Faster Pet Adoption

Strategic Marketing Solutions

by Elsa Velázquez

http://elsa-data-sci.tech

About Me

I ♥ python + **Data Engineering**

College of Engineering & Applied Science
UNIVERSITY OF COLORADO BOULDER
BS Computer Science

+

TEXAS A&M UNIVERSITY
BS Psychology

http://elsa-data-sci.tech

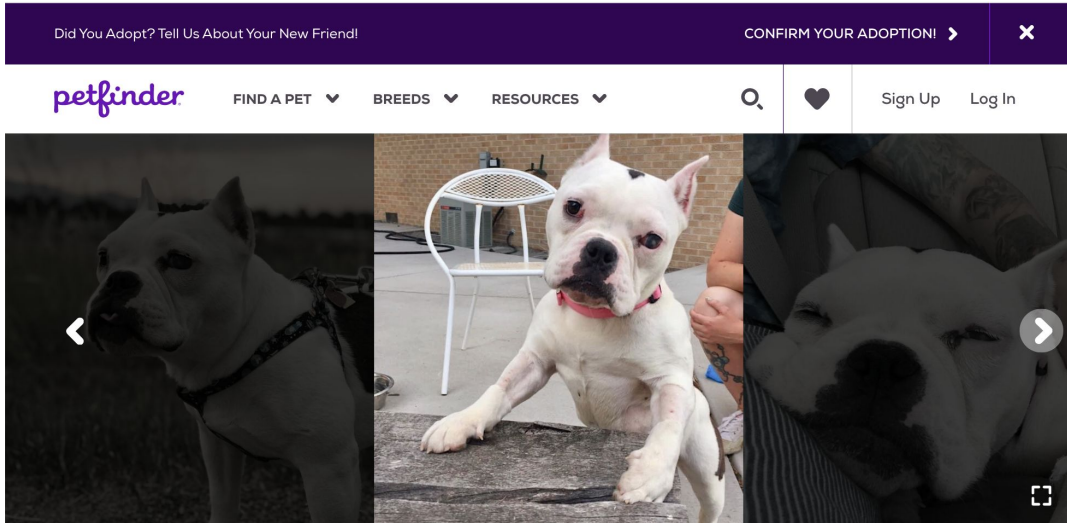
Why This Project

I am a disabled veteran.

My service dogs saved my life.

Now it’s my turn to help save the dogs.

Example Dog Description



Tiny Tina - I need a foster (or forever) hero!

Pit Bull Terrier Mix • Boulder, CO

Puppy • Female • Medium

About

Meet Tiny Tina - I need a foster (or forever) hero!

As Tiny Tina rounds the corner into her final procedures and gets ready to start her adoption journey, we are looking for a foster (or forever) home to hold her paw through it!

After 3 months of loving this girl and nursing her into a whole new dog, Tina's current fosters are hitting the road for new adventures and we are looking for the perfect people to take over their hard work. With one more surgery (on her eye and soft palette) coming up, Tina is so close to her forever finish line, so you will need to be comfortable with a bit of nursing this sweet meatball.

Tina loves being active as much as she loves napping and she does well with other pups, but can be a bit protective of her food/toys, which is understandable given her horrible past. You will have full support from the Underdogs team and we like to think we are a pretty fun bunch :)

If you're interested in fostering Tina or giving her a forever home, you can email kate@underdogsrescue.org or fill out the foster application on our website. And please share this if you know someone you think would be perfect for her!

http://elsa-data-sci.tech

My Data Pipeline

Extract Get the data to my computer.

Transform Clean the data.

Example of 1 dog record:

[{"animal": {"id": 48555420, "organization_id": "IL599", "url": "https://www.petfinder.com/dog/ruff-48555420/il/batavia/starfish-animal-rescue-il599/?referrer_id=5957d654-0b8d-4a02-bbae-6c7dd49e1074",
"type": "Dog", "species": "Dog", "breeds": {"primary": "Terrier", "secondary": "Pug", "mixed": "True", "unknown": "False"}, "colors": {"primary": "Brindle", "secondary": "Brown / Chocolate", "tertiary": "None"}, "age":
"Young", "gender": "Male", "size": "Small", "coat": "Short", "attributes": {"spayed_neutered": "True", "house_trained": "True", "declawed": "None", "special_needs": "False", "shots_current": "True"}, "environment":
{"children": "None", "dogs": "True", "cats": "None"}, "tags": "[]", "name": "Ruff", "description": "Meet Ruff, Ruff is a smaller breed - probably around 10 lbs at 6-7 months old and a complete mix...",
"organization_animal_id": "None", "photos": {"small":"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=100",
“medium":https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=300",“large":"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=600",
"full":"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733"}, {"small":"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/1/?bust=1595364732&width=100", "medium":
"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/1/?bust=1595364732&width=300", "large": "https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/1/?bust=1595364732&width=600", "full":
"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/1/?bust=1595364732"}, {"small": "https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/2/?bust=1595364733&width=100", "medium":
"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/2/?bust=1595364733&width=300", "large": "https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/2/?bust=1595364733&width=600", "full":
"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/2/?bust=1595364733"}], "primary_photo_cropped": ll":"https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=300",
"medium": ttps://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=450",
"large": "https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733&width=600", "full": "https://dl5zpyw5k3jeb.cloudfront.net/photos/pets/48555420/3/?bust=1595364733"}, "videos": "[]",
"status": "adopted", "status_changed_at": "2020-08-16T18:21:00+0000", "published_at": "2020-07-21T20:52:42+0000", "distance": "None", "contact": {"email": "starfishanimalrescuer@gmail.com", "phone": "None",
"address": {"address1": "Batavia", "address2": "None", "city": "Batavia", "state": "IL", "postcode": "60510", "country": "US"}}, "_links": {"self": {"href": "/v2/animals/48555420"}, "type": {"href": "/v2/types/dog"},
"organization": {"href": "/v2/organizations/il599"}}}},

Load Put the data in a SQL database.

http://elsa-data-sci.tech

Natural Language Processing (NLP)

My analysis showed
words mattered more than pictures.

Hence, NLP to get dogs adopted faster.

Predicting the Likelihood of Adoption

My dataset is divided into 2 groups.

