Introduction to Data Science: Capstone Project Presentation

• • •

by Kirill Pervov

Contents:

Biodiversity for the National Parks: Dataset Description

The original dataset 'species_info.csv' has **4** columns:

- category (*e.g.* Mammal);
- scientific_name (*e.g.* Bos bison);
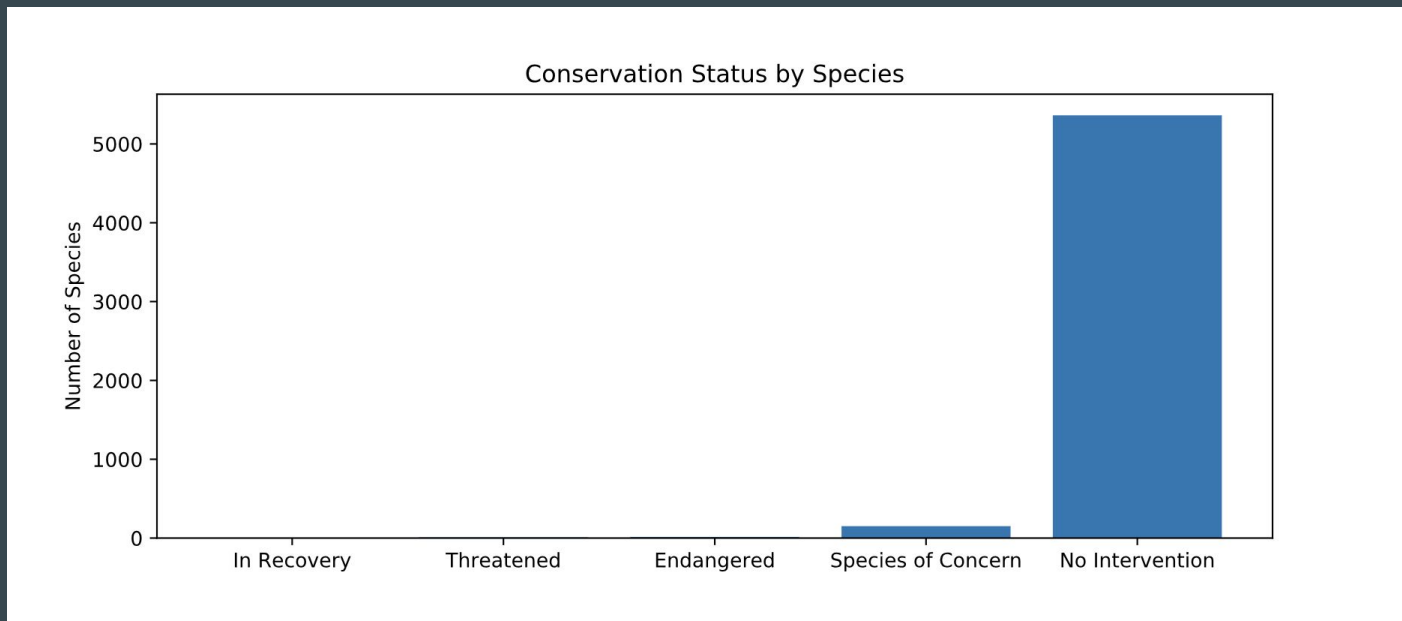- common_names (*e.g.* Bison);
- conservation_status (*e.g.* In Recovery).

It contains **5824** entries, **5541** of which are unique.

Biodiversity for the National Parks:
Dataset Description

Over 80% of the unique entries contain data on plant species.

Less than 4% of presented species are somehow endangered (see the figure 1)

Biodiversity for the National Parks: Dataset Description

Figure 1: Conservation Status by Species

Biodiversity for the National Parks:
Endangered status of different categories of species

In order to check if a species' category affects its conservation status the following steps were taken:

- the original dataset was transformed into the pivot-table, containing the number of protected and unprotected species in each category, as well as the percentage of protected species;
- the resulting data was included in a contingency table;
- a chi-squared test was performed for each pair categories, as well as for all animals vs all plants.

