Pet analytics

Predicting adoption speed of pets from their online profiles

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 Syst Appl 204:117596

Why is it important?

Digital Shift: With increased online presence, understanding how profiles influence adoption helps shelters.

Data Availability: Access to data and predictive analytics enables shelters to analyze adoption patterns effectively.

Animal Welfare: Quick adoptions improve animal welfare and alleviate shelter overcrowding.

1. Study Objective:
   - Improve animal shelter operations by predicting adoption speeds, reducing euthanasia risk, and enhancing overall adoption rates.
2. Methodology Highlights:
   - Uses text mining techniques for analyzing animal profiles.
   - Employs diverse machine learning models, including ensemble modeling.
3. Key Findings:
   - Ensemble model combining predictions performs the best alongside with gradient boosting.
   - Non-textual decision trees outperform textual models.
4. Practical Implications:
   - Enables the identification of pets with lower adoption speed, empowering shelter owners to implement new strategies for their successful adoption.

Algorithm Comparison Table

Method	ASE(no-text)	ASE(with text)
Regression	1.251	1.234
LARS	1.258	1.247
Decision tree(ProbF)	1.178	1.183
Descision tree(Variance)	1.176	1.179
Gradient Boosting	1.155	1.168
Neural Network	1.177	1.185
HP Forest	1.248	1.266
Ensemble		1.155

