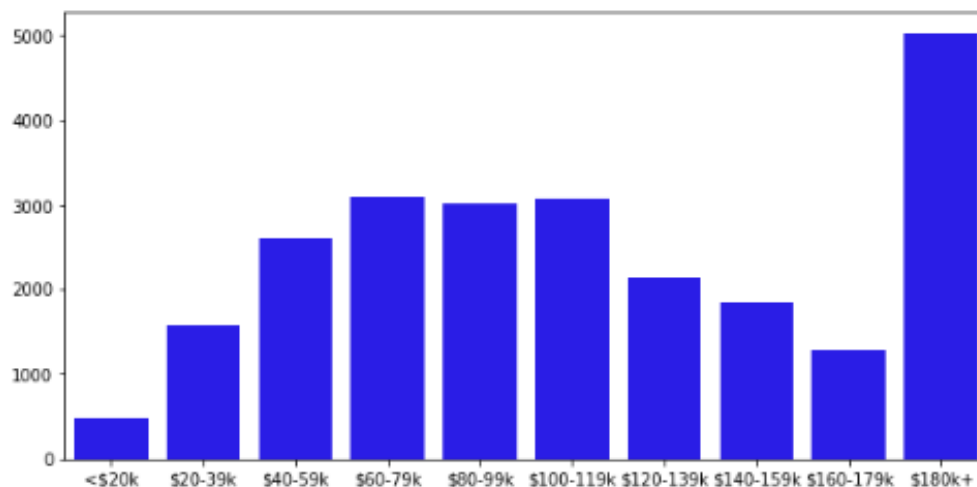


Question 1 – Does owner income influence the health conditions reported in the study?

Our study sought to analyze the greatest influences in a dog's health, and to determine whether any of these factors might present as valuable predictors for whether a dog may experience increased risk for certain health outcomes. Understandably, many of these factors are outside the control of the dog itself, and are instead influenced by the surrounding the environment, which is much more associated with characteristics of the dog owner.

As a result, one of our first questions was whether the income of the dog owner might act as a predictor for health outcomes. It stands to reason that dog owners with higher incomes may be able to afford expensive care for their pets, whether that care involves purchasing more expensive food, veterinary visits, prescription medications and preventatives, etc. However, rather than assume that dogs with higher income owners would be healthier, we simply intended to analyze for differences between income groups in the presentation of health outcomes.

