

1. Introduction

The job market is becoming increasingly competitive, with employers seeking candidates with specific skills and job seekers looking for opportunities that match their skills and interests. Analyzing job posting can provide valuable insights into the job market, helping job seekers, employees and policymakers make informal decisions.

These projects aim to analyze the LinkedIn job posting 2023-24 data set which contains over 13,000+ job postings from various industries and locations. The data set provides a unique opportunity to identify trends in job titles, industry skills and salaries.

By analyzing the LinkedIn job posting 2023-24 dataset we hope to provide valuable insights into the job market and inform decisions for job seekers, employees and recruiters. This project will help others understand the current job market landscapes and identify opportunities for growth and development.

1.1. Purpose

The purpose of our project, which involves analyzing LinkedIn job postings from 2023 related to data analysis, is to uncover hidden insights and trends in the job market. This project aims to provide a comprehensive understanding of the data analyst job landscape, including the most in-demand skills, job requirements, and industry trends. By analyzing these job postings, we can identify patterns and correlations that can inform data-driven decisions in the field of data analysis.

1.2. Objective and scope

1.2.1 Objective

- Identify trends in job titles, industries and skills.
- Analyze salary ranges from different job titles and industries.
- Examine the geographic distribution of job openings.
- Provide recommendations for job seekers, employers and recruiters based on the analysis.

1.2.2 Scope

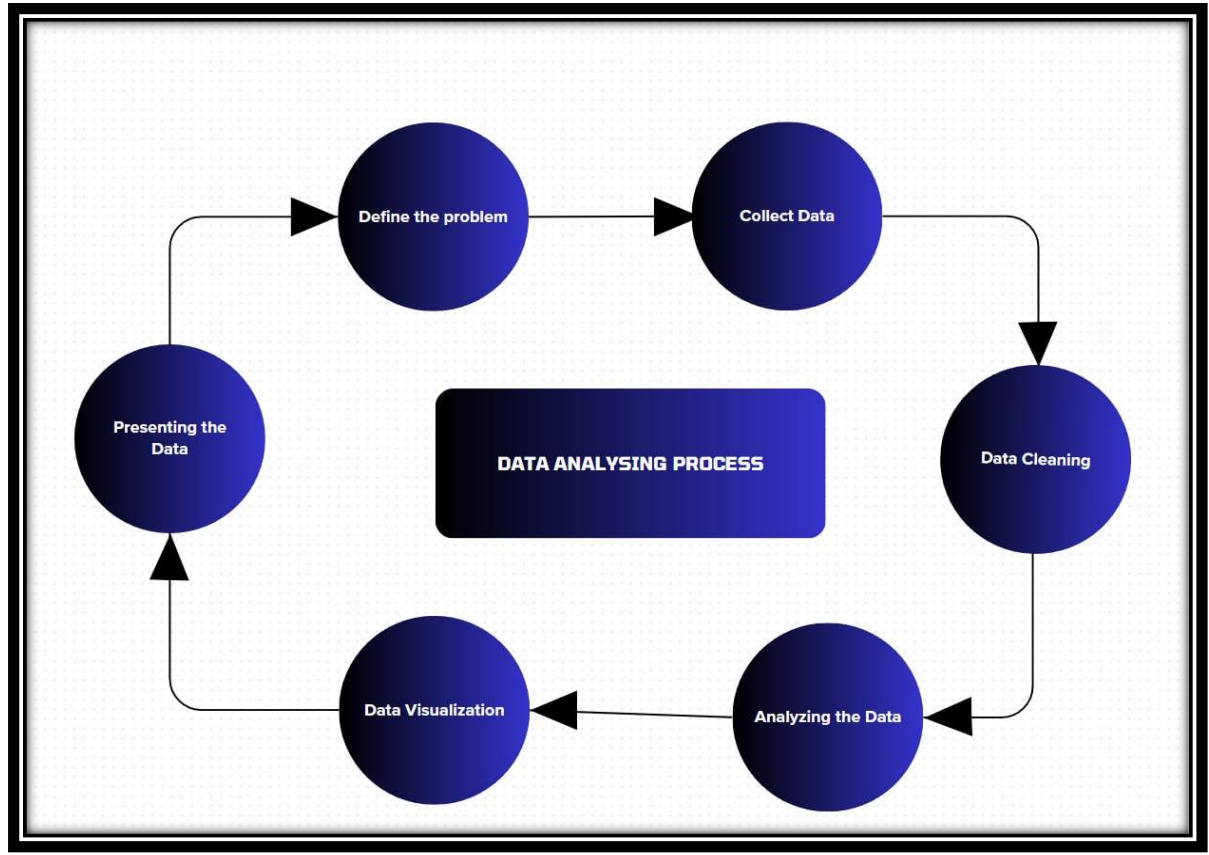
- **Identifying emerging opportunities:** Detecting emerging job roles, high-paying jobs titles and required skills.
- **Provide Actionable insights** offer visualization (graphs charts and plots) and recommendations to users enabling them to make informal decisions.
- **Streamline job search and recruitment:** Develop a user-friendly dashboard catering to the needs of both job seekers and recruiters, providing a centralized platform to accessing and utilizing these insights.

1.3 Theoretical Background definition of problem

The rapid growth of big data and analytics has led to an increased demand for data analysts who can extract insights and drive business decisions. The job market for data analysts is becoming increasingly competitive, with companies seeking professionals who possess a unique combination of technical, business, and soft skills. However, the lack of standardized job descriptions across industries and companies makes it challenging for job seekers, educators to navigate the data analysis job market. The problem addressed in this project is the lack of transparency and consistency in data analysis job postings, which hinders the ability to identify key skills, job requirements, and industry trends.

1.4 Methodology adopted

Data analysis process:



- **Define the Problem:** Here, we gathered the specific question we want to answer from the data set which includes:
 - What are the most common job titles in the data set?
 - What are the most sought-after skills?
 - How does job posting activity vary by job title or location?
 - What are the most common job types being advised?
 - What trends are emerging in job descriptions
 - How salaries range from different job titles.

LinkedIn Job Posting Analysis

- **Collect Data:** Here, the data for this project was collected from the kaggle website, a popular platform for hosting datasets. The linkedin job posting 2023-24 dataset was downloaded in CSV(Comma Separated Values) format, which is widely used for storing and exchanging data.

Source: <https://www.kaggle.com/datasets/arshkon/linkedin-job-postings>

LinkedIn Job Postings (2023 - 2024)

- **Data cleaning:** Data cleaning is an essential step in the data analysis process to ensure that the data is accurate, complete and consistent.
 - Handling missing values
 - Removing duplicates
 - Irrelevant columns or rows
 - Converting data types
 - Imputing missing values
- **Analyzing the data:** The cleaned and preprocessed data was analyzed using descriptive statistics and exploratory data analyzing(EDA) to extract insights and understand the characteristics of the data.
 - Summarize and describe the basic features of data using statistics tools like central tendency, variability and distribution.
 - Explore the data and identify patterns, relationships and correlations.

Details mentioned in dataset

- **company_name:** The name of the company posting the job.
- **title :** The title of the job posting.
- **pay_period:** The pay period for the job (e.g., hourly, monthly, annually).
- **location:** The location of the job posting.
- **formatted_work_type:** The type of work (e.g., full-time, part-time, internship).
- **job_posting_url:** The URL of the job posting.
- **application_url:** The URL for applying to the job.
- **application_type:** The type of application process (e.g., online, in-person).

LinkedIn Job Posting Analysis

- **formatted_experience_level:** The level of experience required for the job (e.g., entry-level, mid-level, senior-level).
- **skills_desc:** A description of the skills required for the job.
- **posting_domain:** The domain or industry of the job posting.
- **currency:** The currency used for the salary.
- **compensation_type:** The type of compensation offered (e.g., salary, hourly, commission).

These categorical variables can be used for analysis, visualization, and modeling tasks, such as:

- Analyzing the distribution of job titles, companies, and locations.
- Identifying the most common skills required for jobs in a particular industry.
- Visualizing the types of compensation offered by companies.

Numerical Variables:

The numerical variables in the dataset are:

job_id: A unique identifier for each job posting.

max_salary: The maximum salary offered for the job.

company_id: A unique identifier for each company.

views: The number of views for the job posting.

min_salary: The minimum salary offered for the job.

applies: The number of applications received for the job.

