Course 1: Foundation of Information

Project (PART A) Report: Job Market Analysis

Introduction:

This report presents the findings obtained from the analysis of information acquired from the job listings part of well-known job portal https://www.glassdoor.com for the position of "**Machine Learning Engineer**" in **India**.

a) Data Collection Method:

Following data extraction from Glassdoor job ads, the data was cleansed to ensure consistency, resulting in a sample size of 750. For the data collection process, implemented web scraping techniques using Selenium and Beautiful Soup and extracted necessary information like company name, position title, location, salary, position type, and skills required. FIG-1 contains important code lines collected from actual code to give an overview of data collection process.

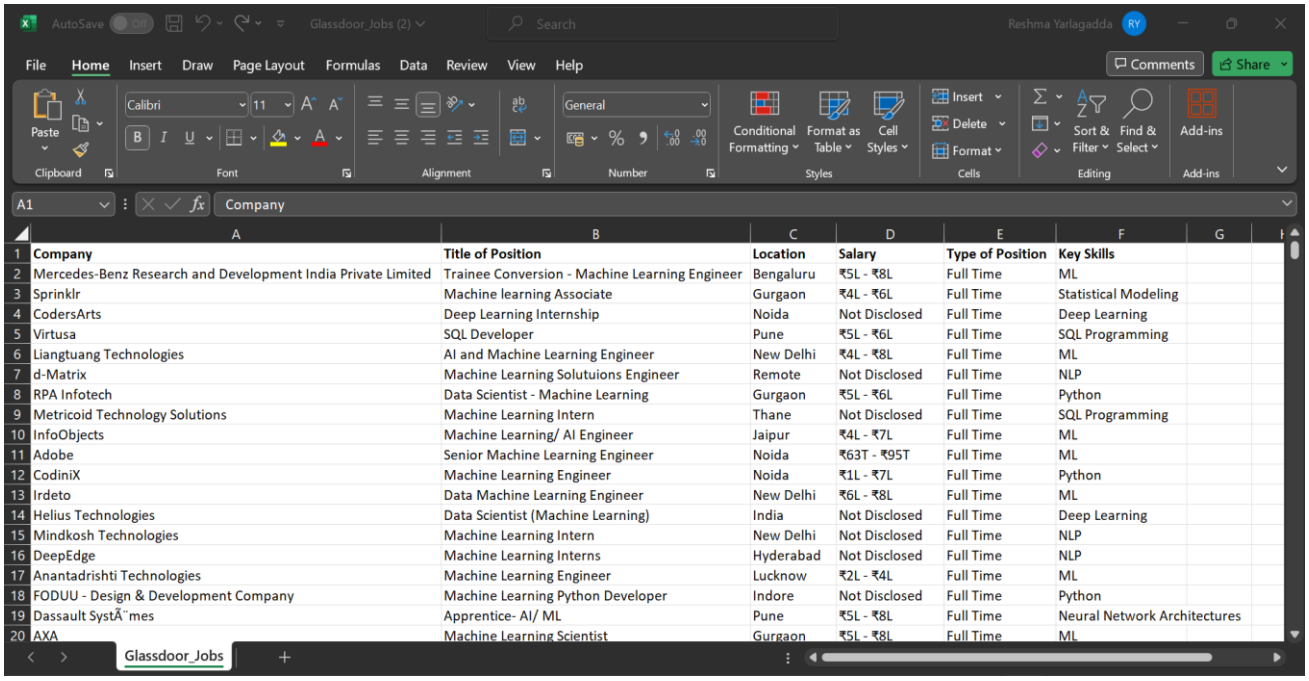
```
[ ]  # Installation of selenium and webdriver_manager
     !pip install selenium
     !pip install webdriver_manager

[ ]  # Main imports (many other imports were taken to extract data)
     from selenium import webdriver
     from bs4 import BeautifulSoup

     # url of glassdoor with search queries Machine learning engineer as position and India as location
     job_website_url = "https://www.glassdoor.co.in/Job/india-machine-learning-engineer-jobs-SRCH_IL.0,5_IN115_KO6,31.htm"
     # Give necessary chrome options
     chrome_options = webdriver.ChromeOptions()
     chrome_options.add_argument('--headless') #many chrome options were added in original code. Just adding one here to give an idea
     # Set up the WebDriver with necessary options
     driver = webdriver.Chrome(options=chrome_options)
     driver.get(job_website_url)
     # Extract HTML source code
     html_source = driver.page_source
     # Create a beautifulSoup instance
     soup = BeautifulSoup(html_source, 'html.parser')
     # Find all job titles
     company_names = soup.find_all('span', class_='EmployerProfile_employerName__Xemli')
```

