	Small category	Count
other	bird, rabbit, hamster	bird	882
		chicken	73
		chinchila	18
		litter, bedding	30
		other	6
		rabbit, hamster	283
	fish, reptile	aquarium	1961
		crustacean	57
		food (fish, turtle)	348
		food (reptile)	89
		other	16
		terrarium	164
	other	bowl	29
		brush, comb, clipper	81
		memorial, tag	39
		other	150
		treatment	118
		waste bag	35