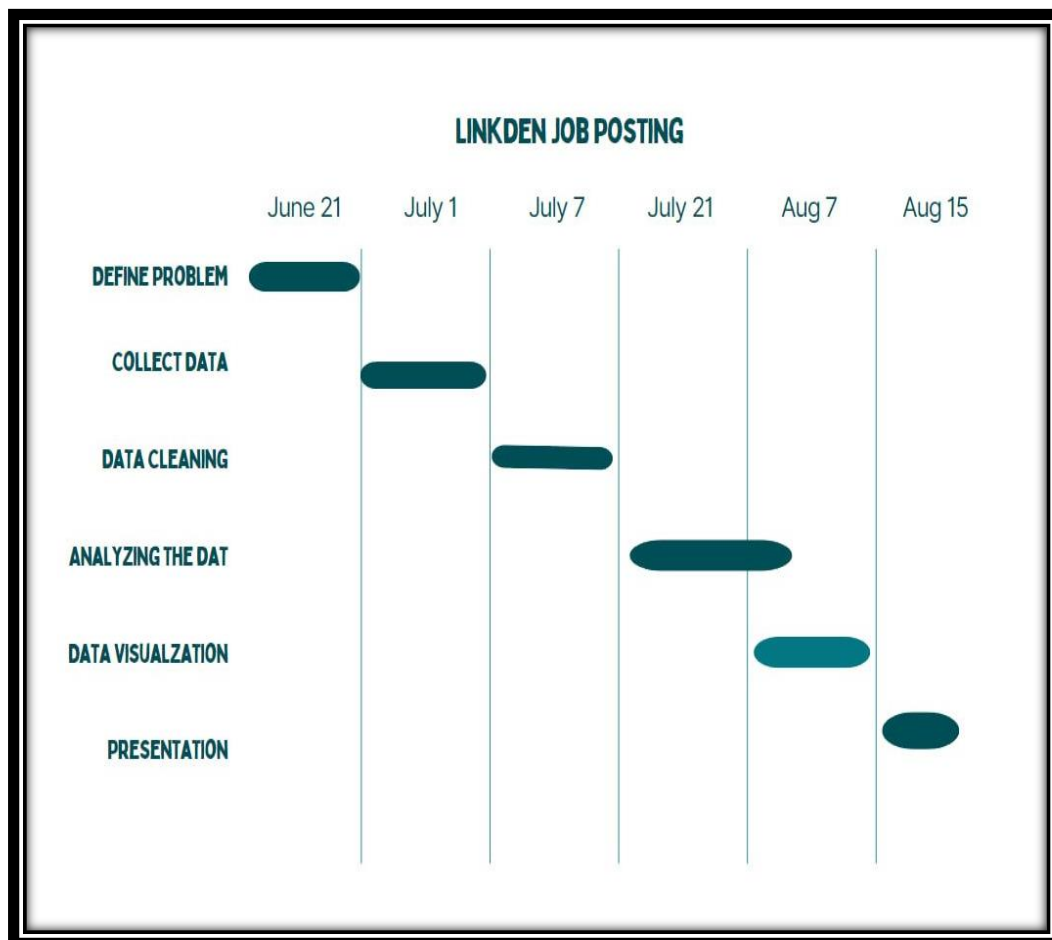
remote_allowed: A binary variable indicating whether the job allows remote work (1 = yes, 0 = no).

- **Data visualization** is a key part of the data analysis process. It involves creating charts, graphs and plots for our data to make it easier to understand and communicate our findings.
- **Presenting the data:** Here, we have used to present our findings in a clear and concise manner using the dashboard using the tool Microsoft Power BI.

1.5 User requirements system planning

1.5.1 Gantt Chart

A Gantt Chart illustrates a project schedule, showing the tasks, dependencies, and timelines for the project. It's a popular and widely used tool in project management, as it provides a clear and concise visual representation of the project plan.



2. Software Requirement Specification

2.1 Analysis Perspectives

The product perspective for a LinkedIn job posting project would involve understanding the needs of both the job seekers and the employers. For job seekers, the product should make it easy to find relevant jobs, apply for them, and manage their job search. For employers, the product should make it easy to post jobs, find qualified candidates, and manage the hiring process.

- **Personalization:** The platform should be personalized to each job seeker, with recommendations for jobs, employers, and networking opportunities.
- **Data and insights:** The platform should provide employers with data and insights about their job postings and candidates, helping them to make better hiring decisions.
- **Customer support:** The platform should provide employers with excellent customer support.

2.2 Analysis Features

- **Decision making:** On the basis of prediction results, it helps to make decisions about job seekers, company and recruiter.
- **It Prediction result will be in the form of graphs, charts and plots**
- **The final report will be in the form of a Dashboard.**
- **Accurate prediction by installing the dataset.**

2.3 Analysis of Existing System

Definition:

- The existing system is a non-computerized system , relying on manual searches and filtering of job postings on LinkedIn.
- SPSS and MS Excel are examples for existing systems.

Limitations:

- Prone to errors and biases
- Time-consuming and labor-intensive
- Lacks insights into job market trends
- Inefficient filtering and searching capabilities

2.4 Analysis Proposed system

To effectively evaluate a data analysis system, it's essential to assess its problem-solving capabilities, data handling prowess, analytical techniques, scalability, user experience, integration, security, cost-effectiveness, and long-term sustainability. A well-suited system should align with organizational goals, handle data effectively, employ appropriate analysis methods, scale to meet growing demands, offer a user-friendly interface, integrate seamlessly with existing systems, protect sensitive data, provide a justifiable return on investment, and offer reliable maintenance and support.

Features:

- Job trend analysis, job titles and locations
- Salary distribution analysis
- Visualization of results using dashboard
- Provides insights into emerging job roles and industriesIdentifies high-paying

2.5 Feasibility Study

A feasibility analysis usually involves a thorough assessment of the operational (need), financial and technical aspects of a proposal. Feasibility study is the test of the system proposal made to identify whether the user needs may be satisfied using the current software and hardware technologies, whether the system will be cost effective from a business point of view and whether it can be developed with the given budgetary constraints. A feasibility study should be relatively cheap and done at the earliest possible time. Depending on the study, the decision is made whether to go ahead with a more detailed analysis.

2.5.1 Technical Feasibility

Technologies which we are going to use in this project are Python Programming Language and their packages. To design the interfaces and forms we will use the Jupyter. Python is a widely used programming language known for its readability, extensive libraries, and strong community support. Pandas for data manipulation, and Matplotlib/Seaborn for data visualization, making it well-suited for tasks such as extracting job data, cleaning and analyzing it, and creating interactive dashboards. Dashboards provide a visually appealing and interactive way to present data insights. These dashboards can be used to explore job trends, analyze job market dynamics, and visualize key performance indicators.

2.5.2 Economic Feasibility

Most of the technologies which are used in this project are open source and available free of cost. As this is a desktop application it doesn't involve hosting and domain registration. The system can be developed with minimum cost so this project is economically feasible. All required software is easily available and free of cost.

2.5.3 Operational Feasibility

Operational feasibility is the measure of how well a proposed system solves the problem and takes advantage of the opportunities identified during scope of definition and how it satisfies the requirements identified in the requirement analysis phase of system development. The interfaces of the software developed are really user friendly and hence there is no possibility for any resistance from users. To understand our system, users should have Data Analysis knowledge and predictive knowledge. Since the prediction results are in graphical format, the user can understand it easily and can make decisions.

2.6 User classes and characteristics:

There is only one module in our system, the **user**. The user should have the Data Analysis knowledge then only he/she can understand the analysis and the prediction results. Users can view the dataset information. User can view the prediction results and can make the appropriate decisions.

- Users can view the dataset information.
- Users can view the prediction report.

2.7 Operating Environment

Software Requirements:

- **Operating System:** Windows, Linux or Mac.
- **Spreadsheet Software:** Microsoft Excel
- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, and Seaborn.
- **IDE:** Jupyter
- **Data Visualization Tools:** Power BI.

Hardware Requirements:

- **RAM:** min 2GB or more
- **CPU:** 2 Ghz or Faster
- **Processor :** i3 processor system or higher
- **Hard Disk:** Min 50 GB free space

2.8 Design Implementation Constraints:

Implementation is a process of ensuring that the information system is operational.

- The dataset should be in .csv format.
- The system should develop in the Python programming Language.
- Limited resources (e.g., computing power, storage), or compatibility issues with different devices and browsers.

2.9 Assumptions and Dependencies:

2.9.1 Assumptions:

- LinkedIn's job posting data is accurate and reliable.
- The analysis must have an interface which is simple enough to understand.
- Users have basic knowledge of data analysis and visualization tools.

2.9.2 Dependencies:

- Accuracy depends on datasets: If the dataset is large, the accuracy is more and if the dataset is small the accuracy is less.
- The system requires integration with other tools or services (e.g., data cleaning, data visualization, data preprocessing etc).

2.10 System Features

A system is an orderly grouping of interdependent components linked together according to a plan to achieve a specific objective.

2.10.1 Functional Requirements:

- **Data Ingestion:** Collect and process LinkedIn job posting data from various sources.
- **Data Analysis:** Apply machine learning and statistical models to identify job trends, patterns, and correlations.
- **Data Visualization:** Provide interactive and dynamic visualizations to help users understand job trends and insights.