Already Adopted

Still In Shelters

http://elsa-data-sci.tech

Effective Advertising

Classify the Description

?

Already Adopted

Still In Shelter

http://elsa-data-sci.tech

Examples of Effective Descriptions

- This happy go lucky furball is looking for furever.

- Sweet affectionate husky, energetic and playful, loves other dogs.

Examples of Ineffective Descriptions

- I am a 2 month old female Boxer mix.
- Seems house trained, not sure about other dogs or cats.
- *Leaving the description blank.*

Word Trends In the Two Groups

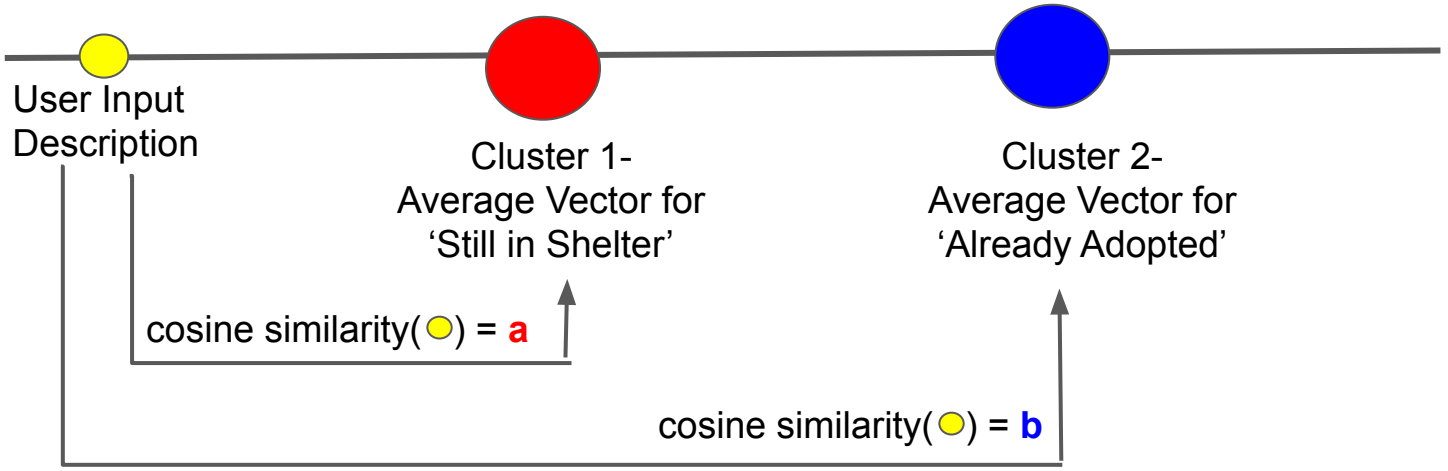
Already Adopted Dogs

Dogs Still In Shelters

http://elsa-data-sci.tech

Term Frequency- Inverse Document Frequency (TF-IDF)

1. K-Means on the TF-IDF matrix with 2 clusters:



2. cosine similarity **a** < cosine similarity **b**

3. The user input is closer to centroid , so the input is classified as more similar to still in shelters

4. Business recommendation- change the description

App Demo

Faster Pet Adoption

Strategic Marketing Solutions

Home | Input a Dog Description | About | Contact

Will your description get the dog adopted?

Please input the dog's description and hit submit.

take this sweet playful puppy home

Submit

This app is intended for dog shelters showcasing adoptable dogs.
It provides immediate feedback on the effectiveness of the dog's description.

http://elsa-data-sci.tech

http://elsa-data-sci.tech

App Results (Ideally)

Faster Pet Adoption

Strategic Marketing Solutions

Home | Input a Dog Description | About | Contact

Will Your Description Get the Dog Adopted?

Your input:

take this sweet playful puppy home

Result:

Yes, keep this description.

http://elsa-data-sci.tech

App Results (Currently)

Faster Pet Adoption

Strategic Marketing Solutions

Home | Input a Dog Description | About | Contact

Faster Pet Adoption

Will Your Description Get the Dog Adopted?

Your input:
blah blah

Multinomial Naive Bayes F1 Score 86% -> the likelihood of adoption as:
More Likely Than Not to be Adopted

TFIDF score: How Close Is Your Description to Dogs that Have Already Been Adopted
manhatten cos_sim euclidean 0 18.917491 0.000000 2.220446e-16 1 18.004218 0.000000 8.215650e-15 2 20.545064 0.076269 3.592138e-01 3 19.626165 0.083636 3.537825e-01 4 20.545064 0.076269 3.592138e-01 5 19.626165 0.083636 3.537825e-01 6 20.078434 0.086029 3.520143e-01 7 19.147070 0.095240 3.451838e-01 8 18.917491 0.000000 2.220446e-16

TFIDF score: How Close Is Your Description to Dogs that Are Still In Shelters
0.0

Sentiment140, based on 1600000 records, classifies the description's sentiment as: matrix has 1600000 rows (documents) and 686637 columns (words) accuracy score of 76%
Negative Sentiment

http://elsa-data-sci.tech

Recommendations

- Make the description personalized.
- Pictures are less distinguishing.
- Don’t leave the description blank.