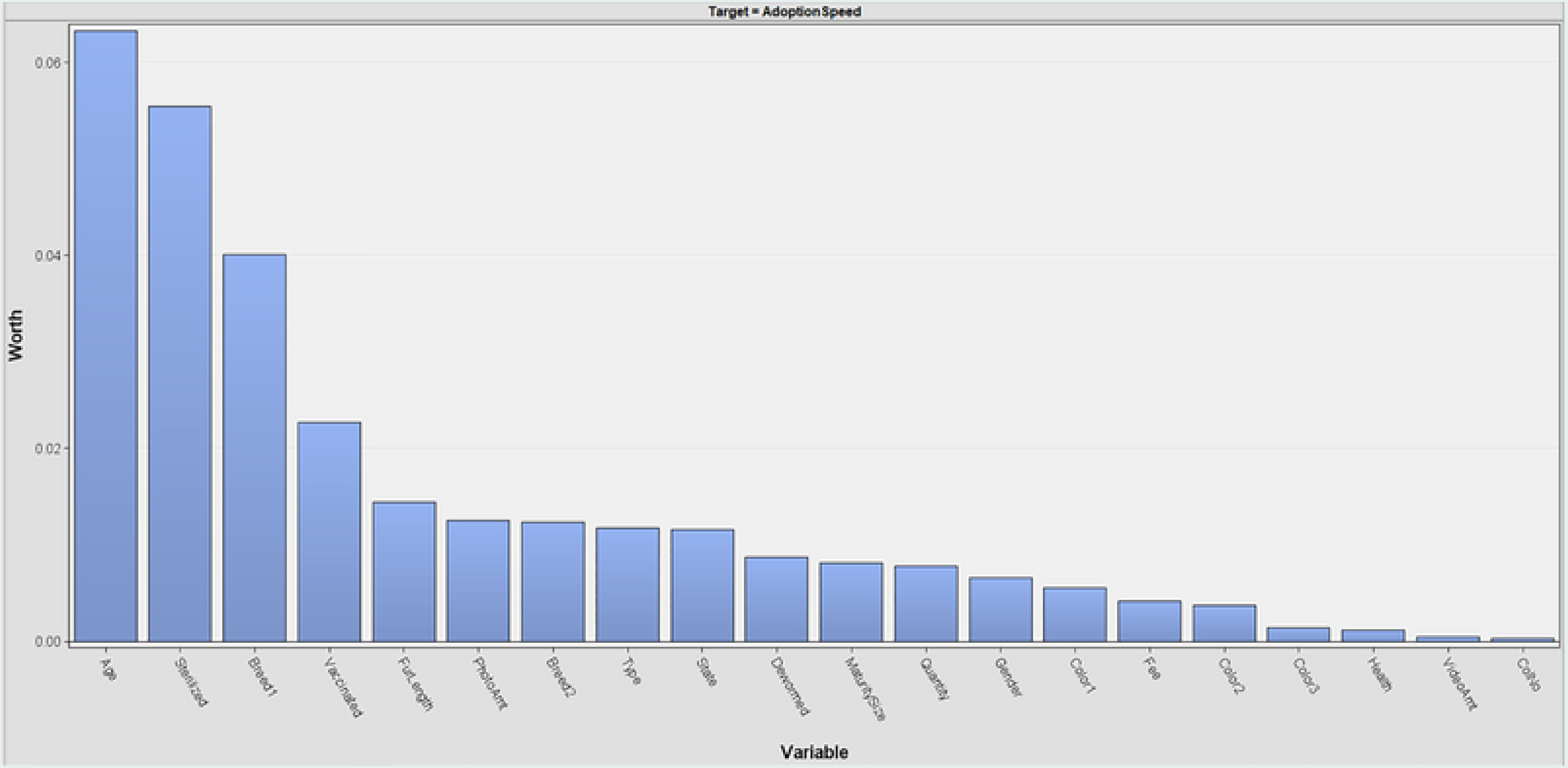
Gradient Boosting

- Gradient Boosting is a machine learning technique used for both regression and classification tasks.
- It works by combining multiple weak learners (typically decision trees) sequentially, where each tree corrects the errors of its predecessors.
- Gradient Boosting iteratively minimizes a loss function by adding new models.