3. System Analysis and design

3.1 Exploratory Data Analysis

Exploratory Data Analysis Is a critica; initial step in the data analysis workflow. It involves using Python Libraries to inspect, summarize and visualize data to uncover trends, patterns, and their relationships.

The main purpose of EDA is to help look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, and find interesting relations among the variables. Data scientists can use exploratory analysis to ensure the results they produce are valid and applicable to any desired business outcomes and goals. EDA also helps stakeholders by confirming they are asking the right questions. EDA can help answer questions about standard deviations, categorical variables, and confidence intervals. Once EDA is complete and insights are drawn, its features can then be used for more sophisticated data analysis or modeling, including machine learning.

Here's breakdown of the key steps in performing EDA with Python:

3.1.1. Importing Libraries:

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

- **pandas (pd):** It provides structures like DataFrames for handling tabular data and offers various functions for manipulation and analysis.

LinkedIn Job Posting Analysis

- **NumPy (np):** Provides functions for numerical operations on arrays.
- **Matplotlib.pyplot (plt):** The foundation for creating visualizations in Python. It offers a wide range of plot types for various data exploration needs.
- **Seaborn (sns):** (Optional) Built on top of matplotlib, Seaborn provides a high-level interface for creating beautiful and informative statistical graphics.
- **Statistics (stats):** It is used to understand the measures of central tendency, position and dispersion and covariance, correlation, association and causation.

3.1.2 Loading the Data:

Read your datasets from its source file into a panda Data Frames. Pandas offers functions like `pd.read_csv()`

```
df=pd.read_csv('/content/drive/MyDrive/finall.csv')
```

df

	job_id	company_name	title	max_salary	pay_period	location	company_id	views	med_salary	min_salary	...	expiry	closed_time	formatted_experience_level	skills_desc	listed_time	posting_domain	sponsored	work_type	currency	
0	390529252	AMETEK	Continuous Improvement Specialist	120000.0	YEARLY	Garden City, NY	8086	4.0	NaN	90000.00	...	1.720000e+12	NaN		Entry level	NaN	1.710000e+12	188100.jobs2web.com	0	FULL_TIME	USD
1	3904955043	Paylocity	Enterprise HCM Account Executive	125000.0	YEARLY	Orlando, FL	24614	14.0	NaN	100000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com	0	FULL_TIME	USD
2	3905223649	Paylocity	Senior Human Resources Business Partner	115000.0	YEARLY	United States	24614	2.0	NaN	98000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com	0	FULL_TIME	USD
3	3905301424	Paylocity	HCM Account Executive	105000.0	YEARLY	Chicago, IL	24614	4.0	NaN	60000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com	0	FULL_TIME	USD
4	3904943759	Paylocity	Implementation Analyst I	60528.0	YEARLY	Lake Mary, FL	24614	24.0	NaN	39520.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com	0	FULL_TIME	USD
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
13262	3905823598	Zillow	Coaching and Development Program Manager	97200.0	YEARLY	United States	13990	23.0	NaN	60800.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	zillow.wd5.myworkdayjobs.com	0	FULL_TIME	USD
13263	3905825885	Zillow	Senior Account Advisor	56.2	HOURLY	United States	13990	5.0	NaN	35.15	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	zillow.wd5.myworkdayjobs.com	0	FULL_TIME	USD
13264	3900084683	Ziply Fiber	Sales Account Executive, Multi-Family Communities	50000.0	YEARLY	Everett, WA	34755301	12.0	NaN	30000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com	0	FULL_TIME	USD
13265	3900084991	Ziply Fiber	Sales Account Executive	50000.0	YEARLY	Cougar d'Alene, ID	34755301	6.0	NaN	30000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com	0	FULL_TIME	USD
13266	3903810782	Ziply Fiber	Sales Account Executive	50000.0	YEARLY	Everett, WA	34755301	4.0	NaN	30000.00	...	1.720000e+12	NaN		Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com	0	FULL_TIME	USD
13267 plus x 27 columns																					

LinkedIn Job Posting Analysis

- **data.head():** View the first few rows of the data.

df.head()

	job_id	company_name	title	max_salary	pay_period	location	company_id	views	med_salary	min_salary	...	expiry	closed_time	formatted_experience_level	skills_desc	listed_time	posting_domain
0	3905299252	AMETEK	Continuous Improvement Specialist	120000.0	YEARLY	Garden City, NY	8086	4.0	NaN	90000.0	...	1.720000e+12	NaN	Entry level	NaN	1.710000e+12	188100.jobs2web.com
1	3904956043	Paylocity	Enterprise HCM Account Executive	125000.0	YEARLY	Orlando, FL	24614	14.0	NaN	100000.0	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com
2	3906223649	Paylocity	Senior Human Resources Business Partner	115000.0	YEARLY	United States	24614	2.0	NaN	98000.0	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com
3	3905301424	Paylocity	HCM Account Executive	105000.0	YEARLY	Chicago, IL	24614	4.0	NaN	60000.0	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com
4	3904948759	Paylocity	Implementation Analyst I	60528.0	YEARLY	Lake Mary, FL	24614	24.0	NaN	39520.0	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	2000recruiting.paylocity.com

- **data.tail():** view the last few rows of the data.

df.tail()

	job_id	company_name	title	max_salary	pay_period	location	company_id	views	med_salary	min_salary	...	expiry	closed_time	formatted_experience_level	skills_desc	listed_time	posting_domain
13262	3905828598	Zillow	Coaching and Development Program Manager	97200.0	YEARLY	United States	13990	23.0	NaN	60800.00	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	zillow.wd5.myworkdayjobs.com
13263	3905825885	Zillow	Senior Account Advisor	56.2	HOURLY	United States	13990	5.0	NaN	35.15	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	zillow.wd5.myworkdayjobs.com
13264	3900084683	Ziply Fiber	Sales Account Executive, Multi-Family Communities	50000.0	YEARLY	Everett, WA	34755301	12.0	NaN	30000.00	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com
13265	3900084691	Ziply Fiber	Sales Account Executive	50000.0	YEARLY	Coeur d'Alene, ID	34755301	6.0	NaN	30000.00	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com
13266	3903810782	Ziply Fiber	Sales Account Executive	50000.0	YEARLY	Everett, WA	34755301	4.0	NaN	30000.00	...	1.720000e+12	NaN	Mid-Senior level	NaN	1.710000e+12	ziplyfiber.zohorecruit.com


3.1.3 Initial Inspection:


Get an overview of the data using

- a. **data.shape():** Get the dimensions of the data (Number of rows and columns).

LinkedIn Job Posting Analysis

`df.shape`

 `df.shape`

 `(13267, 27)`

- **data.info():** Provides information about data types, missing values, and memory usage for each column.

`df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13267 entries, 0 to 13266
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   job_id                                13267 non-null  int64
1   company_name                          13267 non-null  object
2   title                                 13267 non-null  object
3   max_salary                            13267 non-null  float64
4   pay_period                            13267 non-null  object
5   location                              13267 non-null  object
6   company_id                            13267 non-null  int64
7   views                                 13240 non-null  float64
8   med_salary                            0 non-null     float64
9   min_salary                            13267 non-null  float64
10  formatted_work_type                   13267 non-null  object
11  applies                               1846 non-null   float64
12  original_listed_time                  13267 non-null  float64
13  remote_allowed                        1570 non-null   float64
14  job_posting_url                       13267 non-null  object
15  application_url                       12638 non-null  object
16  application_type                       13267 non-null  object
17  expiry                                13267 non-null  float64
18  closed_time                           45 non-null     float64
19  formatted_experience_level             13267 non-null  object
20  skills_desc                           188 non-null    object
21  listed_time                           13267 non-null  float64
22  posting_domain                        13267 non-null  object
23  sponsored                             13267 non-null  int64
24  work_type                             13267 non-null  object
25  currency                              13267 non-null  object
26  compensation_type                     13267 non-null  object
dtypes: float64(10), int64(3), object(14)
memory usage: 2.7+ MB
```


LinkedIn Job Posting Analysis

3.1.4 Data cleaning:

Data cleaning is an essential step in data analysis that involves fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. The goal of data cleaning is to ensure that the data is accurate, complete, and in a usable format for analysis.

- b. Remove duplicate
- c. Handle missing data
 - Identify and handle missing values using methods like `data.isnull().sum()`.

`df.isnull().sum()`



	0
job_id	0
company_name	0
title	0
max_salary	0
pay_period	0
location	0
company_id	0
views	27
med_salary	13267
min_salary	0
formatted_work_type	0
applies	11421
original_listed_time	0
remote_allowed	11697
job_posting_url	0
application_url	629
application_type	0
expiry	0
closed_time	13222
formatted_experience_level	0
skills_desc	13079
listed_time	0
posting_domain	0
sponsored	0
work_type	0
currency	0
compensation_type	0

LinkedIn Job Posting Analysis

- Find and address duplicates with `data.duplicated().sum()`.


`df.duplicated().sum()`

- The `df.drop()` method in Pandas is used to remove rows or columns from a DataFrame (`df`). It can be used to drop a specific row by its index label, or a specific column by its column label.

`df.drop()`

- `ffill`: This line of code fills any missing values in your DataFrame by copying the values from the previous non-missing cell in the same column.

`df.ffill(inplace=True)`



