



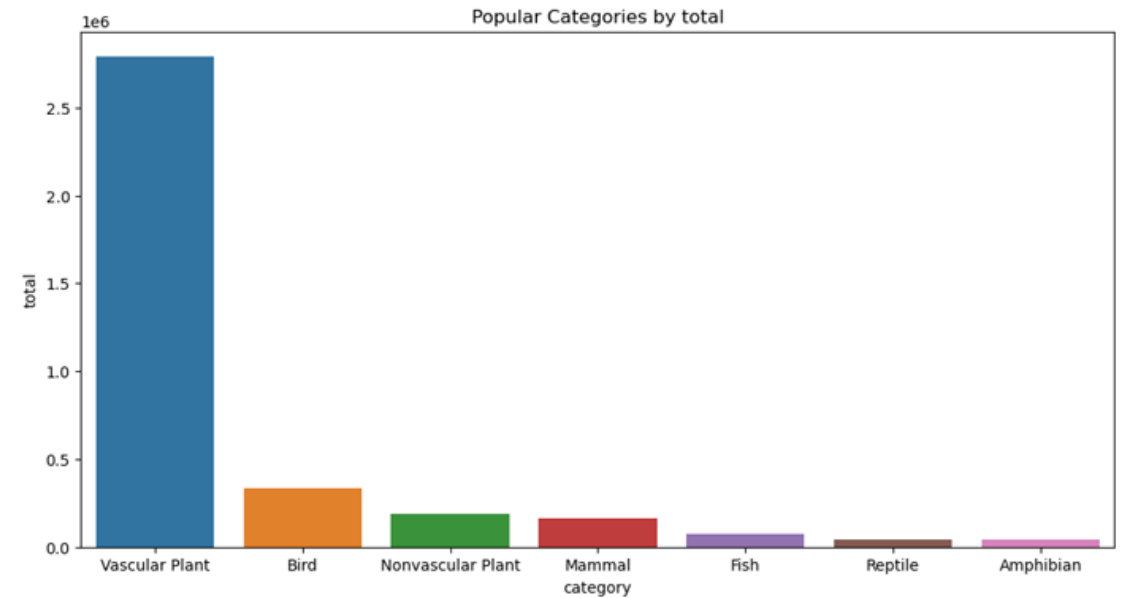
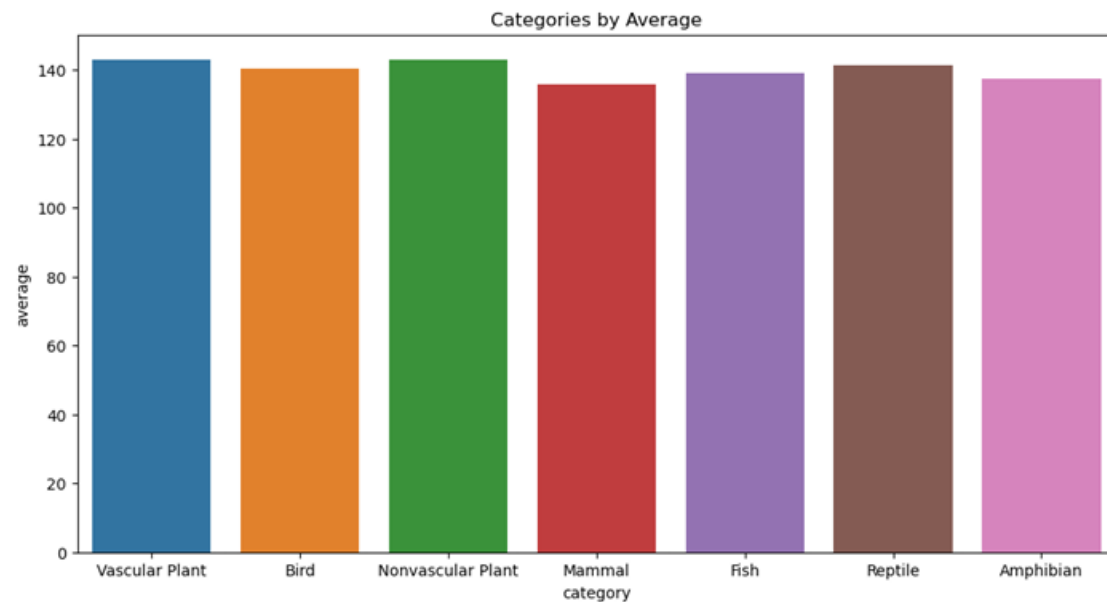
Biodiversity Analysis Findings

Findings Summary

- The mean observation across all categories was the same except for vascular plants and mammals.
- Mean observation across all parks differed, Yellowstone park having the highest.
- Park and category variable have a weak association.
- Category had a moderate association with conservation status.
- There was a weak association between conservation status and parks.
- Mammals had highest count of endangered species, with red wolf being the most endangered with lowest observations.
- Random Forest Classifier was highly accurate in predicting a species by its conservation status with observations being the most significant feature by far.