FIG-1: Code overview

Created a data frame with column names as Company, Title of Position, Location, Salary, Type of Position and Key Skills of extracted information from each page and combined all the data frames into a single data frame and stored it in a CSV file. There are 750 data points that are used for visualization once the data has been cleaned up to assure accuracy and consistency. FIG-2 displays the CSV file containing extracted data opened in Excel.



Company	Title of Position	Location	Salary	Type of Position	Key Skills
Mercedes-Benz Research and Development India Private Limited	Trainee Conversion - Machine Learning Engineer	Bengaluru	₹5L - ₹8L	Full Time	ML
Sprinklr	Machine learning Associate	Gurgaon	₹4L - ₹6L	Full Time	Statistical Modeling
CodersArts	Deep Learning Internship	Noida	Not Disclosed	Full Time	Deep Learning
Virtusa	SQL Developer	Pune	₹5L - ₹6L	Full Time	SQL Programming
Liangtuang Technologies	AI and Machine Learning Engineer	New Delhi	₹4L - ₹8L	Full Time	ML
d-Matrix	Machine Learning Solutuions Engineer	Remote	Not Disclosed	Full Time	NLP
RPA Infotech	Data Scientist - Machine Learning	Gurgaon	₹5L - ₹6L	Full Time	Python
Metricoid Technology Solutions	Machine Learning Intern	Thane	Not Disclosed	Full Time	SQL Programming
InfoObjects	Machine Learning/ AI Engineer	Jaipur	₹4L - ₹7L	Full Time	ML
Adobe	Senior Machine Learning Engineer	Noida	₹63T - ₹95T	Full Time	ML
CodiniX	Machine Learning Engineer	Noida	₹1L - ₹7L	Full Time	Python
Irdeto	Data Machine Learning Engineer	New Delhi	₹6L - ₹8L	Full Time	ML
Helius Technologies	Data Scientist (Machine Learning)	India	Not Disclosed	Full Time	Deep Learning
Mindkosh Technologies	Machine Learning Intern	New Delhi	Not Disclosed	Full Time	NLP
DeepEdge	Machine Learning Interns	Hyderabad	Not Disclosed	Full Time	NLP
Anantadrishti Technologies	Machine Learning Engineer	Lucknow	₹2L - ₹4L	Full Time	ML
FODUU - Design & Development Company	Machine Learning Python Developer	Indore	Not Disclosed	Full Time	Python
Dassault SystÃ¨mes	Apprentice- AI/ ML	Pune	₹5L - ₹8L	Full Time	Neural Network Architectures
AXA	Machine Learning Scientist	Gurgaon	₹5L - ₹8L	Full Time	ML

FIG-2: Extracted data

b) Market Data Visualization:

Visualization is a process of representing complex data sets visually through **charts** and **graphs**, ultimately enhancing understandability of **trends** and **patterns**. As we are dealing with large data set, visualization comes in handy to draw necessary **insights**.

