



Dog Adoptions across the United States

Yenmin Young
Springboard Data Science
Capstone project

Background

- It is better to adopt than buy a dog because there are always dogs in a shelter who need rescuing
- Overpopulation of dogs in the South due to:
 - Less stringent laws around leashing and neutering
 - Families in poverty are unable to keep dogs
 - Less governmental aid to shelters
- Caravans of trucks transport dogs from the South to the Northeast, known as the “Rescue Dog Migration”



Objective

Analyze and visualize the data to understand the movement patterns around dog adoption in the United States

- Guiding questions:
 - Which states export the most dogs and where do they go?
 - Which type of dogs are overrepresented in the shelters?



Data

Three different data files, containing information such as:

- Origin of the dog
- State they are currently listed for adoption in
- Amount of dogs exported and imported in each area
- Features and characteristics about the dog (breed, size, sex, fur color, environment suitability, medical records, etc.)



Disclaimers:

- Only represents a single day of data (9/20/2019). Information has changed since.
- Only includes information from PetFinder.com and not dogs listed on other adoption agencies, websites, or offline transactions.
- Does not tell which dogs are in highest demand, are most popular, or more "adoptable", but rather what is available for adoption on that day in the market.

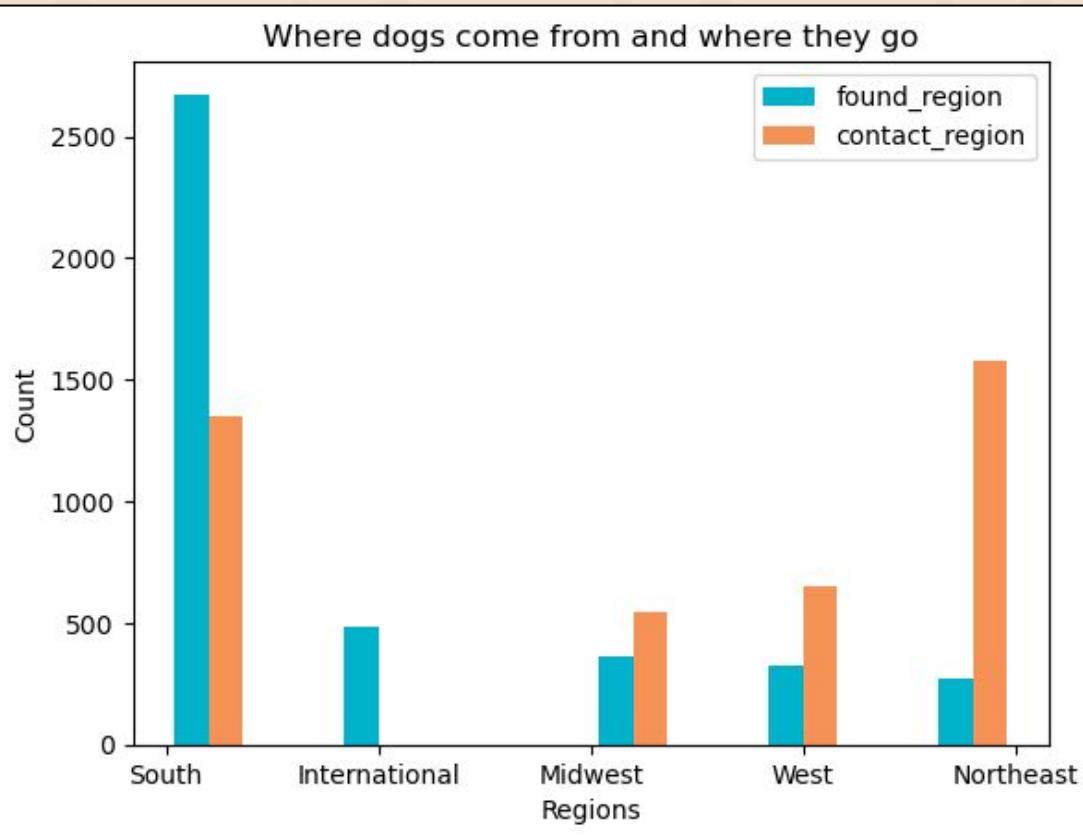




Data Preparation

- Standard data cleaning (dropping redundant information, unique identifiers, duplicates, mostly missing data, or irrelevant to the problem, correcting data types)
- Fixing shifted rows
- Filling missing information through research and reading other values in the row
- Overriding outdated information
- Standardizing spelling, capitalization, formatting, scope of region
- Manually updating incorrect information by reading the description of the dog's story
- Added a 'region' column to generalize patterns across locations

