

DSC 530 Final Project

Chase Lemons

Bellevue University

Introduction

This project consisted of an analysis of the Austin Animal Center Shelter Outcomes dataset from Kaggle. This dataset was about cats and their outcomes. I had always heard that black cats get treated differently than other cats. Hence, I wanted to know, do black cats have the same outcomes as non black cats? Through the data exploratory analysis we determined that black cats do have different outcomes.

EDA

For the exploratory data analysis, we looked at age, color, outcomes, sex and breed. The issues that we faced with this data set is that it was all very much categorical. After committing to this data set there was no turning back. Specifically, it was the Outcomes and color we were the most interested in. So we compared the histograms of outcomes by black and non black cats. We saw some minor differences but nothing that jumped out at me.

Ultimately, we assigned a number to the outcomes and calculated the average outcome for black cats and non black cats. (The assumption to do the average outcome, may be an incorrect assumption and maybe should not have been used.) My thought process on why perform a hypothesis test on the averages of the outcomes, was thinking that if black cats and non black cats had the same outcomes then the average of those numerical outcomes should be equal. Hence when the p value came back as statistically significant, which meant that we reject the null hypothesis of assuming the means would be equal. So we can arrive to the conclusion that black cats do not have the same outcomes as non black cats.

For this project, it would have been better to not have as much categorical data to work with. Maybe I should have looked at time spent at the shelters for black cats versus non black cats. If there was more data from another city or another shelter to test and see if this hypothesis holds true for cats outside of Austin.

Conclusion

Overall, this was a fun project which helped push me to utilize all the skills that were taught this quarter. I feel more confident with Python after this quarter and after this project. And now I know that according to the Austin Animal Center Shelter Outcomes data set from Kaggle, black cats do not have the same outcomes as non black cats.

References

Downey, A. (2015). Think Stats (2nd ed.). Sebastopol, CA: O'Reilly Media.

Schutt, R., & O'Neil, C. (2014). Doing data science. Sebastopol, CA: O'Reilly Media.

Schlegel, A. (2018, February 16). Austin Animal Center Shelter Outcomes. Retrieved November 14, 2018, from <https://www.kaggle.com/aaronschlegel/austin-animal-center-shelter-outcomes-and/version/1>