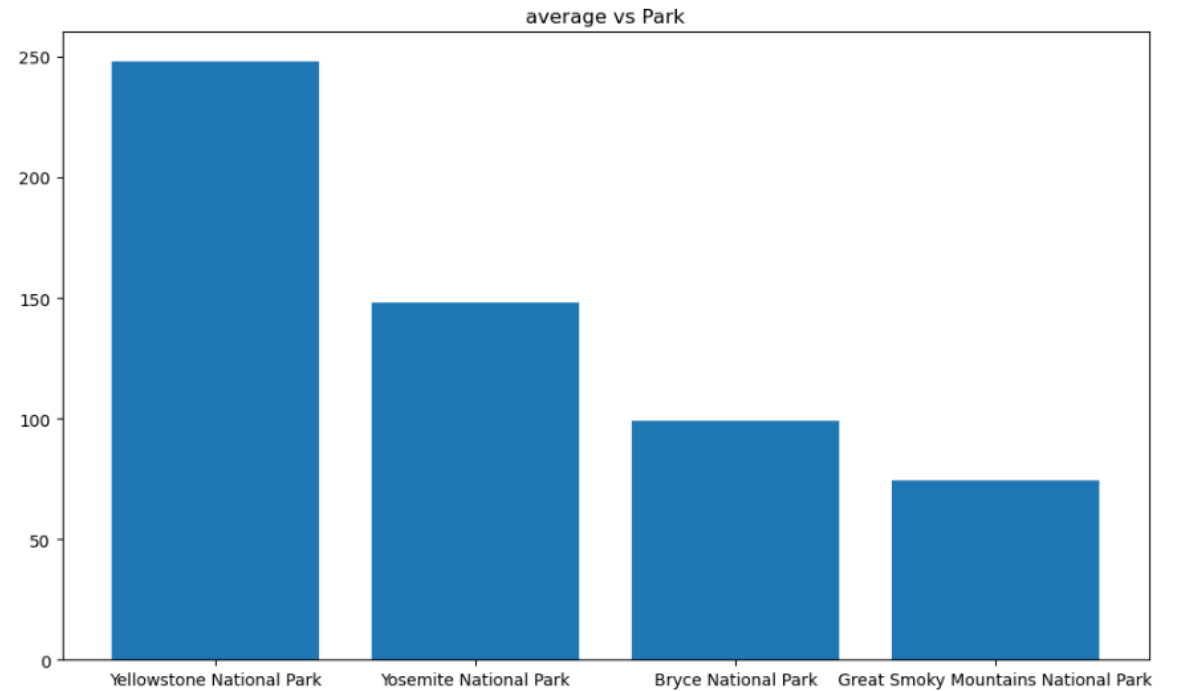
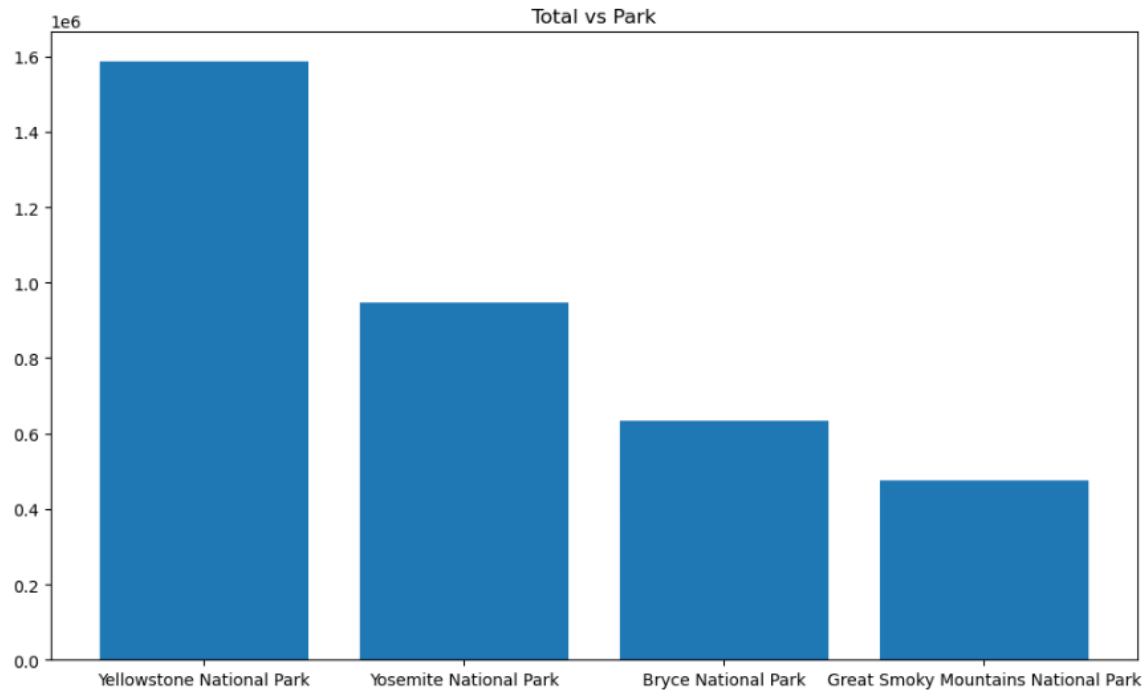
Category Summary

