

SI 507 Final Project Proposal

Lea Wei zhuoqunw@umich.edu

Motivation:

The winter break is coming, my friend and I are planning for our trip. Given the COVID-19 pandemic, we try to avoid visiting big cities for the purpose of social distancing. National park is obvious a good option as it is far away from people. Also, both of us are huge fans of outdoor activities. Thus, to facilitate our trip, I want to design an interactive command line tool to obtain more information about national parks/sites in this project.

*This project is inspired by project 2 in SI507. But it has been updated and improved based on our needs and interests.

Data sources:

1. **National park information on nps.gov, obtained by National Park Service API.** The link to request an API key is <https://www.nps.gov/subjects/digital/nps-data-api.htm>. Since API key is required, the **challenge score is 4**.

Using the API, I can retrieve a lot of useful information of a national site, including state, national site name, url, operation hours, contacts, address, fees, activities, to name a few. However, since I also want to provide a map of the site to the user, yet it is not provided by the API, I will also scrape the website to obtain an url of the map. The website is <https://www.nps.gov/index.htm>. As I have scraped the website in project2, the **challenge score for this data collection task is 1**.

I have successfully retrieved national parks information at Michigan through the API (see the screenshot below). I will retrieve more data and cache it later.

```
1 {"total": "1", "limit": "50", "start": "0", "data": [{"id": "6DA17C86-088E-4B40-8862-7C1B05CF236B", "url": "https://www.nps.gov/acad/index.htm", "fullName": "Acadia National Park", "parkCode": "acad", "description": "Acadia National Park protects the natural beauty of the highest rocky headlands along the Atlantic coastline of the United States, an abundance of habitats, and a rich cultural heritage. At 3.5 million visits a year, it's one of the top 10 most-visited national parks in the United States. Visitors enjoy 27 miles of historic motor roads, 158 miles of hiking trails, and 45 miles of carriage roads.", "latitude": "44.409286", "longitude": "-68.247501", "latLong": "lat:44.409286, long:-68.247501", "activities": [{"id": "89DF0950-D319-4557-A57E-04CD2F63FF42", "name": "Arts and Culture"}, {"id": "FAED7F97-3474-4C21-AB42-FB8768A2F826", "name": "Cultural Demonstrations"}, {"id": "13A57703-BB1A-41A2-9488-53B692EB7238", "name": "Astronomy"}, {"id": "D37A0803-8317-4F04-8FB8-4CF0A272E195", "name": "Stargazing"}, {"id": "7CE6E935-F839-4FEC-A63E-052B1DEF3902", "name": "Biking"}, {"id": "7CE6E935-F839-4FEC-A63E-052B1DEF3902", "name": "Biking"}, {"id": "7CE6E935-F839-4FEC-A63E-052B1DEF3902", "name": "Biking"}]}
```

2. Biodiversity is another interesting topic related to national parks. I would also like to know what kinds of animals and plants we are likely to see at each national site, thus I will display this information in the project as well. I obtained the data about biodiversity from Kaggle.com: <https://www.kaggle.com/nationalparkservice/park-biodiversity>. It is provided by national park service, however, it only contains data for 50+ sites. Therefore, this information will be supplementary and will not be displayed when there is no data.

Two csv files are provided in this dataset. **The challenge score for this data task is 4.**

- Parks.csv: basic information about a park. Includes about 60 entries of data.
- Species.csv: detailed information about species in a park. includes 100k+ entries of data.

I have downloaded the data to my local machine.

Data processing:

1. Join two csv files on national site code.
2. Create dictionaries to collect national parks by state.
3. Create dictionaries to collect detailed information by national site. Information includes national site name, operation hours, contacts, address, fees, activities, map url, etc.
4. Create dictionaries to group species by category (e.g. bird, insect, mammal, plant etc), national park

Data presentation:

I am thinking of using a command line tool, however, this is subject to change. If after learning Flask, I find Flask is better at presenting the data, I may change to Flask.

1. The program will ask the user to input a state of interest and will return a nicely formatted list displaying all national sites names within that state with indices associated at the beginning of each name.

2. The user will then input a number, and the program will display more detailed information of the selected national site. It will also open the webpage containing the map of the site for the user.
3. If applicable, the user will also be presented with plots showing the biodiversity in the selected national park. More specifically:
 - A bar chart showing the number of species under each category, sorted by the number of species
 - A table showing the species' names and their category, commonality of sightings (common, uncommon, occasional etc.) and conservation status (species of concern, endangered etc.)
 - A pie chart showing the percentage of species under each nativeness category (native, not native, or unknown)
 - A bar chart showing the number of species under each conservation status (species of concern, endangered etc.), sorted by the number of species.