```
[ ] df.drop(['med_salary','expiry','listed_time','sponsored','work_type','original_listed_time','closed_time'],axis=1,inplace=True)

[ ] df['remote_allowed'].fillna(0,inplace=True)

[ ] df['skills_desc'].fillna('skills related to field',inplace=True)

df.ffill(inplace=True)
```

3.1.5 Statistics Summary

Statistical analysis in data analysis involves using mathematical techniques to extract insights and meaning from data. It includes descriptive statistics, inferential statistics, exploratory data analysis, inferential modeling, and data visualization to identify patterns, relationships, and correlations, and make informed decisions from data.

- **describe()**– function gives all statistics summary of data

df.describe()

	job_id	max_salary	company_id	views	min_salary	applies	remote_allowed
count	1.326700e+04	1.326700e+04	1.326700e+04	13240.000000	13267.000000	1846.000000	13267.000000
mean	3.899656e+09	8.647448e+04	6.764700e+06	13.018429	59140.483444	7.213976	0.118339
std	6.988683e+06	9.376655e+04	1.801109e+07	145.295141	62574.088524	20.743687	0.323021
min	3.815848e+09	1.000000e+00	1.016000e+03	1.000000	1.000000	1.000000	0.000000
25%	3.894873e+09	3.375000e+01	8.737000e+03	4.000000	25.000000	1.000000	0.000000
50%	3.902834e+09	7.835000e+04	1.627500e+05	4.000000	56100.000000	2.000000	0.000000
75%	3.904959e+09	1.440000e+05	2.615835e+06	6.000000	99894.500000	5.000000	0.000000
max	3.906265e+09	1.000001e+06	1.023689e+08	9975.000000	600000.000000	398.000000	1.000000

Conclusion:

The data shows the descriptive statistics for different job attributes. The maximum salary is 1.000001e+06. The minimum salary is 1.000000. The average salary is 8.647448e+04. The average views for jobs are 13.018429. The average applications for jobs are 7.213976. The average number of companies associated with jobs is 6.764700e+06. The average minimum salary for jobs is 59140.483444. The maximum number of views is 9975. The maximum number of applications is 398. The maximum number of companies associated with a job is 1.023689e+08.

LinkedIn Job Posting Analysis

- **Mean:** Mean is a fundamental concept in data science used to summarize a set of data by finding its average value. It essentially represents the central tendency of the data.

```
mean =df1['max_salary'].mean()  
mean  
[8]:  
86474.4788309339
```

- **Median:** The median is a fundamental statistic used to understand the "center" of a data set. It represents the middle value when the data is arranged in ascending or descending order.

```
mid =df1['max_salary'].median()  
mid  
[9]:  
78350.0
```

3.1.6 EDA Univariate Analysis

Univariate analysis is a fundamental component of Exploratory Data Analysis (EDA) that involves examining individual variables or features of a dataset. This process utilizes graphical methods such as histograms, box plots, and stem-and-leaf plots to visualize the data, providing insights into the distribution of each variable.

LinkedIn Job Posting Analysis

Line plot

Line plots are a useful tool for visualizing time-series data or showing the relationship between variables. To create a line chart using Matplotlib, use the plot() function, which accepts x and y coordinates.

```
aa=df['title'].value_counts().head(20)
```

```
plt.plot(aa.index,aa.values,color='pink',linewidth=2,marker='*',linestyle='dashed',markerfacecolor='blue',markersize=7,alpha=0.4)
```

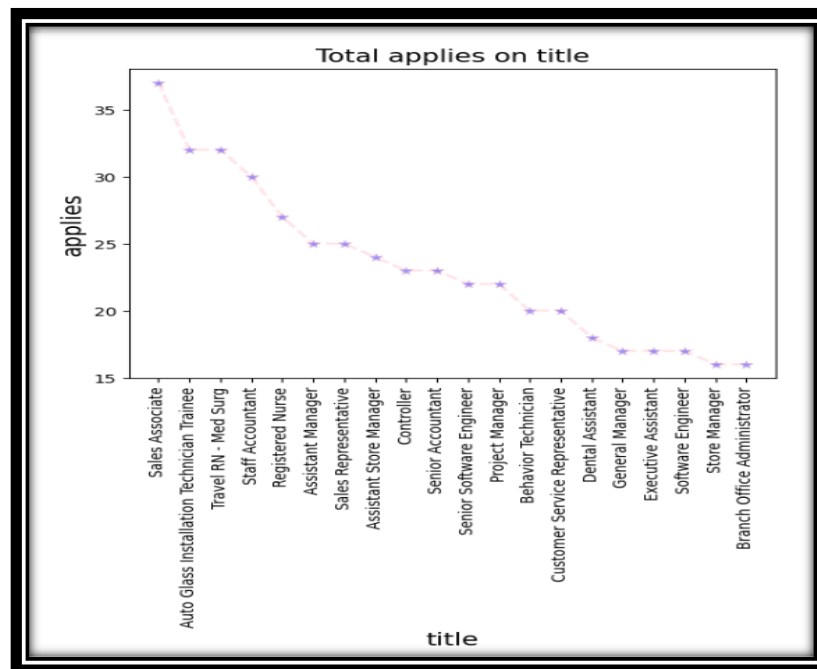
```
plt.xlabel("title",fontsize=15)
```

```
plt.ylabel("applies",fontsize=15)
```

```
plt.title("Total applies on title",fontsize=15)
```

```
plt.xticks(rotation=90)
```

```
plt.show()
```



LinkedIn Job Posting Analysis

Conclusion:

The most popular job title is "Sales Associate" with 37 applications. The least popular job title is "Branch Office Administrator" with 16 applications. The job titles with the most applications are generally those that are more entry-level or more common in the company. The job titles with the fewest applications are generally those that are more specialized or more senior-level.

Pie chart

A pie chart graph provides valuable insights into data analysis by displaying the proportional distribution of data, showing how different categories contribute to the whole.

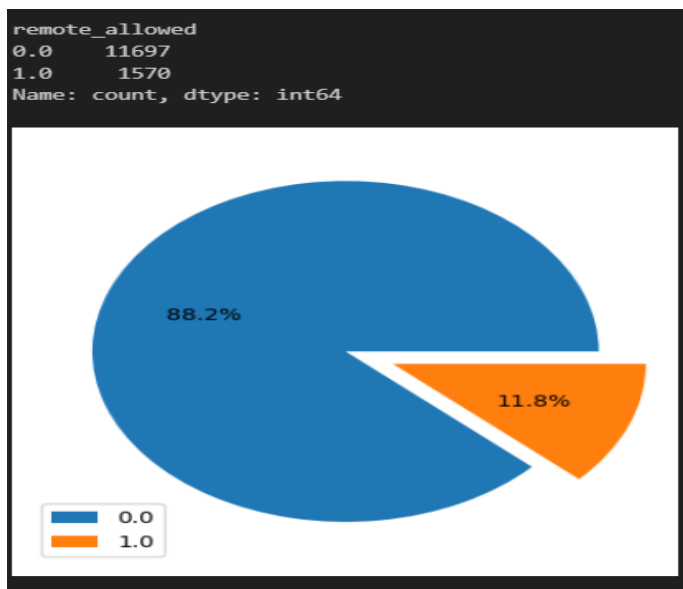
```
re= df['remote_allowed'].value_counts()
```