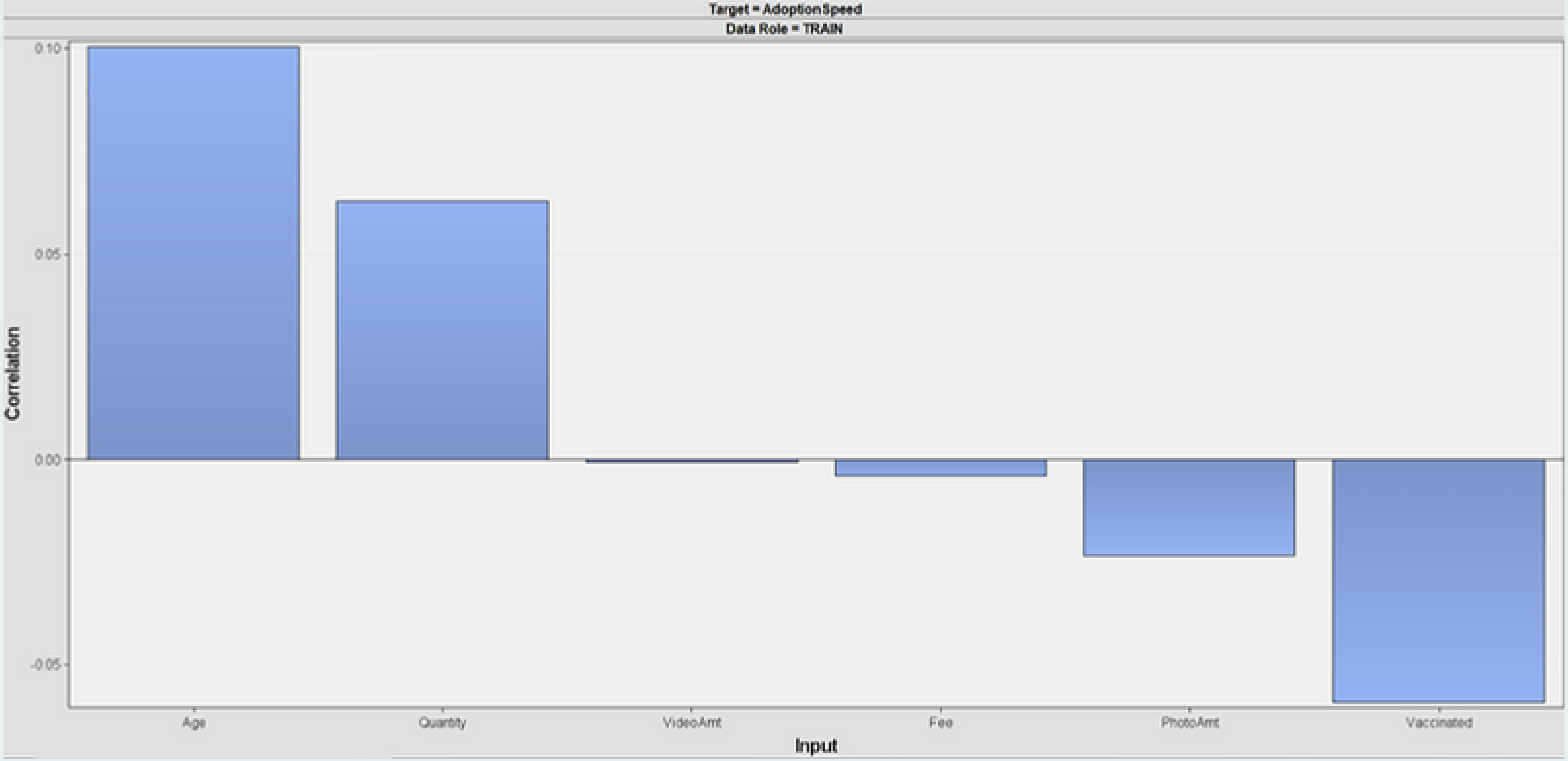
Ensemble

- **Multiple Models:** Train several models, each with a unique perspective on the data.
- **Diversity Matters:** Use different algorithms, data subsets, or training approaches to get broader prediction.
- **Combine Predictions:** Leverage the collective wisdom, Vote, average, or weight predictions based on individual model strengths.
- **Boost Performance:** Achieve higher accuracy, reduce variance, and tackle complex problems compared to solo models.

Variable Worth-Plot



Correlation Plot



Our study

In our study we will test the differences between using tags with low correlation and not using it on a completely different dataset, both checking the validity of the results from the research and improving on it.

