Web crawling for Software Engineeing jobs

OverView:

The goal of this final project:

1. Get all recruiting Software Engineering jobs in America.
2. Show job locations on map.
3. Analyze overall software engineer demand for America and for each state.

Part1: Data Sources

1. Scrap all pages (100 pages, 2500+ Software Engingering jobs provided on this website.) related to Software Engineering and their detail information from [CareerBuilder-SoftwareEngineer](). Use cache when scarping data.

2. Request all company geometric coordinates via Google Place API.

3. Save all the information in file `jobs.sqlite` Save job details in TABLE Jobs; save geometric coordinate of companies in TABLE Company.

   - Columns in TABLE Jobs:
     Id(Primary key, assigned by DB), Title(Name of the job), Job type(full-time/part-time/Intern), Pay($), PayUnit(year/ hour), CompanyId(Foreign key - points to Company), PostDate(XX days ago), JobSnapshot(Key words for this job), JobDescription(Description for this job), JobRequirement(Requirement for this job).
   - Columns in TABLE Company:
     Id(Primary key, assigned by DB), Name(CompanyName), GeoLat(Latitude of this Company), GeoLon(Longitude of this Company)

- Data Source Challenge Score:

  1. Web ApI used before: Google Place API (2 points)
  2. Crawling and scraping multiple pages in a site that haven't used before. (8 points)

Part2: Data Presentation

1. Data presentation (with flask)
   A Flask App where user can choose specific state, computer skills and job type to see all jobs meeting this criteria. Results will be showed (with job titles hover above) on map via Plotly. (Over 50 possible combinations for plotting).
2. Overall Data Analyzation (with plotly):
   Show the average salary by state.
   Analyze the job demand by state.
   Show top 5 common Required Qualifications.
   Generate a WordCloud for SoftwareEngineer based on job descriptions (use wordcloud package).