```
print(re)
```

```
plt.pie(re,autopct='%1.1f%%',explode=[0,0.2])
```

```
plt.legend(labels=re.index,loc='lower left')
```

```
plt.show()
```



LinkedIn Job Posting Analysis

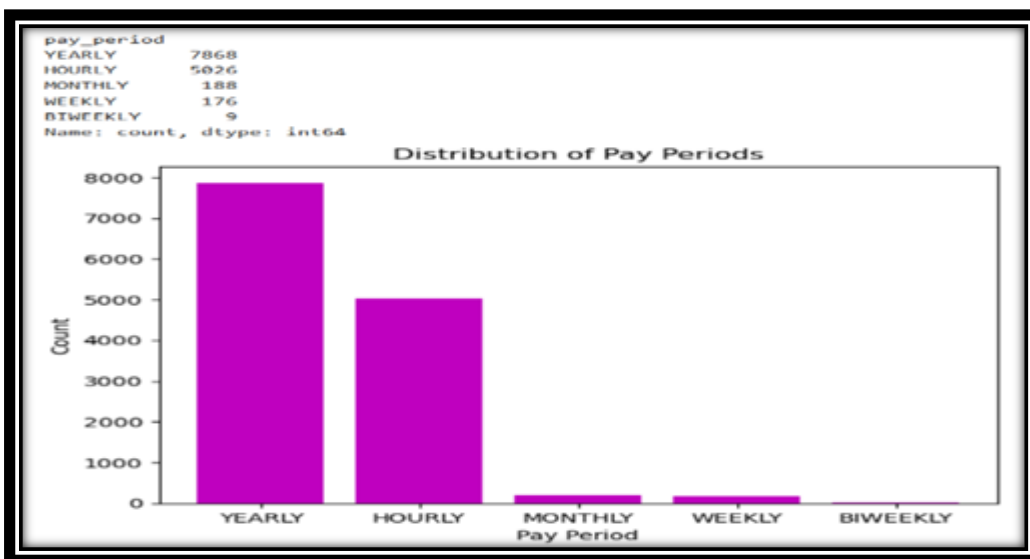
Conclusion:

- The majority of data (88.2%) represents the case where remote access is not allowed.
- A smaller portion of the data (11.8%) represents the case where remote access is allowed.

A bar plot

A bar plot can be used to show the relationship between Categorical variables and continuous variables

```
df1['pay_period'].unique()
value_counts = df['pay_period'].value_counts()
print(value_counts)
plt.bar(value_counts.index, value_counts.values,color='m')
plt.xlabel('Pay Period')
plt.ylabel('Count')
plt.title('Distribution of Pay Periods')
plt.show()
```



LinkedIn Job Posting Analysis

Conclusion:

The majority of employees are paid yearly, with 7868 employees being paid this way. The second most common pay period is hourly, with 5026 employees being paid this way. The remaining pay periods are significantly less common, with 188 employees being paid monthly, 176 employees being paid weekly, and 9 employees being paid biweekly. This tells us that the vast majority of employees are paid on a yearly or hourly basis, with a small number of employees being paid on a more frequent basis.

Box Plot

A box plot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a dataset. It provides a visual summary of key statistical measures such as the median, quartiles, and potential outliers within the data.

#Box Plot of Salary by Experience Level(main graph)

```
plt.figure(figsize=(8, 6))
```

```
sns.boxplot(x='formatted_experience_level', y='max_salary', data=df2)
```

```
plt.title('Salary Distribution by Experience Level')
```

```
plt.xlabel('Experience Level')
```

```
plt.ylabel('Max Salary')
```

```
plt.show()
```


LinkedIn Job Posting Analysis



Conclusion:

The boxplot shows the distribution of maximum salary for entry-level and mid-senior level positions. We can see that the median maximum salary for mid-senior level positions is higher than that of entry-level positions. Additionally, the range of maximum salaries for mid-senior level positions is also wider, suggesting that there is a greater variation in maximum salary for these positions. This indicates that experience level is a significant factor in determining maximum salary.

3.1.6 EDA Bivariate Analysis

Bivariate Analysis helps to understand how variables are related to each other and the relationship between dependent and independent variables present in the dataset.

LinkedIn Job Posting Analysis

Stem plot

A stem plot is a graphical representation of data that provides a quick and easy way to explore the distribution of values

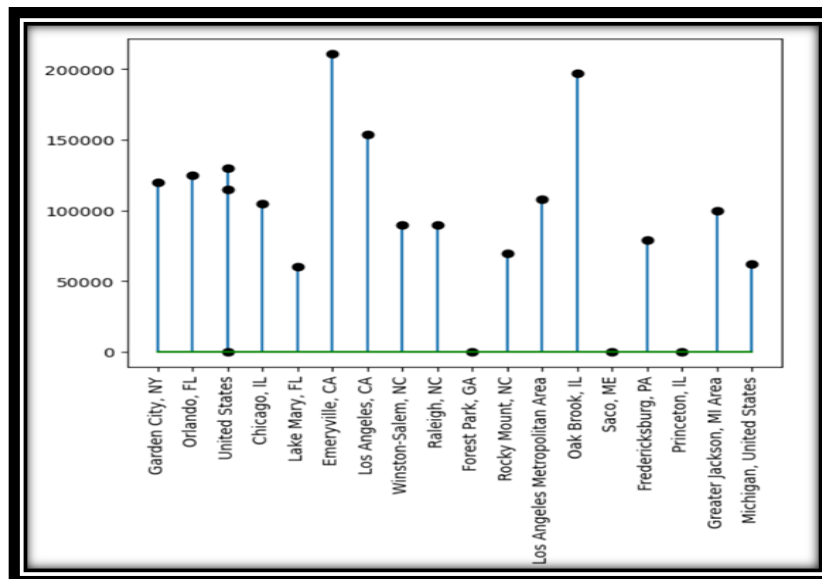
```
x=df2['location']
```

```
y=df2['max_salary']
```

```
plt.stem(x,y,markerfmt='ko',bottom=0,basefmt="g",label="python")
```

```
plt.xticks(rotation=90)
```

```
plt.show()
```



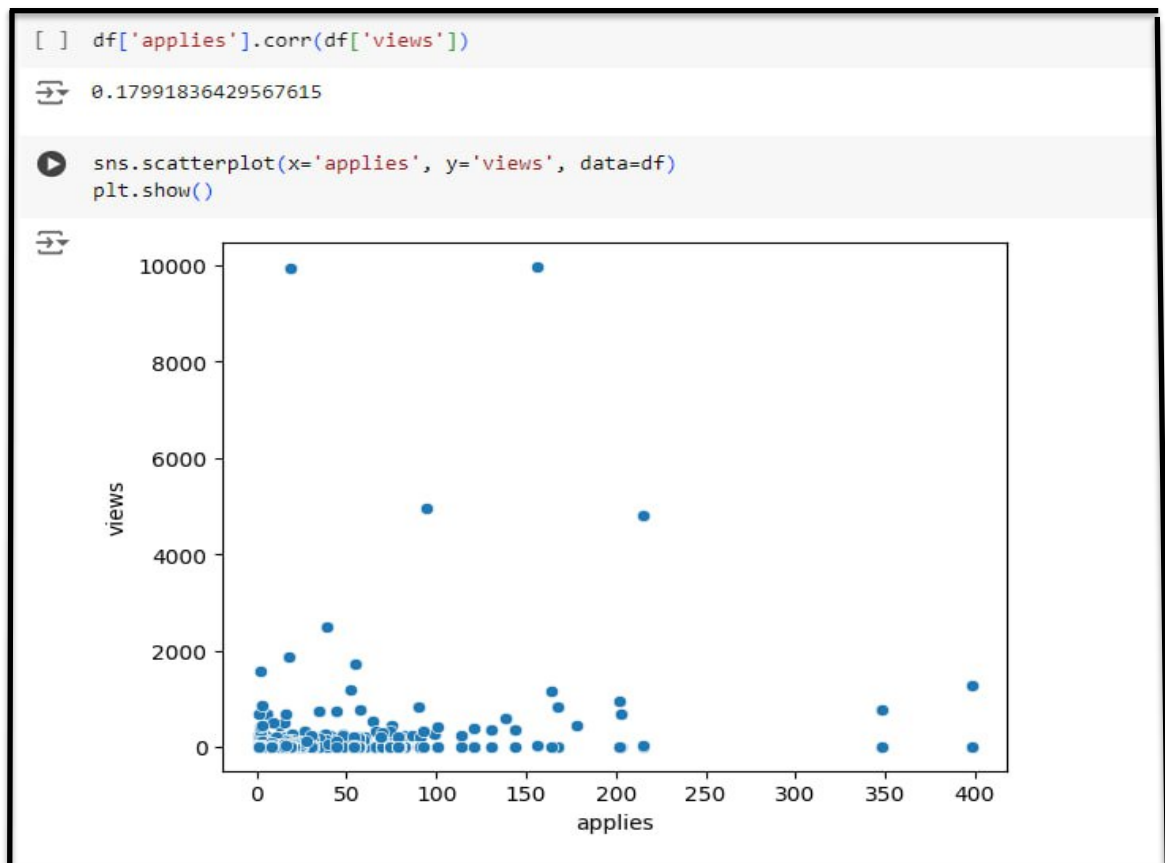
Conclusion:

The graph shows the maximum salary for different locations. The location with the highest maximum salary is Emeryville, CA, followed by Los Angeles, CA and Los Angeles Metropolitan Area. It is interesting to note that the maximum salary in the United States is significantly lower than in some specific locations, such as Emeryville and Los Angeles.

LinkedIn Job Posting Analysis

Scatter Plot

A scatter plot is a graphical representation that displays the relationship between two continuous variables. It's a two-dimensional plot where each data point is represented by a dot or marker on the graph



Conclusion:

The scatter plot shows a weak positive correlation between the number of applies and views. This means that as the number of applies increases, the number of views tends to increase as well, but the relationship is not very strong. There are a few outlier points with very high view counts but relatively low application numbers. These might represent popular job postings with high visibility but lower applies.

LinkedIn Job Posting Analysis

USD DATA

The data shows that the most job postings are in the United States (719), followed by New York, NY (680). This suggests that the job market is strong in those areas, with a high demand for workers. The data is consistent with the idea that certain locations tend to have more job opportunities than others. So, we analysis on United States because they have strong job market and high demand for workers.

