The Science of Reddit

With Libraries, APIs & Python

By Isaac Park and Jayam Shah

Google Slides: https://bit.ly/2GhnU41

```
def introduction():
    print(“Hello World”)
```

Course Schedule

1. Hi
2. Setup (if EWS works for students)
    a. Python, Jupyter, PRAW, Pandas installation with miniconda
    b. Cloning the sample repository with Git
3. More detailed notes and demo code on Github
    a. https://github.com/Parkkeo1/SciReddit-CS-Sail
4. Introduction to Python & Jupyter Notebook
5. Using the Reddit API with PRAW
    a. Reddit account & API key
    b. Pulling various data from subreddits, threads, accounts, etc
    c. Basic analysis & manipulation with Pandas
6. Stuff for if we have time left

Setup

- Log onto the EWS computer using the provided login.
- Open a browser window and go to:
  https://conda.io/miniconda.html
- Download the 64-bit bash installer for Python 3.6 for Linux.
  - Keep track of where you downloaded it to.
- Right-click on desktop screen and ‘Open Terminal’
- In terminal type:

```
cd <wherever the file downloaded to>
```

- For example: `cd Downloads/`

Setup

- Once you have navigated to the directory where the file is, type in terminal:

```
ls
bash Miniconda3-latest-Linux-x86_64.sh
```

- Follow along the prompts. This may take a while (that's fine).
- After the installation has finished, type in terminal:

```
conda install jupyter
conda install pandas
pip install praw
```

Setup

- After you've finished installing the libraries, it's time to open Jupyter Notebook and clone the starting code from github.

```
cd ..
cd Desktop/
git clone https://github.com/Parkkeo1/SciReddit-CS-Sail.git
cd SciReddit-CS-Sail/
jupyter notebook
```

- Once the browser window has opened, navigate to the Lessons folder. We'll tell you which .ipynb file to open.

Python & Jupyter Notebook

- Python is an interpreted language
  - Runs the code as it reads it
  - User-friendly syntax
  - Highly applicable to real-world programming
- **Jupyter**: web application for live-running Python code
  - Lets you run snippets at a time
  - Great for debugging and presentation
  - We'll be using Jupyter Notebook to live demo our code

The Power of Libraries in Python

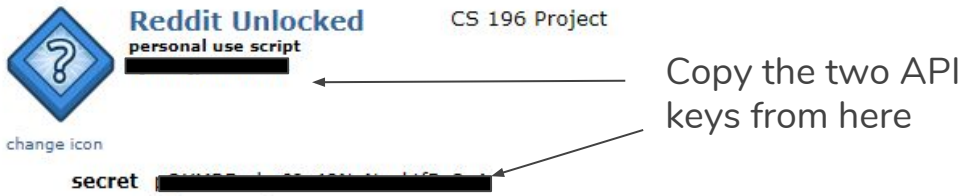
- **Libraries**: pre-written, reusable code designed for specific purposes.
- Some popular Python libraries:
  - NLTK - natural language processing
  - Scikit-learn - regression & basic machine learning
  - **Pandas** - data analysis
  - Flask - web applications
- We'll be using the "Python Reddit API Wrapper" (**PRAW**) library to interact with the Reddit API.
  - Why do we use it?

Onto actual code & programming

- Official documentation & tutorials, for reference:
  - **PRAW**: http://praw.readthedocs.io/en/latest/index.html
  - **Pandas:** http://pandas.pydata.org/pandas-docs/stable/
- Plan from now on
  - **Setup**: API access using your Reddit account
  - We'll be demo-ing code and going through PRAW, Pandas and their details using Jupyter Notebook
  - All our code will be available after this course at: https://github.com/Parkkeo1/SciReddit-CS-Sail

Reddit API - PRAW Non-code Setup

- Authentication
  - Login to a reddit account (make a dummy one if you don't have one). **Go to**: preferences -> apps -> create app
  - Select "script" & set "redirect uri" to http://localhost:8080
  - http://praw.readthedocs.io/en/latest/getting_started/authentication.html#script-application



- Using PRAW, in Jupyter Notebook

Reddit API - Data Collection

- We can provide the API with our authentication info and request data
  - Pass in a subreddit and some clarifying info
  - Get back a bunch of JSON data that describes each of the posts that match our criteria
- Organize this incoming data using pandas
- Filter out any unwanted data through pandas

Reddit API - Data Analysis

- Go through pandas data structure and aggregate (collect) the data we're interested in
  - This can involve a lot of different things
- Perform calculations (if applicable) on aggregate data
- Visualize or send off the processed data to wherever you need it

Reddit API - How reddit bots work

- The bots you see on Reddit all use the API (if they're written in Python, they're most likely also using PRAW)
  - Article summary bot: NLP on data pulled from the thread's news link
  - RemindMeBot: sets up reminders with PMs to users
  - CompileBot: parses posts with code and runs it in a online compiler
- Basic pseudocode for a passive bot (parses existing posts, not called)

```
for [post] in [subreddit]:
    post_data = *read post into data structure*
    if (post_data.contains(event_trigger)):
        *do my job*
```

Stuff for if we still have time

- Making your own reddit bot (if we can get EWS to work)
  - Look for the bot-starting-code.ipynb file in the starting repo
  - Read the basic instructions and experiment with the code
  - If posting, do it to: https://www.reddit.com/r/sailsp18/
- Looking at some real-life applications that use the Reddit API
  - https://snoopsnoo.com/ - reddit account analytics
  - http://www.redditinsight.com/ - subreddit tracking & analytics