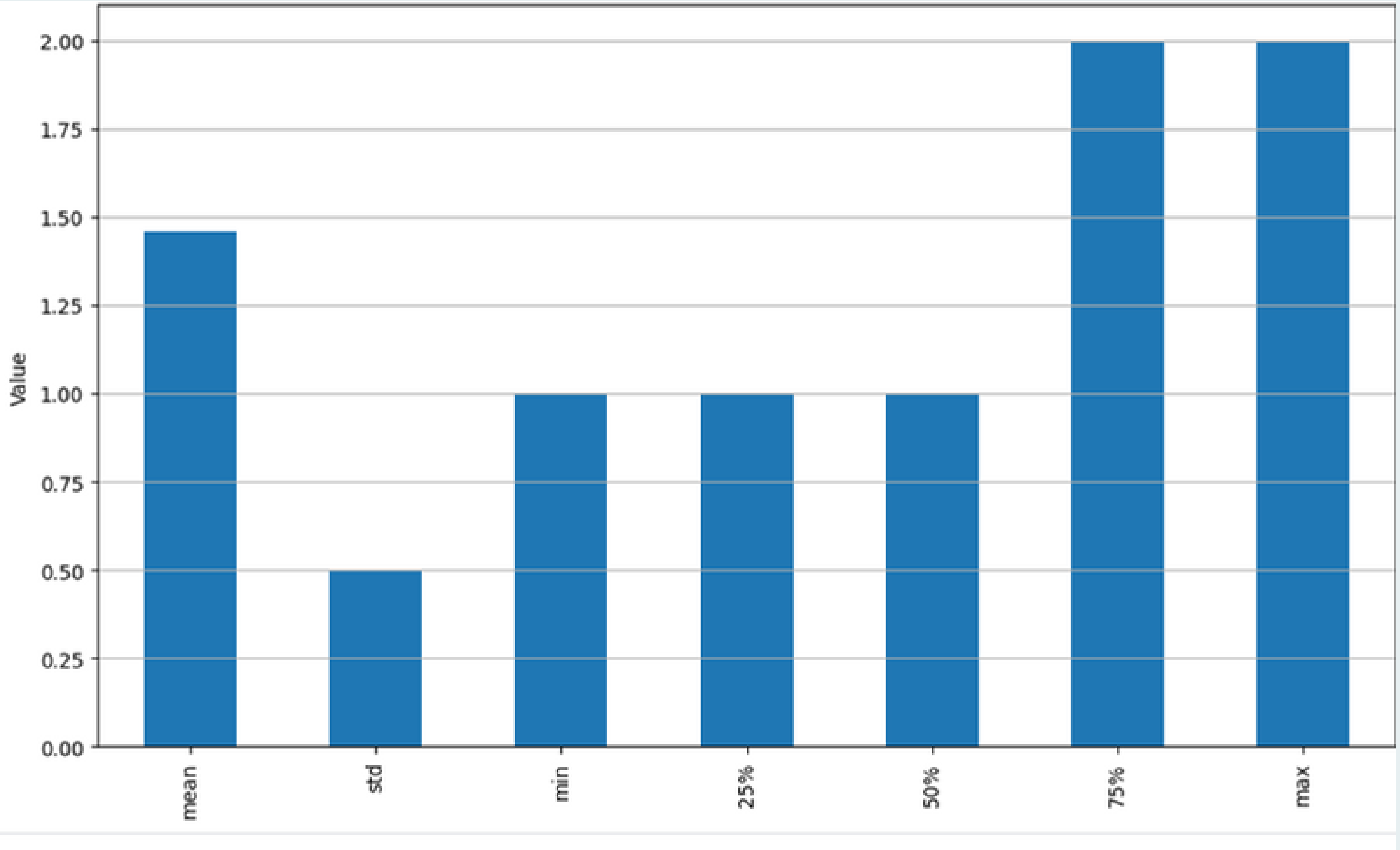
Dataset

```
Type                int64
Name               object
Age                 int64
Breed1              int64
Breed2              int64
Gender              int64
Color1              int64
Color2              int64
Color3              int64
MaturitySize        int64
FurLength           int64
Vaccinated          int64
Dewormed            int64
Sterilized          int64
Health              int64
Quantity            int64
Fee                 int64
State               int64
RescuerID          object
VideoAmt            int64
Description        object
PetID              object
PhotoAmt          float64
AdoptionSpeed       int64
dtype: object
```

