**Structure**

1. Introduction: introduce the problem, why interesting

One of the most reported and observed bird species in the United States is the Northern Cardinal (*Cardinalis cardinalis*). This species is a particularly popular ‘backyard’ species and is widely known by people across the United States (1). This specific bird is especially interesting to study on a quantitative basis, as with high levels of observation frequencies, and large volumes of interactions with people and urban development, data analysis can unveil how populations of people and their behaviors impact the frequency of Northern Cardinals throughout US counties. Northern Cardinals have seemingly become dependent on certain human development features, such as frequenting specific neighborhoods and houses to feed from, as they are generally safe and easy to find food for seed-eating birds. This project will dive into the research question: how does urban development and population influxes in United States counties impact observation levels of Northern Cardinals. This research project will focus on the year 2022 and look at the County-Level Rural-Urban Continuum Codes and Urban Influence Codes, County-Level Economic Typology, Natural Rate of Change, and County Immigration Rates. As Northern Cardinals are commonly observed as ‘backyard’ bird species, this project hypothesizes that a more developed county sees a higher level of Northern Cardinal observations. This suggests a greater environmental responsibility for humans to take care of local urban environments, as some bird species are adapting to these developing environments.

(1)<https://birdwatchinghq.com/common-birds-in-the-united-states/>

2. Data: present the dataset, describe key features

This project contains two datasets from different sources and were merged. The first dataset comes from eBird, Cornell University Ornithology Lab’s digital bird observation collection project. Upon request, Cornell granted access to the data sets. From here, data on Northern Cardinal observations in all counties in the United States were selected. The dates selected were all observations in January, April, July, and October in 2022. This is because the data files are so large, so for efficiency, the datasets were split up. Additionally, this helps avoid seasonal observation biases, in case this species as strong seasonality shifts. A total of 982,667 observations were collected for this project. The cleaning process involved narrowing the columns down. The data was aggregated by county level, and kept the sum of observation counts, average duration in minutes spent on the observation, and the effort distance in km that the observation took place over. The additional columns were kept to see how features of the observation are associated with observation counts.

3. Model:

(a) Present the model you estimate, argue for your model choice

(b) Discuss your variables, process of feature engineering. (In words, put all needed graphs, estimation into the appendix if needed.)

(c) Show core results. Interpret what you got precisely.

4. Generalization and external validity (robustness check)

(a) Show some robustness / alternative models.

5. Causal interpretation / main summary

(a) Summarize your findings. Discuss room for a causal interpretation.

6. Conclusion

(a) Conclude and make business / policy comments / recommendations