Predicting Pet Adoption Speeds

W207 Summer 2023

Lucy, Nicole, Bailey, Alberto, & Erik

Project Repository Link:
https://github.com/UC-Berkeley-I-School/mids-207-final-project-summer23-Rueda-Sambrailo-Herr-Liu-Kuehl

Pet Adoption culture

Motivation: Animal illness and euthanization in shelters

6.3 million

pets are surrendered to US animal shelters, annually

250%

spike in **adoption rates** in early-mid 2020 (COVID-19)

1.5 million

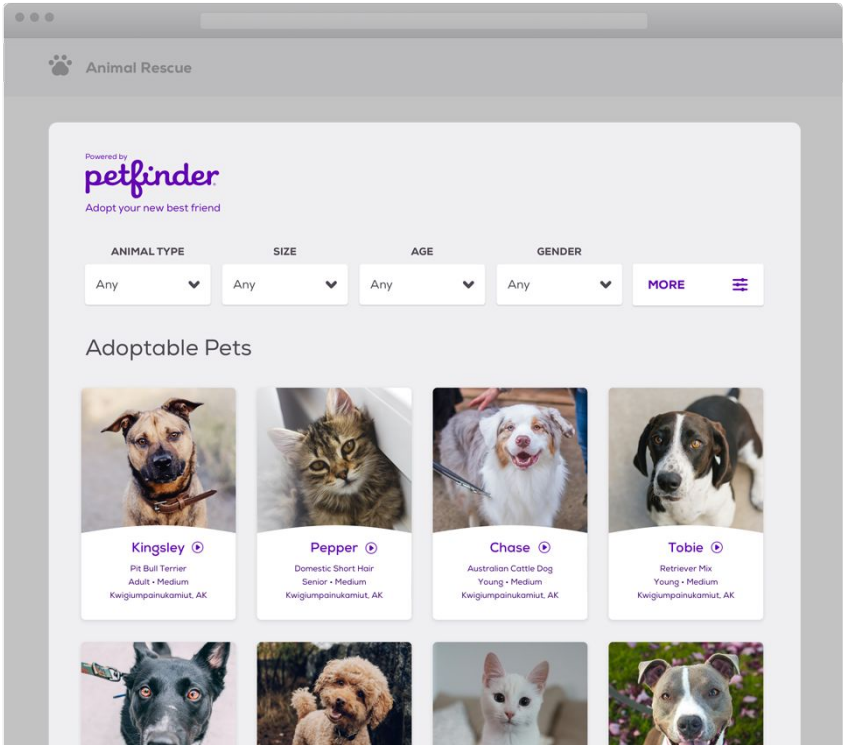
dogs and cats in shelters get **euthanized**, annually

Objectives and Research Questions

Research Objective: Can we use machine learning to predict adoption speed?

Secondary Question: Can we use those predictions to advise shelters on how to increase adoption?

Example Application: Adoption postings

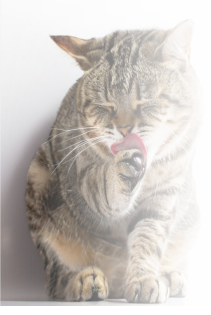


Data: Overview

PetFinder.my Adoption Prediction Dataset (Kaggle)

	Type (Source)	Size	Total Features	Key Features
Pet Profile Features	Tabular	(14993, 24)	24	Type, Breed, Age, Gender, Health, State (location) Adoption Speed
Pet Image Metadata	JSON (Google Vision API)	14652 pets	14	Face Annotation, Label Annotation, Image Properties
Pet Description Sentiment Analysis	JSON (Google NL API)	14442	7	Sentiment Score, Sentiment Magnitudes, Languages