Table below contains the summary of extracted information

Title	Location	No. of Positions	Avg. Salary	Type of Position	Top Skills Required
Machine Learning Engineer	India	750	6 Lakhs per Annum	Mostly Full Time	ML, Python, NLP

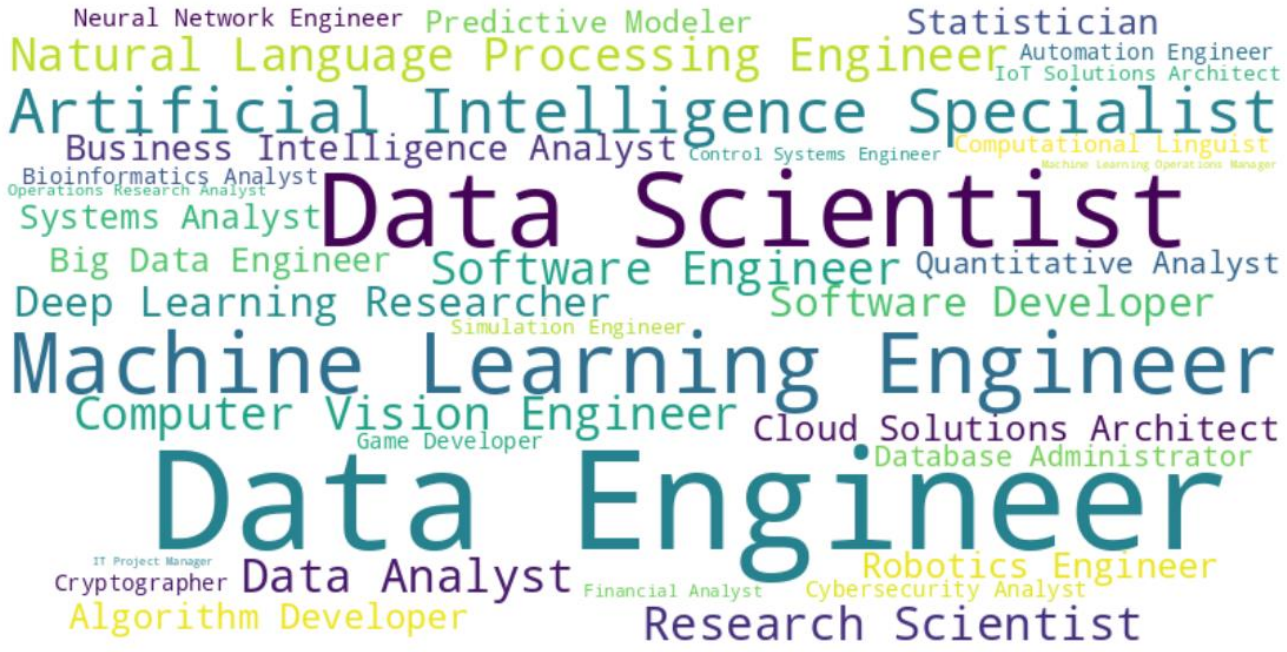
Visualizations are created for the data present in the CSV file of extracted information using a combination of bar charts, pie charts and histogram to make them visually appealing and effective in facilitating data interpretation and decision-making.

Analysis of titles of positions available:

Using a word cloud turned out to be an engaging and dynamic way to visually represent the most commonly occurring job titles by representing them as bigger and bolder terms.

It is evident from the word cloud below that '**Machine Learning Engineer**', '**Data Scientist**', '**Data Engineer**' are most **in-demand** positions as they appear bigger and bolder.

Various positions, including "**Systems Analyst**," "**Research Scientist**," "**Data Analyst**," and others, imply that the field is **multidisciplinary**.



Analysis of locations where the roles are offered:

Analysis of geographic distribution of job opportunities using a count plot proved to be an effective way by delivering insightful information about the places where different roles are available. This analysis helps job seekers in understanding the regional dynamics of the job market today.

The higher bars could be a reflection of **Gurgaon's** and **Delhi's** thriving **Machine Learning** and **technology communities**. There is probably a concentration of tech companies, startups, and businesses that need Machine Learning talent in these areas.