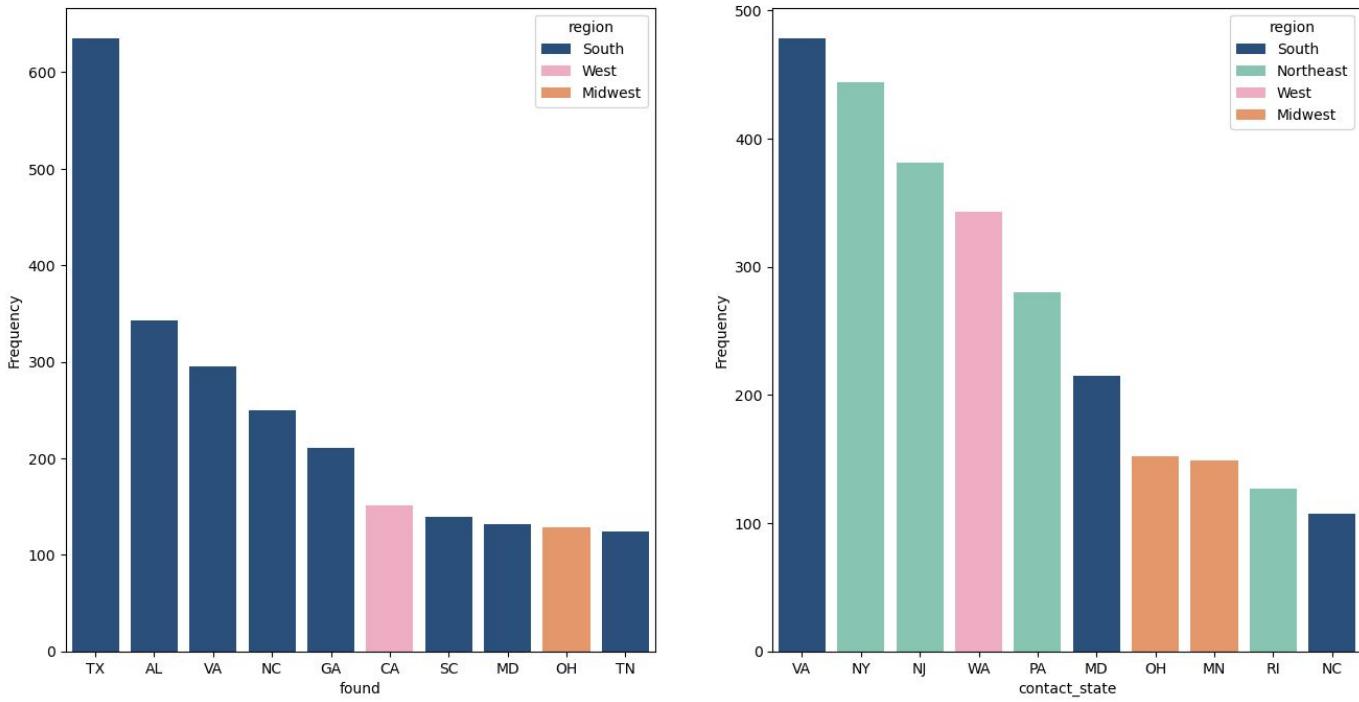




Most dogs are found in the South and transported to shelters across the United States, especially the Northeast.