Data: Pipeline



Raw tabular data

Parse sentiment & image metadata

Updated tabular data

Train/test split

Tabular EDA

Image EDA

Sentiment EDA

Feature Engineering & Selection

Text: Embeddings

Nominal/ordinal: Binary/OH Encode

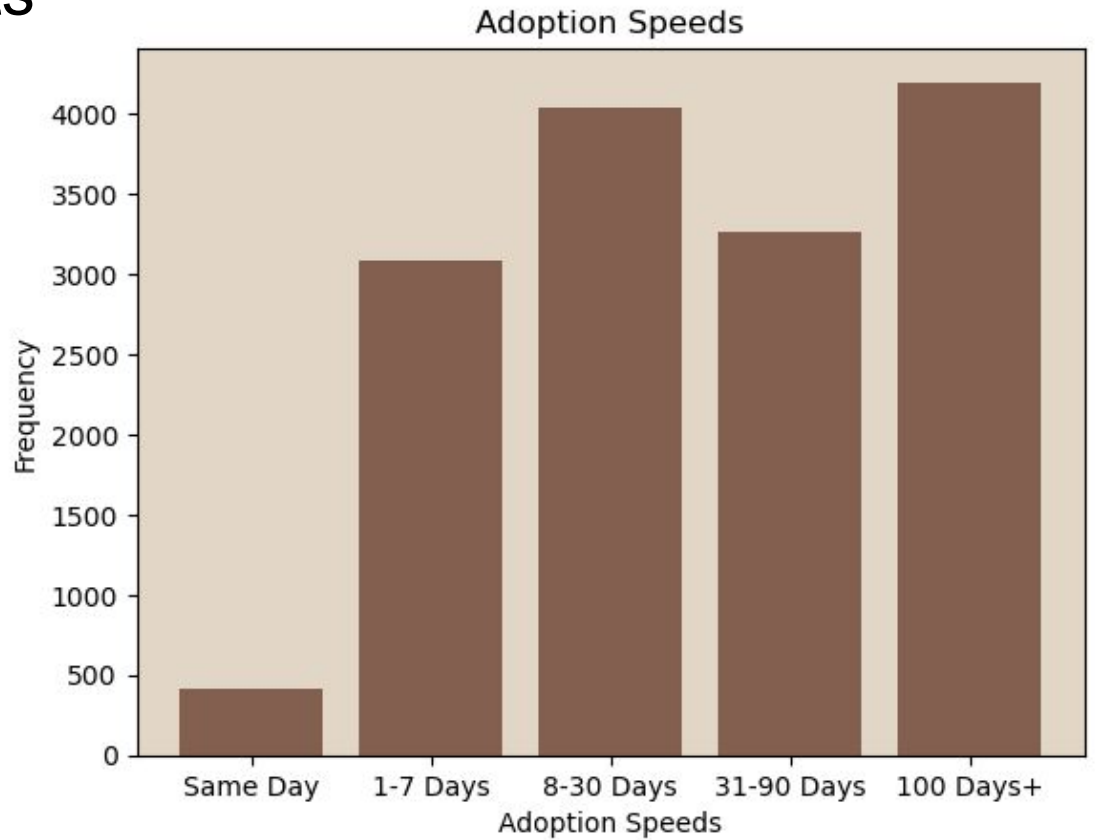
Numeric: Standardize

Baseline Model

Model Development & Performance Metric Selection

Model Evaluation

Data: Summary Stats



Data: Summary Stats

Median Adoption Speeds

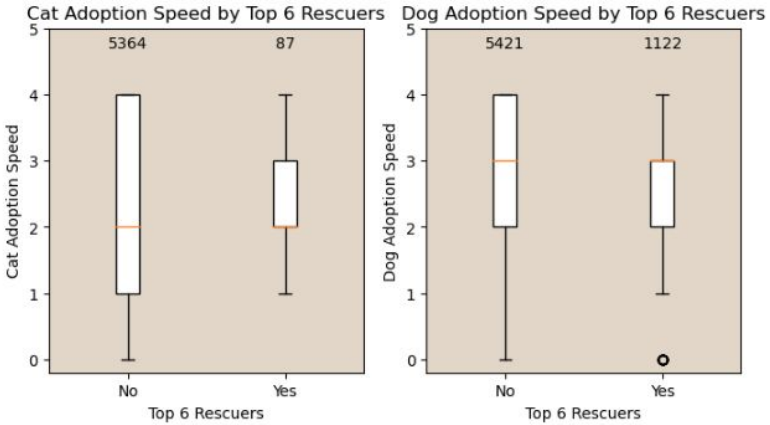
faster for. . .	than . . .
cats	dogs
long-haired dogs	short-haired dogs
specified name *(Miu Miu)*	unspecified *(kitten)*
dog of specified breed	mixed breed
black cat	black, yellow & white cat

Data: Summary Stats

Pet Rescuers

4,800 Rescuers → **~12,000** Adoption Records

6 Rescuers → **1,200+** Adoption Records



Data: Feature Engineering

Feature Additions

- **RescuerCount** (qty of rescues)
- **State Population & Median Income** (for adoption locations)
- **Breed Groups** (by American Kennel Club)
- **Guessed Age** (age was likely guessed)

