Project Biodiversity in National Parks

Interpretation of National Park Service data on threatened species in different parks.

Description of the data

To carry out this analysis, I used the following data :

species_info.csv - contains data about different species and their conservation status and observations.csv - holds recorded sightings of different species at several national parks for the past 7 days.

Issues raised in this analysis.

In the course of my work I have tried to answer the following questions:

1. Are certain types of species more likely to be endangered?
2. Are the differences between species and their conservation status significant?
3. Which species were spotted the most at each park?

To answer these questions, I used the following libraries:

Pandas

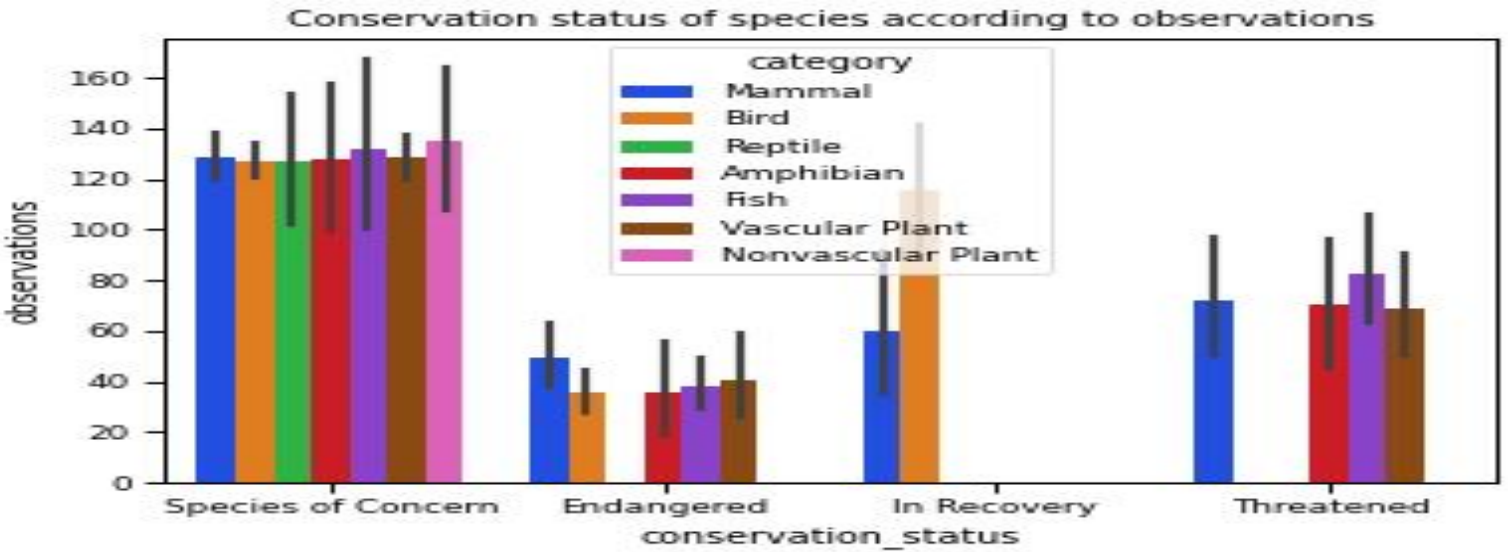
Numpy

Scipy

Matplotlib

Seaborn

The following graph is an answer to the first question:



The graph above shows that mammals are most at risk.

I used the Chi-square test to answer the second question

Chi-square test

To test for an association between two categorical variables, we can use a chi-square test. The null hypothesis for a chi-square test is that there is no association between the variables and the alternative hypothesis is that there is an association between the variables. A Chi-Square test can be implemented in Python using the chi2_contingency() function from scipy.stats.

The following code puts into practice what has been explained above.

