

<https://www.austintexas.gov/department/aac/faq>

To start, we had 111,157 total roles

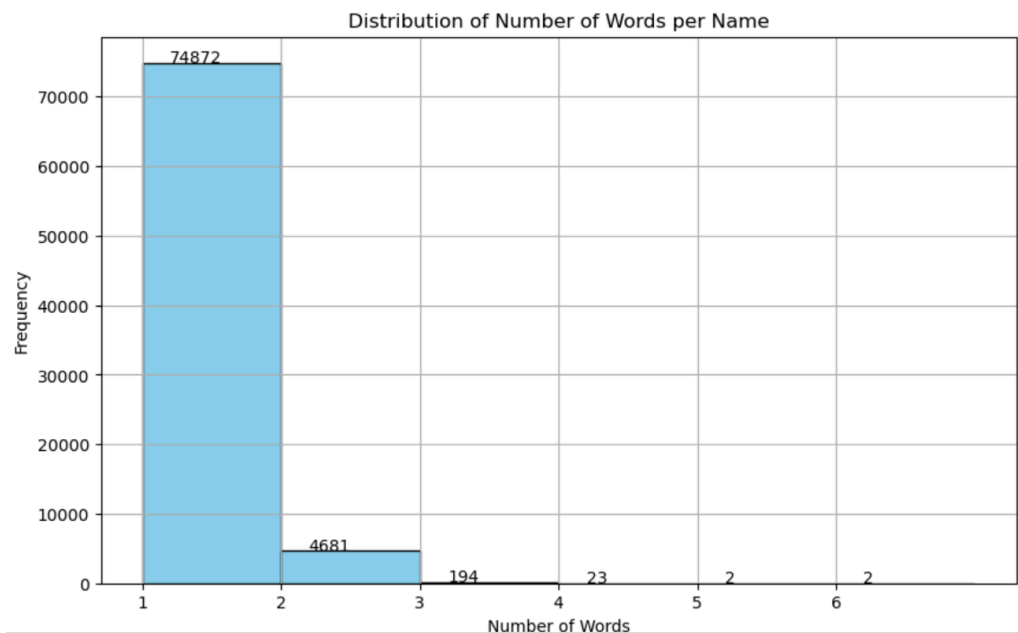
	id	name	intake_time	found_location	intake_type	intake_condition	animal_type	sex_upon_intake	age_upon_intake	breed	color
count	111157	79774	111157	111157	111157	111157	111157	111155	111156	111157	111157
unique	111157	24185	78769	48762	6	19	2	5	48	2440	569
top	A706918	Luna	09/23/2016 12:00:00 PM	Austin (TX)	Stray	Normal	Dog	Intact Male	2 years	Domestic Shorthair Mix	Black/White
freq	1	481	49	19071	82163	95010	61378	41060	16192	25361	11620

outcome_time	date_of_birth	outcome_type
111157	111157	111157
95134	8181	5
04/18/2016 12:00:00 AM	05/05/2014	Adoption
28	85	55044

Features to Explore

- **ID**
 - Get rid this attribute
- **Name**
 - Some names are accidentally animal IDs
 - Some have *s in front
 - Why? -> consider calling AAC to ask
 - The asterisk simply means that the pet didn't come in with a name so we named the pet here at the shelter!
 - Some are random numbers?
 - LOTS of empty names
 - Some animals have the grams they are weighed in → maybe this is an interesting feature
 - Featuring Engineering Idea
 - Maybe we can say if name is a “fruit / food” name,
 - We can't use sentiment analysis but we can see if Names that resemble inanimate objects vs names for people affect these rates

- We can check the names against a list of “most popular cat/dog names” to determine how “unique” the name is
- We should replace empty names with “NO_NAME”
- Some names need to be cleaned - look at parentheses and impute with what makes sense
- **FACTS**
 - **There are 79,774 names (31,383 missing values)**
 - **24,185 are unique names**
 - **The most common name was Luna (occurring 481 times)**
 - **There are 29,595 pets with a * in their name**
 - Named by the Austin Animal Shelter
 - **46 pets that have parentheses attached to their names**
 - Providing extra context
 - **Distribution of number of words per name**
 - Some of these include the parentheses



- **3,907 names are in the food.csv file**

```

name
Oliver      123
Ginger      114
Honey       105
Peanut       88
Sugar        77
...
Ice Milk      1
*Steamer      1
Baby Cakes    1
Rumplestilskin 1
*Nutter Butter 1
Name: count, Length: 1058, dtype: int64

