**Project code:**

Github repo: <https://github.com/qmkazo/507final>

Readme.md and requirements.txt are provided in github repo.

**Data sources**

Origin <https://www.usnews.com/best-colleges>

<https://www.mapquest.com/>

The raw data is HTML of the two websites.

The program can storage the raw data into json files.

In this program, Beautifulsoup is used to crawl data from websites. Once proper condition is provided, the program can automatically fetch data from usnews.com. Colleges data will be fetched, storage into database and processed into instances. The program will then use the instances to build figures and graphs.

Caching is used in this program. A file called “final\_cache.py” will storage the data fetched from websites. When the data are used for a second time, the program will use cache file rather than fetching again.

The data consists of 3 parts.

1.states

2.College Attribute includes rank, name, address, zipcode, in-state tuition, out-of-state tuition, url and enrollment.

3.locations near colleges

Attribute include name category address area

**Database**

**Interaction and Presentation Options**

The program can fetching data from

First, users type a state in command line prompt. Then a list of top colleges will be shown.

This program uses plotly to present data. File called “my\_plotly.py” is the graph module of this program. Main file will import this file and use the functions to make graphs and figures. Bar graphs and tables can be presented by this program.

**Demo video link**: https://youtu.be/zGKCZsPZsHg