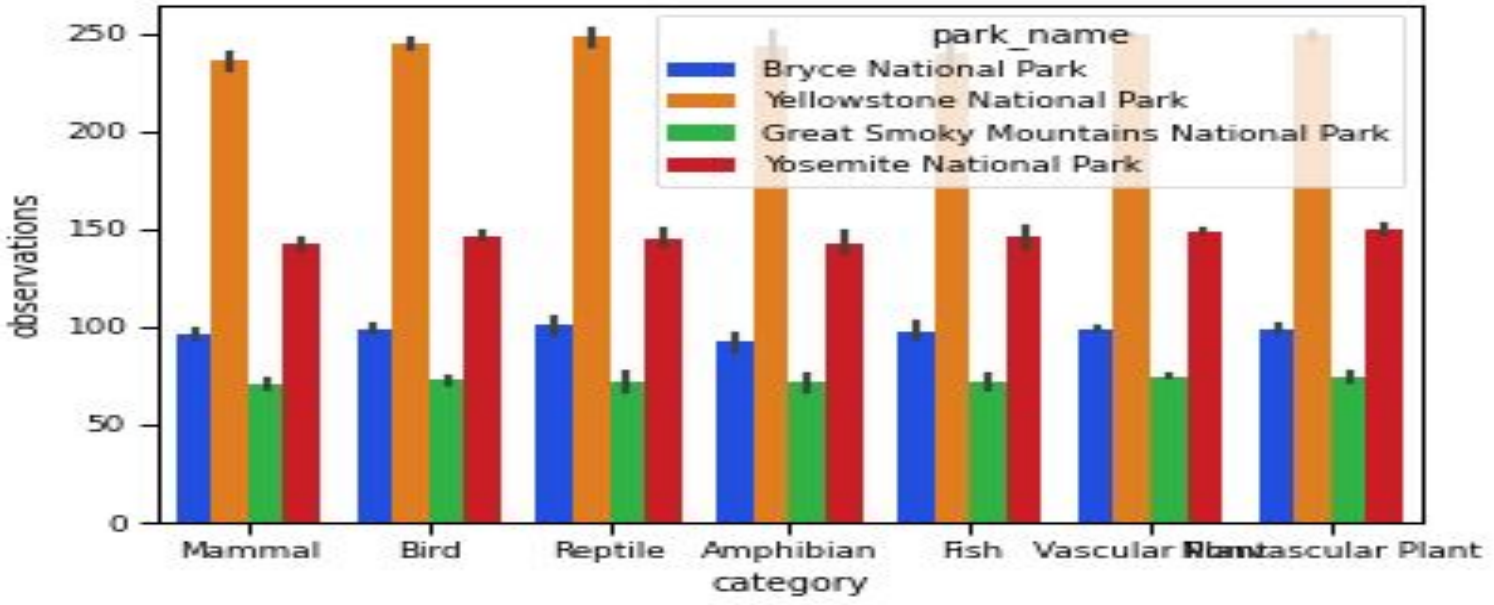
```
from scipy.stats import chi2_contingency
import math
# create contingency table
ab_contingency = pd.crosstab(df1_df2.category, df1_df2.conservation_status)

# run a Chi-Square test
chi2, pval, dof, expected = chi2_contingency(ab_contingency)
print(pval)
```

```
2.7594442674479634e-49
```

The p-value obtained is less than 0.5, so the differences between the species and their conservation status are significant.

The graph below is an answer to the last question



category	Amphibian	Bird	Fish	Mammal	Nonvascular Plant	Reptile	Vascular Plant
park_name							
Bryce National Park	128	130	105	152	135	87	133
Great Smoky Mountains National Park	85	115	87	97	108	110	95
Yellowstone National Park	238	266	253	261	268	245	260
Yosemite National Park	162	178	149	169	142	162	169

The table below and the graph above answer the last question.

Recommendation

The National Park Service should make considerable efforts to ensure that the species listed below are not further endangered.

1. Mammals that are most at risk
2. Vascular plants
3. Fish
4. Amphibians and birds

Thank you very much