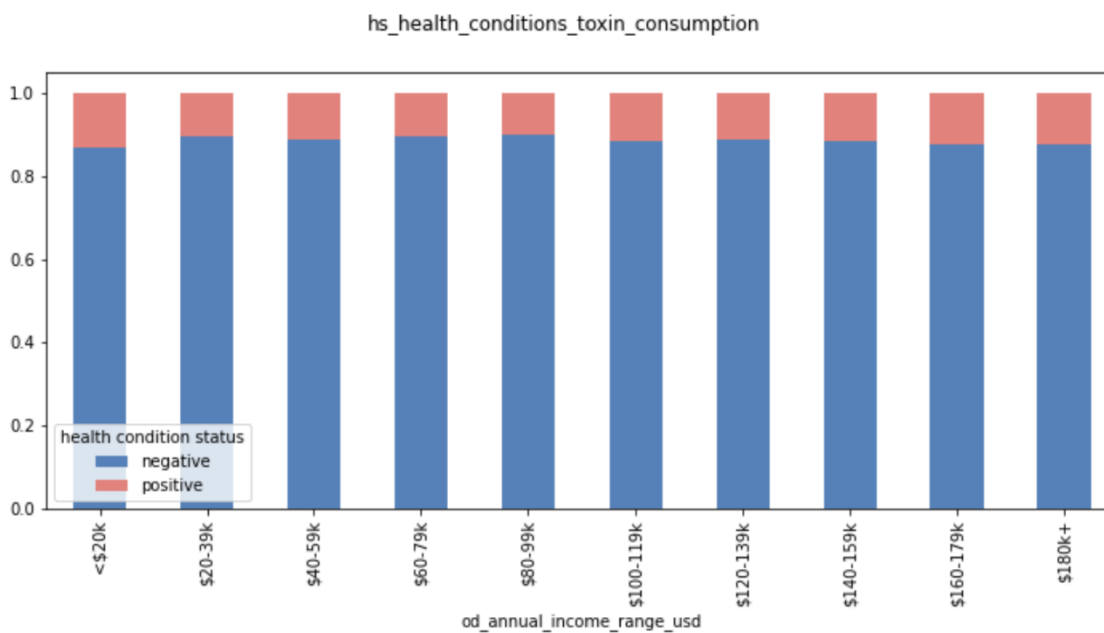
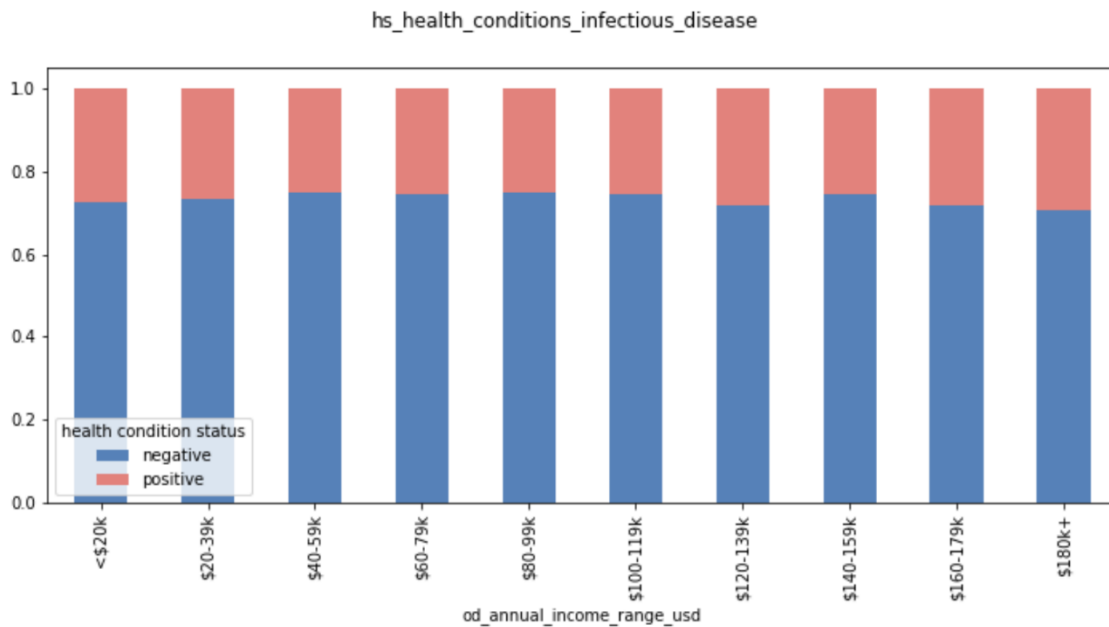
To better understand the presentation of income in our data set, we needed to understand the distribution of different income groups. The below histogram describes the frequency of each group within the data. Records from owners who preferred not to answer were excluded ($n = 3,357$ out of 27,541 total records). While there is almost a normal distribution of income groups, it's important to note that the most frequent income groups was that of the owners making the highest income. This suggests that the data is skewed towards owners of a higher income group, rather than having an even distribution of owner income within the study.



Our next step was to perform a Chi Squared analysis in our owner data set to determine whether each of the health outcomes reported in the data might be more prevalent in certain owner income groups than others. For each income group, we sought to analyze whether reported health conditions followed a pattern, for example, whether certain health outcomes became more common as owner income increased or decreased. Although most health conditions did not return a significant p-value ($> .05$), two specific health conditions did identify differences among groups that were statistically significant. These included infectious disease ($p = .0000383$) and toxin consumption ($p = 0.04$).