```

- **3,890 animals have a number in them** (will rename to "NO_NAME")
-

• Found Location

- Sometimes have valid addresses, sometimes just say a city "Manor, TX"
- Featuring Engineering Idea
 - Translate to longitude and latitude (use ChatGPT)
 - Use perplexity AI to translate coordinates - we can maybe use proximity
 - Maybe an in_Austin column
 - **Made a city_name column + a street address name**
- **FACTS**
 - **20,565 pets were NOT found in Austin**
 - **1,049 pets were found outside Austin Animal Shelter jurisdiction**
 - About half of those who came from outside jurisdiction came from Owner Surrender

These are the Intake Types Among the Animals Found Outside Jurisdiction

Owner Surrender	539
Stray	327
Public Assist	169
Abandoned	13
Euthanasia Request	1

Name: count, dtype: int64

intake_condition	Aged	Injured	Medical	Neonatal	Neurologic	Normal	Nursing	Pregnant	Sick
intake_type									
Abandoned	0	0	0	3	0	9	0	0	1
Euthanasia Request	0	0	0	0	0	1	0	0	0
Owner Surrender	0	17	1	0	0	508	0	0	13
Public Assist	0	0	0	0	0	168	0	0	1
Stray	1	30	2	12	1	246	17	1	17

- These are the counts of the animals who came from outside the jurisdiction and the condition they were in

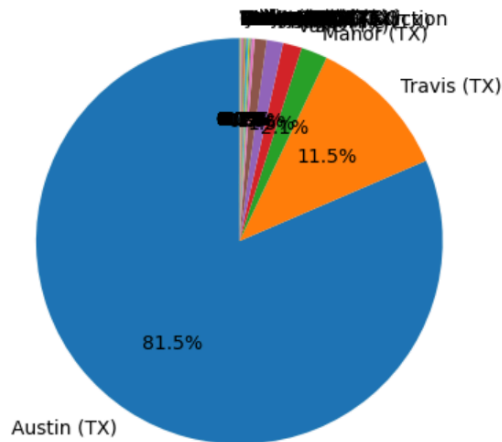
These are the Intake Conditions Among the Animals Found Outside Jurisdiction

Normal	932
Injured	47
Sick	32
Nursing	17
Neonatal	15
Medical	3
Pregnant	1
Aged	1
Neurologic	1

Name: count, dtype: int64

■ **Distribution of city names** (didn't get all of outside jurisdiction {missing 2})

Austin (TX)	90578
Travis (TX)	12741
Manor (TX)	2322
Valle (TX)	1671
Pflugerville (TX)	1462
Outside Jurisdiction	1047
Leander (TX)	257
Vista (TX)	161
Park (TX)	125
Lakeway (TX)	118
Creedmoor (TX)	118
Jonestown (TX)	98
Ridge (TX)	77
Hays (TX)	68
Rock (TX)	63
Bastrop (TX)	57
Cave (TX)	52
Webberville (TX)	43
Williamson (TX)	42
Valley (TX)	13
Hills (TX)	12
Venture (TX)	10
Caldwell (TX)	10
Burnet (TX)	5
Rollingwood (TX)	4
Hill (TX)	1
Blanco (TX)	1
Leanna (TX)	1



- City Name v Intake Type is more or less consistent with overall demographics

■ **There are 3,430 pets found on highways**

3430

These are the Intake Types Among the Animals Found on Highways intake_type

Stray 2715

Abandoned 403

Public Assist 231

Owner Surrender 81

Name: count, dtype: int64

These are the Intake Conditions Among the Animals on Highways intake_condition

Normal 2991

Injured 161

Sick 96

Nursing 81

Neonatal 61

Aged 11

Other 11

Pregnant 6

Feral 6

Medical 5

Behavior 1

Name: count, dtype: int64

●

- **There are 1,869 pets found somewhere on Interstate Hwy 35**

These are the Intake Types Among the Animals Found on the Interstate Hwy 35 intake_type

Stray	1621
Public Assist	163
Owner Surrender	54
Abandoned	31

Name: count, dtype: int64

These are the Intake Conditions Among the Animals on the Interstate Hwy 35 intake_condition

Normal	1649
Injured	105
Sick	55
Neonatal	26
Nursing	18
Aged	6
Feral	4
Medical	2
Other	2
Pregnant	1
Behavior	1

Name: count, dtype: int64

- There are 1,183 pets found on Levandar Loop
- There 378 pets found on Mopac
- There are 19,059 pets that were found on a corner
 - “And,” “and,” “&” in column
 - Cats - match dists

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
found_location					
False	22283	644	1690	1851	18022
True	2612	93	221	257	2106

•

- Dogs

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
found_location					
False	23345	252	1306	10919	11786
True	6804	52	232	3572	3110