Vascular plant was most popular species, however on average all similar by count of observations.



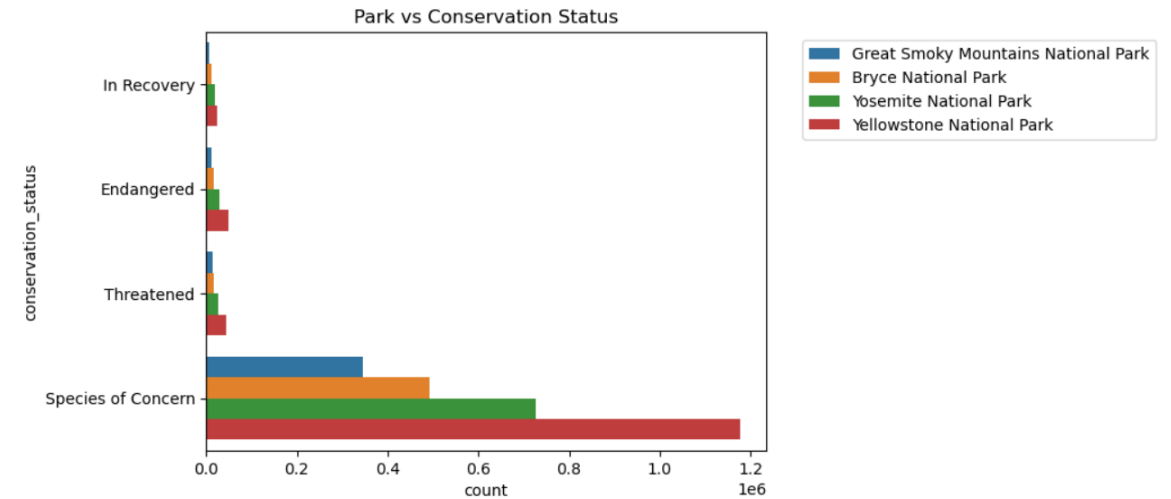
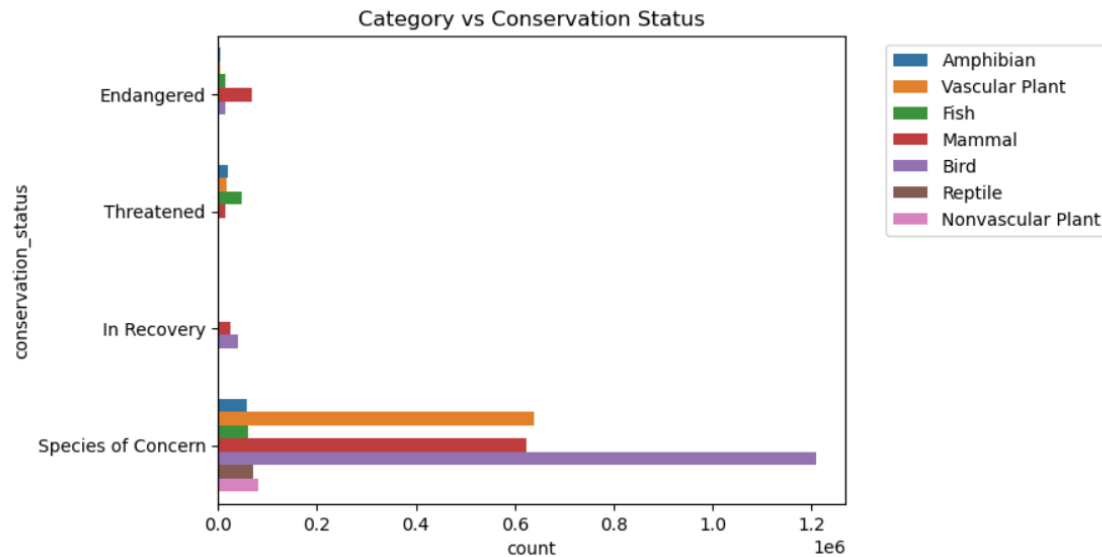
Park Summary

