The connection between personal, physiological characteristic and the probability of adoption of pets from the shelter in a family

I have chosen such topic because nowadays, there is a big problem of overfilling the shelters, especially in Ukraine. I hope, my research and his results will help correctly make ads on shelter websites, to make faster adoption of any pet and they all to find their own family.

Implementation and results:

I have data set of an adopted pet from different shelters. There are such features like Type (Cat or Dog), Name, Age, Breed1, Breed2, Gender, Color1, Color2, Color3, MaturitySize, FurLength, Vaccinated, Dewormed, Sterilized, Health, Quantity, Fee, State (place), RescuerID, VideoAmt, Description, PetID, PhotoAmt, AdoptionSpeed.

For my research, I do not need all this feature so via method drop() in pandas I deleted such features like State, Name, RescuerId, VideoAmt, Description, PhotoAmt, PetID.

After that, I checked my data on missing values via method .isnull().sum() . Thanks to my luck there no missing value.

Then I used Z-score to find all outliers. While calculating the Z-score we re-scale and center the data and look for data points which are too far from zero. These data points which are way too far from zero will be treated as the outliers. In most of the cases a threshold of 3 or -3 is used i.e if the Z-score value is greater than or less than 3 or -3 respectively, that data point will be identified as outliers. In my case, there were 2178 outliers.

Then I use Recursive Feature Elimination to know feature ranking. I have got that:

Type -> 1,
Age -> 6,
Breed1 -> 7,
Breed2 -> 9,
Gender -> 1,
Color1 -> 3,

Color2 -> 5,
Color3 -> 4,
MaturitySize -> 1,
FurLength -> 1,
Vaccinated -> 1,
Dewormed -> 1,
Sterilized -> 1,
Health -> 1,
Quantity -> 2,
Fee -> 8

From these results, we can see that people make their decisions on of the type of pet, gender, maturity size, fur length, vaccinated, dewormed, sterilized and health. It was weird for me that breed didn't have a big influent but as we see later in shelters there is a more mixed breed and this why breed do not have such an impact.

After that, I divided my data on cats and dogs to see the amount of each in shelters.

From overall dataset, dogs have 54.23% and cats have 45.77%. Now we can say that according to this dataset there are more dogs in a shelter than cats.

Than I divided dogs and cats on high and low adoption speed. After that, I divide my data into cats and dogs to see the amount of each in shelters.

From overall dataset, dogs have 54.23% and cats have 45.77%. Now we can say that according to this dataset there are more dogs in a shelter than cats.

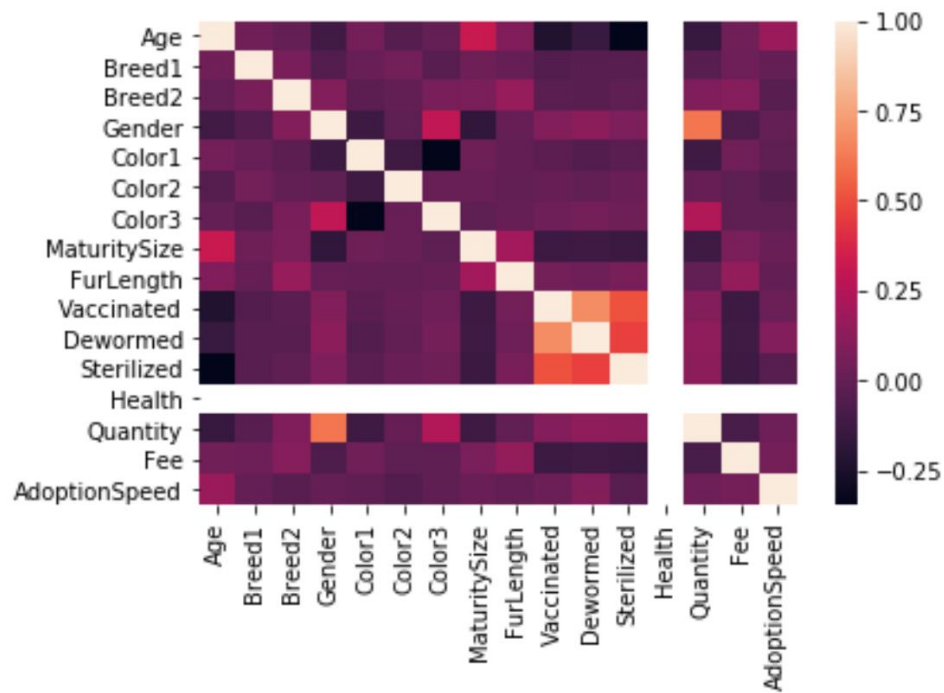After that, I wanted to know the correlation coefficient of each group. For this, I used "seaborn" library.

