National Park Service Species Conservation Analysis

Download: Jupyter Notebook file, ReadMe

Abstract: 5,824 dataset that includes 4 features.

Data Set Characteristics:	Multivariate	**Number of Instances:**	5824	**Area:**	Ecology
Attribute Characteristics:	N/A	**Number of Attributes:**	4	**Missing Values:**	N/A

Source:
https://content.codecademy.com/programs/intro-data-analysis/biodiversity.zip

Data Set Information:
Species dataset represents species conservation including Scientific Name, Common Name, Conservation Status, and Species Category.

Data variables include:

Species Data Set

Label	Data Type	Description
'category'	String	Species scientific category
‘scientific_name’	String	Species scientific name
‘common_name’	String	Common name of species
‘conservation_status’	String	Status of species on conservation list

Data Summaries Calculated:

- Frequency

Units for Data Summaries and Variables:

- The species category of an animal
- The scientific name of an animal
- The common name of an animal

- The status of an animal

Data Transformations:

- Data was downloaded from https://content.codecademy.com/programs/intro-data-analysis/biodiversity.zip
- File “species.csv” was imported into Jupyter Notebook.
- Identify number of different species, categories, and conservation statuses.
- Re-code NA values to 'No Intervention' for ease of readability and analysis.
- Recode values into categories "protected" and "not protected".
- Calculated percentage "protected" category.
- Merged dataframes for Mammal species of Sheep observations across 7 days for each national park.
- Identified significance of likelihood for endangered conservation status between species.
- Identified required sample size for testing special intervention program for hoof and mouth disease.