•

• Intake Type

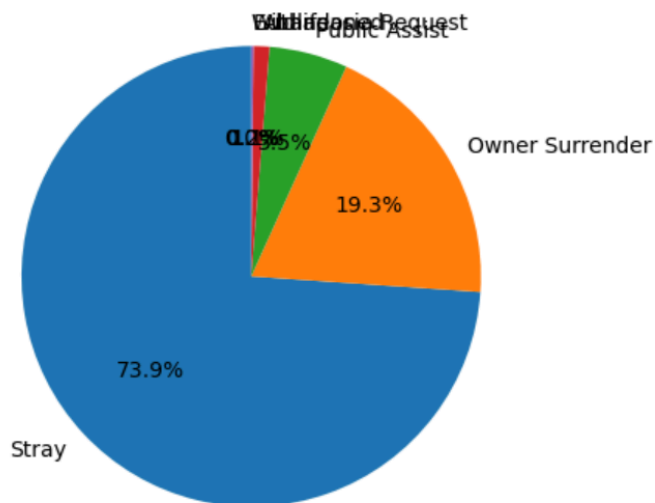
○ FACTS

- Most of the intake types are taking in as Stray, occurring (82,163 times)
- These are the counts of the types of intakes

```
intake_type
Stray      82163
Owner Surrender  21447
Public Assist  6131
Abandoned  1233
Euthanasia Request  182
Wildlife    1
Name: count, dtype: int64
```

- These are the percentages of how they dominate the data

```
intake_type
Stray      0.739162
Owner Surrender  0.192943
Public Assist  0.055156
Abandoned  0.011092
Euthanasia Request  0.001637
Wildlife    0.000009
Name: count, dtype: float64
```



- There are 0 missing values
- There is 1 value for Wildlife → I recommend we just remove it

- Though we don't know if the test set might have a lot of wildlife or something
- Euthanasia request is still really small but i think we should track it as well
- No additional cleaning recommended

✓ 0.05

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
intake_type					
Abandoned	777	5	8	112	331
Euthanasia Request	8	3	143	7	21
Owner Surrender	13870	148	587	1158	5684
Public Assist	1136	27	200	3896	872
Stray	39253	858	2511	11425	28116
Wildlife	0	0	0	1	0

• Intake Condition

○ FACTS

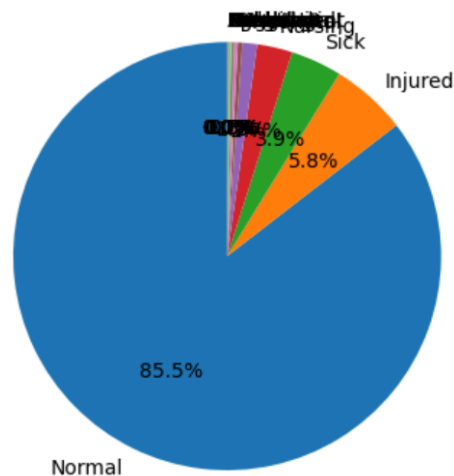
- There are some categories that can be combine - we can combine Unknown with Other (maybe),

```
array(['Normal', 'Injured', 'Pregnant', 'Neonatal', 'Sick', 'Nursing',
      'Aged', 'Unknown', 'Med Attn', 'Medical', 'Other', 'Feral',
      'Behavior', 'Med Urgent', 'Space', 'Agonal', 'Parvo', 'Neurologic',
      'Congenital'], dtype=object)
```

- I don't know what **'Behavior'** means? Maybe combine with Feral?
- What does **'Space'** mean? There are only 2 in that category, i think we can move it
- Medical stuff
 - Med Attn, Med Urgent (< 20 pets), and Agonal (1 pet) are very similar but deal with different severities
 - Parvo → contagious disease
 - Congenital - comes with diseases present
- No missing values
- The same even when you segmented by intake condition

intake_condition	
Normal	95010
Injured	6394
Sick	4295
Nursing	2957
Neonatal	1240
Aged	373
Medical	298
Other	247
Pregnant	111
Feral	104
Med Attn	48
Behavior	42
Unknown	12
Neurologic	10
Med Urgent	7
Parvo	5
Space	2
Agonal	1
Congenital	1

- Could combine unless we want to mention to urgency / severity of some of the medical conditions
 - Maybe separate into medical severity categories (severely injured or permanently disabled animals less likely to be adopted)
- Majority of set is normal on intake