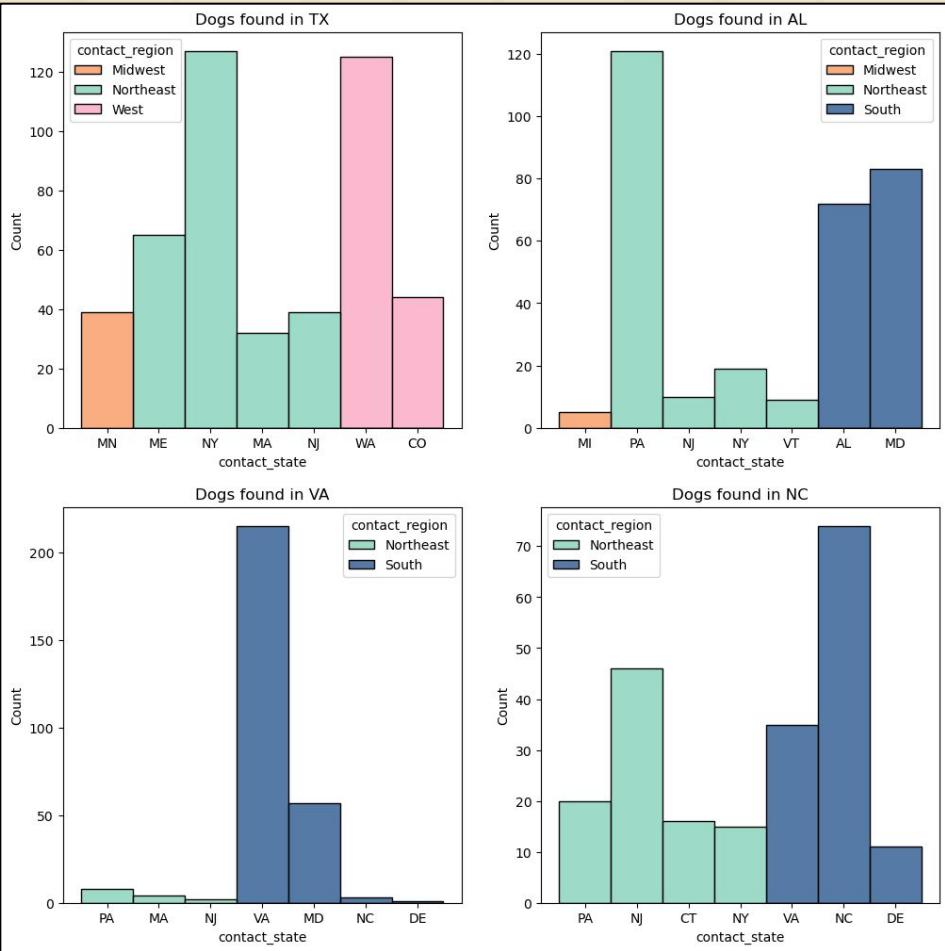


Top 10 states of where the dogs were found and where they're now located



Detailed breakdown by state of where dogs are found (mainly the South, especially TX), and where they are listed for adoption (mainly the Northeast)





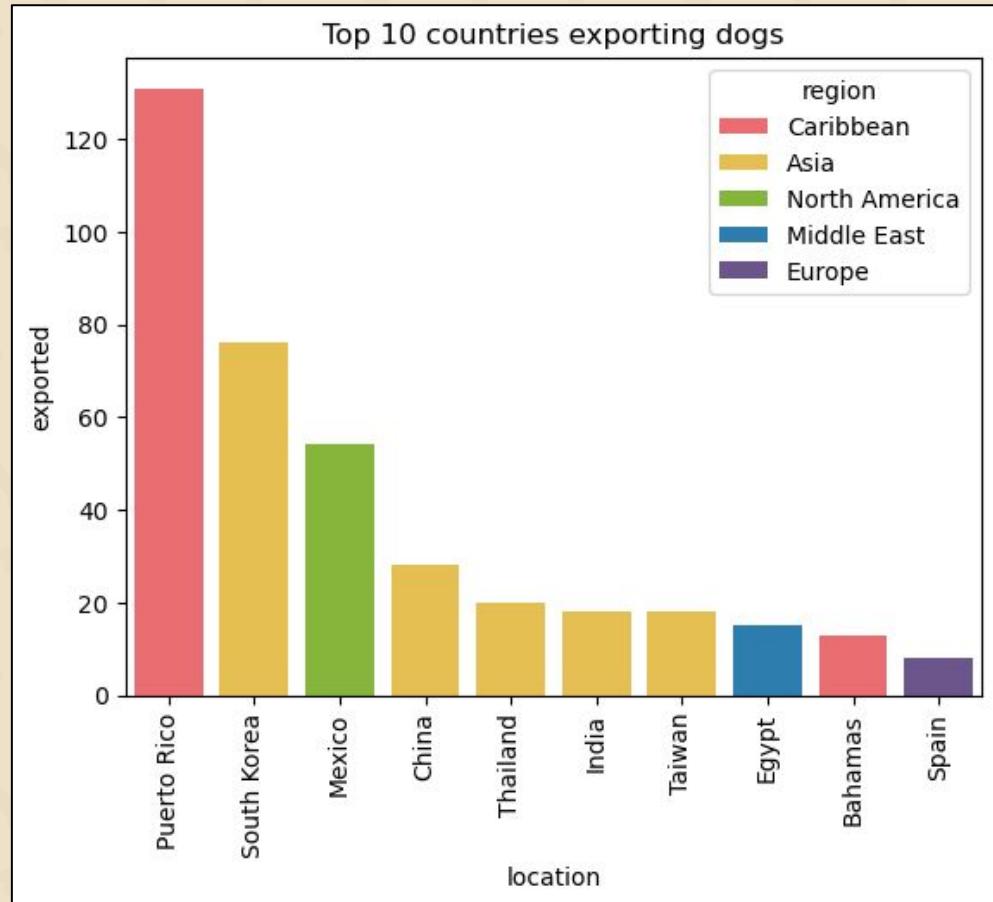
A closer look into the states where most dogs are found:

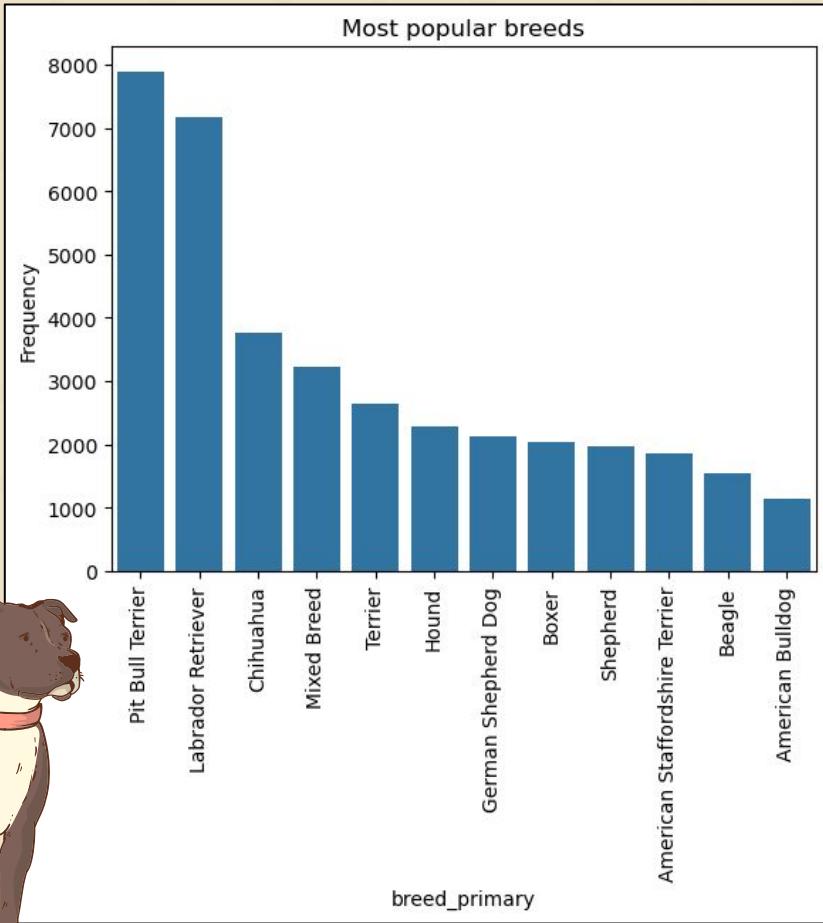
- Some states (like TX and AL) transport their dogs to other states far away
- Possible reasons:
 - Partnerships between rescue organizations
 - Adoption caravan's "rescue route" hosting dog adoption events along the way up to ME
- Some states (like VA and NC) mainly supply themselves - their dogs stay local



Some dogs are imported internationally:

