Part A : Analysis of job market data to explore the open positions related to the jobs in the field of data or machine learning in a particular region.

To Collect job openings for the role of **Data Scientist in Canada**. I have done extensive research and selected the
Website:**https://www.simplyhired.ca/search?q=data+scientist&l=Canada&job=JjTVCNPaWP9D3wvhEw0EizLLDxCG68FHv5na0otV38rVZAkiJwUH3Q** ;
To take advantage of the options, I have selected the role of Software Engineer - Data Scientist . In a country like Canada, I believe there will be a lot of Software Engineer - Data Scientist roles available to collect and make a wise decision. With the higher samples we can make structured analysis using Tabulation and graph-Plotting.

   a) **Data collection method** (include the sample size & process of data collection) [Marks 3]

I have utilised **Webscarping and python** to extract the **first 20 job openings** and to represent them in a tabulated format.
To Scrape the data out of **'Simplyhired'** I imported Python libraries like Requests, Pandas, BeautifulSoup, Matplotlib and Seaborn.
Request : It is used to request data from Simplyhired
Pandas : It has functions for analysing, cleaning, exploring, and manipulating data that was scraped
BeautifulSoup : It is used for web scraping purposes to pull the data out of HTML
Matplotlib : Is a comprehensive library for creating static, animated, and interactive visualisations in Python
Seaborn : It provides a high-level interface for drawing attractive and informative statistical graphics

First, the **Response Code** was verified as **'200'** to make sure there were no issues interacting with the server. Following that I used the function **BeautifulSoup** to extract the data from the website and stored it in a variable. Then, I used the content to inspect the elements and attributes to check what kind of data can be scraped. Subsequently, I used the inspect option on the website after carefully analysing the structure of the website. After I found the tags, elements and attributes of the HTML content, I scraped the first page of the website, which had **20 job listings**.
To create a structured analysis, **Job Title, Location and Company Name** were considered. The Job Title was easy to scrape as it was placed well in the HTML code. But I encountered trouble when fetching the location of the company. The HTML had placed it twice, enclosed between 'span' tags. Due to which the output returned duplicate items. Resolution to this was not easy as I could not use functions such as **set()** to kick the duplicates nor could I use a for loop. Reason can be explained with an example; the output had [toronto, toronto, ottawa, ottawa, toronto, toronto] kicking duplicates would mean that the output would look like this [toronto, ottawa] which will be incorrect as there might be many jobs focused in toronto. After a lot of thinking, I came up with a rudimentary, python-101 idea to kick even/odd items from the list using their indexes.
After representing the data in a **Dataframe** format using **Pandas** I exported it and then imported it back to use **Data Wrangling** technique. I performed some Data cleaning to remove some texts from the Job titles so that they could be grouped well when I come back for them while doing pictographic analysis.
I took the advantage to split the location to City and State to get a better analysis. Further, I dropped the Location column and added **City** and **State** as new columns.
The data set was ready for some comprehensive, animated, and interactive visualisations using **MatPlot**. I used **countplot** and **histoplot** to represent and achieve better understanding.

b) **Market Data Visualization [Marks 10]**

The country of **Canada** is widely popular for living comfort and job availability. For years now Canada is a hotspot for immigrants to taste success in various fields. One particular field that is now booming is the field of **Information Technology**.

Information Technology has a lot of roles, one particular field of study is **Data Science**. Data science is the study of data to extract meaningful insights for business. It is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyse large amounts of data. This analysis helps data scientists to ask and answer questions like what happened, why it happened, what will happen, and what can be done with the results. People who have procured this technological prowess are called Data Scientists.

Let us dig deeper and use some of these multidisciplinary tools/techniques to attain some understanding of the Data Scientist Job available in Canada.

Here is a Tabular representation of 20 Data Scientist Jobs available in various parts of canada.

Job Title	Company Name	City	State
Data Scientist	Bank of Canada	Ottawa	ON
Data Scientist	TD Bank	Toronto	ON
Data Scientist	Longo's	Vaughan	ON
GIS and Data Analyst	Town of Ajax	Lake	ON
REPORTING & ANALYTICS SPECIALIST	Manitoba Hydro	Winnipeg	MB
Data Scientist	Intact	Calgary	AB
RES O 27R - Data Scientist	BC Public Service	Abbotsford	BC
Data Scientist	Faire	Toronto	ON
Data Scientist	Canada Post - Postes Canada	Ottawa	ON
Research Data Scientist	Canadian Light Source Inc.	Saskatoon	SK
Marketing Data Scientist	Scene+	Toronto	ON
Data Scientist - Commercial Analytics	Veeva Systems	Toronto	ON
Data Scientist	Takeda Pharmaceutical	Toronto	ON
Data Analyst	Thames Valley District School Board	London	ON
Data Scientist	Cisco Systems	Toronto	ON
Data Scientist	IBM	Toronto	ON
Data Scientist	Power Grid Innovations Inc.	North Vancouver	BC
Data Scientist	Morgan Stanley	Montréal	QC
Data Scientist	Home Trust Company	Toronto	ON
data analytics consultant	Mindshare Canada ULC	Toronto	ON

The table represents all the Data Science or some roles like data science.