Describe

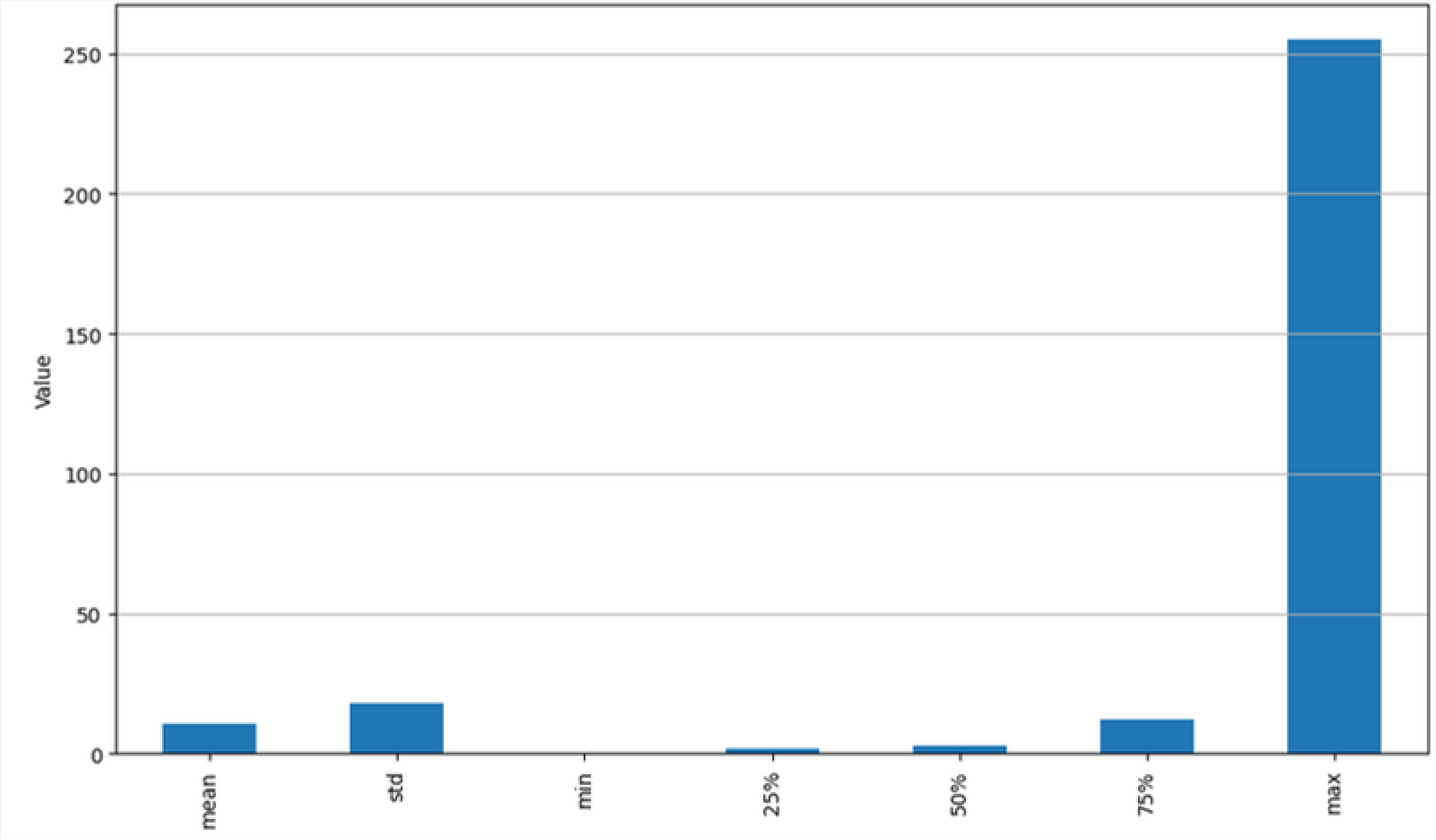
	Type	Age	Breed1	Breed2	Gender	Color1	Color2	Color3
count	14993.000000	14993.000000	14993.000000	14993.000000	14993.000000	14993.000000	14993.000000	14993.000000
mean	1.457614	10.452078	265.272594	74.009738	1.776162	2.234176	3.222837	1.882012
std	0.498217	18.155790	60.056818	123.011575	0.681592	1.745225	2.742562	2.984086
min	1.000000	0.000000	0.000000	0.000000	1.000000	1.000000	0.000000	0.000000
25%	1.000000	2.000000	265.000000	0.000000	1.000000	1.000000	0.000000	0.000000
50%	1.000000	3.000000	266.000000	0.000000	2.000000	2.000000	2.000000	0.000000
75%	2.000000	12.000000	307.000000	179.000000	2.000000	3.000000	6.000000	5.000000
max	2.000000	255.000000	307.000000	307.000000	3.000000	7.000000	7.000000	7.000000

