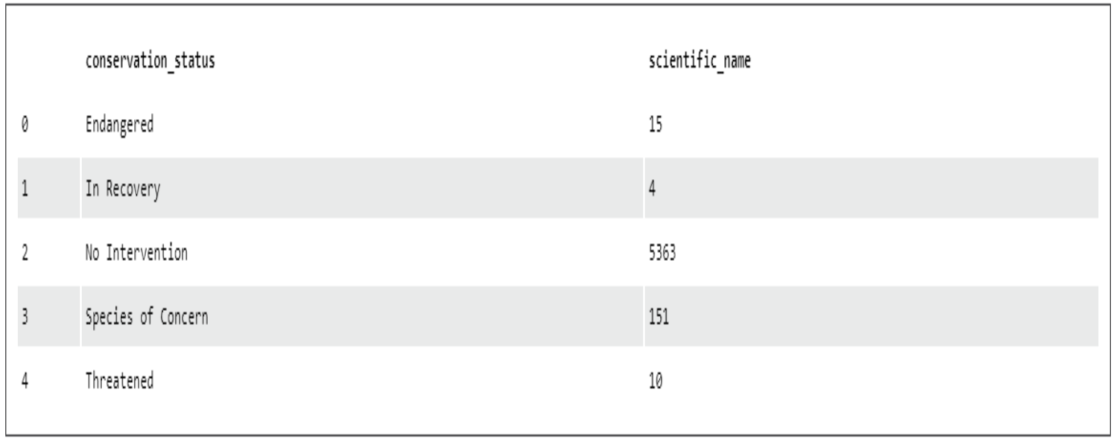
Yaqoub Al Qaoud

IDA Capstone Project

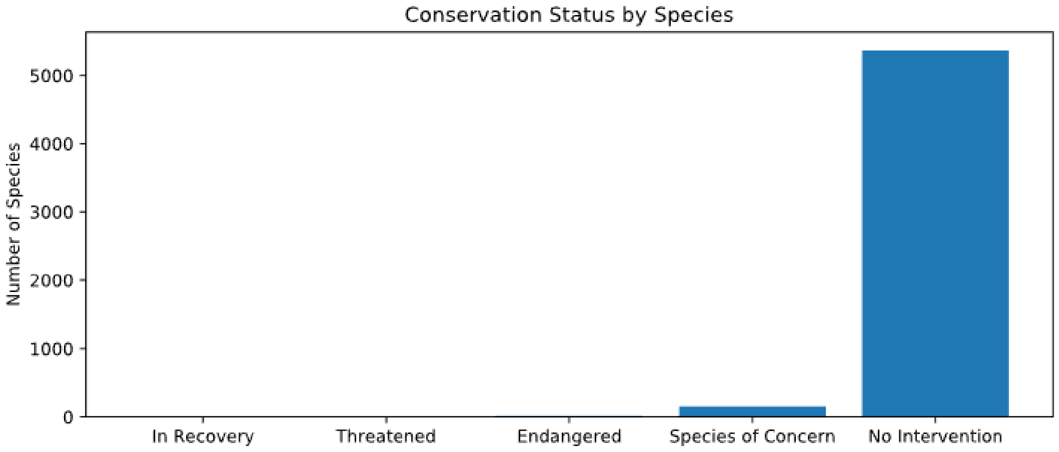
Biodiversity for the National Parks

Description of the Species Data

- After minor organizing of our species.csv data file, we came to realize that by grouping by conservation status, we had 15 unique scientific names of species under the “Endangered” status, 4 under “In Recovery”, 151 under “Species of Concern”, and 10 under “Threatened”.
- However, since we had some names under no status “Null Fields” we renamed all nulls to be “No Intervention”, and found that 5363 unique scientific names lies within that status.