Contact Me

www.elsa-data-sci.tech

www.linkedin.com/in/elsa-data-sci/

www.github.com/elsaVelazquez/

elsa.velazquez@colorado.edu

I ♥ python

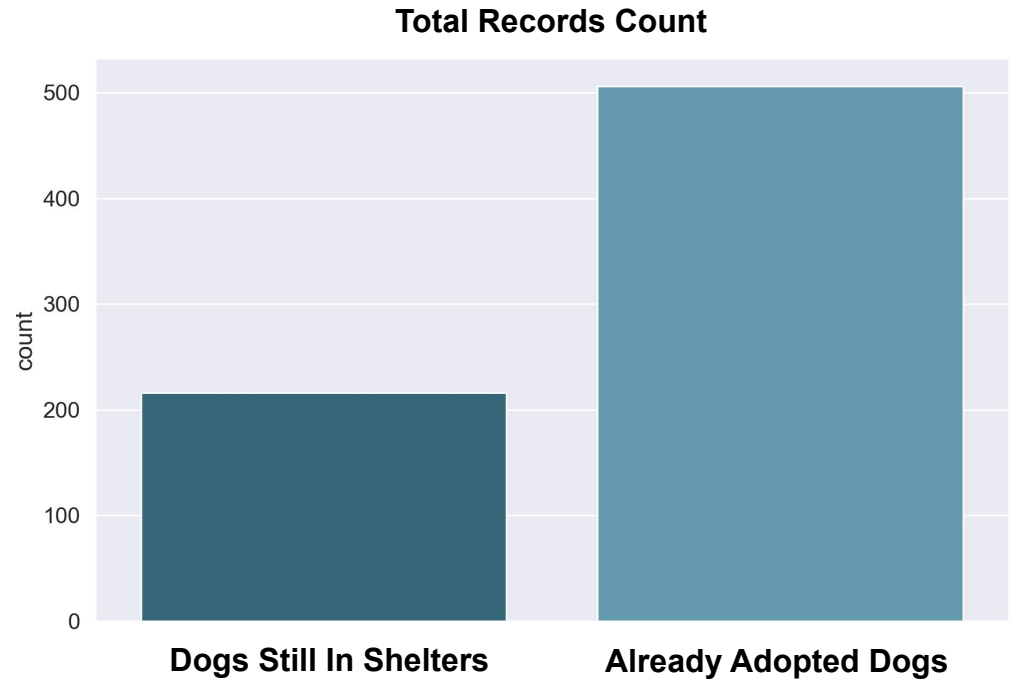
Appendix- by Slide Number

18. *The Dataset*
19. *Turning Words Into Numbers*
20. *Tuning the Model*
21. *Why Not Only Multinomial Naive Bayes*
22. *Sentiment Analysis*
23. *Why Not Images (slides 23 - 27)*
28. *Why Not Age, Gender, Size or Color*
29. *A Window In Time- Time Series*
30. *Initial Exploratory Data Analysis*
31. *Initial Hypothesis*
32. *EDA Using PySpark*
33. *My Credentials*
34. *Gratitude*

http://elsa-data-sci.tech

The Dataset

- Imbalanced Dataset
- 22 records imputed with “None”



Turning Words Into Meaningful Numbers

Term Frequency- Inverse Document Frequency (TF-IDF) Matrix

TF- Term Frequency
Compensates for document length.

$$TF = \frac{\text{Total times } \textit{word} \text{ appears in the document}}{\text{Total number of words in the document}}$$

IDF: Inverse Document Frequency
Rarer words are more important.
ex: i*s, be, a, the,* are less important.

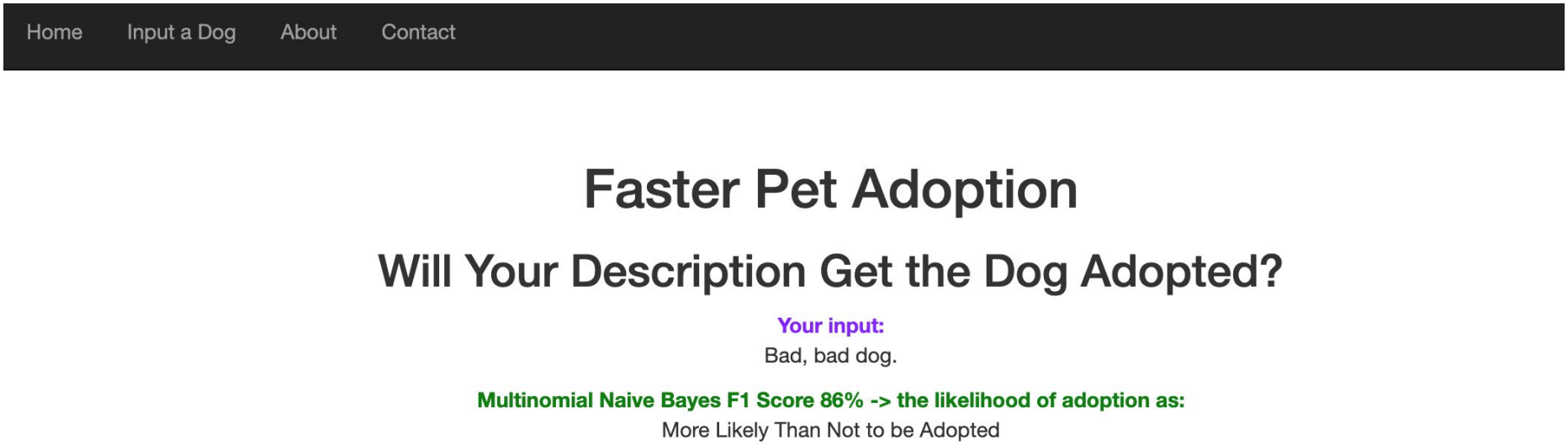
$$IDF = \frac{\text{Total number of documents}}{\text{Number of documents containing the } \textit{word}}.$$

Tuning the Model

- Bag of Words per Cluster
- Stopwords=
  English words + loving, sweet, friendly, dog
- LaPlace Smoothing