However, one drawback to the use of Chi Squared analysis is that it can return a significant p-value due to any differences present in any of the analyzed groups, regardless of whether the difference is great enough to suggest further analysis. This was the case with our analysis of own analysis, as a normalized

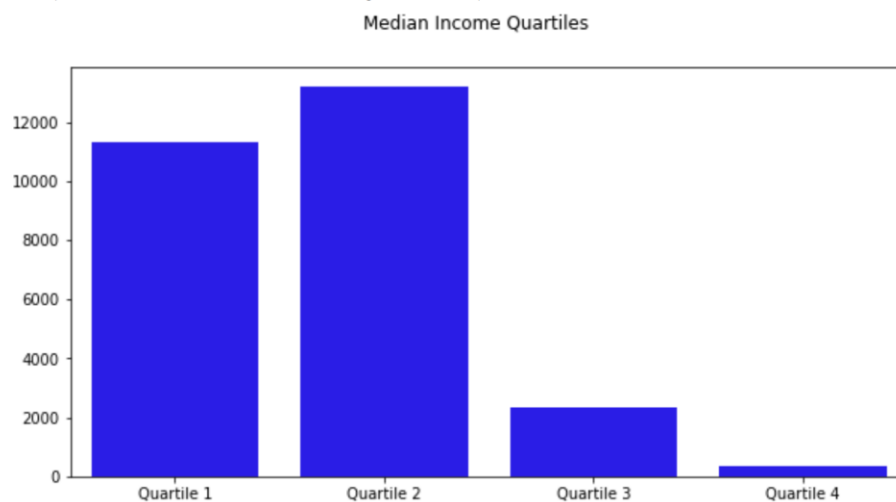
table reporting the percentage within groups reporting these health conditions was particularly small. Overall, this suggests that there does not appear to be a significant difference within income groups and the prevalence of health conditions reported in the data. The following plots demonstrate the distribution of records reporting the presence of infectious disease or toxin consumption across owner incomes. While there are differences among the different income groups, there does not appear to be any pattern within income groups or specific groups demonstrating a marked difference in the reporting of the health condition.



As can be seen in the above charts, the incidence of infectious disease and toxin consumption shows very little variation within the reported income groups. This suggests that owner income itself is not

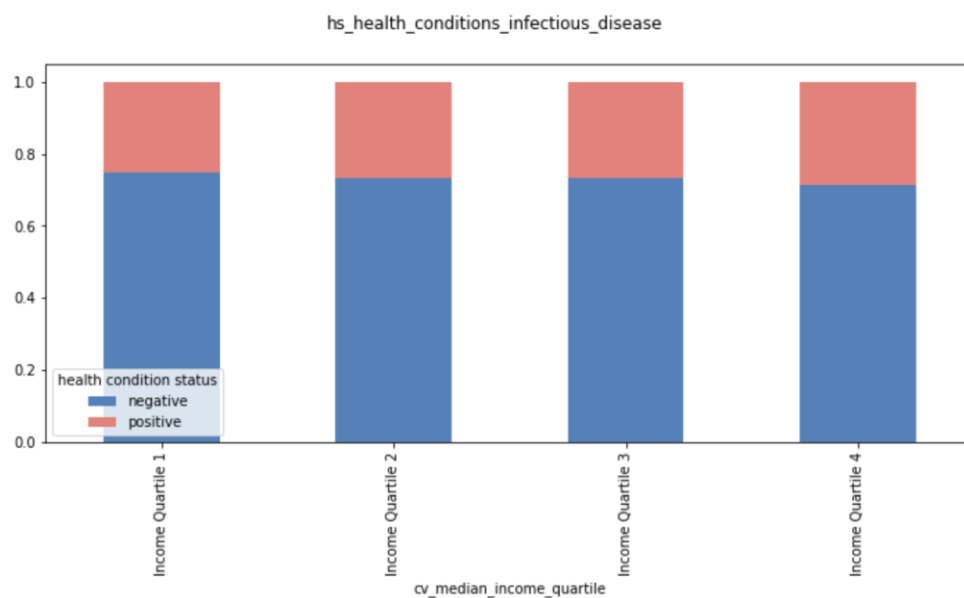
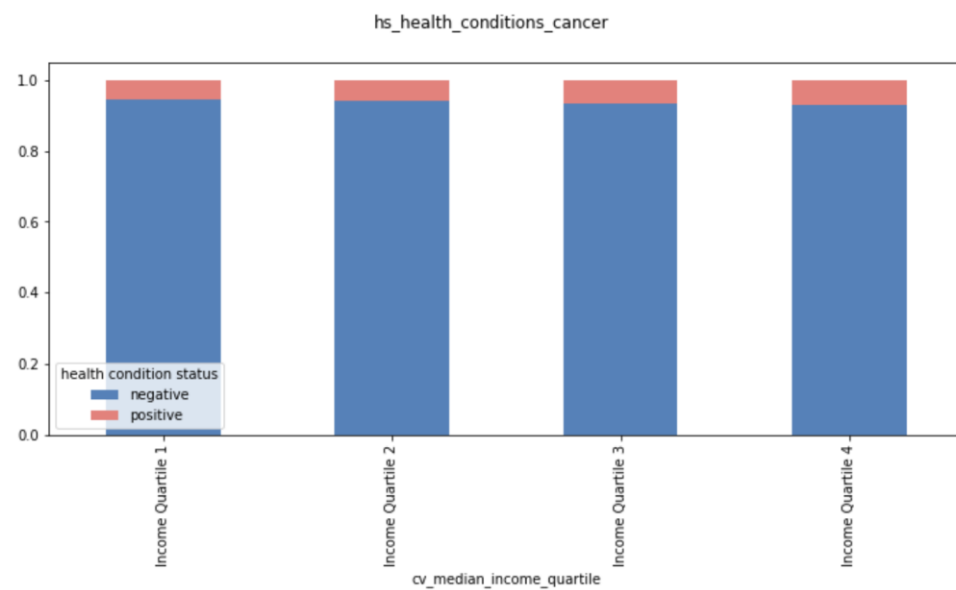
likely a good predictor of the incidence of the different health conditions reported in the data. As an alternative to owner wealth, we next looked to the median income of the neighborhood where the dog lived to determine if there might be any relationship between the wealth of the neighborhood and the dog health.

