After exploring different features of the 311 calls, we would like to build up a predictive model to predict how long a case may take based on animal type, season, borough, agency, etc.

We created a new variable indicating what kind of animal was mentioned in each case. Before digging into the model, we first looked at the features of complaints related to different animals and their relationship between other key variables. The following two circos plot show the relationship between animal types and the season, as well as the relationship between animal types and the county (borough). Please note that we did not include dogs and pests in these plots. One reason is that they make up a very big part of all animal-related cases. Including them would make the other animals’ portion invisible. The other reason is that dog and pest complaints are almost equally spread across all seasons and boroughs. Besides, they are analyzed in detail in the previous part.

From the above two graphs we can observe some interesting facts. For example, in the left graph, we can see that bees-related complaints seldom happen in winter. Most of them are during summer. This is align with the fact that bees are more active in spring and summer, hence people complain more about illegal bees/beekeepers during these seasons. Another interesting fact from the right graph is that animal-related complaints in Manhattan are almost all about wildlife and pigeons. On the other hand, the complaint types in the other boroughs are more diverse.

Next we would like to fit several decision trees based on some of the key variables from the 311 complaints. We picked the case length as our dependent variable. However, as we would like to make the model a classification model, we created a ‘CaseLength’ variable with 6 categories: Same Day, In 3 Days, In a Week, In 2 Weeks, In a Month, Over a Month.

The following graphs show several decision trees with different combinations between the key independent variables. The color of the labels are ranged from dark green to dark red, indicating the time from short to long. The bandwidth of the edges represent the proportion of cases falling into that category.

The first Decision Tree takes Animal as the only predictor. We can see that if the case is about wildlife, then it will be solved within a day; if the case is about cats and dogs, it will take shorter time (3 days) too; however, when the animal involved is a bit unusual, the case will take about a month to close.

The second Decision Tree takes animal and borough into consideration. We can also observe similar pattern as the above: cases related wildlife, cats and dogs will take less time, while the others will take longer. Also, we can see that cases in Bronx and Staten Island are predicted to take longer time.

The third Decision Tree studies animal and seasons. We can clearly see in this tree that winter cases tend to have longer processing time. It is reasonable to assume so because the holiday season and bad weather conditions.

The fourth Decision Tree shows the predictions based on animal and agency. It is very clear that DOB, DOHMH and HPD generally take longer time to close cases. This can be related to the nature of the kind of cases they are dealing with.

Besides, we also made some other decision tree models. They are shown in the Appendix.

We also tried including more than 2 variables. They were definitely going to be good to increase the prediction accuracy. However, the tree was very large and was hard to be clearly visualized here.

So from the above Decision Tree graphs, we know that with the information about the animal type, borough, agency and seasons, we can have a general idea about how long a case may take.