Why Not Only Multinomial Naive Bayes

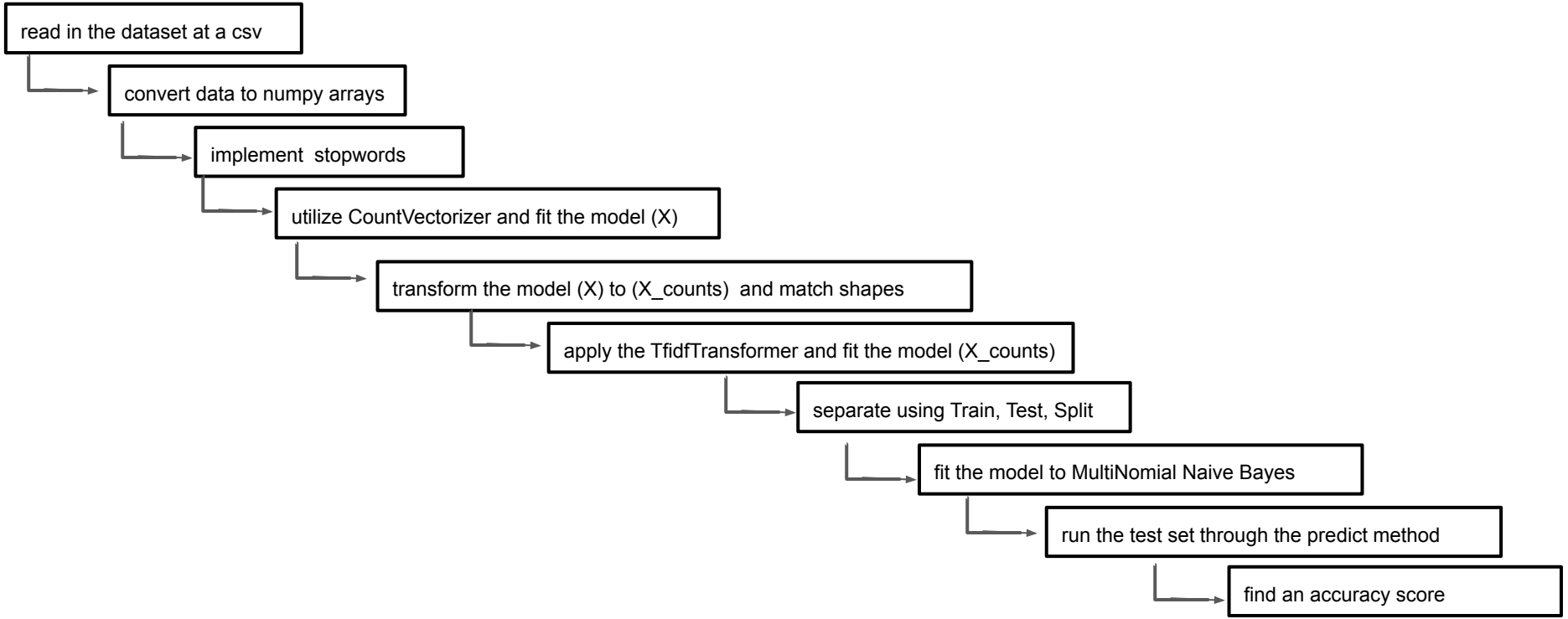
- F1 Score of 83 %
- Non-intuitive results



http://elsa-data-sci.tech

Sentiment Analysis

- Use sklearn's: MultinomialNaiveBayes, feature_extraction, CountVectorizer, TfidfTransformer, f1_score, model_selection

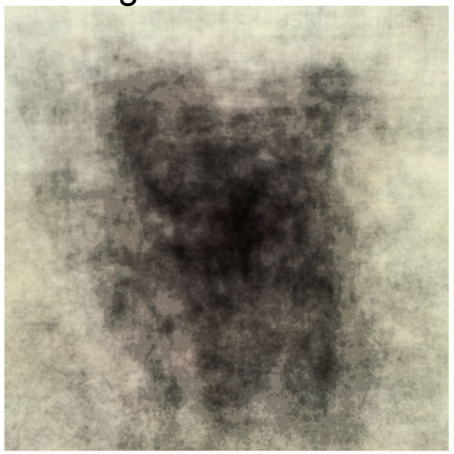


Why Not Images

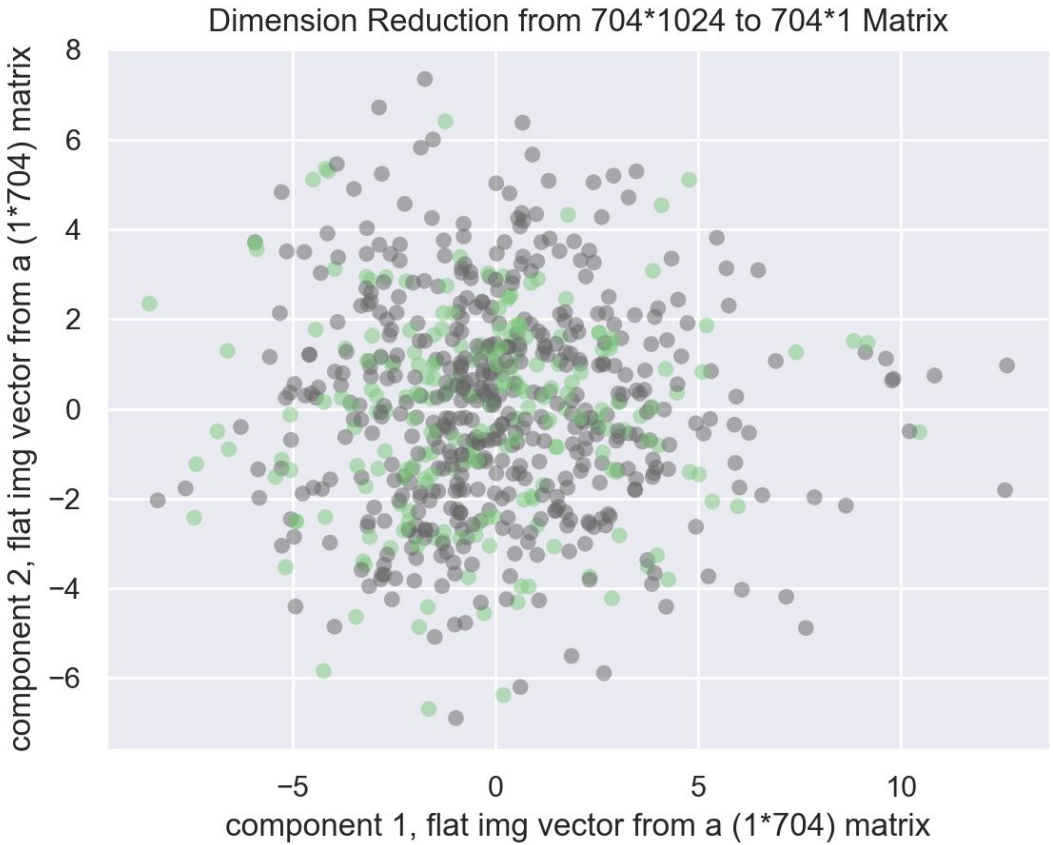
- The 2 average images did not show much difference.
- I investigated further (next slide).