- This is intake type v. intake condition

intake_condition	Aged	Agonal	Behavior	Congenital	Feral	Injured	Med Attn	Med Urgent	Medical	Neonatal	Neurologic	Normal	Nursing	Other	Parvo	Pregnant
intake_type																
Abandoned	0	0	0	0	0	19	0	0	12	74	0	1017	61	1	0	0
Euthanasia Request	29	0	0	0	0	12	0	1	0	0	0	65	0	1	0	0
Owner Surrender	85	1	16	0	6	523	12	2	45	102	1	19493	204	75	0	20
Public Assist	30	0	12	0	1	155	0	0	31	27	0	5592	96	36	0	4
Stray	229	0	14	1	97	5685	36	4	210	1037	9	68843	2596	134	5	87
Wildlife	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

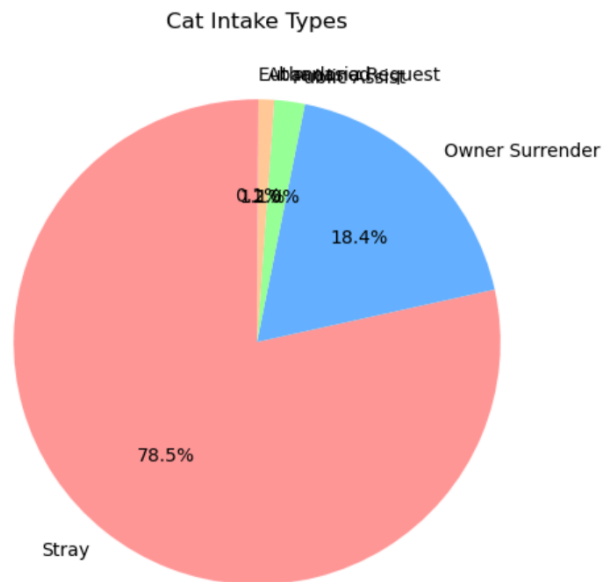
Space	Unknown
2	0
0	0
0	4
0	2
0	6
0	0

- THIS IS IT SEGMENTED BY ANIMAL

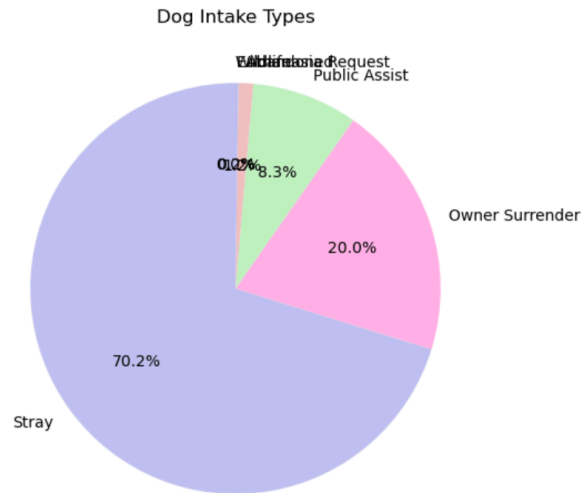
```

CAT INTAKE TYPES
intake_type
Stray          39063
Owner Surrender 9142
Public Assist  1009
Abandoned      523
Euthanasia Request 42
Name: count, dtype: int64

```



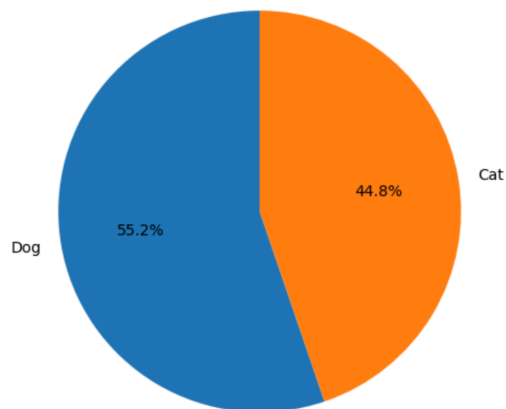
```
DOG INTAKE TYPES
intake_type
Stray          43100
Owner Surrender 12305
Public Assist  5122
Abandoned      710
Euthanasia Request 140
Wildlife        1
Name: count, dtype: int64
```