- Puerto Rico:
 - Overabundance of stray dogs (called 'satosh')
 - Similar reasons to the U.S. South (poverty, lack of governmental support, less neutering)
- South Korea
 - Organizations aimed to prevent dogs from being included in cuisine

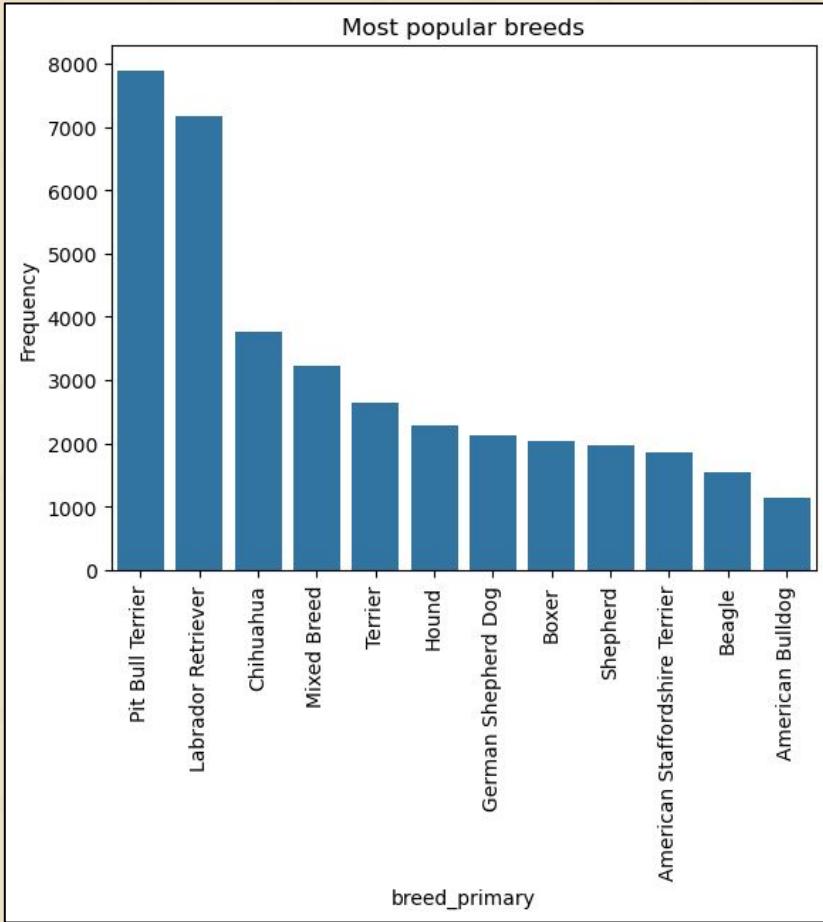




Overabundance of Pit Bulls:

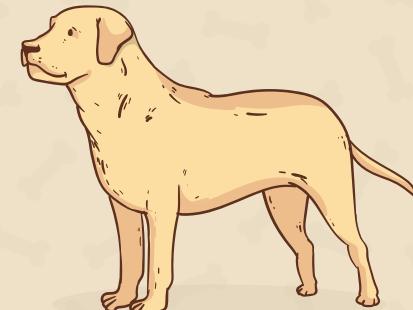
- Demand for guard dogs and fighting dogs
- Unregulated backyard breeders looking for profit
- Abandoned by unprepared owners, or were unprofitable in the ring
- Misrepresented as 'aggressive', scaring potential adopters.
- Housing and city regulations banned certain 'dangerous' breeds





