INFO 6210  
Gathering, Scraping, Munging and Cleaning Data

Assignment 1

At the beginning of the assignment, I chose “Biodiversity in National Parks” dataset after searching lots of datasets from Kaggle. Here is the hyperlink <https://www.kaggle.com/nationalparkservice/park-biodiversity/data>.

There are two tables in the dataset: **parks.csv** and **species.csv**. Before querying tables, I need to clean the data. I have done the parks.csv’ s cleaning in the **parks.ipynb** and species.csv in the **species.ipynb**. I got a new table “**species1.csv**” after cleaning species.csv.

Then, I downloaded mySQL through brew in terminal and started to create a database named “diversity” and insert tables into the database. This part was done in the **CreateUser&Read.ipynb**.

Finally, I came up with 5 use-cases:

1. Count The number of species in each state;

2. Count the number of endangered species in each state and list them in descending order;

3. Count the number of species of the maximum park;

4. Count the species of mammals in each state and list them in descending order；

5. Count the number of native species in each state and list them in descending order.

I queried my tables with 5 use-cases involving a join between two tables successfully. It was finished in **usecase. ipynb.**

All of files have “markedown” or comments to make a deeper explaination.