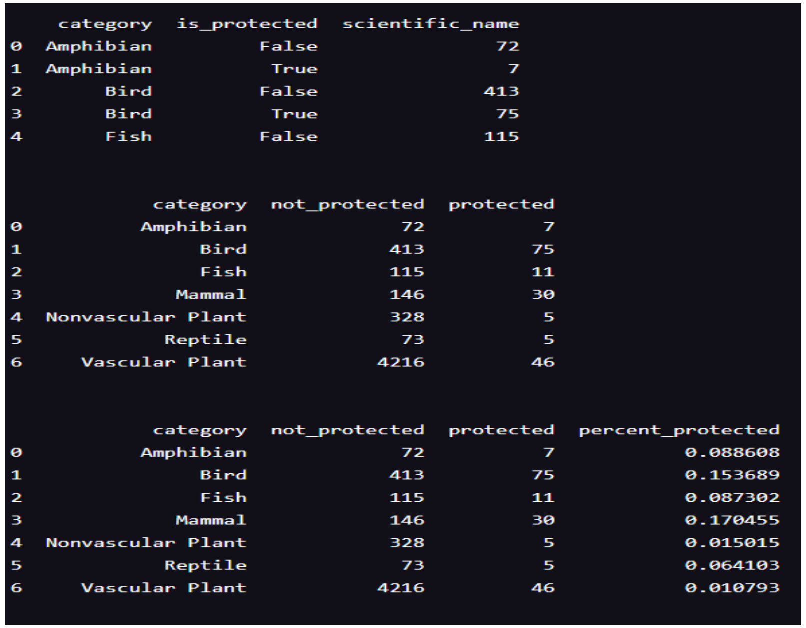
	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10



Conservation Status by Species

Description of the Species Data (Continued)

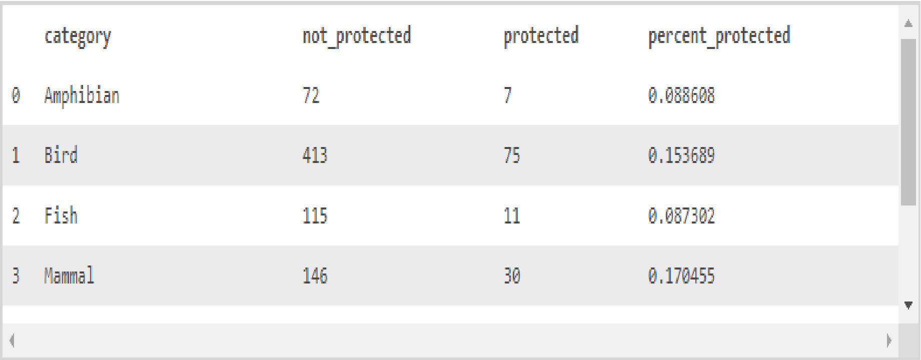
- Next, we added a new column to the data (“Is Protected”) to which has one of two values, True or False. True being under the “No Intervention” status, which basically mean that it is not endangered.
- After that, we grouped by both “Category” AND “Is Protected”, and counted how many unique scientific species names falls under each category and each value of True/False under the new column.
- Also, we created a pivot of the table with the “Is Protected” values as the columns, and renamed **True** to be “Protected”, and **False** to be “Not Protected”.
- Finally, we added a new column to our pivot table, “Percent Protected”, which shows the percentage of protected scientific names out of the total scientific names under each category.



```
    category  is_protected  scientific_name
0  Amphibian         False               72
1  Amphibian          True                7
2       Bird         False              413
3       Bird          True               75
4       Fish         False              115


            category  not_protected  protected
0          Amphibian             72          7
1               Bird            413         75
2               Fish            115         11
3             Mammal            146         30
4  Nonvascular Plant            328          5
5            Reptile             73          5
6     Vascular Plant           4216         46


            category  not_protected  protected  percent_protected
0          Amphibian             72          7           0.088608
1               Bird            413         75           0.153689
2               Fish            115         11           0.087302
3             Mammal            146         30           0.170455
4  Nonvascular Plant            328          5           0.015015
5            Reptile             73          5           0.064103
6     Vascular Plant           4216         46           0.010793
```



