# JOB SCRAPPY

Web Scrap Project

Frankie Law

2024

# Project Description

## Problem Statement:

The overall condition of the job market caused by the influx of refugees & immigrants, the great resignation, business closures and downsizing made job seeking process increasing problematic.

This in turn puts strenuous demand on HR and hiring managers to process copious amounts of applicants. With an abundance of Internet resources such as Indeed, LinkedIn, and Monsters. Job seekers have greater access to jobs than ever before, however, the cavate is the free accessibility available and convenience that attracts too many applicants makes job hunting difficult.

According to Glassdoor, on average, each corporate job posting attracts approximately 250 resumes, while only 4-6 will get called to interview, and one will get the job. If conventional job search method is flooded with job applicants. An alternative strategy is to seek out jobs that is not common from the eyes of the public.

This project is sets out to find hidden job postings online that is often overlooked, hidden beneath an ocean of saturated job posts.

## Possible Impact of Project:

By thoroughly searching the web for job opportunities, one will gain access to the overall condition of the job economy and to gain a broader scope of different employment availablities.

# Project Scoping

## Dataset(s):

This project uses web scrapping as its primary source of data and stores the parsed information into a database for future analysis. Each job board and company will require a unique set of web scrapping techniques to extract information.

Canadian Job Bank consist of 142,000 jobs as of this writing.

* Ontario (46,031)
* Northwest Territory (216)
* Quebec (15,865)
* New Brunswick (2,148)
* Manitoba (3,560)
* Alberta (34,923)
* British Columbia (34,923)
* British Columbia (34,923)
* Newfoundland (1,820)
* Nova Scotia (3,014)
* Nunavut (112)
* Prince Edward Island (389)
* Saskatchewan (4,064)
* Yukon (336)

## Methods:

Unique web scrap program for individual websites as each websites are constructed differently. The challenge here is to maintain a large set of programs and maintain its integrity as website improves and modify.

## Milestones:

* Create initial program to extract total number of job postings from each of the provinces of Canada.
* Extract all job postings related to each provinces. Completed all minor provinces with job posting less than 10,000 posts.
* Provinces that have more than 10,000 job postings are a big problematic as the website requires manual update for new job postings to show. Provinces that have large number of jobs will be required to separate by cities.
* Parse all HTML job posts and turn it into a text file or CSV format for analysis.
* Store all cities information into its dedicated database to streamline program execution and prepare the next phase of development which is to automate the whole process on a monthly basis.