

ReadMe :

Group Members:

Mukunda Aithal (maithal@andrew.cmu.edu)

Saumya Bansal (saumyab@andrew.cmu.edu)

Pranay Muppala (pmuppala@andrew.cmu.edu)

Yashika Goyal (ygoyal@andrew.cmu.edu)

Installation instructions:

After unzipping the D2_Team2 zip file please ensure you have the following files in the same directory

Main_program.py, city_info.py, job2.py, review_scraping.py(glassdoor scraping) final.csv(glassdoor reviews), Population_UnitedStates-Citywise_2019(Downloaded population data) and final_jobs2.csv(initial jobs list, has 32 jobs and additional jobs can be pulled from the program)

Execution:

The main program is called 'main_program.py' and is the only program you have to run and you can follow the steps in the Walkthrough to have a look at the features

Walkthrough:

Please refer this video link <https://youtu.be/N9ApePisJwc> for the program walkthrough

Notes:

- 1) The final.csv file has glassdoor reviews which are not dynamically scraped due to the nature of glassdoor webpage. Every company review webpage has a unique code that starts with E. like <https://www.glassdoor.com/Reviews/PwC-Reviews-E8450.htm>.
<https://www.glassdoor.com/Reviews/TikTok-Reviews-E2230881.htm>
This code is different for every company so dynamically pulling from the jobs we get was very challenging
- 2) The final_jobs2.csv file has the initial jobs pull that we did and we can pull additional data from the console when prompted but we filter it to software roles, so scraping 1000 posts doesn't guarantee 1000 jobs added to list