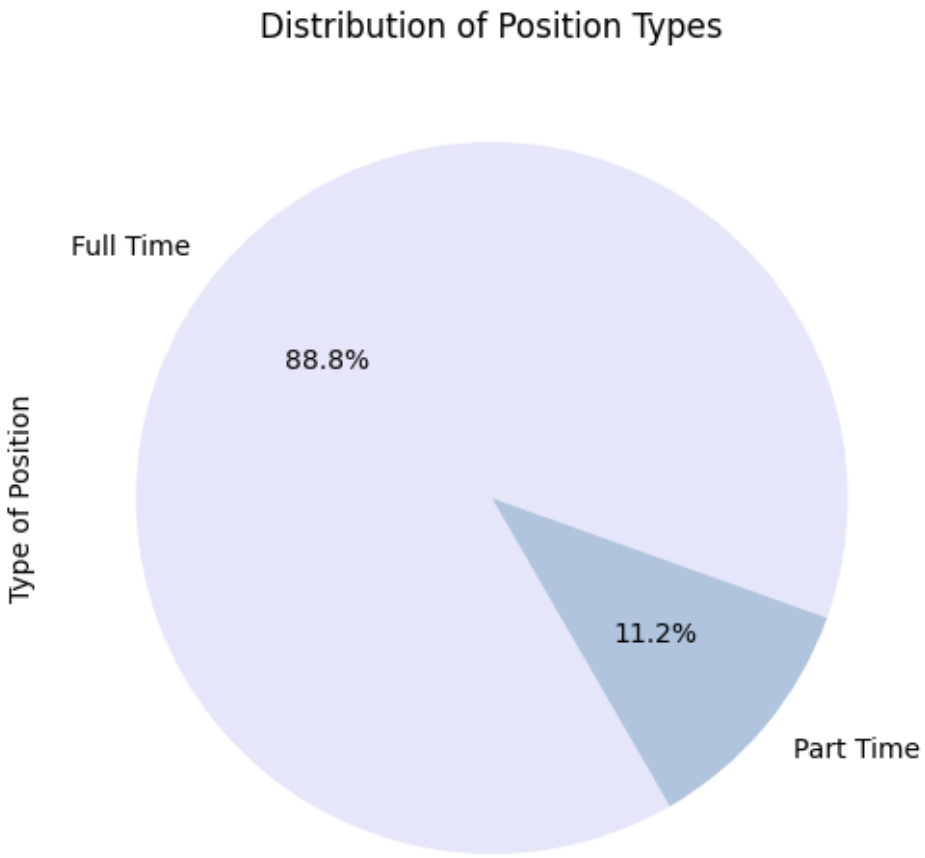
The lower bar in **Lucknow** indicates that there may be **room for growth** in the **Machine Learning** job market. It might also suggest a slower employment market or a concentration on particular industries.

The availability of **remote** positions in **Machine Learning** widens the talent pool beyond geographical boundaries. Companies can choose from a wide range of qualified data scientists around the nation, encouraging a more inclusive hiring process.

Employers and **data science talent** find cities with intermediate bar heights, like **Pune**, **Hyderabad** to be **appealing**. Employers can access competent local professionals, while job seekers may find a balance between **lower competition** and a **growing environment.**

Analysis of type of positions:

A pie chart is employed in this analysis with an aim to provide a clear, colourful and engaging representation of the distribution of type of positions.