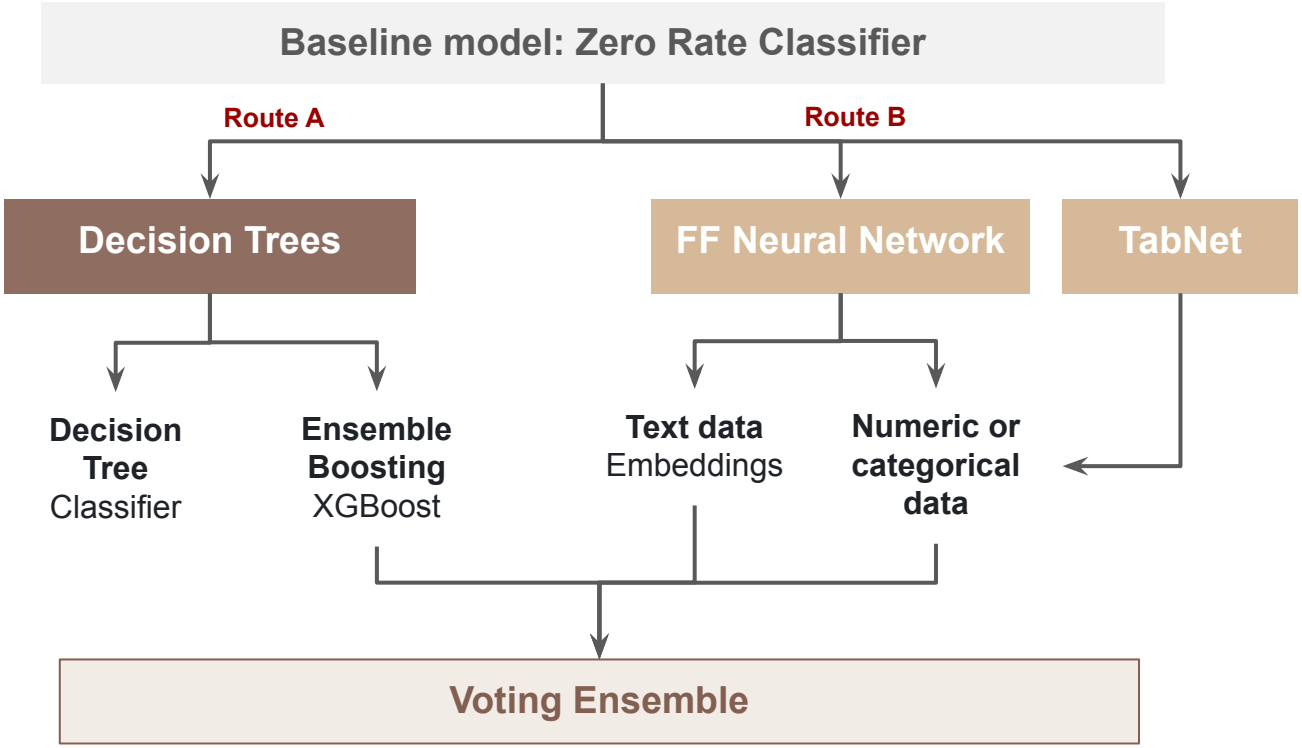
Feature Transformations

Binning	Binary Encoding	Multi-Hot Encoding	Standardize
• Age • Quantity	• Breeds • State	• Colors • Gender	• Age • Fee

Addressing Nulls

Re-classifying & Balancing Labels
(Adoption Speed)