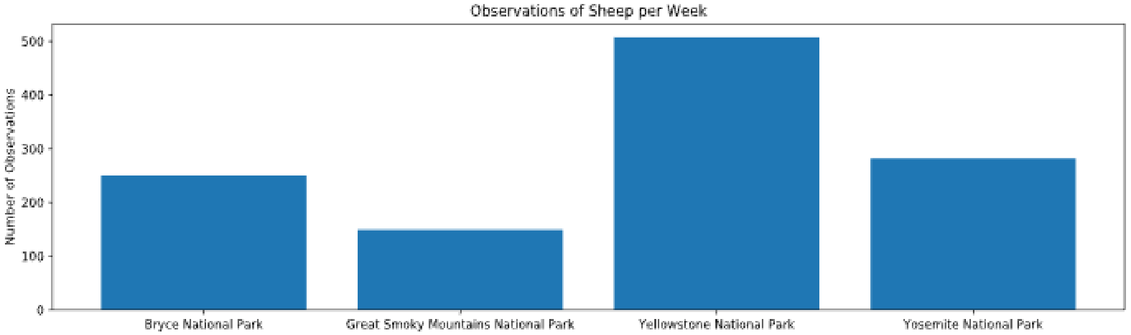
	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455

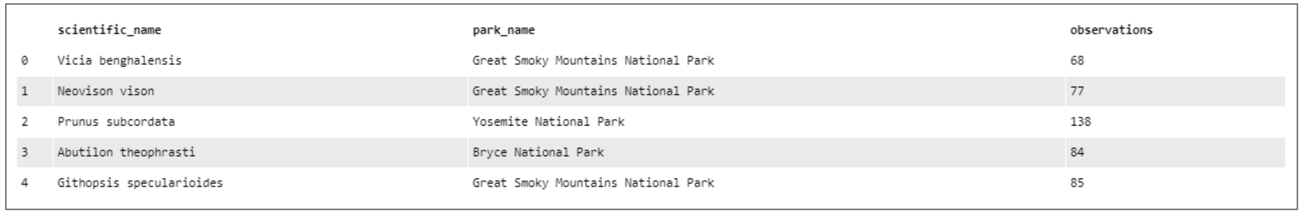
Significance Testing

- After that, we were asked to answer the following question: “are certain types of species more likely to be endangered?”
- Our data shows that 17% of Mammals are protected, while protected Birds shaped 15% of their category.
- Specifically, we wanted to check whether Mammals are more likely to be endangered than Birds based on that difference, or maybe it was just due to chance (null hypothesis).
- As we had categorical data, and we had more than 2 variables, the proper test to go for is the Chi-Squared.
- Our contingency testing with the Chi-Squared shows a **pval** of ~68.76%, which weakens our by 68.76% confidence, and have us accept our null hypothesis.
- We then did the same test but with Reptiles and Mammals, and got a **pval** of ~3.84%, and hence were strongly confident in accepting our significance test and rejecting the null hypothesis.

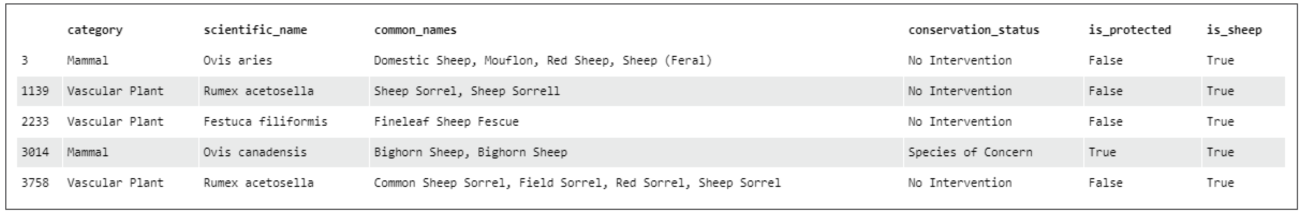
Description of the Observations Data

- The **observations.csv** data file contains the scientific names of different species in each park, and the number of observations of each specie in each park.
- However, we were interested in the sheep species movements across the different national parks.
- Since the **observations** table only had the scientific names, it was difficult to know for sure whether those names represent sheep or not. So, we merged both species.csv and observations.csv and selected the rows that had the word “sheep” in the common names of **species**, and then selected the “Mammal & Sheep” only.
- Finally, we grouped the merged table by “Park Name”, and summed the total observations of all Mammal Sheep in each park.