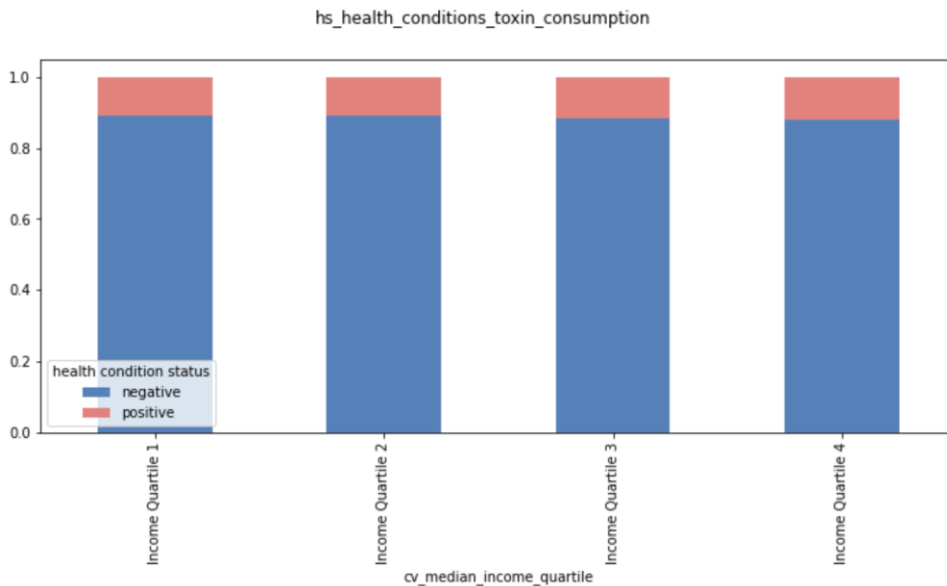
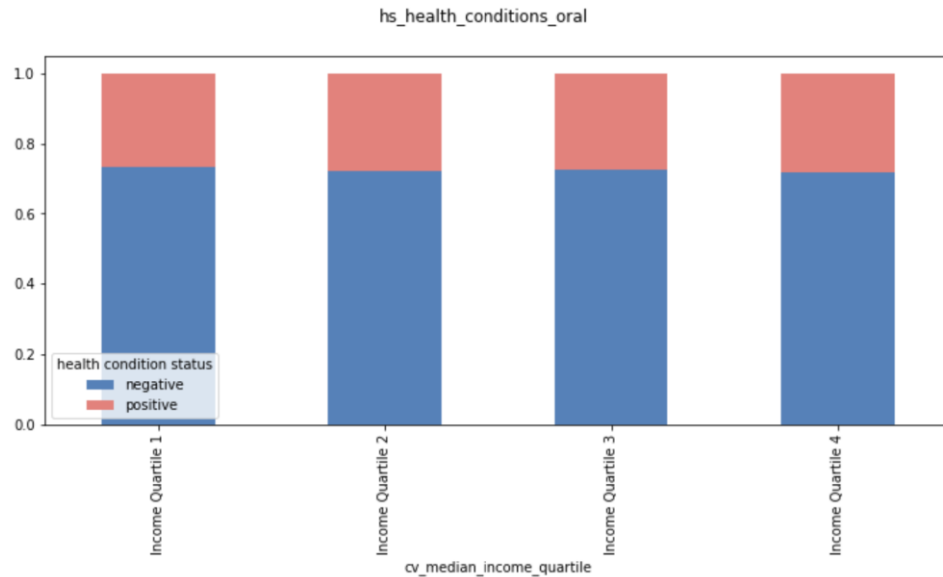
Rather than use the raw median income values in the data set, we decided to group median income values (reported according to 2020 Census Tracts based on the owner's address) into quartiles and use the same chi squared analysis to determine if there might be any differences in prevalence of health conditions within the different quartiles. Our first step in the analysis was to understand the frequency of the different median income quartiles in the data:



The above chart describes a different picture from the view of owner income, which indicated that the largest group of responses came from those in the top income group. Instead, in the case of the income quartiles, the largest income group (Quartile 4) is now the least frequent group in the data. The vast majority of responses come from Quartiles 1 and 2, which are the lower median income quartiles.

To analyze whether certain health conditions may be more prevalent in certain income quartiles, we performed a Chi-Squared analysis on the grouped data. Similar to our Chi-Squared analysis against owner income, we identified differences within groups that exceeded a significant p-value. However, these differences were very small, despite being differences among the median income quartiles. The health conditions which returned significant results included cancer ($p = 0.00004$), infectious disease ($p = 0.000$), oral health conditions ($p = 0.00576$), and toxin consumption ($p = 0.00477$). Again, while our Chi Squared analysis identified differences within groups, the differences within groups were largely negligible and did not truly suggest a relationship between the median income of the neighborhood and the health of the dogs in the study.





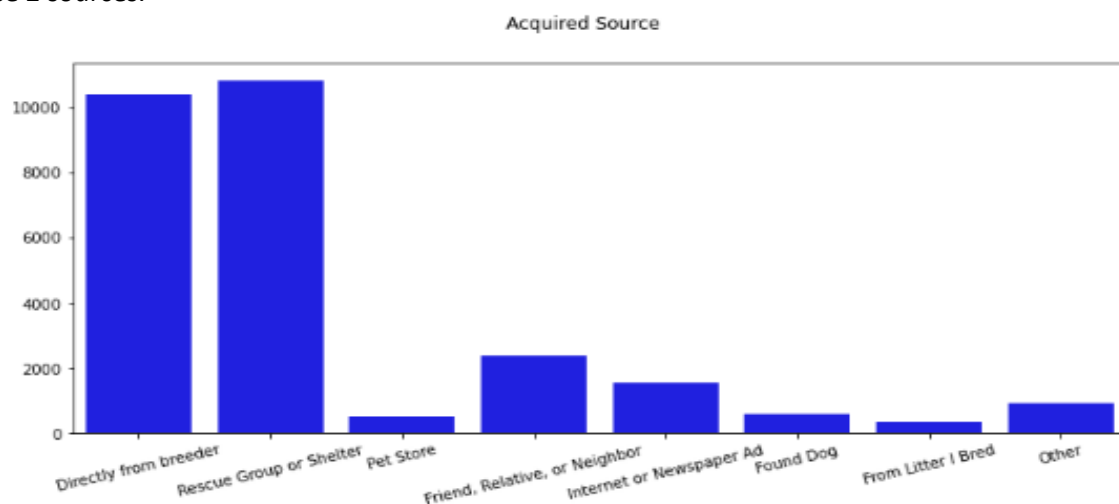
Ultimately, it would appear that the prevalence of the different health conditions in our data set seems pretty evenly distributed across owners of all income groups, and even different median incomes. This is in contrast to our initial hypothesis that dogs in wealthier incomes would be healthier than dogs owned by owners with lower incomes, or that dogs living in more affluent neighborhoods would be healthier.