• Animal Type

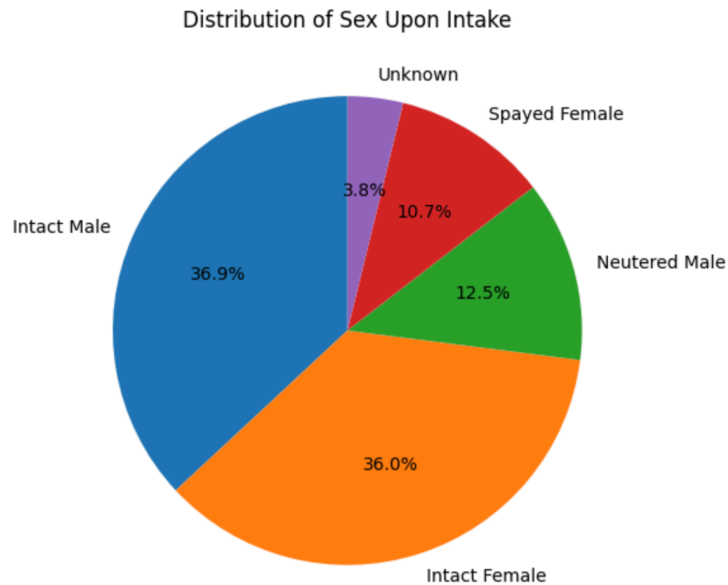
- Pretty good animal type distribution

Distribution of Animal Type



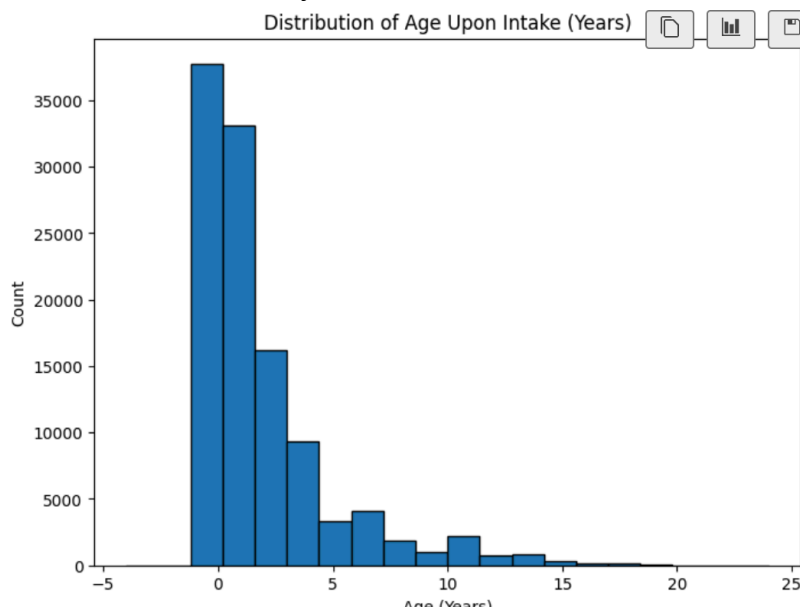
• Sex upon Intake

- One record has sex upon intake missing, I picked “Unknown” to fill it in. Only one record, so our choice here would not affect the result.



• Age upon Intake

- In days, weeks, months, years stored as strings
- A few negative values for some reason
- One that is “0 year”

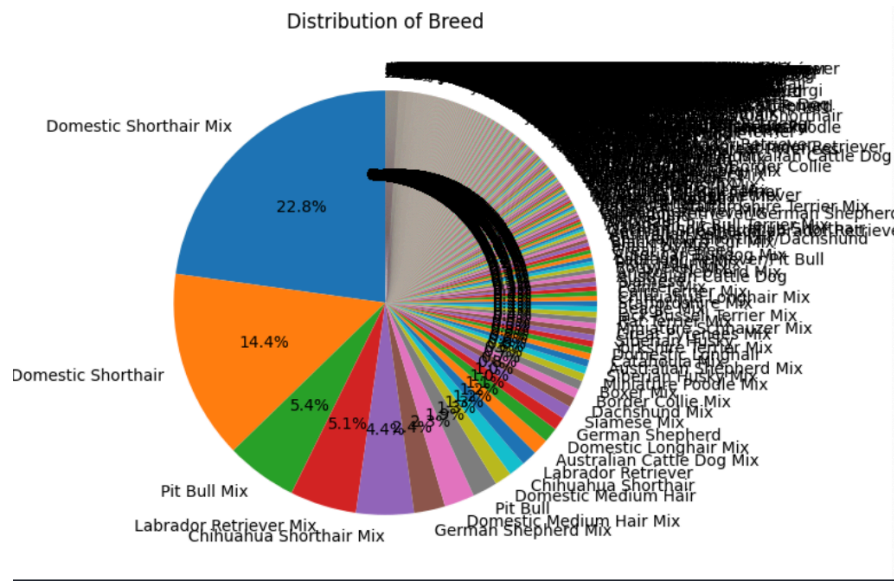


- Could consider changing age upon intake to be a boolean flag of sorts, like “age upon intake < 5 years”

• Breed

- LOTS of breeds that only appear once. 1111 records have a breed that no other record has. 2330 records have a breed that AT MOST 2 other records share (so 3 in total)

- Breed may not be a great predictor since nearly half of the records have a breed that takes up 1% or less of the total data.
- Add isMix column
- Only use breeds that have at least 1% of records