Average Already Adopted

Average Still In Shelters

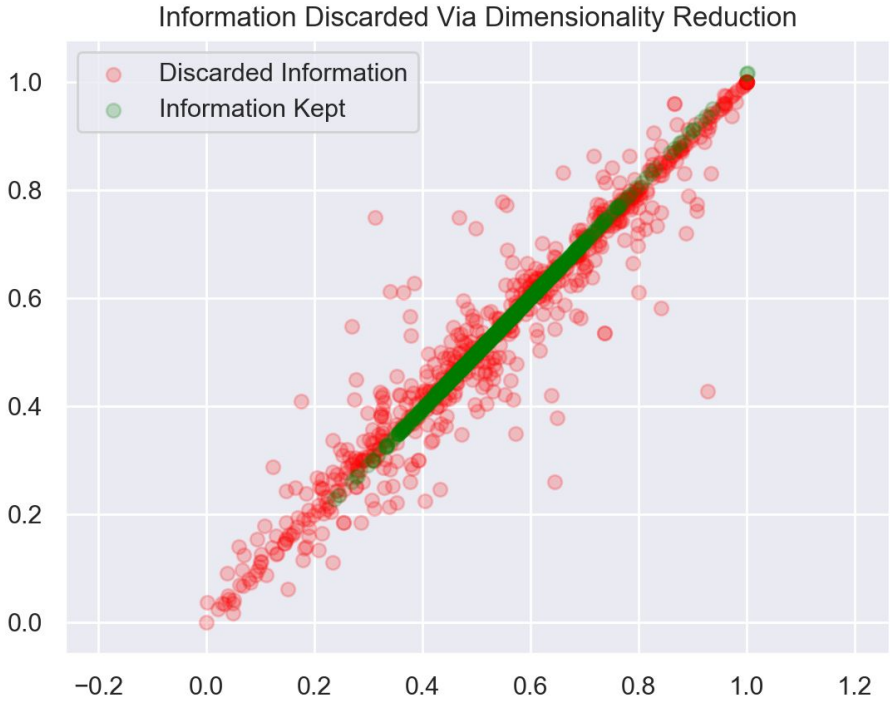


Principal Component Analysis On Images



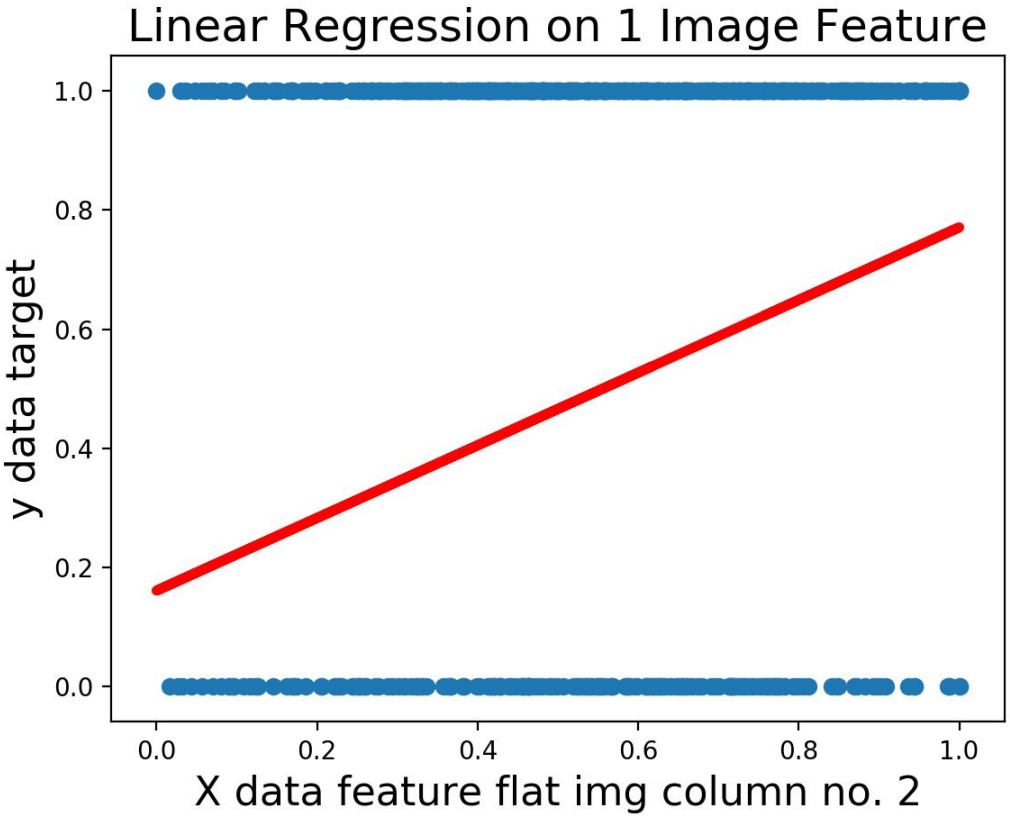
- No distinct clusters means there is no signal.