That said, there are many potential confounds in our data set. For example, this analysis simply looks into the prevalence of dog health conditions across different income groups. It is possible that dogs of certain income groups may encounter certain health conditions at an earlier age compared to other income groups. This would be an interesting follow up study, as the current analysis focused on whether the health condition was reported at all for a given dog. Similar to humans, as dogs age they are likely to experience some deleterious health conditions, although it is possible that these health conditions may be encountered differently based on environment, care, experience, etc.

Question 2 – Does the “source” of the dog influence the health conditions reported in the study?

In addition to investigating wealth as a predictor of dog health, we also decided to investigate whether the source of a dog might act as a predictor of the health conditions reported in the study. It is possible that dogs spending time in a shelter, or have been abandoned, may have been at greater risk for health conditions, or may have received less care than dogs that had been acquired at a young age directly from a professional breeder. Thus, we hypothesized that dogs acquired from certain sources might present the incidence of certain health conditions at a greater rate.

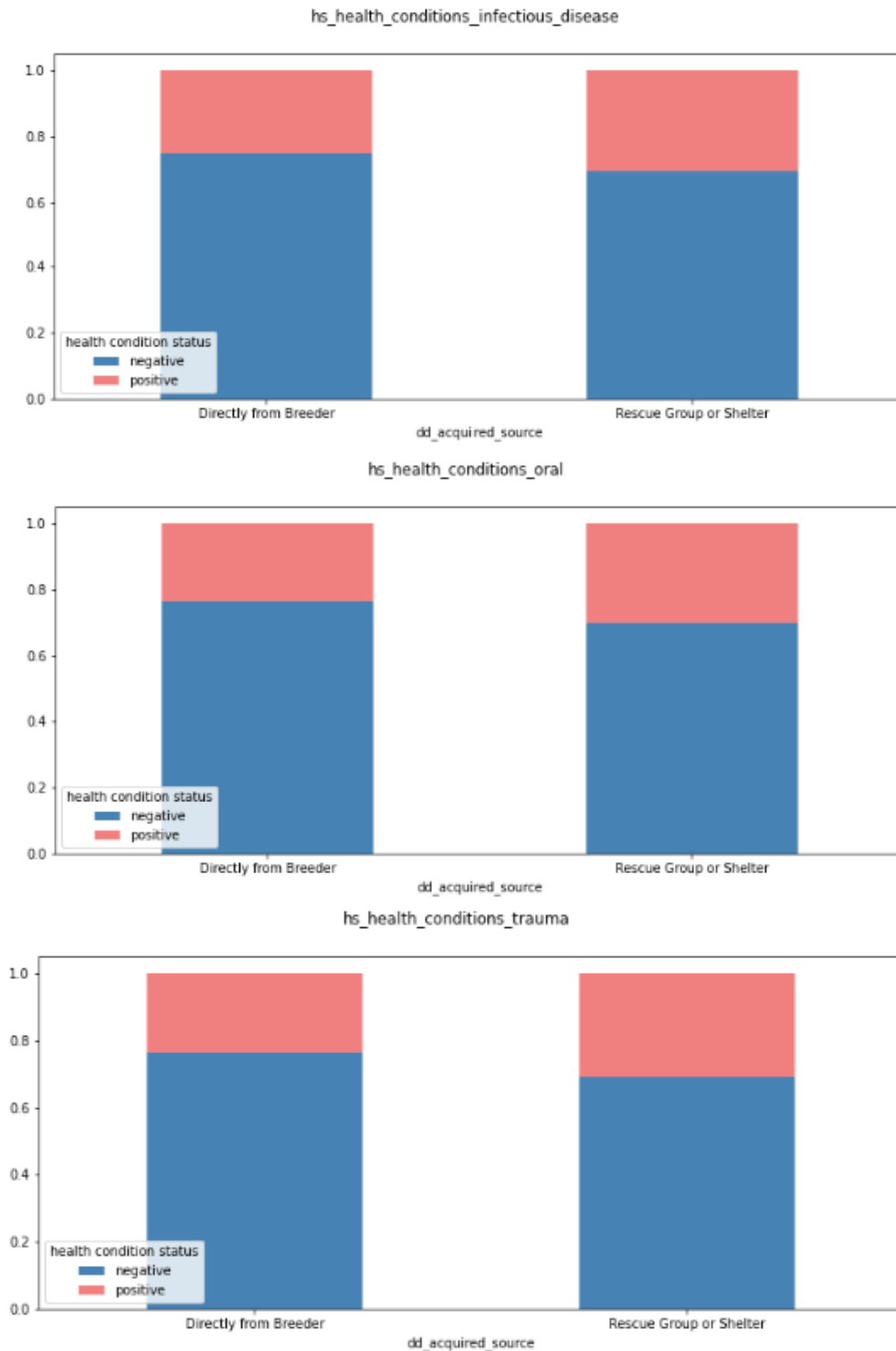
Our first step was to analyze the frequency of the different sources reported in the data. It quickly became clear that the dogs in the data set were typically acquired from one of two sources – either directly from a breeder, or through a rescue group or shelter. All other groups were present in the data at such a lower frequency that we decided to focus our analysis on dogs acquired through either of these 2 sources.