```
df['location'].value_counts()
```

	count
location	
United States	719
New York, NY	680
Seattle, WA	231
Los Angeles, CA	229
San Francisco, CA	212
...	...
Lake Buena Vista, FL	1
Wesley Chapel, FL	1
Plainfield, IN	1
Timmonsville, SC	1
Holtsville, NY	1

2591 rows × 1 columns

dtype: int64

Conclusion:

It shows the frequency count for locations. The table shows the top 5 most frequent locations and some additional locations that have a count of 1. The table suggests that the most common locations are in the United States, with the top locations being United States, New York, NY, Seattle, WA, Los Angeles, CA, and San Francisco, CA.

LinkedIn Job Posting Analysis

- Identify the top 20 title in the usd

```
popular_title=usd[['title','company_name','formatted_work_type','pay_period']].value_counts().head(2)
print(popular_title)
```

title	company_name	formatted_work_type	pay_period	
(US) Solutions Analyst - Clinical - 1 year Contract	PointClickCare	Contract	HOURLY	4
VP, HR Business Partners & Organizational Effectiveness (US)	PointClickCare	Full-time	YEARLY	4
Software Support Specialist	Storable	Full-time	YEARLY	3
Strategic Account Executive	Outreach	Full-time	YEARLY	3
Inside Unlicensed Sales Agent	National General	Full-time	YEARLY	3
(US) Software Implementation Consultant - 12 Month Contract - Acute & Payer	PointClickCare	Contract	HOURLY	3
Industry Sales Executive	Oracle	Full-time	YEARLY	3
(US) Software Implementation Consultant - Level 2 - Financial	PointClickCare	Full-time	YEARLY	3
(US) Sr. Customer Success Manager, Practice Groups - Value Based Care	PointClickCare	Full-time	YEARLY	3
Data Center Technician	Oracle	Full-time	HOURLY	3
Applied Researcher - Machine Learning	PointClickCare	Full-time	YEARLY	2
Product Marketing Specialist	EnergyHub	Full-time	YEARLY	2
Associate Director, Professional Services - Post Acute Care	PointClickCare	Full-time	YEARLY	2
Associate Medical Director - Clinical Performance Team	Included Health	Full-time	YEARLY	2
Senior Backend Engineer, Treat Team	Included Health	Full-time	YEARLY	2
Lead Engineer, QA	Majesco	Full-time	YEARLY	2
Banking Product Support - Deposit Origination Analyst	FIS	Full-time	YEARLY	2
Senior AI Deep Learning Engineer - REMOTE	Perficient	Full-time	YEARLY	2
C# Developer, Payments	Versapay	Full-time	YEARLY	2
LPS Controls Engineer I	FORTNA	Full-time	YEARLY	2

Name: count, dtype: int64

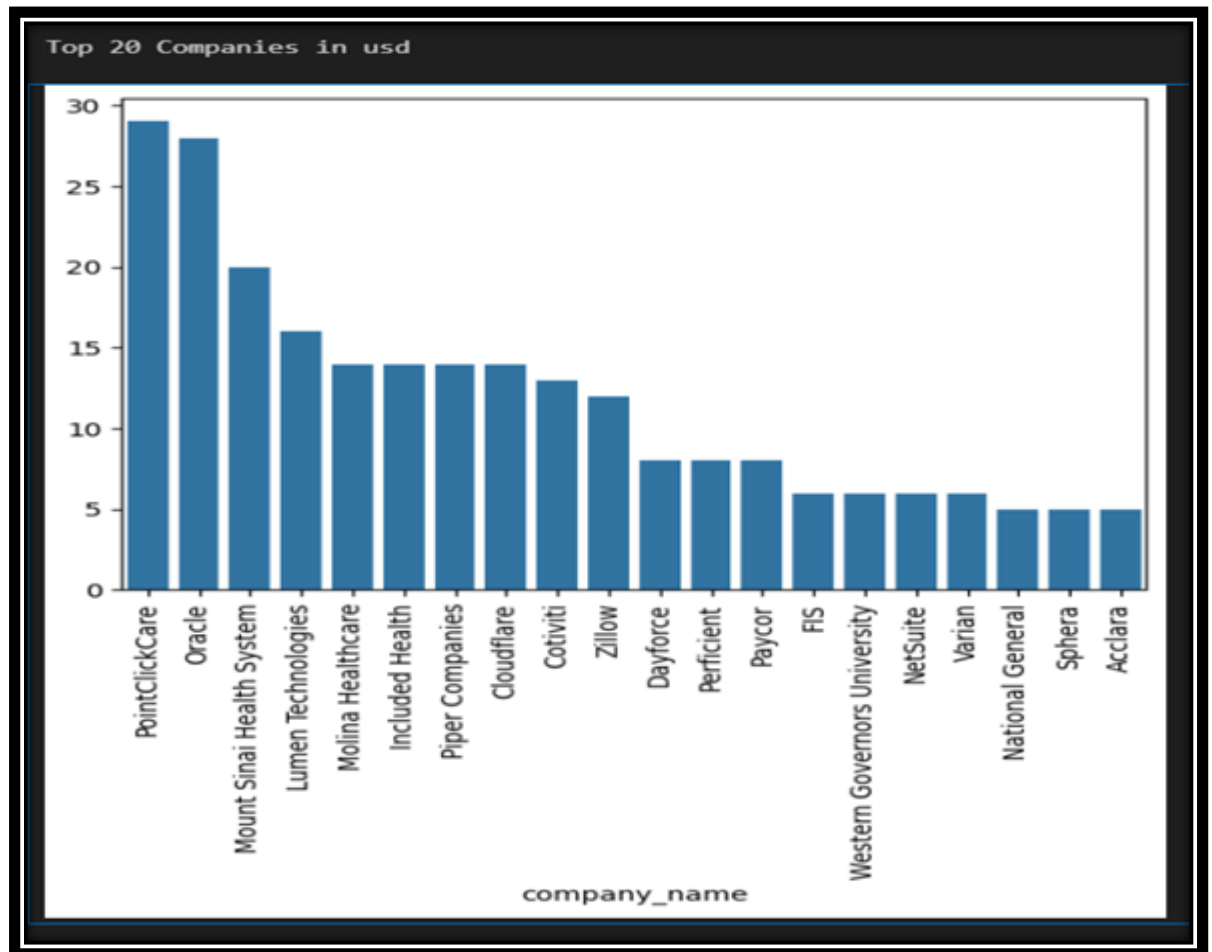
Conclusion:

The top 2 title in the United State are 'Software Support Specialist' and 'Data Center Technician'. Their formatted work type is full time. the dataset are paid on a yearly basis. There are only a small number of people who are paid on an hourly basis.

LinkedIn Job Posting Analysis

- Top 20 Companies in United State

