# **Web Scraping and Data Analysis Report**

## **Project Title: LinkedIn Job Scraper**

### **1. Introduction**

The objective of this project was to scrape job postings from LinkedIn based on specific search criteria such as job title, location, time posted, and remote work options. The extracted job data was stored in a structured format for analysis and visualization. This project aimed to provide insights into job market trends for Python Developer positions in Toronto.

### **2. Objectives**

* Develop a web scraper to extract job postings from LinkedIn.
* Store the scraped data in a structured format (CSV).
* Analyze job trends, demand, and competition using visualizations.
* Present meaningful insights from the collected data.

### **3. Methodology**

#### **3.1 Data Collection**

* The script used requests and BeautifulSoup libraries to retrieve job postings.
* User-defined search criteria such as job title, location, and posting time were used to construct the LinkedIn job search URL.
* The script fetched job listings by sending requests and parsed the response using BeautifulSoup.
* Pagination was implemented to scrape multiple pages of results.

#### **3.2 Data Processing**

* Extracted details including job title, company name, job URL, date posted, number of applicants, and location.
* Data was cleaned and structured into a Pandas DataFrame.
* Missing values were replaced with N/A.

#### **3.3 Data Storage**

* The structured job data was exported to a CSV file, named dynamically based on search parameters and date.

#### **3.4 Analysis & Visualization**

* Job trends were analyzed by counting occurrences of different job titles.
* A bar chart displayed the most frequent job titles.
* A line chart showed the distribution of job postings over time.
* A scatter plot visualized the relationship between the job posting date and the number of applicants.

### **4. Key Insights**

#### **4.1 Job Market Overview**

* **Total Jobs Scraped:** 412 job postings for Python Developer roles in Toronto.
* **Most Common Job Titles:**
  + Senior Full Stack Developer (12 postings)
  + Full Stack Engineer (11 postings)
  + Senior Backend Engineer (7 postings)
  + Senior Python Developer (4 postings)
  + Back End Developer (4 postings)
* **Most Competitive Positions:**
  + Some job postings had significantly higher applicant counts, indicating high competition for certain roles.

#### **4.2 Job Posting Trends**

* The number of job postings fluctuated over time, with noticeable peaks indicating periods of increased hiring demand.
* The analysis of job posting trends over months showed a steady increase in demand for Python Developers.

#### **4.3 Applicants’ Distribution**

* The number of applicants per job varied widely, with some postings attracting over 100 applicants, while others had very few.
* The scatter plot indicated that jobs posted more recently had fewer applicants compared to older listings.

### **5. Tools & Technologies Used**

* **Programming Language:** Python
* **Libraries:**
  + requests – Fetch job postings from LinkedIn.
  + BeautifulSoup – Parse and extract data from LinkedIn job listings.
  + Pandas – Structure and analyze the data.
  + Matplotlib – Generate visualizations.

### **6. Challenges & Solutions**

| **Challenge** | **Solution** |
| --- | --- |
| LinkedIn restrictions on web scraping | Used random user-agent headers and request retries to avoid detection. |
| Handling missing data | Used N/A placeholders for missing job details. |
| Pagination for job listings | Implemented a loop to scrape multiple pages of job results. |
| Variation in job posting structures | Used conditional checks to extract job details reliably. |

### **7. Conclusion**

This project successfully demonstrated the application of web scraping to analyze job market trends. The collected data provided valuable insights into the demand for Python Developers, competition levels, and hiring patterns. The methodology can be extended to other job roles or platforms for broader analysis.