Unrelated but I also found that the dataset has one test record (as in, the dataset owners put in a record to test the dataset out)

	id	name	intake_time	found_location	intake_type	intake_condition	animal_type	sex_upon_intake	age_upon_intake	breed	color	outcome_time	date_of_birth	outcome_type
60146	A830333	Test	03/03/2021 06:30:00 PM	1234 Test Street in Austin (TX)	Stray	Normal	Dog	NaN	NaN	Kuvasz	Unknown	10/12/2022 02:51:00 PM	10/12/2022	Return to Owner

Rebecca's slay data exploration block ^ - - - -

● Color

- We need to standardize this
 - Primary Color / Other Color → We split by primary color
 - Has_secondary color

● Intake Time

- Hrs that was most popular among intakes
- Months of the year with outcome times

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
intake_month					
1	0.510633	0.006224	0.029824	0.193465	0.259855
2	0.497060	0.008031	0.035136	0.192170	0.267604
3	0.478174	0.008404	0.030607	0.194305	0.288510
4	0.464219	0.008904	0.030010	0.162361	0.334506
5	0.497355	0.012941	0.032962	0.119639	0.337104
6	0.504808	0.013341	0.034826	0.106558	0.340466
7	0.486806	0.012127	0.030268	0.122429	0.348370
8	0.510520	0.008006	0.025865	0.121831	0.333778
9	0.491390	0.009578	0.024752	0.135708	0.338571
10	0.489606	0.007845	0.032653	0.144244	0.325652
11	0.494180	0.005880	0.033961	0.167647	0.298332
12	0.519423	0.007146	0.031181	0.189814	0.252436

- Years vs outcome time

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
intake_year					
2013	0.412185	0.005446	0.070116	0.172226	0.340027
2014	0.392536	0.006628	0.063279	0.169481	0.368076
2015	0.413506	0.008946	0.051382	0.173145	0.353020
2016	0.475714	0.009749	0.025679	0.178795	0.310063
2017	0.476282	0.008314	0.023718	0.177928	0.313758
2018	0.492046	0.009294	0.017158	0.173995	0.307507
2019	0.483834	0.011104	0.017903	0.147379	0.339779
2020	0.522070	0.010480	0.028263	0.150842	0.288346
2021	0.584573	0.010789	0.023065	0.108879	0.272693
2022	0.588070	0.008819	0.022171	0.105463	0.275478
2023	0.616616	0.009273	0.017912	0.081047	0.275152
2024	0.610746	0.013916	0.021453	0.086200	0.267685

- Hrs vs outtake

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
intake_hour					
0	0.331878	0.008734	0.078603	0.314410	0.266376
1	0.242424	0.020202	0.070707	0.515152	0.151515
2	0.186667	0.000000	0.053333	0.666667	0.093333
3	0.120000	0.100000	0.120000	0.460000	0.200000
4	0.416667	0.000000	0.104167	0.312500	0.166667
5	0.325000	0.025000	0.050000	0.400000	0.200000
6	0.414414	0.022523	0.067568	0.162162	0.333333
7	0.350322	0.020782	0.070757	0.158337	0.399802
8	0.376553	0.021955	0.064623	0.167357	0.369511
9	0.427419	0.016513	0.049539	0.217742	0.288786
10	0.488286	0.009961	0.035418	0.190371	0.275964
11	0.504948	0.008022	0.026641	0.129809	0.330581
12	0.515896	0.008427	0.022983	0.127043	0.325651
13	0.515314	0.009218	0.026985	0.131876	0.316607
14	0.509701	0.007744	0.033142	0.134899	0.314514
15	0.510957	0.008475	0.027906	0.133025	0.319637
16	0.507109	0.009167	0.030683	0.147053	0.305987
17	0.510236	0.007693	0.027122	0.150085	0.304864
18	0.480860	0.007531	0.025416	0.184343	0.301851
19	0.452418	0.013261	0.034321	0.265991	0.234009
20	0.348269	0.014257	0.067210	0.303462	0.266802
21	0.401349	0.008432	0.052277	0.295110	0.242833
22	0.340659	0.015385	0.090110	0.274725	0.279121
23	0.307692	0.016194	0.046559	0.344130	0.285425

- Weekend v. time

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
is_weekend					
0	0.504870	0.009173	0.031013	0.142409	0.312536
1	0.464197	0.009981	0.031078	0.171493	0.323251

- Outcome Time**

- Maybe we can use the day we got the data set as the “outcome time”?
- **Months animals were adopted**
-