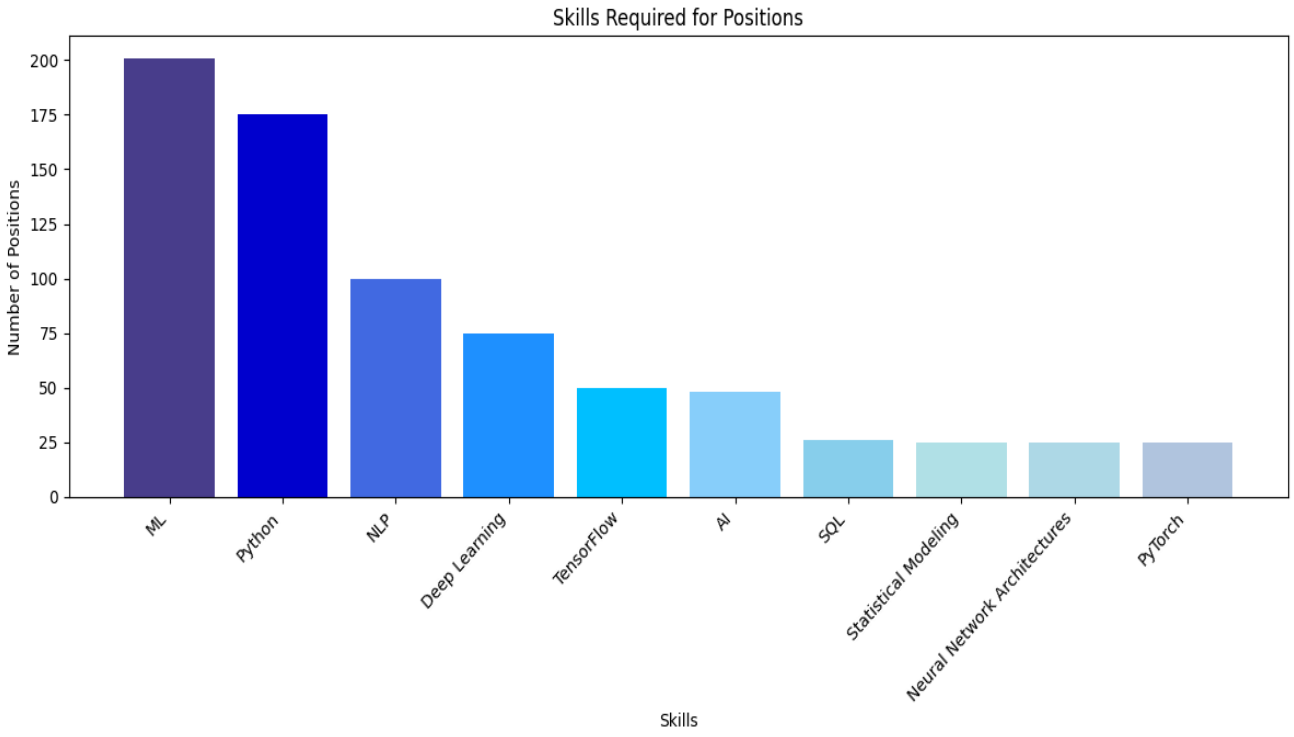


The "**Full-time**" slice looks considerably **bigger** than the "**Part-time**" slice, indicates that the vast majority of the dataset's **machine learning** positions are **full-time** ones. This could suggest that machine learning work is rigorous and ongoing.

People looking for part-time work should have a clear understanding of the distribution and be aware as there aren't as many options available as there are for full-time jobs.

Analysis of key skills required:

To illustrate the skills needed for various positions, a bar chart has been utilized. Each bar in this graphic representation represents a different skill, and the height of the bar shows how many positions require a particular skill.



A **strong** need for **machine learning expertise** is shown by the notably **longer bar** for "**ML**". This would suggest that knowledge of machine learning methods, algorithms, or applications is necessary for most of the positions related to **Machine Learning Engineer**.

The **higher bar** for "**Python**" suggests that **mastery** of this **programming language** is required as it is widely utilized in data science and machine learning, employers in the dataset value having this talent.

A **variety of bars** show a varied skill landscape **beyond** "**ML**" and "**Python**," each of which denotes a unique competence that is sought after in the roles that have been evaluated.

c) Your Ideal Job:

My Ideal job is '**Machine Learning Engineer**'. My interest in becoming a machine learning engineer stems from its capacity to simplify and optimize intricate procedures.

Its **main attraction** lies in its ability to **evaluate information**, **predict outcomes**, and **make decisions automatically**. This allows me to tackle real-world issues and boost productivity in various sectors.

At a **coding event** in **India**, I learned about **Deloitte India's** significant **machine learning** work, which piqued my **curiosity**. Hearing about their work in person greatly motivated me to contribute to practical solutions in the industry.

The **Relevant skills required** to become a potential **Machine Learning Engineer** are **machine learning with Python, Deep learning, web scraping with Selenium**, and **NLP**. I'm gradually becoming proficient in them and also honing my **Pandas data manipulation** skills and exploring **cloud platforms**.

With these abilities, I hope to join a team at **Deloitte** in **India** as **Machine Learning Engineer**, the company that inspired me in the first place, contributing to cutting-edge machine learning solutions.