For dogs with high adoption speed:
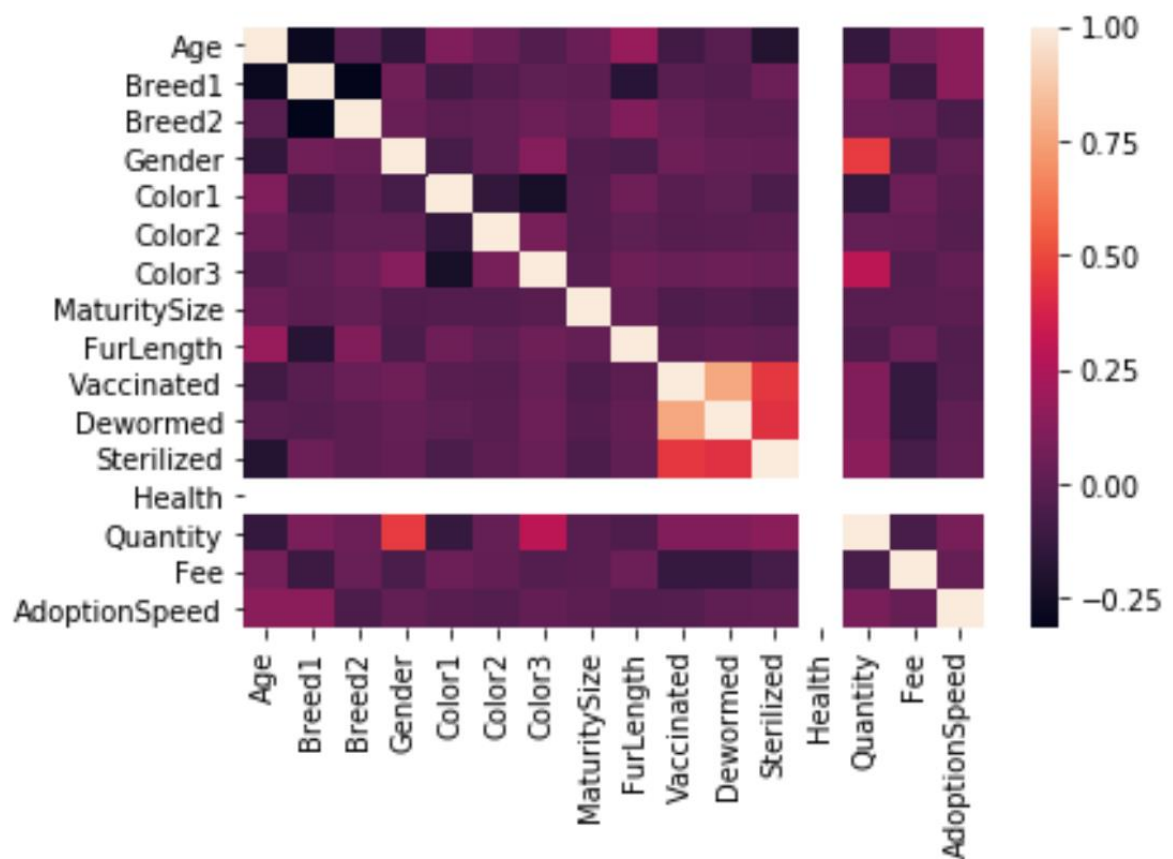
For dogs with low adoption speed:
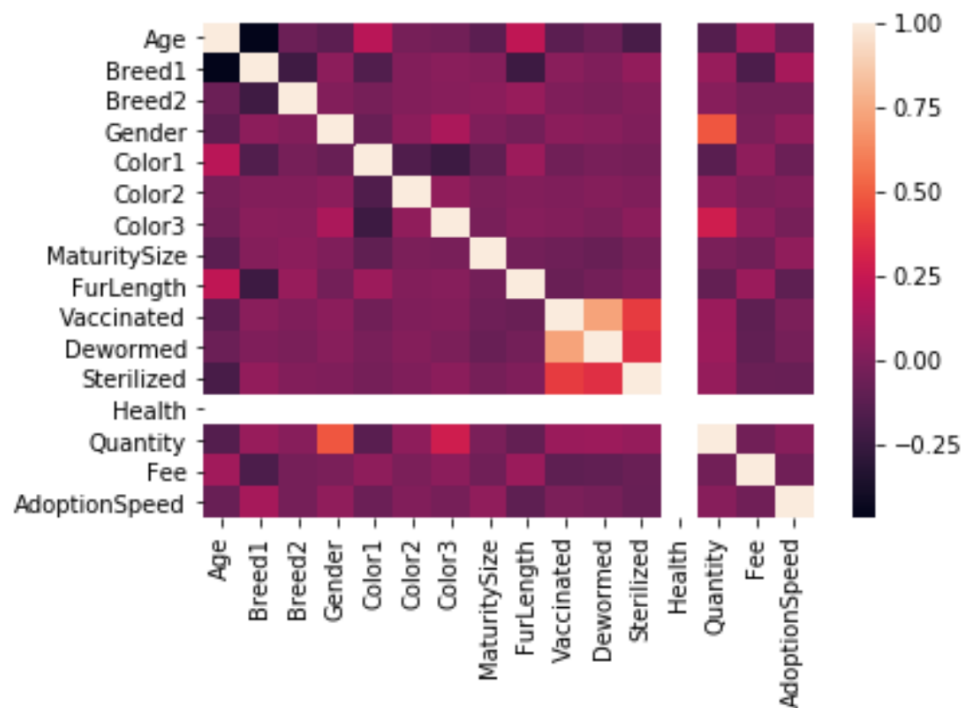
For cats with high adoption speed :

For cats with low adoption speed :

*All another percentages are in the poster