Yellowstone National Park has most observations and on average each park differed.



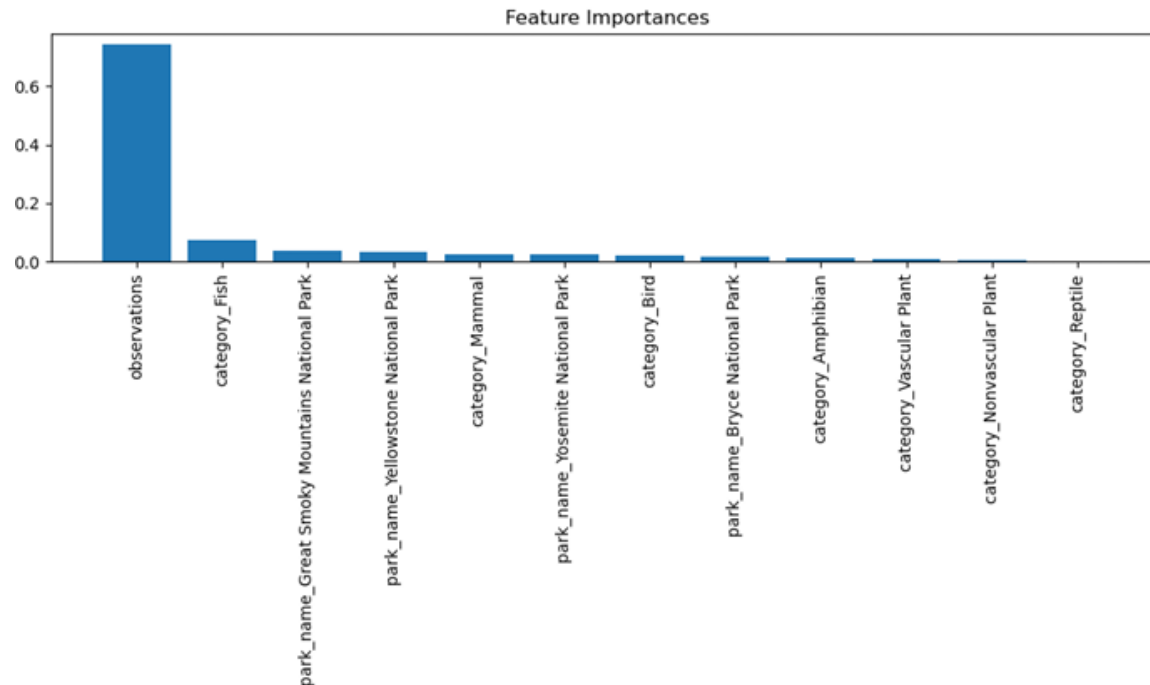
Conservation Status Summary

- Mammals were most endangered, and further analysis showed red wolf having lowest count.
- Yellowstone National Park had highest count of endangered species.



Prediction Model Summary

- Random Forest Model was 97% accurate in predicting conservation status.
- Observation count had significance of 0.74, most significant by far.
- Table sums precision for each conservation status.



Accuracy: 0.97

precision

Endangered	0.83
In Recovery	0.85
Species of Concern	0.99
Threatened	0.99

Conclusions

- Data was incomplete, had a lot of NaN values in endangered species. Bootstrap method was used to simulate more accurate data.
- More categories and data would be needed for further analysis.
- A time variable would be useful for modelling.
- From data given decent insights were made and prediction model was highly accurate. However, one variable accounted for most.