SI507
Final Project Proposal (Apr 11th, Revision)
Xi Li

Due to that the web (Medium) I chosen for the first time required me to scrap data from infinite scrolling page, which was a little out of scope, as well as it showed different display mode in various browsers (infinite scrolling in Chrome, a limit of about 50 posts per page in Firefox), I decided to change the data source to WWF, a more static website: https://www.worldwildlife.org/ and deal with wildlife data.

Data Source: World Wild Life official website

The data source I use now are 100 pages with all WWF species details whose url information can be found in species catalog inside the WWF website. Inside the page there are contents including: general name, its scientific name, height, weight, length, population, habitat, place where you can find it, conservation status and some facts articles.

In this project, I will scrap the data from these species' details page and get a database of their information. The goal of this revised project is to understand those species current status including their basic features, population, their distribution and so on to get a better idea of how one can do to help protect those wildlife with help of data.

Method: Crawling & Scraping

I will start from the home page and find Species Catalog to get access to each species details page. 10 Fields include: name, status, population, scientific name, height, weight, length, habitats, place, and general habitat which is not as specific as "habitat".

Challenge Score: 8

Information Displayed:

(1) a series of bar charts of species population, based on their conservation status (like "vulnerable chart" or "endangered")

(2) a map showing where these 100 species are

(3) (optional*) a chart showing their height/weight comparison (need to deal with raw data) based on their conservation status

Presenting Tool: Graph on Plotly

The final presenting work will be several charts based on the species data shown on Plotly.