```
c_count=usd['company_name'].value_counts().head(20)sns.barplot(x=c_count.index,y=c_count.values)plt.xticks(rotation=90)plt.show()
```



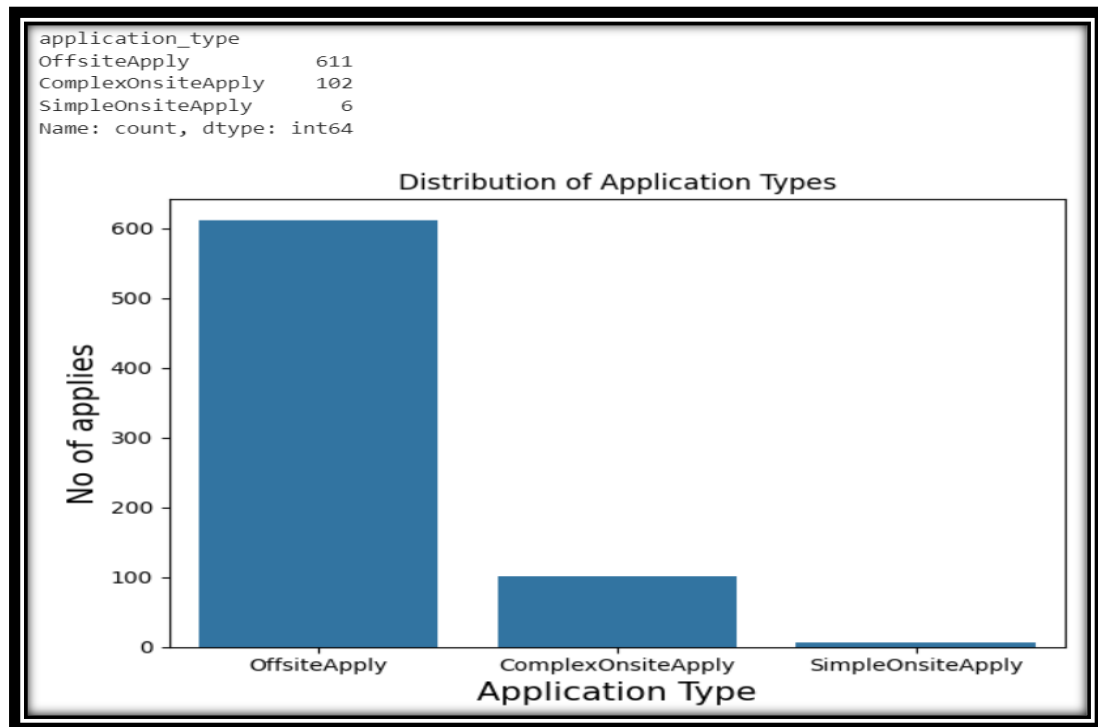
Conclusion:

The graph shows the top 20 companies in terms of market capitalization. PointClickCare is the company with the highest market capitalization, followed by Oracle and Mount Sinai Health System. The graph shows a clear trend of decreasing market capitalization as you move down the list of companies. This suggests that the top companies in the market are significantly larger than the rest of the companies.

LinkedIn Job Posting Analysis

- Distribution of Application type in United State

```
application = usd['application_type'].value_counts()
print(application)
sns.barplot(x=application.index, y=application.values)
plt.xlabel('Application Type',fontsize=15)
plt.ylabel('No of applies',fontsize=15)
plt.title('Distribution of Application Types')
plt.tight_layout()
plt.show()
```



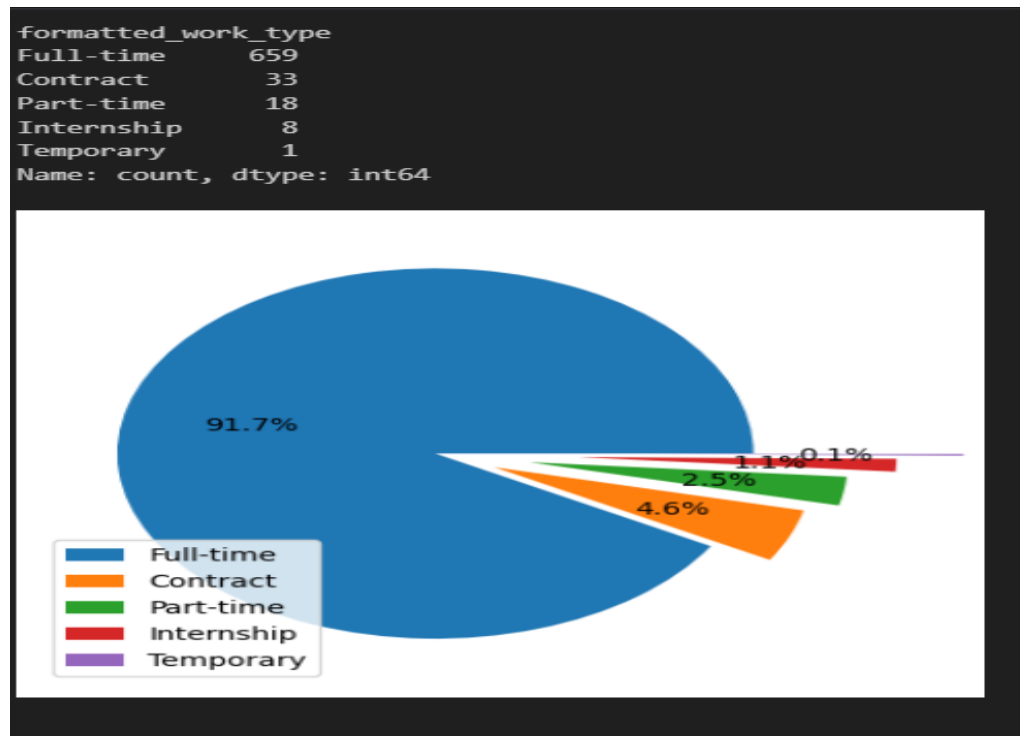
Conclusion:

The graph shows the distribution of application types. OffsiteApply is the most common type, followed by ComplexOnsiteApply and SimpleOnsiteApply. This suggests that most applicants apply for offsite positions, with a smaller number applying for onsite positions. The number of applicants for SimpleOnsiteApply is very low.

LinkedIn Job Posting Analysis

- Formatted work type in United State

```
v=usd['formatted_work_type'].value_counts()
print(v)
plt.pie(v,autopct='% 1.1f%%',explode=[0,0.2,0.3,0.45,0.66])
plt.legend(labels=v.index,loc='lower left')
plt.show()
```



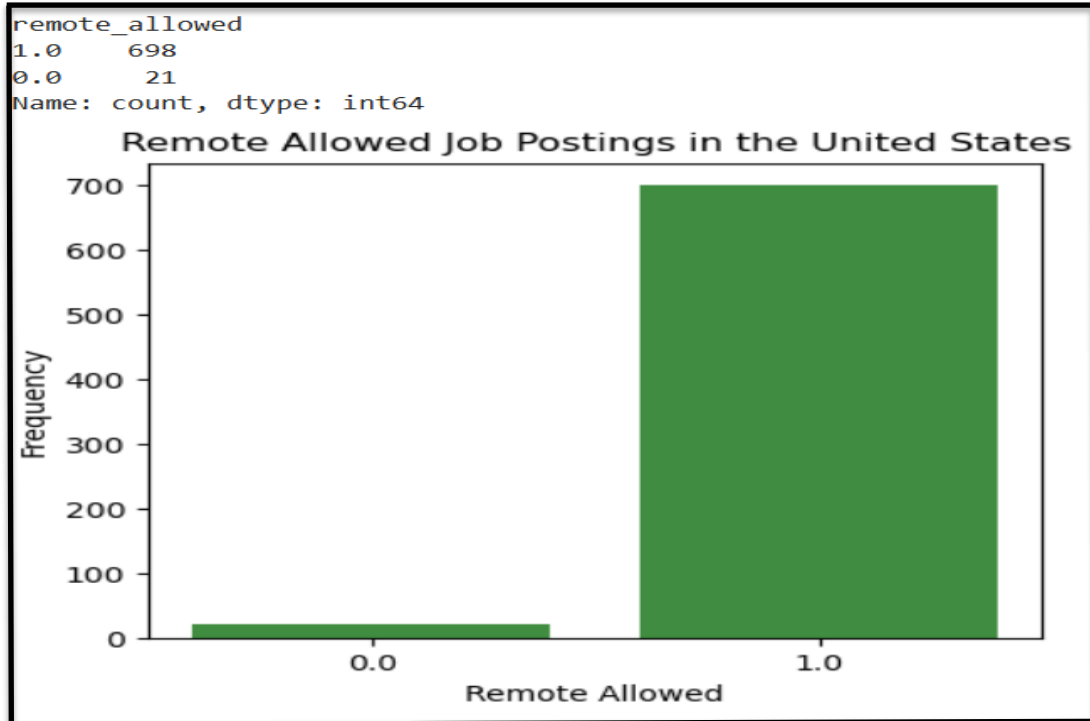
Conclusion:

The pie chart shows the distribution of work types. Full-time work is the most common type, accounting for 91.7% of the data. Contract work is the second most common type, accounting for 4.6% of the data. Part-time work accounts for 2.5% of the data, followed by internship work at 1.1%. Temporary work is the least common type, accounting for just 0.1% of the data.

LinkedIn Job Posting Analysis

- Remote allowed job posting in United State

```
plt.figure(figsize=(5, 4))
rem = usd['remote_allowed'].value_counts()
print(rem)
sns.barplot(x=rem.index, y=rem.values, alpha=0.8)
plt.title('Remote Allowed Job Postings in the United States')
plt.xlabel('Remote Allowed')
plt.ylabel('Frequency')
plt.show()
```



Conclusion:

The bar chart shows that the majority of job postings in the United States allow remote work. Specifically, 698 postings allow remote work, while only 21 do not.

LinkedIn Job Posting Analysis

- **Pay period in United State**

```
usd['pay_period'].unique()

value_counts = usd['pay_period'].value_counts()

print(value_counts)

