



Alexandria University  
Faculty of computer Science and  
Data Science  
Department of Data Science  
Academic year 2024-2025

*14/4/2025*  
*Data Science Tools*  
*Dr.Mohammed Abd-*  
*Alhaveez*

## *Team members:*

23012172	محمد سعد ادم بابكر
23011666	محمد عثمان ابراهيم عثمان
23011965	منذر علي جمعة علي
23012281	محمد اسامة موسى احمد
23011998	مروان عبدالله ابراهيم

### Target Website:

[Glassdoor.com](https://www.glassdoor.com)

### Project Description:

This project aims to scrape job listings and company reviews from Glassdoor to analyze trends in the job market and employee sentiment across industries. The focus will be on extracting information related to job titles, company names, locations, salary estimates, job descriptions, and review ratings. The goal is to provide insights into the most in-demand roles, companies with the highest employee satisfaction, and how job offerings vary across locations.

### Types of Data to be Collected:

- Job titles
- Company names
- Locations
- Job descriptions
- Company ratings
- Review titles and review texts
- Date of posting/review

### Goal of the Project:

- Understand job demand trends and identify high-demand job roles by location and industry.
- Analyze employee sentiment using review ratings and review text analysis (e.g., most mentioned positive/negative keywords).
- Compare average salaries across companies and roles (if available).
- Visualize trends and findings in a clear, interactive format.

### Approach Overview:

#### 1. Data Extraction:

Use web scraping libraries like BeautifulSoup, Scrapy, or Selenium to gather job postings and reviews from Glassdoor.

## **2. Data Cleaning & Processing (with Regex):**

- Remove duplicate or incomplete entries.
- Use Regular Expressions to extract:
  - Locations (City, State)
  - Salary ranges (convert to numeric format)
  - Clean review texts (remove special characters, extract sentiment-related keywords)

## **3. Data Analysis:**

- Compute statistics such as average salary per role, average rating per company.
- Identify frequently mentioned pros and cons in reviews.
- Find top hiring companies and cities with the most job postings.

## **4. Data Visualization:**

- Bar charts for top companies by job postings and rating.
- Word clouds for commonly mentioned review terms.
- Heatmaps showing job distribution across cities.

## **5. Data Storage:**

- Store final cleaned dataset in MongoDB for efficient access and future querying.

## **Bonus – Streamlit App:**

- Create an interactive dashboard where users can filter results by company, job title, location, or rating.

**Our greatest glory is  
not in never failing,  
but in rising every  
time we fail.**

**- Confucius**