Type



2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Value
mean
std
min
25%
50%
75%
max

Age

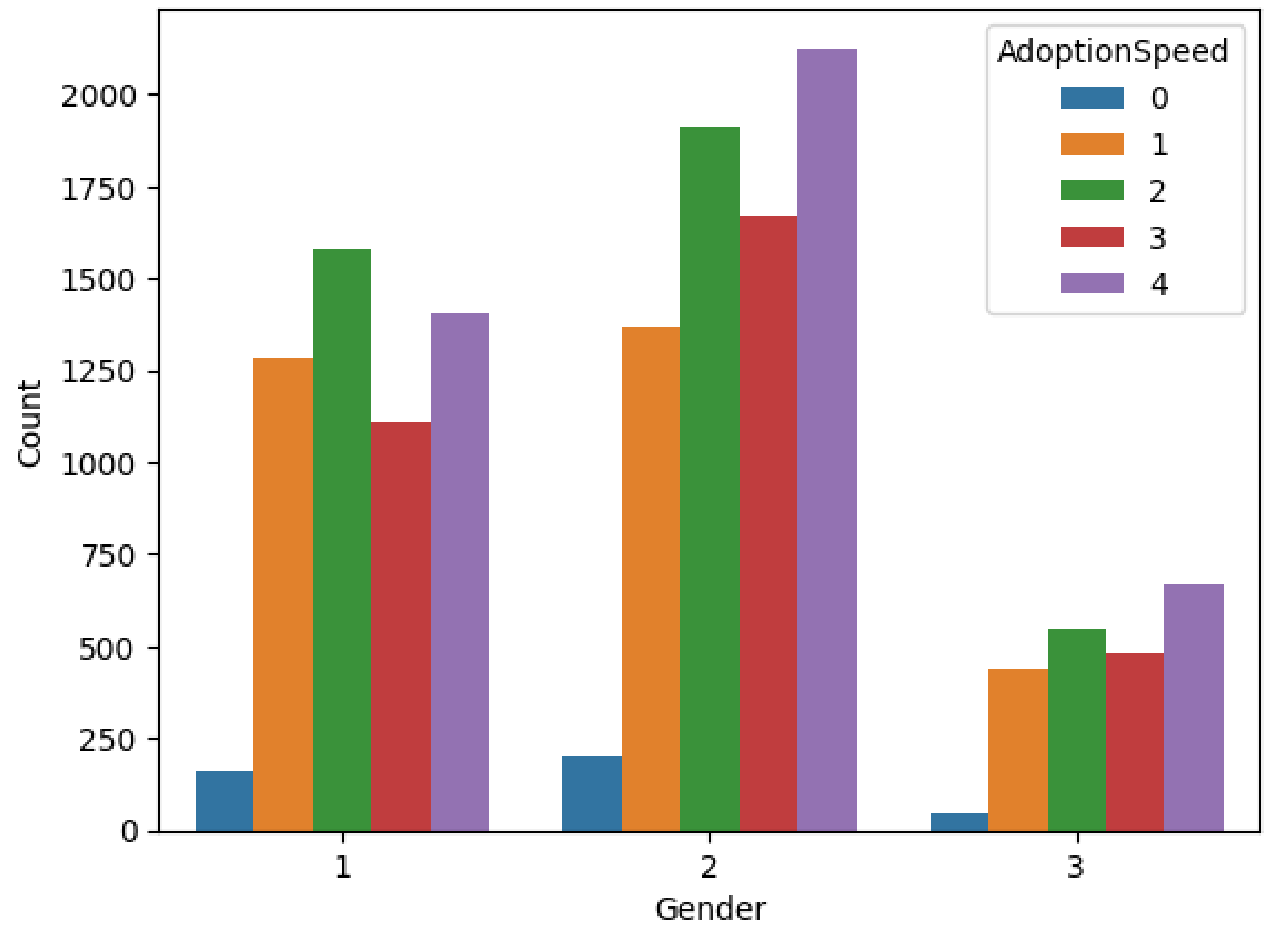


Pre-Processing

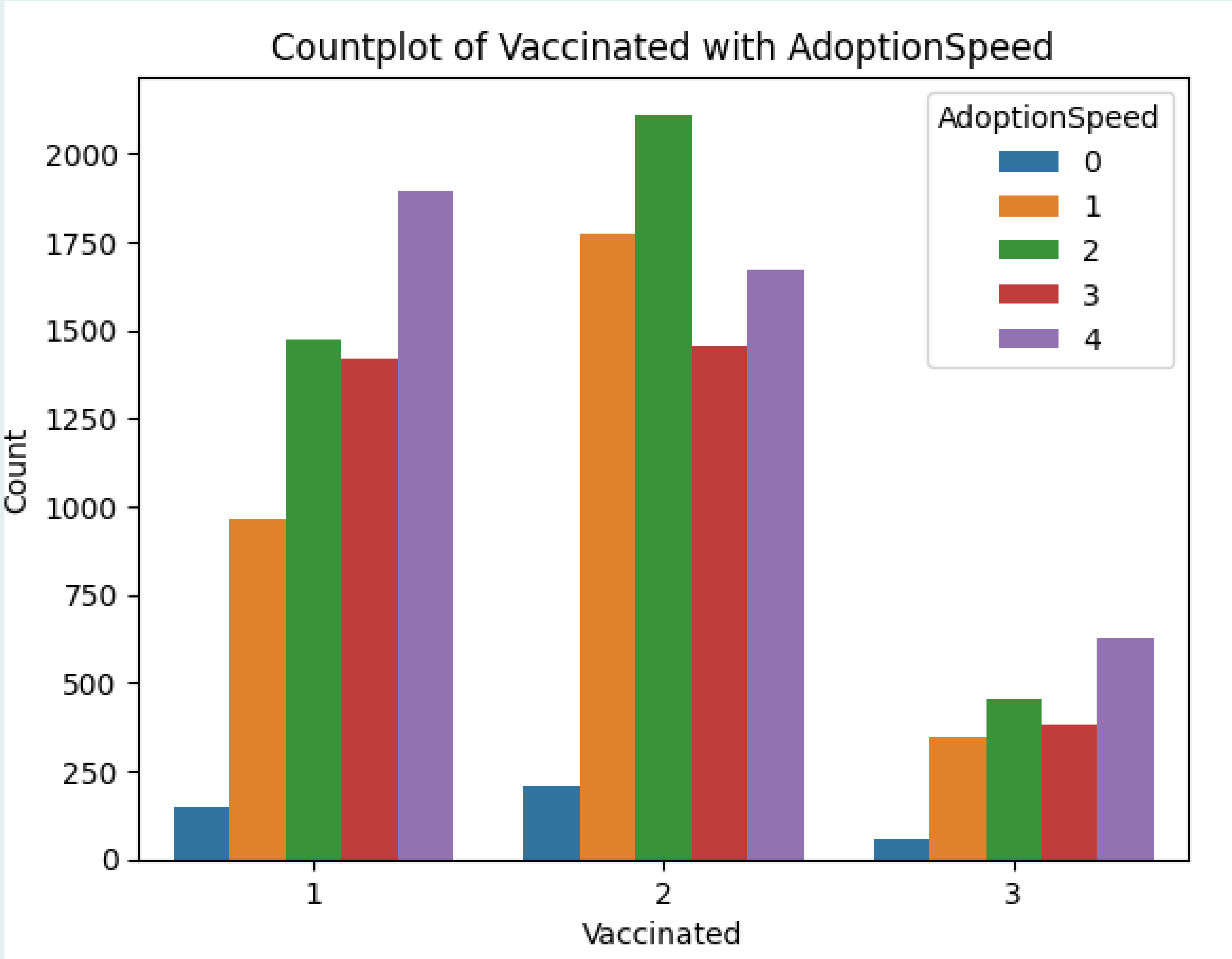
```
empty values:
`Type`: 0
`Name`: 1265
`Age`: 0
`Breed1`: 0
`Breed2`: 0
`Gender`: 0
`Color1`: 0
`Color2`: 0
`Color3`: 0
`MaturitySize`: 0
`FurLength`: 0
`Vaccinated`: 0
`Dewormed`: 0
`Sterilized`: 0
`Health`: 0
`Quantity`: 0
`Fee`: 0
`State`: 0
`RescuerID`: 0
`VideoAmt`: 0
`Description`: 13
```