Dimensionality Reduction Results



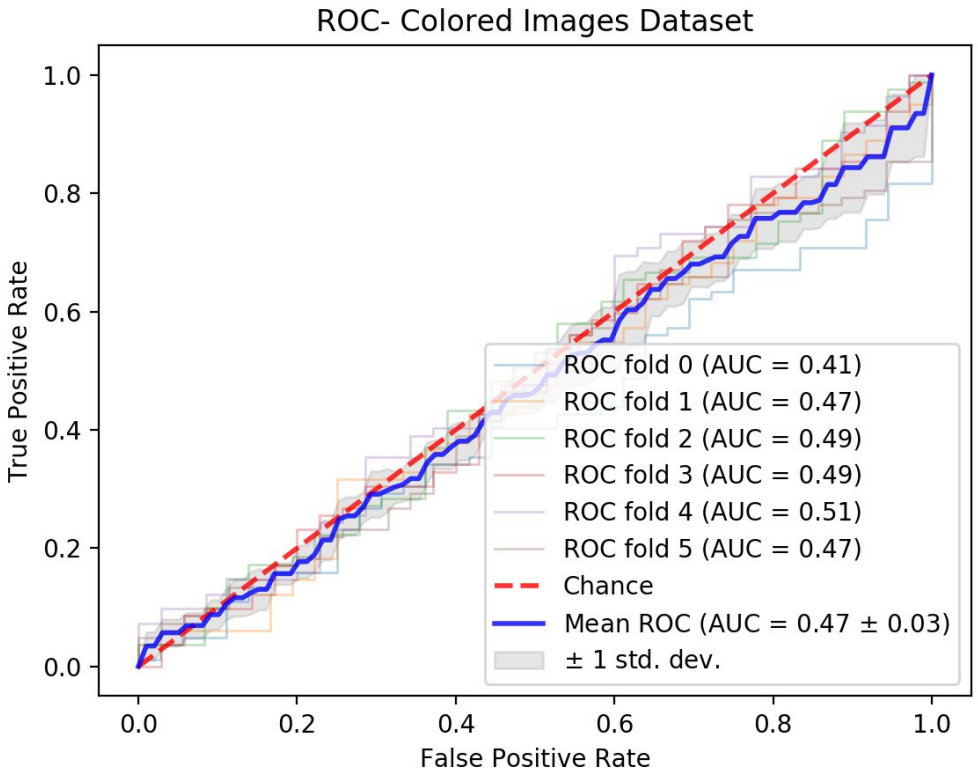
- The principal components did not yield significant data compared to keeping all components.
- There was no clear signal that was being missed in the discarded components.

Linear Regression by Feature



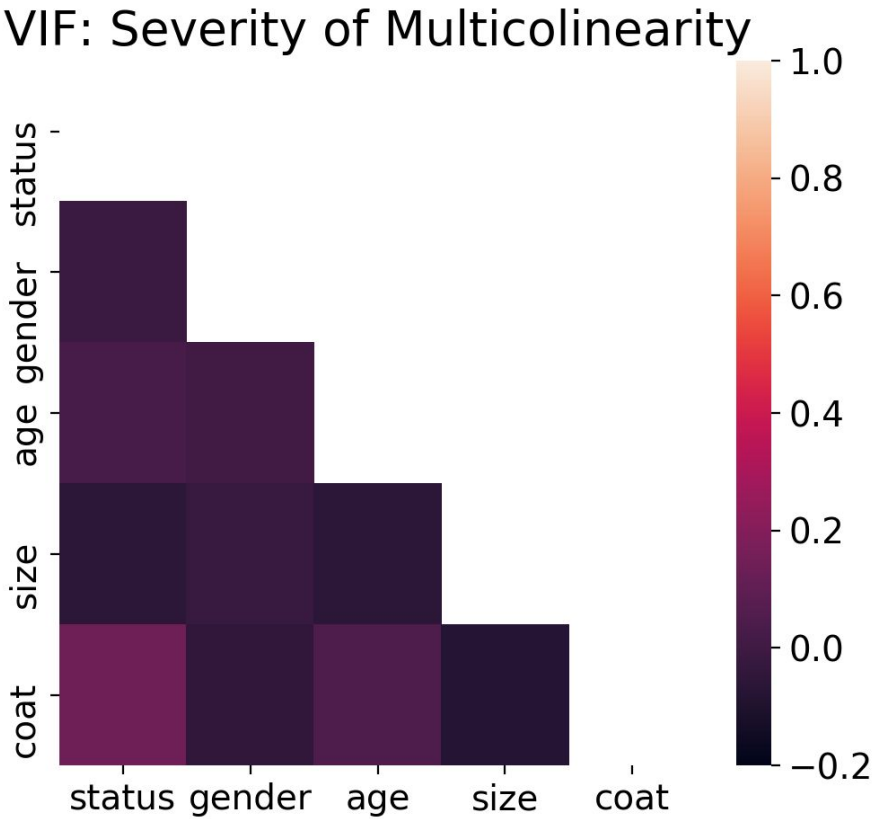
- No distinct signal.

Receiver Operator Characteristic (ROC) Curve



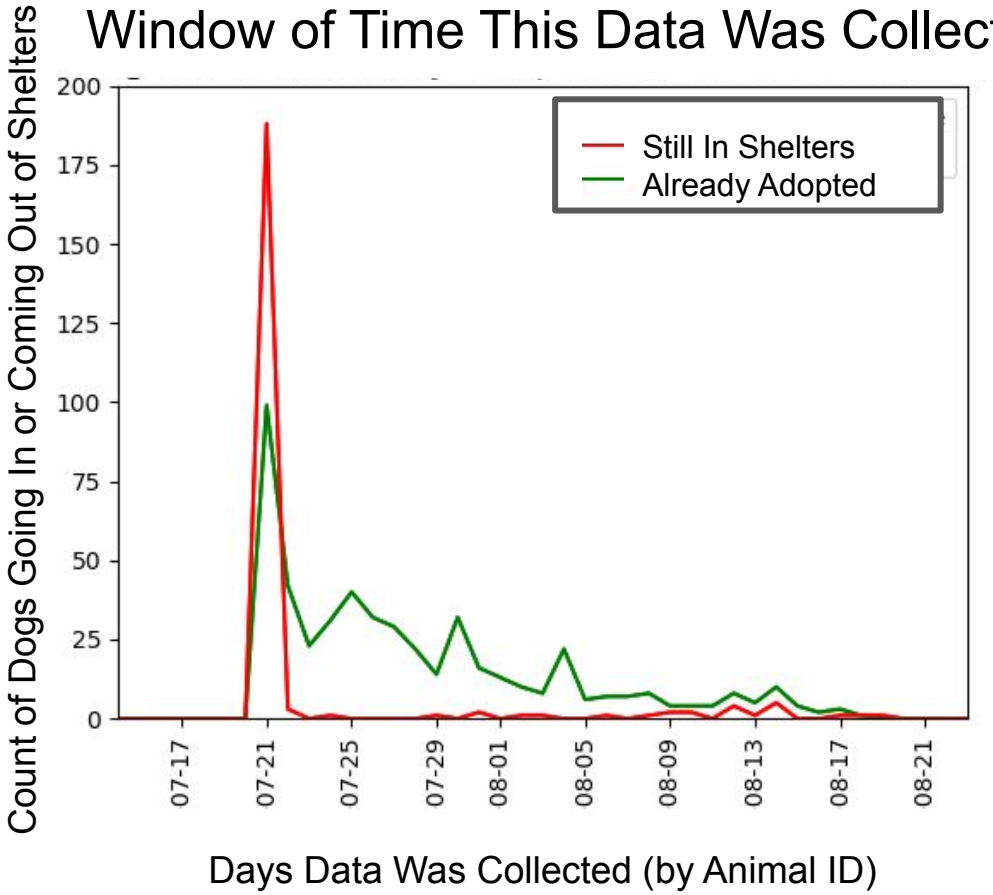
- No chance for random error because data was randomly sampled 5 different ways, and still showed no signal.