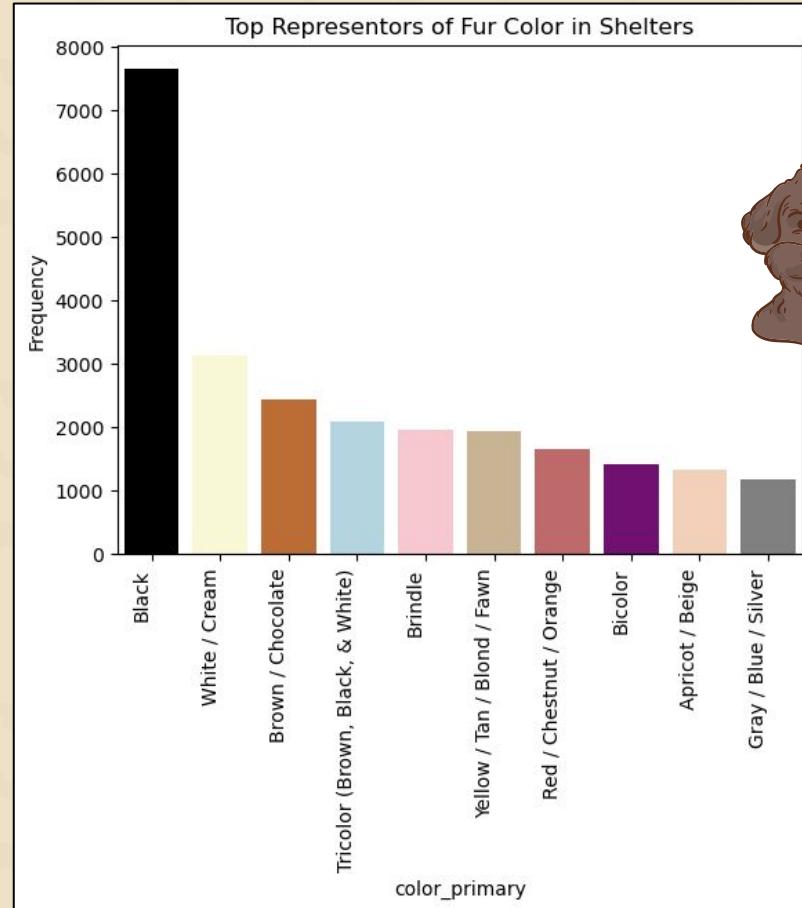
Overabundance of Labrador Retrievers:

- Seen as an “easy starter pet” for families
- High demand leads to overbreeding
- High energy levels and needs a lot of attention
- Surrendered by uncommitted families



A vast majority of dogs in shelters have black fur.

- Black dogs are adopted at lower rates than their lighter shelter mates
- Possibly due to:
 - negative perception of the color black
 - (in)visible in dimly lit shelters
 - difficult to distinguish between other black dogs



Predicting where a dog came from

Goal: predict which region a dog came from based on the location of the shelter listing

Data: dogTravel.csv

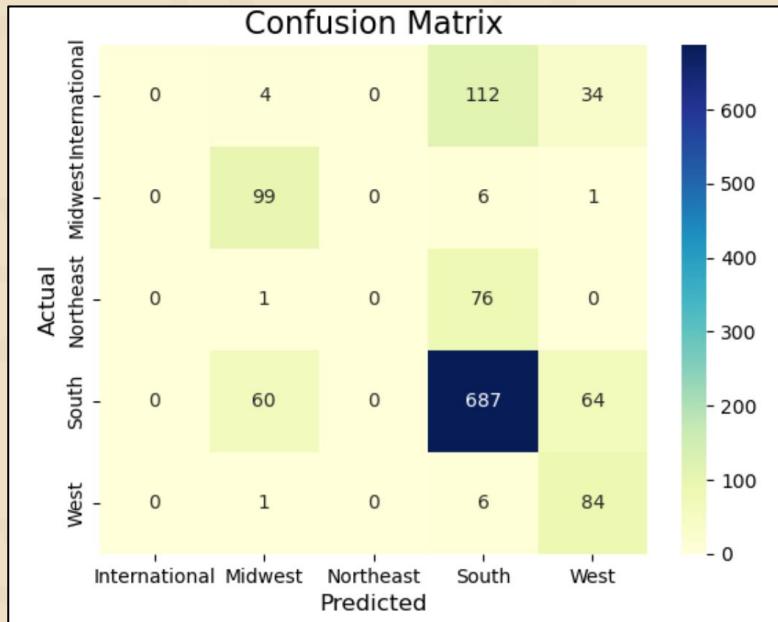
Features: the region that the dog came from, one-hot encoded

Models: suitable for multiclass categorical datasets (the prediction could be any of the four regions (Northeast, Midwest, West, South))

- Logistic regression
- Support Vector Machine
- Decision Tree
- Random Forest
- K-Nearest Neighbors.