- Date of Birth**

- Age upon adopted
- **Months animals were born**

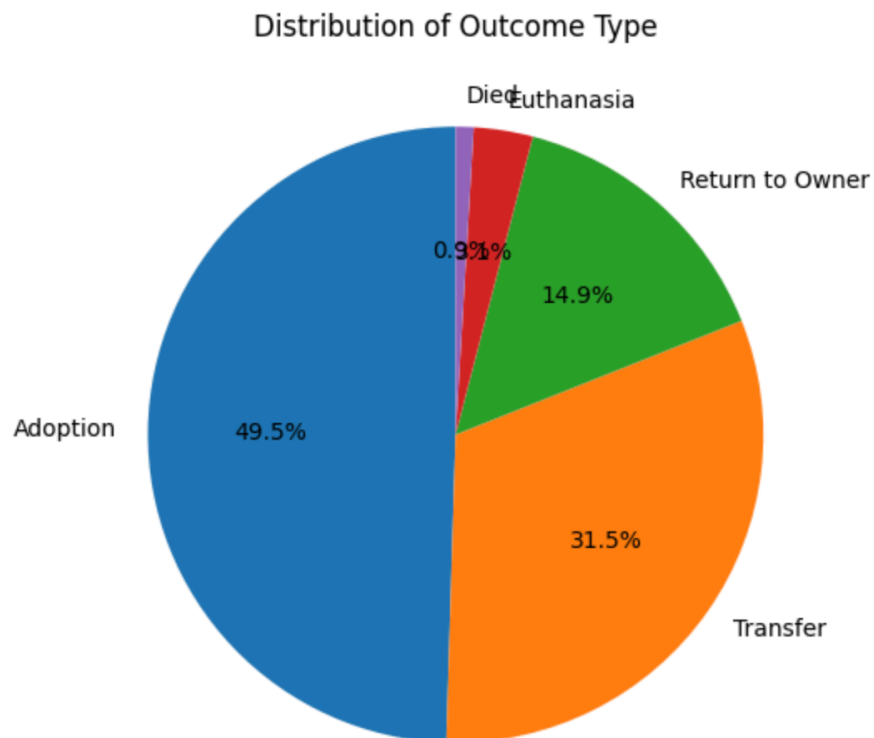
- Dog Breed Characteristics**

- Used a dog_breed_csv file to indicate:

- The size (big, small, very large)
- Lifespan (will ask chatgpt instead of taking it from the dataset)
- Dog breed group

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
Dog Breed Group					
Companion Dogs	0.405271	0.006888	0.021960	0.259932	0.305949
Herding Dogs	0.558189	0.004266	0.019673	0.197440	0.220431
Hound Dogs	0.466780	0.003407	0.020869	0.235094	0.273850
Sporting Dogs	0.528398	0.005218	0.020752	0.203762	0.241869
Terrier Dogs	0.471703	0.007208	0.019487	0.256540	0.245061
Working Dogs	0.482133	0.005426	0.030309	0.267727	0.214406

Outcome type



- Rare classes, died and Euthanasia

animal_type	outcome_type	count	percentage
Cat	Adoption	24894	50.01
Cat	Transfer	20128	40.44
Cat	Return to Owner	2108	4.23
Cat	Euthanasia	1911	3.84
Cat	Died	737	1.48
Dog	Adoption	30148	49.12
Dog	Transfer	14894	24.27
Dog	Return to Owner	14490	23.61
Dog	Euthanasia	1538	2.51
Dog	Died	304	0.50

Cats in top 10 colors: 72.97%

Dogs in top 10 colors: 53.37%

🐱 Cat Outcomes for Top 10 Colors			
	outcome_type	count	percentage
0	Adoption	18223	50.17
1	Died	543	1.49
2	Euthanasia	1313	3.61
3	Return to Owner	1493	4.11
4	Transfer	14752	40.61
5	TOTAL	36324	100.00

🐶 Dog Outcomes for Top 10 Colors			
	outcome_type	count	percentage
0	Adoption	16130	49.24
1	Died	191	0.58
2	Euthanasia	809	2.47
3	Return to Owner	7537	23.01
4	Transfer	8089	24.69
5	TOTAL	32756	100.00

🐱 Cat Outcomes (NOT in Top 10 Colors)			
	outcome_type	count	percentage
0	Adoption	6671	49.58
1	Died	194	1.44
2	Euthanasia	598	4.44
3	Return to Owner	615	4.57
4	Transfer	5376	39.96
5	TOTAL	13454	100.00

🐶 Dog Outcomes (NOT in Top 10 Colors)			
	outcome_type	count	percentage
0	Adoption	14018	48.98
1	Died	113	0.39
2	Euthanasia	729	2.55
3	Return to Owner	6953	24.30
4	Transfer	6805	23.78
5	TOTAL	28618	100.00