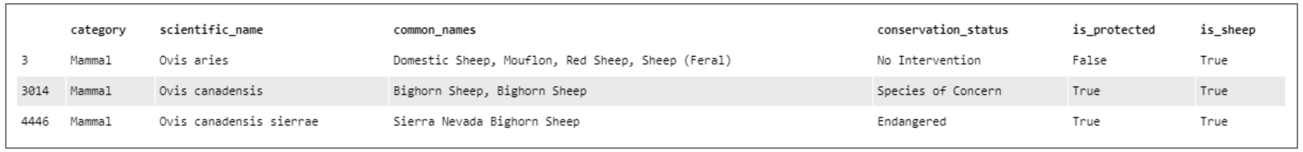




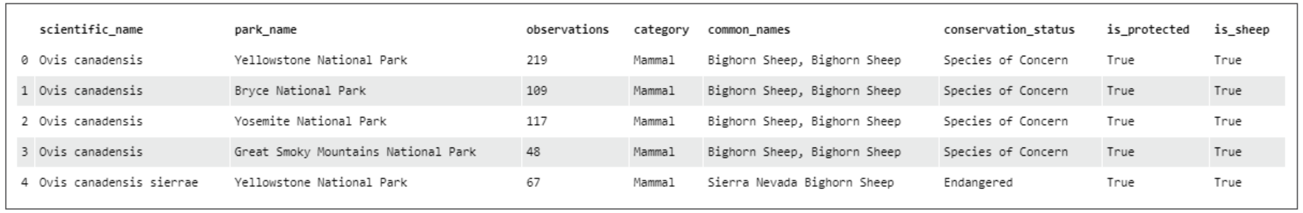
	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85



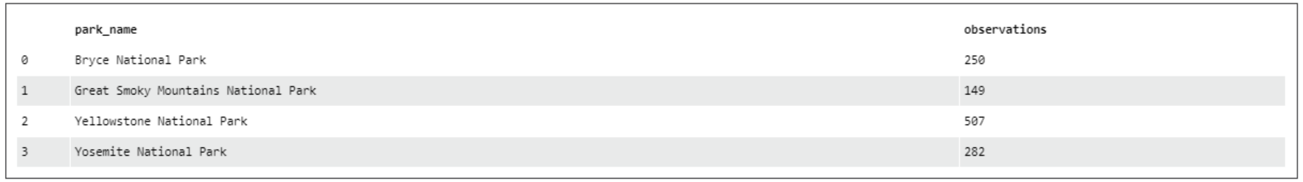
	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
1139	Vascular Plant	Rumex acetosella	Sheep Sorrel, Sheep Sorrell	No Intervention	False	True
2233	Vascular Plant	Festuca filiformis	Fineleaf Sheep Fescue	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3758	Vascular Plant	Rumex acetosella	Common Sheep Sorrel, Field Sorrel, Red Sorrel, Sheep Sorrel	No Intervention	False	True



	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True



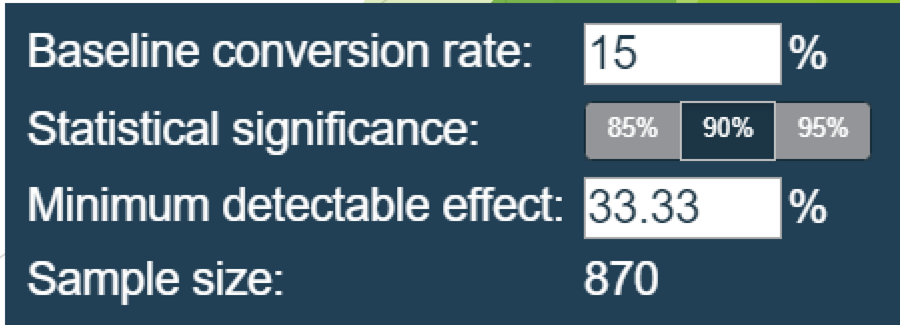
	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True



	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Sample Size Determination:
Foot & Mouth Disease

- Our lovely scientists want to test a program that has been implemented in Park Rangers at **Yellowstone National Park**, and they want to see if this program is working by detecting at least 5% reductions of the disease affections.
- The scientists recorded last year that 15% of the sheep at **Bryce National Park** have foot and mouth disease, which represents our **Baseline.**
- Given this, we were able to figure out our **Minimum Detectable Effect of 33.33% ➔ 5% / 15%.**
- After setting the level of confidence at 90%, our calculator showed that we need a sample size of 870 to be sure that a 5% (or more) decrease in observed cases of the disease was significant.



Baseline conversion rate: 15 %

Statistical significance: 85% [✓ 90%] 95%

Minimum detectable effect: 33.33 %

Sample size: 870