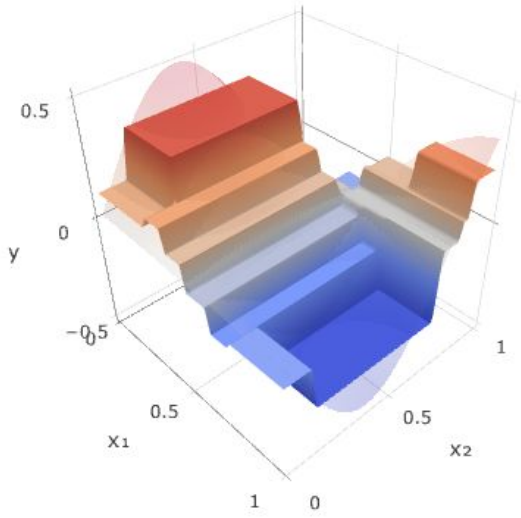
Modeling Approach



Decision Tree vs. XGBoost

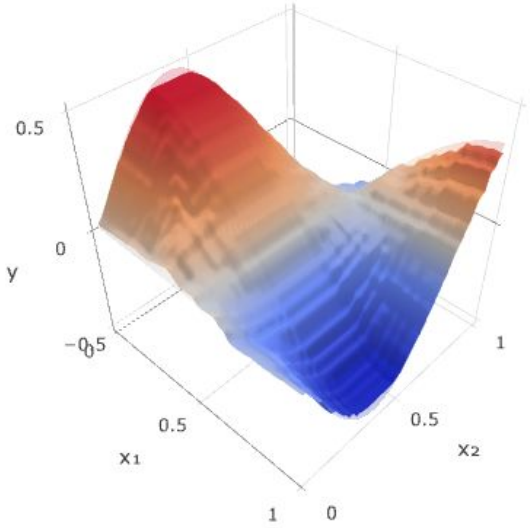
Single Decision Tree

- one tree
- high feature influence



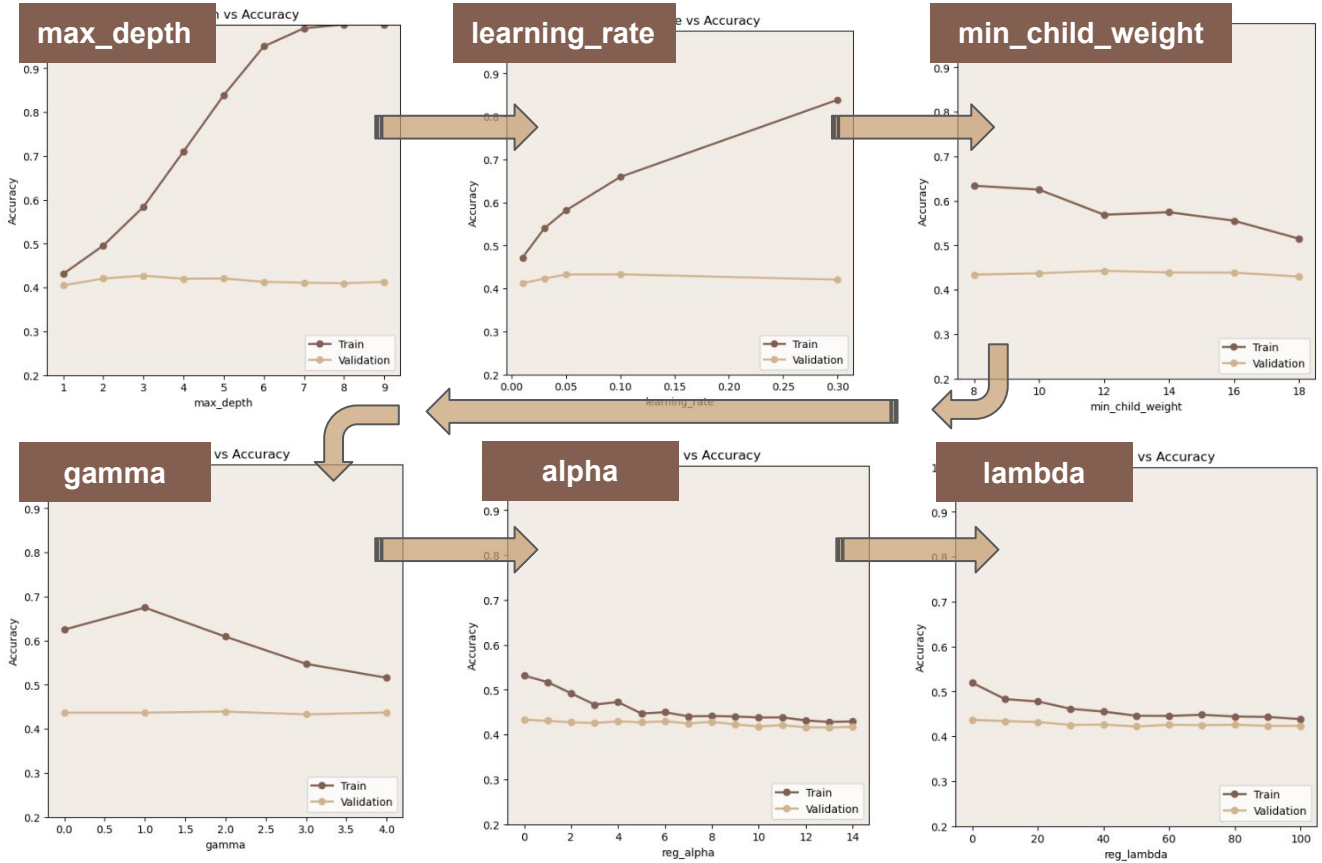
XGBoost

- 100’s of trees
- very prone to overfitting



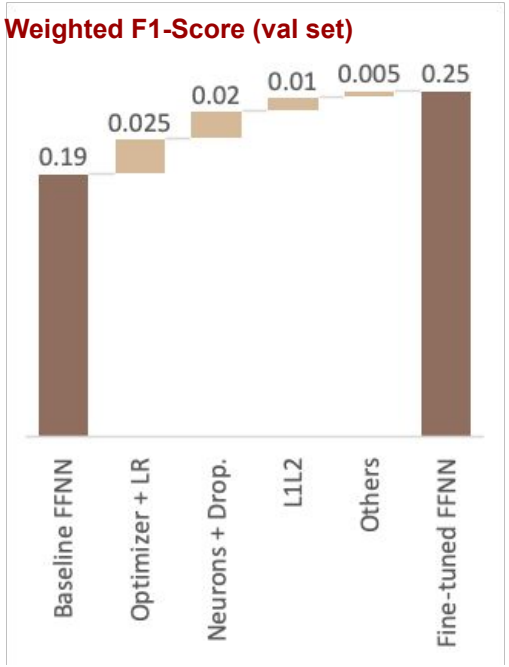
source: https://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html

XGBoost: Model & Hyperparameters



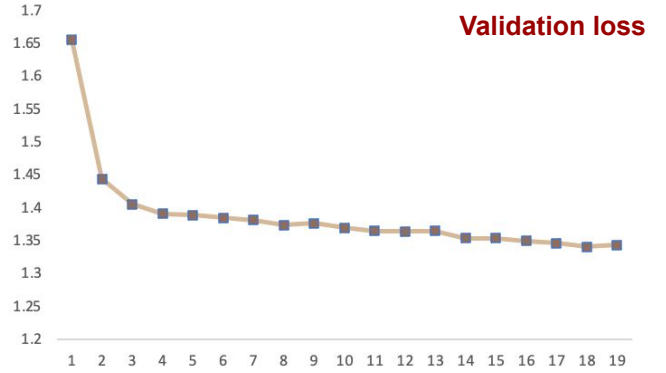
Deep Neural Networks - Architecture & Hyperparameters

Feed-Forward NN (w/out text data)



❌ *Features, activation function, batch normalization, weight decay*

TabNet (w/out text data)



Voting Ensemble (FFNN + TabNet)

	precision	recall	f1-score
0	0.33	0.59	0.43
1	0.31	0.10	0.16
2	0.34	0.24	0.28
3	0.47	0.57	0.51
accuracy			0.38
macro avg	0.36	0.38	0.34
weighted avg	0.37	0.38	0.35

Summary of Results

	Val Set: weighted-F1 score	Test Set: weighted-F1 score
Baseline Majority Predictor	0.10	0.12
Decision Tree	0.38	0.36
XGBoost	0.41+	0.41+
Feedforward Neural Network	~0.35	~0.30
Transformers: TabNet	0.29	0.29
Ensemble: FFNN + TabNet + XGBoost	**~0.40**	**~0.40**

Key Takeaways from Models

FFNN: suffering from overfitting and feature interference

	Val Set: weighted-F1 score	Test Set: weighted-F1 score
FFNN (numeric features only)	0.37	0.25
FFNN (text description only - embedding)	0.31	0.33
FFNN (text embeddings + numeric)	0.34	0.16

