DAMG 6210 Data Management and Database Design

Name: Darshana Rokade

NUID: 002705752

Notes:

To build a database with fields for location, pay, job kind, etc. for those looking for work in finance. Created a Twitter bot to scrape the web using languages like SQL and Python for data visualization and to extract data.

The Bot pulls up job listings and adds them to the database of open positions. Additionally, it gathers keyword data from Twitter and adds it to the database, making it simple for users to search for jobs using keyword searches. In the database, JSON files eliminate redundant data and save relevant information. Data that was scraped was stored in CSV files.

This database also displays a job seeker's eligibility and matches them with a suitable employer. The jobs database suggests opportunities based on users' interests and openings from various job websites, like Indeed.com, Glassdoor, etc. This database contains positions in the financial, educational, biotech, artificial intelligence, analyst, and other fields. This includes a CSV file with Glassdoor company ratings. The file has an employee title and a link for reviews. Firms including Prodigy Finance, Morgan Stanley, Wells Fargo, and the Bank of England are among them.

The Indeed table's Job ID and the jobs URLs’ Primary Key are both utilized. To lessen data redundancy and implement practical use cases, normalization is done in this database. There is a single source of data from which web scraping and web API are conducted.

There are two tables: Twitter and Indeed. The database that contains fields like location, pay, work role, etc. is created by these two tables. To maintain data completeness and retain data quality, normalization is done on tables.