Predicting where a dog came from



Model	Accuracy	Precision	Recall	F1 score	ROC_AUC	mean CV score
Logistic Regression	0.70445	0.70445	0.70445	0.70445	0.92038	0.70693
Support Vector Machine	0.70445	0.70445	0.70445	0.70445	0.92038	0.70693
Decision Tree	0.70445	0.70445	0.70445	0.70445	0.92042	0.70693
Random Forest	0.70445	0.70445	0.70445	0.70445	0.92050	0.70693
K-Nearest Neighbors	0.70445	0.70445	0.70445	0.70445	0.83638	0.68846

Best model: Random Forest Model (barely)

Confusion Matrix:

- Great at predicting Midwestern, Southern, and Western dogs
- Couldn't identify any International or Northeastern dogs.



Predicting the breed of a dog

Goal: predict the breed of the dog based on the traits

Data: allDogsDescription.csv

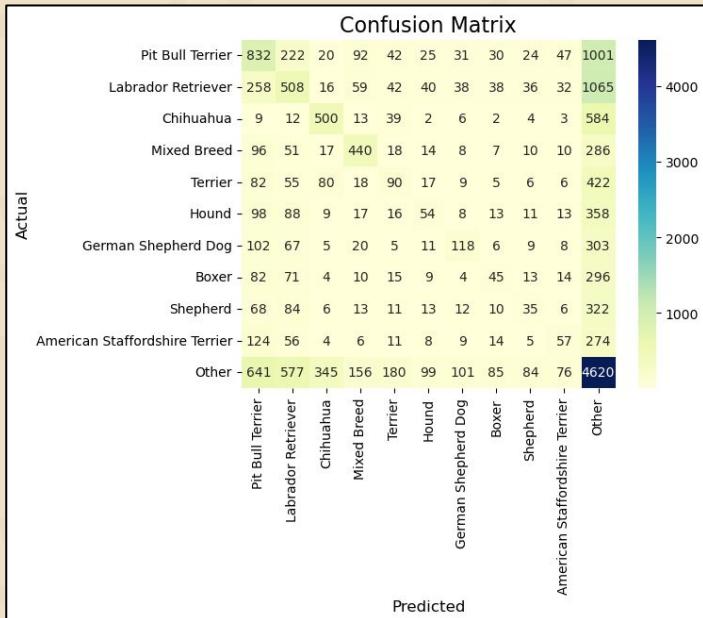
Features: age, sex, size, coat, fixed, house trained, vaccinated, good in an environment with children and other dogs, shelter location

Models: suitable for multiclass categorical datasets

- Logistic regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors.



Predicting the breed of a dog



Model	Accuracy	Precision	Recall	F1 score	ROC_AUC	mean CV score
Logistic Regression	0.418803	0.418803	0.418803	0.418803	0.849813	0.295243
Decision Tree	0.389204	0.389204	0.389204	0.389204	0.748191	0.268122
Random Forest	0.418689	0.418689	0.418689	0.418689	0.823618	0.305397
K-Nearest Neighbors	0.351230	0.351230	0.351230	0.351230	0.730177	0.240484

Best model: Random Forest Model

Overall, not terribly accurate, but better than random guessing.



Takeaways

- Stray dogs mainly come from the South and are exported to other states throughout the U.S. (especially to the Northeast where supply is low)
- Some states, like Virginia and North Carolina, mainly keep their dogs local. Other states, like Texas and Alabama, heavily rely on other states to take in their stray dogs.
- Overabundance of pit bulls, labrador retrievers, and black dogs due to selective overbreeding, bias, and inability to properly care for dogs.



What can be done?

Some actionable steps:

- Educate potential adopters about how to properly care for a dog to prevent surrender
- Provide more resources to neuter dogs in Texas
- Campaign against the discrimination of pit bulls and black dogs





Future Research

- Interactive chart of weighted vector arrows overlaid on top of a U.S. map, indicating the flow of dogs between states, to demonstrate a clearer picture of the movement of dogs across the U.S.
- Investigate what traits make a dog more adoptable. Collect and analyze data with more extensive knowledge on the dog and its background, such as:
 - Appearance (size, color, fur length)
 - Age, sex (puppy, adult, old)
 - Behavior and personality
 - Medical background (already neutered or has a sponsored neutered scheduled, vaccinated, diseases, etc.)
 - Length of time in shelter
 - Reasons for entering the shelter