Biodiversity for the National Parks:
Endangered status of different categories of species

The following insights were discovered:

- the highest percentage of endangered species is among mammals and birds: **17%** and **15.3%** respectively;
- mammals are more likely to be endangered than reptiles (p-value < 0.05);
- animals are much more likely to be endangered than plants (**15.6%** vs **1.1%** of endangered species, p-value <0.05).

Biodiversity for the National Parks: Recommendations for conservationists

Taking into account all data provided, the following recommendations seems reasonable:

- animals generally require more conservation efforts than plants;
- though mammals are more likely to be endangered than reptiles, additional investigation is required to assess if certain categories of animals need more conservation efforts than others, and in which proportion.

Mouth and Foot Reduction Effort: Sample Size Calculation

The values used in the calculation were the following:

- baseline conversion: **15%** (last year data from Bryce National Park);

- minimum detectable effect: **33.3%** (5 / 15 * 100);

- confidence level: **90%**.

Result: **870**.

Mouth and Foot Reduction Effort: Sample Size Calculation

The values used in the calculation were the following:

- baseline conversion: **15%** (last year data from Bryce National Park);

- minimum detectable effect: **33.3%** (5 / 15 * 100);

- confidence level: **90%**.

Result: **870** (sheeps). The weekly number of sheeps observations in each national park is plotted in figure 2.

Mouth and Foot Reduction Effort: Sample Size Calculation
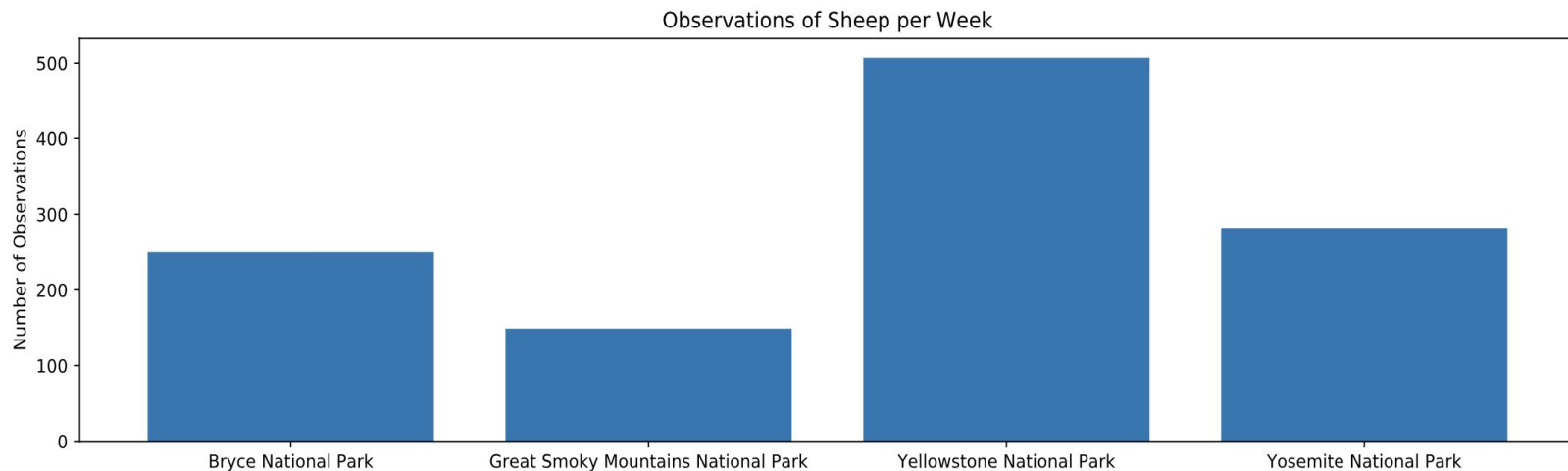
Figure 2: Observations of Sheep per Week (by Park)

Thank you for your attention and this awesome course!