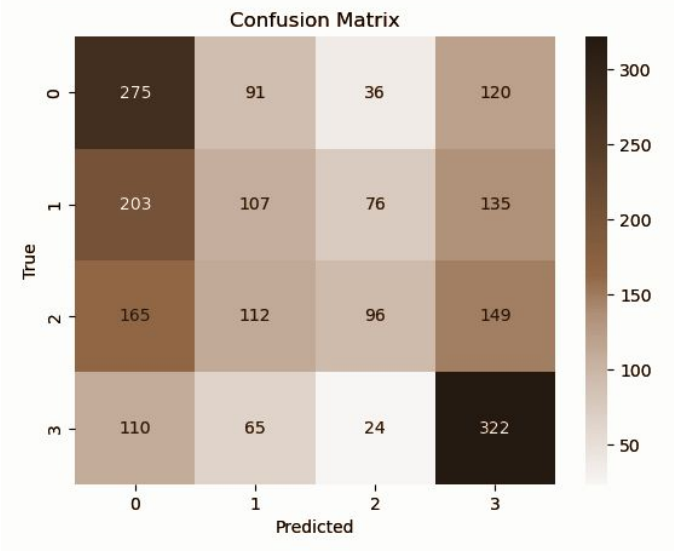
Ways to improve model performance

Feature engineering	Model depth and width	Regularization	Learning Rate and Optimizer

Conclusion

Research Objective: Can we use machine learning to predict adoption speed?

Secondary Question: Can we use those predictions to advise shelters on how to increase adoption?

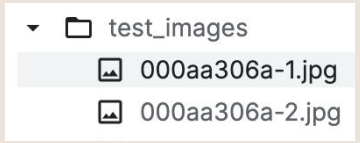


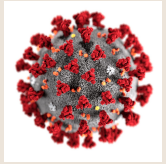
Information gain ranking from decision tree

Features	Feature Importances
RescuerCount	0.334191
Age	0.286694
isGeneric_Breed	0.133757

Limitations and Future Work

Limitations





Future Work

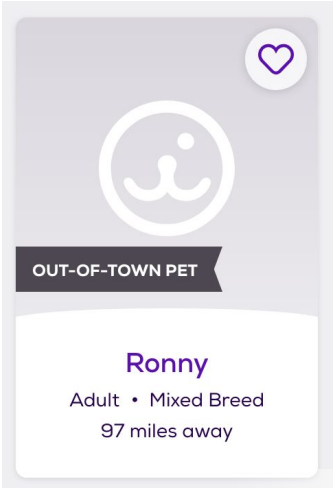
Image processing techniques

Collect recent data and include dates / timestamps

Refine key features to use for decision trees

Fairness in ML

- **Accessibility of data and images** (variability in image quality, demographic / geographic representation)
- **Image processing / Description bias** (embeddings, image metadata)
- **Privacy** of veterinary medical records



Impact: Inaccurate recommendations for adoption postings could negatively impact eligible pets.

Contributions

All team members worked on all stages of the project collaboratively, sharing the work streams in an equal and effective way

Data processing / Feature Engineering

- **Bailey** worked on the image files, **Nicole** on the sentiment files, **Erik and Alberto** split the numeric features and **Lucy** worked on incorporating additional data and putting everything together

Modeling

- **Bailey** and **Erik** worked on Decision Trees
- **Nicole**, **Lucy** and **Alberto** worked on FFNN and Transformers

Slides

- Divided equally according to the previous work done

References

1. Ho, J., Hussain, S., & Sparagano, O. (2021). Did the COVID-19 pandemic spark a public interest in pet adoption?. *Frontiers in Veterinary Science*, *8*, 647308.
2. Zadeh, A., Combs, K., Burkey, B., Dop, J., Duffy, K., & Nosoudi, N. (2022). Pet analytics: Predicting adoption speed of pets from their online profiles. *Expert Systems with Applications*, *204*, 117596.
3. https://worldanimalfoundation.org/advocate/pet-adoption-statistics/
4. https://www.americanhumane.org/fact-sheet/animal-shelter-euthanasia-2/
5. https://www.petfinder.com/tools-widgets/custom-pet-list/getting-started/
6. https://www.washingtonpost.com/business/2022/01/07/covid-dogs-return-to-work/
7. https://raw.githubusercontent.com/MoH-Malaysia/covid19-public/main/static/population.csv
8. https://storage.googleapis.com/dosm-public-economy/salaries_state_sex.csv
9. https://www.hhhstopeka.org/adopt
10. https://www.cleveland19.com/2021/09/27/pandemic-pets-are-facing-euthanization-after-being-surrendered-due-overcrowding/
11. https://www.kaggle.com/competitions/petfinder-adoption-prediction/discussion/88773
12. https://www.kaggle.com/c/petfinder-adoption-prediction/discussion/89042