From the **Listed companies** it is clear that the job role is highly versatile. One can make an observation that the **Banking** sector, **IT**, **Finance** sector and **Government** jobs all have a need of data scientists. Which is great as many Computer science engineers are allocated to only IT sectors. This will be highly beneficial to people who like to take on different aspects of work ensuring better job satisfaction.

The data clearly suggests that the State of Ontario has the highest need of data scientists. It is conclusive from the data. The supporting reason would be that the state of Ontario is relatively warmer as compared to other parts of the county and has the highest population.

Here is a Pie chart representing the concentration of jobs against Cities.

Count of State
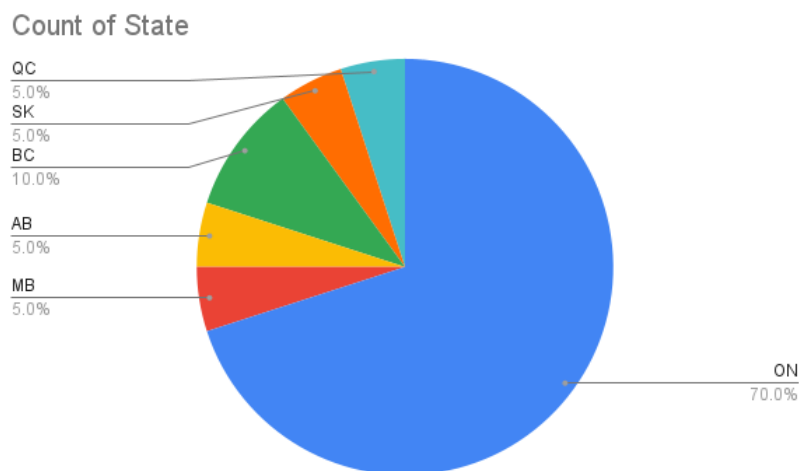
The data shown in the Pie chart is clear that 70% of the job offers come from Ontario.

Exploring Further, In the state of Ontario, Toronto has the highest openings. The city is well connected and is an IT hub, it is the house of 80% of all the Tech giants in the country. The data represented below has some Insights supporting the query.

Job Title Distribution by City
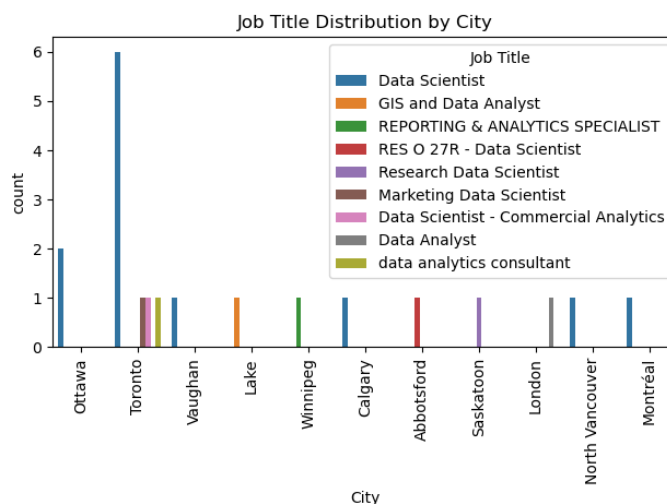
Two cities, Ottawa, the capital and Toronto have the most openings, which is what the data show.

The data must hold true for other job roles and should look more so the same. This analysis was centred keeping demographics in mind and is focused on location.
By this research, It is conclusive that Toronto is the IT capital of Canada and there is more potential for more hires in this field. This suggests the city is far from saturation, also, as Ottawa is upcoming there are many possibilities for one to secure a role and explore options.

3

b) **Your ideal job** (job title and relevant skills required) [Marks 2]

I have always been fascinated by **Artificial Intelligence Engineer - Machine Learning/Deep Learning.** I believe that we live in a world that is drifting toward making our lives more comfortable. There is no limit to what comfort can mean when it comes to Machine Learning.

Just recently, I learnt about how Artificial Intelligence and Machine Learning has vastly benefited the experience of vision for the visually impaired. That is soulful and most grateful aspect of working for a cause. Furthermore, there are limitless possibilities for AL/ML in this limitless world. I truly hope and wish, I can learn the knowledge and make an effort to better the world.

Some of the skills the Job demands are Good understanding of Deep Learning tools like **Tensorflow /Keras/ Pytorch Libraries**- Good understanding and experience in traditional machine learning including **SVM**, **logistic regression** etc - Experience in implementing Machine Learning Models in different programming languages - **Python, C/C++** - Experience in writing readable and good code in Python/C/C++- Hands-on experience with open source tools such as Kaldi, **Pytorch-Kaldi** and any of the end-to-end ASR tools such as **ESPNET** or **EESEN** or **DeepSpeech Pytorch** is desirable.- Good understanding of open-source text-to-speech conversion tools such as **Tacotron-2, DeepVoice-3** or similar.- Hands-on experience with open source tools such as PyAudioAnalysis or voice box or similar is needed to observe the speech applications such as speaker diarization, silence removal, etc.

Sub Domain :- **Neural Networks**- Machine Learning- **NLP** (Natural Language Processing)- Robotics- Speech Processing

This Job comes with a lot of learning, research (not to mention the good pay) and is set to push candidates to their limits. I intend to take on this challenge and excel in it.

Date: 12th August, 2023
Name: Viren Sasalu