Karen Bates
Uniqname: karbates
SI 507 Final Project Documentation

Project Code:

- https://github.com/karbates/final-project-covid19.git
- In the above repository is a README file with instructions on how to run the program, interact with it, and what API keys are needed. In addition to the below packages that are leveraged, the Twitter API and NewsAPI also require a user to have an API key, both of which are in my secrets.py file, which is hidden by the .gitignore file.
- This program uses Flask (render_template, request), requests, Beautiful soup, OAuth1, sqlite3, and plotly.

Data Sources:

COVID-19 Tracking API

- URL(s):
  - https://covidtracking.com/api/states/daily.json
  - https://covidtracking.com/api/states/info.json
- Format: JSON
- I set up a function to call the "states daily" extension of the COVID Tracking API to collect data on a certain state (when the state is input as a parameter). This information is then cached and formatted to display the most pertinent information (positive cases, hospitalized cases, recovered cases, and deaths) to the user in an HTML page using Flask. I created a function that generates a unique key for the API call that includes the date of the call so the cache will be used if a call for that state was already made in that same day. However, since the information is updated every day, a new call will be issued if there is not information for the most recent day to ensure the dashboard is kept up-to-date.
- The "states info" extension of the API returns more static information about a state, including its respective health department and affiliated websites, and the state health department's Twitter handle. I set up the program to call the extension and write it to a CSV which populations the database.
- The COVID Tracking team began collecting data from Washington State on 2/28/2020. There is one record for each state since the first day that state reported a confirmed case of COVID-19. As of 4/19/2020, calling the "states daily" filter and including all states, returns 2,508 dictionaries with information by date and state. When one state is searched, the number of dictionaries returned is around 50, depending on when the first cases were found in the state. The "states info" call returns one dictionary for each state, which is synthesized into 51+ rows (some territories and DC are included) that populate the database.

Kaiser Family Foundation

- URL for the at-risk population information: https://www.kff.org/other/state-indicator/adults-at-higher-risk-of-serious-illness-if-infected-with-coronavirus/
- Format: HTML
- I also downloaded CSV's from the following links since I was unable to scrape the JavaScript that contained the tables with the information (my original plan):

    - Obesity rates by state: https://www.kff.org/other/state-indicator/adult-overweightobesity-rate-by-gender/
    - Number of hospital beds by state: https://www.kff.org/other/state-indicator/total-hospital-beds/
    - Number of ICU beds by state: https://www.kff.org/other/state-indicator/icu-beds/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D
- I scraped the page at the first link for the percent of the adult population that is at an increased risk for severe illness or death should they contract COVID-19 due to their health status. After scraping, I organized the information in a dictionary and wrote the dictionary to a csv. The csv was then used to create a dataset in the database.
- There are only 102 pieces of data pulled from the tags (50 states plus one territory and 51 percentages) that are zipped into 51 rows. While it's small, I felt that it was important information that may help contextualize each states' situation so I wanted to include it.
- I also downloaded the information on obesity rates, hospital beds, and ICU beds for each state since I was not able to scrape the information. I wrote the CSV's into tables in the database.
- In total, from the scraped page and the three downloaded files, there are 4 sets of 51 rows of information, with 2-4 columns of information. While I recognize this is a small amount, I felt that it was important contextual information and I wanted to include it since it was my original intent to scrape this information.

Twitter

- URL: https://api.twitter.com/1.1/search/tweets.json
- Format: JSON
- I'm using the Twitter API to pull recent Tweets from the CDC and from the selected state's health department. If a user opts to see the Tweets, they are displayed alongside the COVID-19 case counts table. The Tweets are being cached and a request is being made through the cache if it is available.
- The maximum number of Tweets Twitter returns is 500 but it's also limited to Tweets from the last 30 days. I'm estimating that the number of records returned is between 100 and 500 for Tweets from the CDC. I'm using the count parameter to call and display 5 Tweets to not overwhelm the page.
- The information I'm displaying from the Twitter call's is just the Tweet itself.

NewsAPI

- URL: https://newsapi.org/v2/top-headlines
- Format: JSON
- I'm calling the NewsAPI for recent top headlines relating to COVID-19 and displaying them on a separate HTML page the user has the option to navigate to. Returns from the API vary in quantity.
- From the API call, I'm saving the article headline, the author, and the URL so that a user could navigate to the article if they wanted to.

Database:

- The database includes five tables, four of them from Kaiser Family Foundation information, and one from the COVID-19 Tracking API. The Kaiser Family Foundation related tables include information on the at-risk population in a state, the proportion of obese individuals in the state, hospital beds, and ICU beds. The fifth table includes each state's health department Twitter account, additional websites for information on COVID-19 in the state, and the state abbreviation.
- All of the tables can be joined on state names.

