Delaney Brown
CS3 Case Study: *Indeed.com* Job Posting Patterns and Predictions Over Time
GitHub Link: https://github.com/delaneybrown23/CS3-Indeed-Case-Study
DS 4002
April 28, 2025

Case Study Hook

You might be wondering right now what your life will look like after college. Searching for a job can be an overwhelming endeavor, and you might be unsure of what you want to do, where you want to be, or where you want to apply, especially given the plethora of online job application platforms. However, what if you could be one step ahead and 'see the future' of the job market for the next five years, allowing you to predict where and in which sectors there will be the most job growth and where you might consider applying? That's where this case study comes in.

Indeed.com is the most utilized and popular website for the job search globally, with the highest number of hires of all online platforms, consumer-driven features, and AI implementations that allow for a highly customized job feed seeking to produce matches [1]. With the sheer magnitude of job postings on *Indeed.com* and the relevancy of obtaining insights into the future job market, particularly for soon-to-be college graduates, this project's motivation drew from the FRED's "Job Postings on Indeed" datasets, which chronicles the change in percentage of *Indeed* job postings daily, or the 'index,' from February 1, 2020 to present, allowing for seasonal adjustment utilizing past data from 2017-2019 [2]. This case study specifically focuses on U.S. data. The FRED data draws from the *Indeed Hiring Lab* GitHub repository "Indeed Job Postings Index," where the primary datasets utilized in this case study are located, including "aggregate_job_postings_US.csv," "job_postings_by_sector_US.csv," "metro_job_postings_us.csv," and "state_job_postings_us.csv" [4]. These datasets contain the daily job posting data for the U.S. across all job sectors, and the daily postings categorized by job sectors, metro cities, and states, respectively [4,7,8,9,10]. The project seeks to utilize these datasets to develop predictions and forecasts regarding future *Indeed* job postings for the next five years, focusing on the aggregate job market, the top ten U.S. metro cities and job sectors with the highest number of posts using the Python SARIMAX forecast model, which specializes in seasonal time series data with both autoregressive and seasonal autoregressive elements [19,21].

These datasets have been downloaded and are located in the DATA folder of the case study repository. The scripts to reproduce the results of this case study are located in the SCRIPTS folder. Relevant articles and sources for further reading about SARIMAX and the "*Indeed* Job Postings Index," the rubric for this case study, and references for the original project are located in the MATERIALS folder [24,25]. Follow the steps in the README and rubric to reproduce the project's resulting outputs with the scripts. Hopefully, the conclusions drawn from this case study will provide helpful information for your job search. Thanks for reading!