Why Not Age, Gender, Size or Color



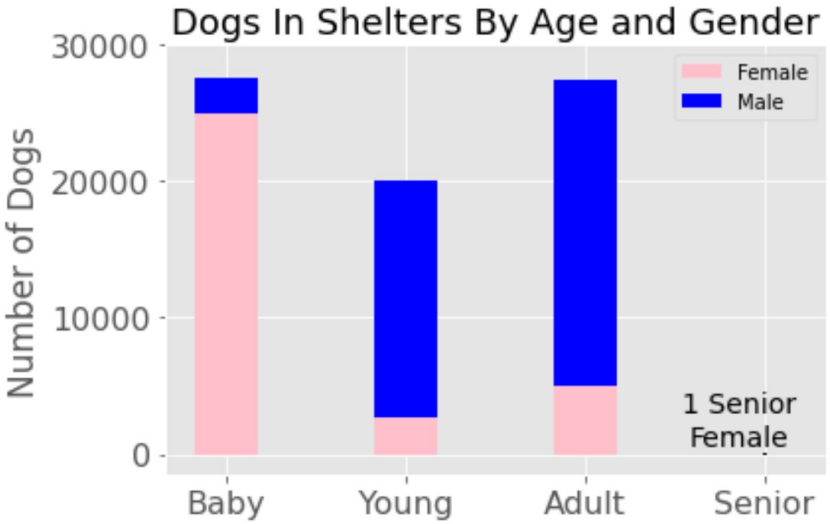
- **Variance Inflation Factor** (VIF) shows the ratio of the overall model variance is equal to the variance of the model per individual variable.

A Window In Time- Time Series



- COVID may have impacted these results.
- The data was collected immediately post-COVID international lock-downs.
- Many reliable news sources indicated dogs were being adopted at higher rates than pre-COVID.

Initial Exploratory Data Analysis (EDA)

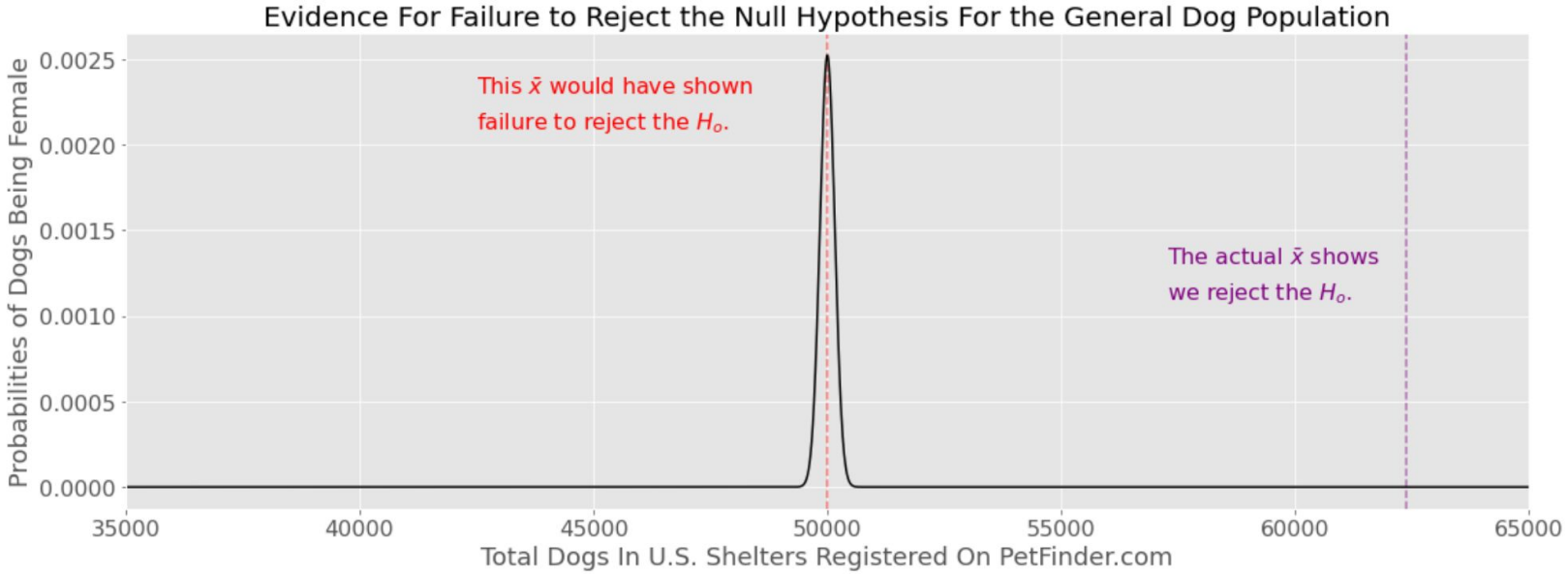



```
+------+------+-----+
|gender|   age|Count|
+------+------+-----+
|  Male| Adult|30001|
|Female| Young|22378|
|Female|  Baby|22376|
|Female| Adult|17623|
|  Male|  Baby| 5001|
|  Male| Young| 2620|
|Female|Senior|    1|
+------+------+-----+
```