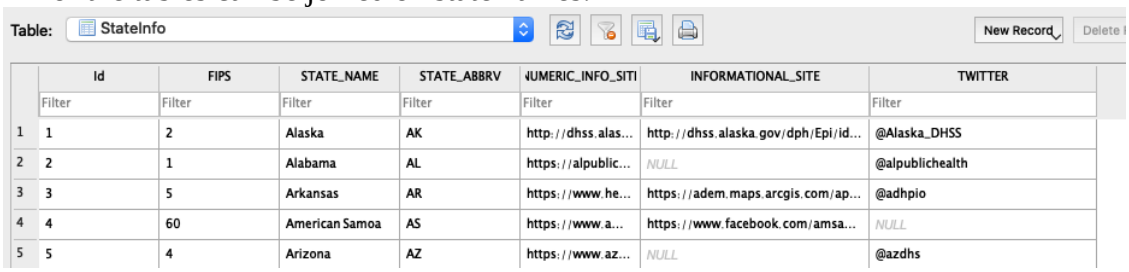

Table: StateInfo

	Id	FIPS	STATE_NAME	STATE_ABBRV	UMERIC_INFO_SITI	INFORMATIONAL_SITE	TWITTER
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	2	Alaska	AK	http://dhss.alas...	http://dhss.alaska.gov/dph/Epi/id...	@Alaska_DHSS
2	2	1	Alabama	AL	https://alpublic...	NULL	@alpublichealth
3	3	5	Arkansas	AR	https://www.he...	https://adem.maps.arcgis.com/ap...	@adhpio
4	4	60	American Samoa	AS	https://www.a...	https://www.facebook.com/amsa...	NULL
5	5	4	Arizona	AZ	https://www.az...	NULL	@azdhs

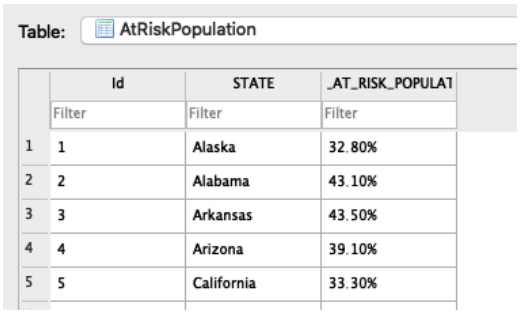

Table: AtRiskPopulation

	Id	STATE	_AT_RISK_POPULAT
	Filter	Filter	Filter
1	1	Alaska	32.80%
2	2	Alabama	43.10%
3	3	Arkansas	43.50%
4	4	Arizona	39.10%
5	5	California	33.30%

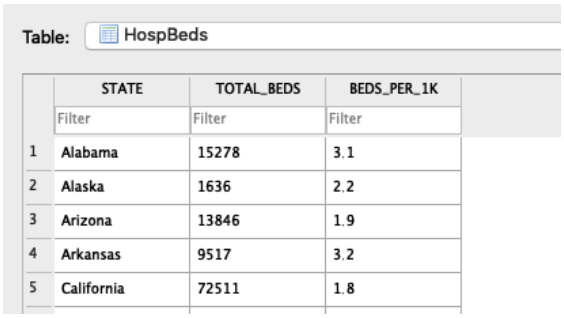

Table: HospBeds

	STATE	TOTAL_BEDS	BEDS_PER_1K
	Filter	Filter	Filter
1	Alabama	15278	3.1
2	Alaska	1636	2.2
3	Arizona	13846	1.9
4	Arkansas	9517	3.2
5	California	72511	1.8

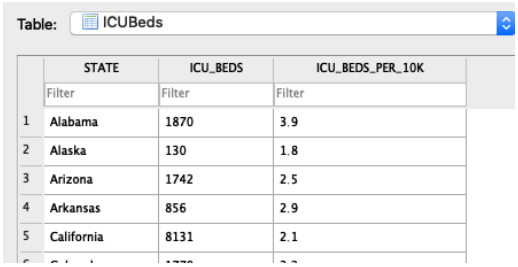

Table: ICUBeds

	STATE	ICU_BEDS	ICU_BEDS_PER_10K
	Filter	Filter	Filter
1	Alabama	1870	3.9
2	Alaska	130	1.8
3	Arizona	1742	2.5
4	Arkansas	856	2.9
5	California	8131	2.1

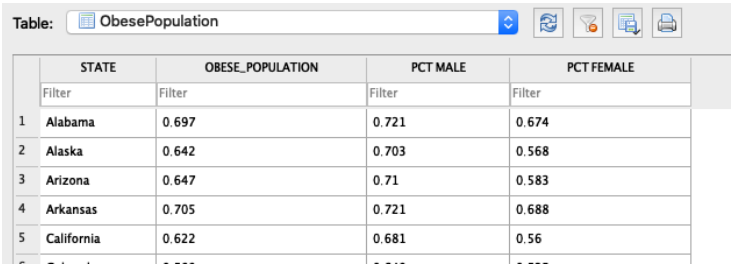

Table: ObesePopulation

	STATE	OBESE_POPULATION	PCT MALE	PCT FEMALE
	Filter	Filter	Filter	Filter
1	Alabama	0.697	0.721	0.674
2	Alaska	0.642	0.703	0.568
3	Arizona	0.647	0.71	0.583
4	Arkansas	0.705	0.721	0.688
5	California	0.622	0.681	0.56

Interaction:

- At the home page of the app, a user can select the state from a drop down menu that they would like to see more COVID-19 information on. They also have the option to select Tweet's from the CDC, Tweet's from the respective state's health department, and contextual health status and hospital bed information for the state. Upon submitting the form, the user is taken to a page displaying the information they selected. If they want, they can also navigate to a page with headlines relating to COVID-19 and plotted results for the number of confirmed cases, hospitalized cases, recovered cases, and deaths as a result of COVID-19.
- Loom link to view the final presentation: https://www.loom.com/share/17fe455b6b88411aa86da97017dca975