

TEAM KYQ

Yimeng Qiu, Yawei Li, Kelsey Anderson

PROJECT OVERVIEW

We created a web application to perform a cost-benefit analysis of attending graduate school for prospective students, given their current experience and intended field of work.

Although a large amount of content exists online about expected wages, it is largely geared towards helping job seekers negotiate a job offer or raise. Similarly, there is information about educational requirements for various jobs but the cost of that education is generally removed from such data.

The decision-making process prospective students engage in is a personal one. They weigh their chances of working up a job ladder against the boost that investment in education can provide.

We aim to address the knowledge gap between the estimated cost of higher education and its long term benefits for students by providing personalized information about projected wages with or without a degree for an individual.

SOFTWARE OUTLINE

The Invest in U software contains two data specific collection modules, a module to bring it all together as a feature table for predictive modeling, a prediction model module and a Flask application module for the user interface.

Modules

All contents are found under the **project** package on the **kjanderson-qym-yaweili** gitlab repository.

- **ui.py** - our main file for loading the app package and starting the Flask web framework for our user interface.
- **predict_salary.py** - the model for accepting data about the experience, education, and wages and producing a predicted optimal choice.
- **feature_table.py** - the functions used to retrieve previously collected data and bring it together for use in the predict_salary module.
- **bls_parser.py, onet_scraper.py** - scraper scripts. See 'Data Sources' below.
- Other unmentioned scripts worked as utilities for scripts above. Read comments inside if interested.

Data Sources

- **O*Net Online - Career pathways**. We used two data sources from the Occupational Information Network (O*NET) (a grant sponsored project of the U.S. Department of Labor/Employment and Training Administration).
 - We downloaded a list of 1100 occupations with titles, career clusters, and career pathway information as a csv file from a [search page](#).
 - We then developed a web scraper to add education level, required work experience from the details pages under the links from the search page referenced above.
- **US Bureau of Labor Statistics - Wage estimates**. We used the [BLS API](#) to scrape [modeled wage estimates](#) for each job by the state to supplement our career data.
- **US News - Graduate Programs**. We downloaded the name, location, and tuition info for top graduate programs from [US News rankings](#).

START DIRECTIONS

1. In terminal, run `bash install.sh` from project root folder
2. Then `http://127.0.0.1:5000` should appear in a browser window.

If it does not, please verify the flask server is running and refresh the page.

3. Make selections on the web form and press submit to see model predictions for wage and educational outcomes.
4. Press Test buttons to run data collection process demos.

Troubleshooting: If the install shell script fails to start the UI in your default browser, after installing all required packages from requirements.txt, execute `flask run`.

EVALUATION OF ACCOMPLISHMENTS

Aim

This project is aimed to help the decision of whether one should acquire more level of education by providing a cost-benefit analysis.

Actual

Given user background input (intended field, highest education, work experience, state to work in), it returns the following:

- The expected annual wage with the current background, calculated by our model.
- The expected hourly wage and annual wage with educational level at Master's level.

The users are then better informed to make their decisions.

Limitations

- The prediction model could be further optimized. Because the focus of this course is not machine learning, we put also our focus elsewhere.
- We had to abort the plan to provide graduate school recommendations due to failure to debug that.