	gender	breeds	age	Number
0	Female	(True, Chihuahua, None, False)	Baby	9998
1	Male	(False, Dachshund, None, False)	Adult	9998
2	Male	(True, German Shepherd Dog, None, False)	Baby	4999
3	Male	(False, Hound, None, False)	Adult	4999
4	Female	(True, Maltese, Chihuahua, False)	Baby	4999
5	Female	(False, Chihuahua, None, False)	Young	4999
6	Male	(True, Labrador Retriever, Hound, False)	Adult	4999
7	Male	(True, Australian Shepherd, None, False)	Adult	4999
8	Female	(False, Rottweiler, None, False)	Young	4999
9	Female	(True, Chihuahua, Mixed Breed, False)	Young	4999
10	Female	(True, Jack Russell Terrier, Chihuahua, False)	Baby	4758
11	Female	(False, Pit Bull Terrier, None, False)	Adult	2622
12	Female	(False, German Shepherd Dog, None, False)	Baby	2620
13	Male	(False, Terrier, None, False)	Young	2620
14	Female	(True, Terrier, None, False)	Young	2379
15	Male	(False, Pit Bull Terrier, None, False)	Adult	2
16	Male	(True, Pit Bull Terrier, Pointer, False)	Adult	1
17	Female	(True, Pit Bull Terrier, None, False)	Adult	1
18	Female	(True, Chihuahua, Whippet, False)	Young	1
19	Male	(True, Labrador Retriever, Newfoundland Dog, F...	Adult	1
20	Female	(True, Hound, Labrador Retriever, False)	Young	1

Initial Hypothesis- Failure to Reject Null

- Initial hypothesis was that we would see differences by gender.
- The initial analysis showed there was no reason to explore by gender.



EDA Using PySpark SQL

```
root
 |-- _links: struct (nullable = true)
 |    |-- organization: struct (nullable = true)
 |    |    |-- href: string (nullable = true)
 |    |-- self: struct (nullable = true)
 |    |    |-- href: string (nullable = true)
 |    |-- type: struct (nullable = true)
 |    |    |-- href: string (nullable = true)
 |-- age: string (nullable = true)
 |-- attributes: struct (nullable = true)
 |    |-- declawed: string (nullable = true)
 |    |-- house_trained: boolean (nullable = true)
 |    |-- shots_current: boolean (nullable = true)
 |    |-- spayed_neutered: boolean (nullable = true)
 |    |-- special_needs: boolean (nullable = true)
 |-- breeds: struct (nullable = true)
 |    |-- mixed: boolean (nullable = true)
 |    |-- primary: string (nullable = true)
 |    |-- secondary: string (nullable = true)
 |    |-- unknown: boolean (nullable = true)
 |-- coat: string (nullable = true)
 |-- colors: struct (nullable = true)
 |    |-- primary: string (nullable = true)
 |    |-- secondary: string (nullable = true)
 |    |-- tertiary: string (nullable = true)
 |-- contact: struct (nullable = true)
 |    |-- address: struct (nullable = true)
 |    |    |-- address1: string (nullable = true)
 |    |    |-- address2: string (nullable = true)
 |    |    |-- city: string (nullable = true)
 |    |    |-- country: string (nullable = true)
 |    |    |-- postcode: string (nullable = true)
 |    |    |-- state: string (nullable = true)
 |    |-- email: string (nullable = true)
 |    |-- phone: string (nullable = true)
```

```
 |-- description: string (nullable = true)
 |-- distance: string (nullable = true)
 |-- environment: struct (nullable = true)
 |    |-- cats: boolean (nullable = true)
 |    |-- children: boolean (nullable = true)
 |    |-- dogs: boolean (nullable = true)
 |-- gender: string (nullable = true)
 |-- id: long (nullable = true)
 |-- name: string (nullable = true)
 |-- organization_animal_id: string (nullable = true)
 |-- organization_id: string (nullable = true)
 |-- photos: array (nullable = true)
 |    |-- element: struct (containsNull = true)
 |    |    |-- full: string (nullable = true)
 |    |    |-- large: string (nullable = true)
 |    |    |-- medium: string (nullable = true)
 |    |    |-- small: string (nullable = true)
 |-- primary_photo_cropped: struct (nullable = true)
 |    |-- full: string (nullable = true)
 |    |-- large: string (nullable = true)
 |    |-- medium: string (nullable = true)
 |    |-- small: string (nullable = true)
 |-- published_at: string (nullable = true)
 |-- size: string (nullable = true)
 |-- species: string (nullable = true)
 |-- status: string (nullable = true)
 |-- status_changed_at: string (nullable = true)
 |-- tags: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- type: string (nullable = true)
 |-- url: string (nullable = true)
 |-- videos: array (nullable = true)
 |    |-- element: string (containsNull = true)
```

- Initial Schema
- Explored data using PySpark/ SQL.

My Credentials

SOFTWARE

- Python, SQL, C, JavaScript, HTML, CSS
- GitHub, Perforce
- Trello, Jira
- Fluent in Spanish
- Blockchain Tech
- Quantum Computing Theory
- Cyber Security

EDUCATION

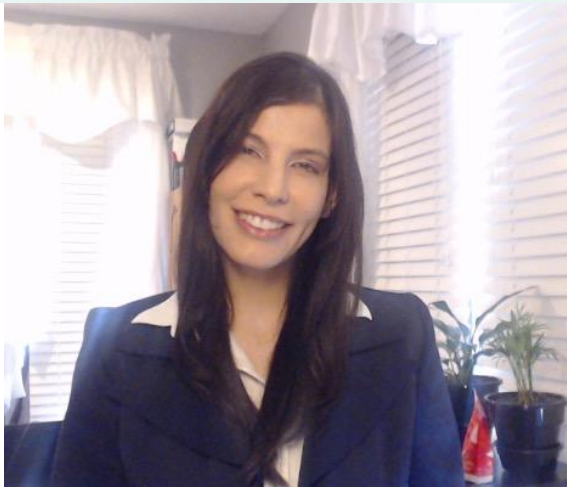
- B.S. Computer Science, TBP Honor Society, CU Boulder
- M.Ed. Bilingual Education, UTEP
- A.A.S. Web Development, PTK Honor Society, Seattle Central
- B.S. Psychology, Texas A&M

EMPLOYMENT

- Seagate Technology
- Curve10 Engineering Firm
- CU Boulder Research Assistant in NLP Project
- Freelance Web Developer > 10 years
- Elementary School Teacher in Texas and Colorado
- US Navy Cryptology and Electronics
- Research Specialist at GHC Ctr. Health Studies

http://elsa-data-sci.tech

Gratitude



Thank you

PetFinders.com
for the data.

http://elsa-data-sci.tech