We dropped 'Name', 'RescuerID', 'PetID', 'Description' because we deemed them not important for predictions

Gender to adoption speed graph

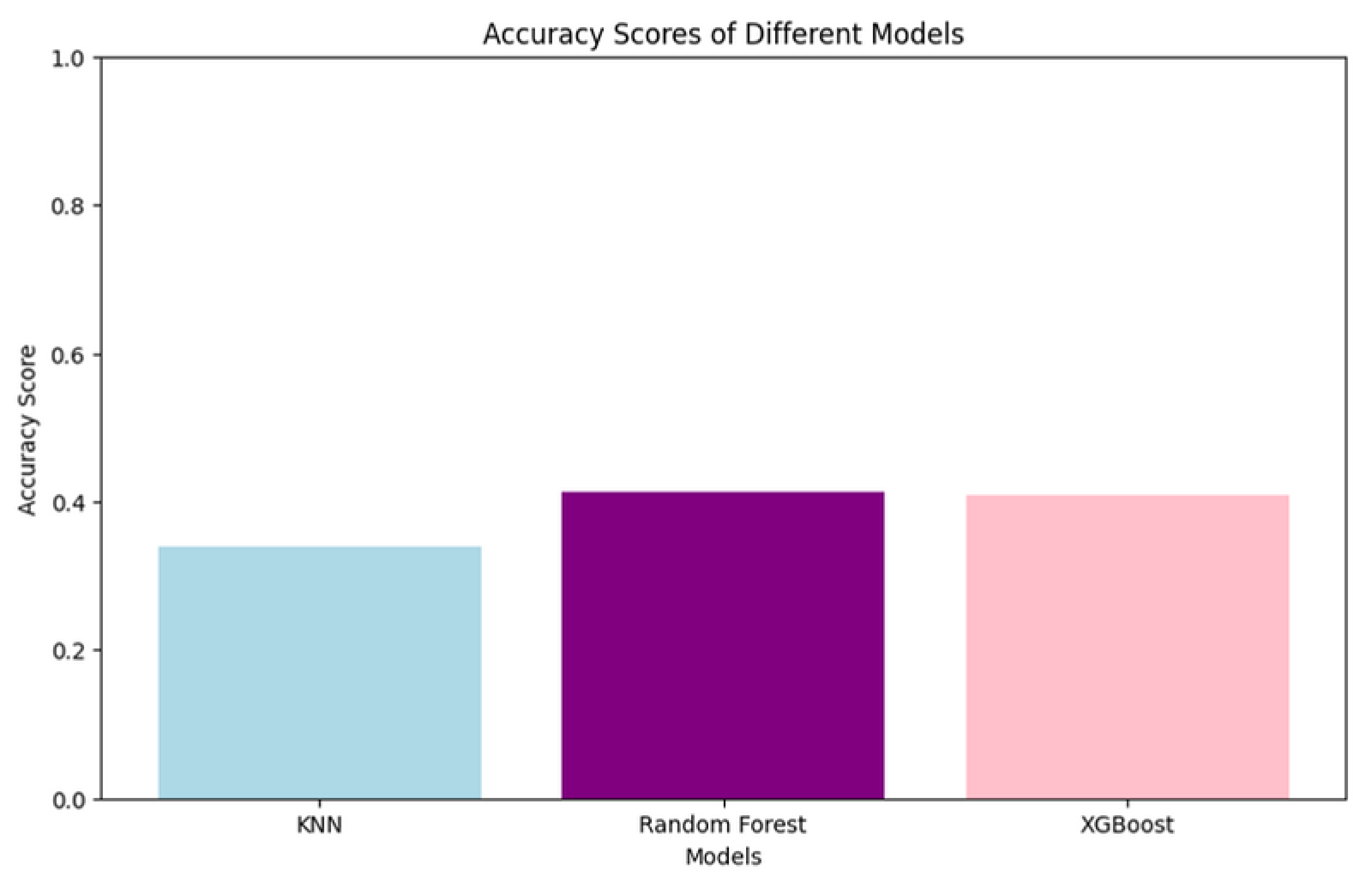


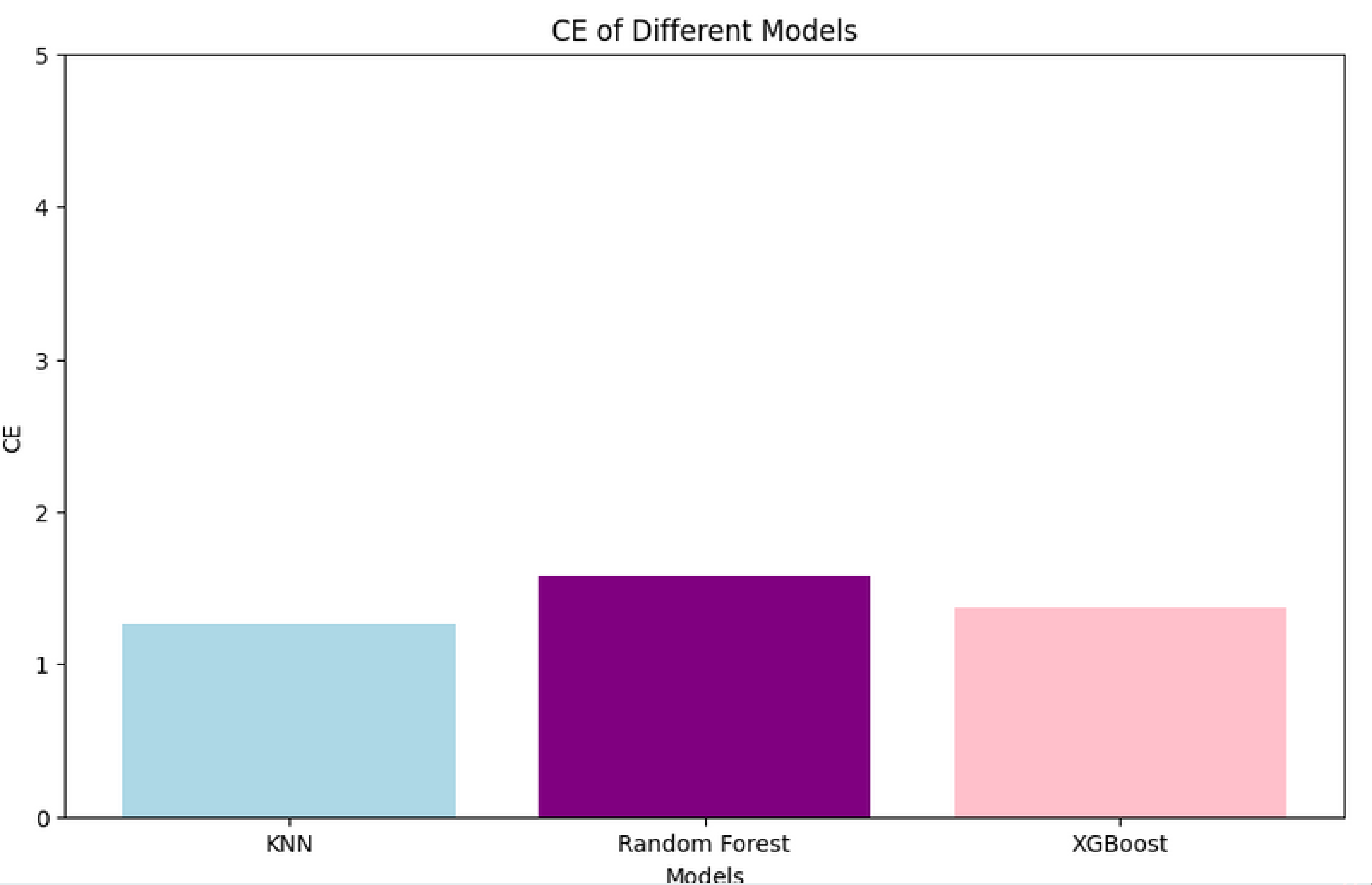
Vaccinated to adoption speed graph



Correlation

For this dataset we didn't find a big correlation for features and AdoptionSpeed, the best was 0.11 for Breed 1 and 0.1 for Age and all the other features have a worse correlation.


Accuracy Scores of Different Models
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy Score
KNN
Random Forest
XGBoost
Models


CE of Different Models
CE
5
4
3
2
1
0
KNN
Random Forest
XGBoost
Models

Plan for future work

- Analyze correlation between the features and adjust them accordingly
- Find the best set of features using feature selection algorithms like RFECV
- Find the best set of hyper-parameters
- Use PCA
- Run chosen models again to see if we get better results