Most popular cat colors: color

Brown Tabby	11694
Black	11034
Orange Tabby	5508
Blue	3145
White	2938
Blue Tabby	2937
Tortie	2479
Calico	2363
Torbie	1846
Cream Tabby	1382
Lynx Point	769
Gray Tabby	543
Seal Point	522
Flame Point	325
Orange	311
Gray	300
Black Tabby	291
Black Smoke	214
Lilac Point	153
Tortie Point	134

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
color					
Agouti	13	0	0	0	6
Apricot	1	0	1	0	0
Black Brindle	0	0	1	0	1
Black Smoke	133	1	5	5	70
Black Tabby	107	4	15	6	159
Black Tiger	2	0	1	0	1
Blue Cream	42	0	6	5	26
Blue Point	57	1	5	5	33
Blue Smoke	10	0	0	2	3
Blue Tiger	0	0	0	0	1
Brown	27	2	11	8	42
Brown Brindle	0	0	1	0	1
Brown Merle	0	1	0	0	1
Brown Tiger	6	0	3	0	11
Buff	4	1	2	3	16
Calico Point	31	0	2	3	24
Chocolate	14	1	1	1	11
Chocolate Point	49	0	3	4	28
Cream	42	6	6	8	71
Cream Tiger	1	0	1	0	0
Fawn	1	0	1	0	1
Flame Point	186	5	6	23	105
Gray	53	8	34	31	174
Gray Tabby	165	15	34	25	304
Lilac Point	71	0	7	5	70
Lynx Point	428	12	26	45	258
Orange	78	9	35	24	165
Orange Tiger	2	0	1	0	1
Pink	0	0	0	0	1
Red	0	0	2	1	0
Red Tick	1	0	0	0	0
Sable	1	0	0	0	1
Seal Point	246	8	29	35	204
Silver	9	0	1	1	8
Silver Lynx Point	10	0	0	2	14
Silver Tabby	66	0	4	6	57
Tan	2	0	4	2	8
Tortie Point	81	3	4	13	33
Tricolor	1	0	2	0	2
Yellow	2	0	2	1	2

outcome_type	Adoption	Died	Euthanasia	Return to Owner	Transfer
color					
Agouti	68.4	0.0	0.0	0.0	31.6
Apricot	50.0	0.0	50.0	0.0	0.0
Black Brindle	0.0	0.0	50.0	0.0	50.0
Black Smoke	62.1	0.5	2.3	2.3	32.7
Black Tabby	36.8	1.4	5.2	2.1	54.6
Black Tiger	50.0	0.0	25.0	0.0	25.0
Blue Cream	53.2	0.0	7.6	6.3	32.9
Blue Point	56.4	1.0	5.0	5.0	32.7
Blue Smoke	66.7	0.0	0.0	13.3	20.0
Blue Tiger	0.0	0.0	0.0	0.0	100.0
Brown	30.0	2.2	12.2	8.9	46.7
Brown Brindle	0.0	0.0	50.0	0.0	50.0
Brown Merle	0.0	50.0	0.0	0.0	50.0
Brown Tiger	30.0	0.0	15.0	0.0	55.0
Buff	15.4	3.8	7.7	11.5	61.5
Calico Point	51.7	0.0	3.3	5.0	40.0
Chocolate	50.0	3.6	3.6	3.6	39.3
Chocolate Point	58.3	0.0	3.6	4.8	33.3
Cream	31.6	4.5	4.5	6.0	53.4
Cream Tiger	50.0	0.0	50.0	0.0	0.0
Fawn	33.3	0.0	33.3	0.0	33.3
Flame Point	57.2	1.5	1.8	7.1	32.3
Gray	17.7	2.7	11.3	10.3	58.0
Gray Tabby	30.4	2.8	6.3	4.6	56.0
Lilac Point	46.4	0.0	4.6	3.3	45.8
Lynx Point	55.7	1.6	3.4	5.9	33.6
Orange	25.1	2.9	11.3	7.7	53.1
Orange Tiger	50.0	0.0	25.0	0.0	25.0
Pink	0.0	0.0	0.0	0.0	100.0
Red	0.0	0.0	66.7	33.3	0.0
Red Tick	100.0	0.0	0.0	0.0	0.0
Sable	50.0	0.0	0.0	0.0	50.0
Seal Point	47.1	1.5	5.6	6.7	39.1
Silver	47.4	0.0	5.3	5.3	42.1
Silver Lynx Point	38.5	0.0	0.0	7.7	53.8
Silver Tabby	49.6	0.0	3.0	4.5	42.9
Tan	12.5	0.0	25.0	12.5	50.0
Tortie Point	60.4	2.2	3.0	9.7	24.6
Tricolor	20.0	0.0	40.0	0.0	40.0
Yellow	28.6	0.0	28.6	14.3	28.6