To analyze differences between these groups, we again performed a chi-squared test to determine whether any of the health conditions reported in the study were more frequent in either group. Although certain health conditions returned significant p-values but negligible differences between the two sources, a few interesting results were identified:

- Infectious disease was reported more frequently in adopted dogs (30%) than dogs acquired from a breeder (25%).
- Oral health conditions were reported more frequently in adopted dogs (30%) than dogs acquired from a breeder (24%).
- Trauma was reported more frequently in adopted dogs (31%) than dogs acquired from a breeder (24%).

Our chi-squared test results indicate that for certain health conditions, there seem to be significant differences in reported incidence between dogs acquired from a breeder vs. those acquired from a rescue group or shelter. The below charts show the normalized difference in frequency between dogs acquired from a breeder vs a shelter for dogs impacted by infectious disease, oral health conditions, or trauma.



Due to these identified differences, we attempted to build a Machine Learning model to predict whether a given dog may experience one of these health conditions. We began with Trauma, which showed the greatest difference in reported frequency between adopted dogs (31%) and dogs acquired through a breeder (24%). As a first step, we reviewed the different dog demographic attributes in the data set for any potential correlation with the presence of reported trauma.

Unfortunately, dog demographic attributes did not show much sufficient correlation with trauma. This suggests that dog demographic information in the data set does not have much of a relationship with the presence of trauma, even though trauma is more present in adopted dogs than dogs acquired through a breeder. The dog demographic variables showing the strongest absolute value in correlation coefficient include the following:

Variable	Correlation with Trauma
dd_alterate_recent_residence5_weeks	-0.666667
dd_alterate_recent_residence4_weeks	-0.346410
dd_alterate_recent_residence3_weeks	0.334807
dd_activities_search_and_rescue	-0.171320
dd_birth_year	-0.147815

The top 3 variables showing the strongest correlation with the presence of Trauma have to do with the amount of time a dog has spent an in alternate residence, though still with the same owner. The fourth variable reports whether a dog's regular activity involves search and rescue efforts, while the fifth variable report the actual year that a dog was born. These data points do not seem to be indicative of whether a dog is adopted or acquired through a breeder, nor would they seem helpful in predicting whether a dog that has been adopted or picked up from a breeder would subsequently report trauma as a health condition.