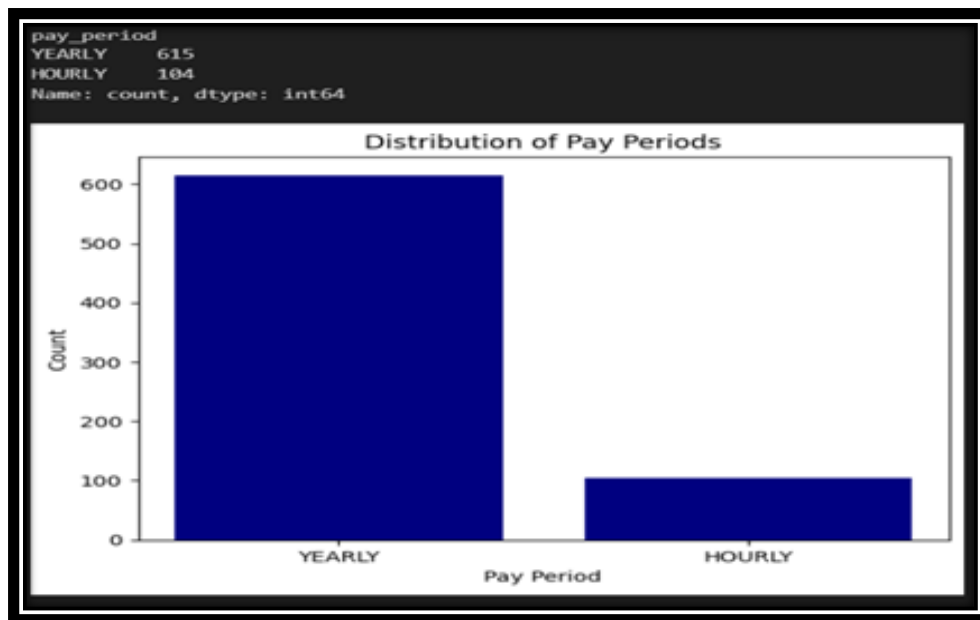
plt.bar(value_counts.index, value_counts.values,color='navy')

plt.xlabel('Pay Period')

plt.ylabel('Count')

plt.title('Distribution of Pay Periods')

plt.show()
```



Conclusion:

The graph shows that the vast majority of individuals in the dataset are paid on a yearly basis(615). There are only a small number of people who are paid on an hourly basis(104). This suggests that the dataset is likely skewed towards individuals who work in more traditional jobs, such as salaried positions.

LinkedIn Job Posting Analysis

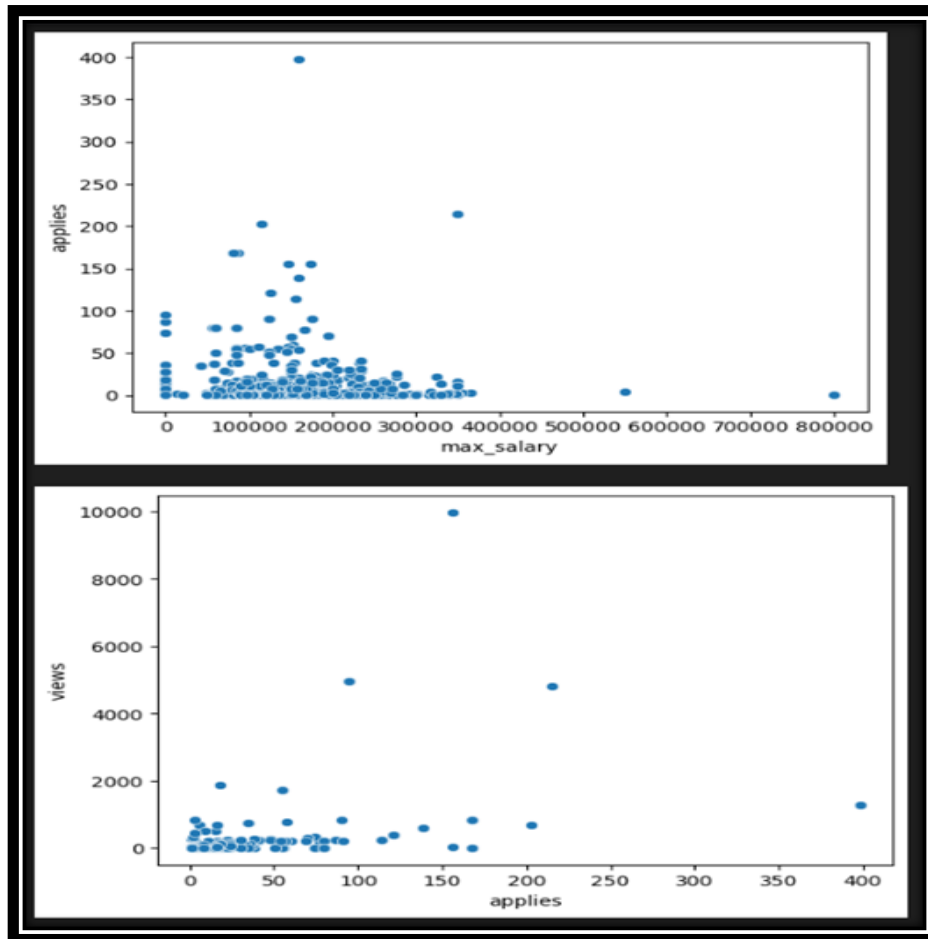
- Maximum salary on Applies and Views on Applies in United State

```
sns.scatterplot(x='max_salary', y='applies', data=usd)
```

```
plt.show()
```

```
sns.scatterplot(x='applies', y='views', data=usd)
```

```
plt.show()
```



Conclusion:

Maximum Salary vs Applies: The scatter plot shows that there is no clear relationship between the maximum salary offered for a job and the number of people who apply for it.

Applies vs Views: The scatter plot shows that there is a weak positive relationship between the number of people who apply for a job and the number of views the job listing receives. This suggests that jobs that receive more views tend to have a slightly higher number of applicants.

LinkedIn Job Posting Analysis

- Top 10 largest max_salary, company name, title, formatted_experience_level, application_type, pay_period, applies , views , remote_allowed , formatted_work_type, skills_desc

```
usd.nlargest(10,'max_salary')[['max_salary','company_name','title','formatted_experience_level','application_type','pay_period','applies','views','remote_allowed','formatted_work_type','skills_desc']]
```

	max_salary	company_name	title	formatted_experience_level	application_type	pay_period	applies	views	remote_allowed	formatted_work_type	skills_desc
10636	800000.0	Alliant Insurance Services	Structured Settlement Account Manager	Associate	OffsiteApply	YEARLY	1.0	6.0	1.0	Full-time	skills related to field
6602	550000.0	Gigamon	Vice President, Enterprise Sales - Americas	Executive	OffsiteApply	YEARLY	4.0	5.0	1.0	Full-time	skills related to field
568	366000.0	Cloudflare	Sr Director, Global People Services & Analytics	Director	ComplexOnsiteApply	YEARLY	3.0	2.0	1.0	Full-time	skills related to field
573	355000.0	Cloudflare	Head of Corporate Finance (Sr. Director / Dire...	Director	ComplexOnsiteApply	YEARLY	3.0	6.0	1.0	Full-time	skills related to field
782	350000.0	Calm	VP, Health Plan Account Management	Executive	OffsiteApply	YEARLY	215.0	4807.0	1.0	Full-time	skills related to field
6704	350000.0	Alpine Immune Sciences, Inc.	Senior Medical Director, Clinical Development...	Director	OffsiteApply	YEARLY	1.0	5.0	1.0	Full-time	skills related to field
6787	350000.0	Outreach	Strategic Account Executive	Mid-Senior level	ComplexOnsiteApply	YEARLY	16.0	44.0	1.0	Full-time	skills related to field
6788	350000.0	Outreach	Strategic Account Executive	Mid-Senior level	ComplexOnsiteApply	YEARLY	11.0	26.0	1.0	Full-time	skills related to field
6789	350000.0	Outreach	Strategic Account Executive	Mid-Senior level	ComplexOnsiteApply	YEARLY	11.0	20.0	1.0	Full-time	skills related to field
6984	347500.0	Merck	Pipeline & Biomarker Global Senior Director Me...	Director	OffsiteApply	YEARLY	2.0	4.0	1.0	Full-time	skills related to field

4. System Implementation

4.1 System Implementation

Definition:

System implementation is a process of ensuring that the information system is operational. It involves constructing a new system from scratch. Constructing system from the existing one.

4.1.1 Installation Steps:

- Here, we first install **Python 3.8 software**:
 - Go to the official Python website ([www.Python.org](https://www.python.org)) and download the latest version of Python 3.8 for your operating system(Windows Mac OS or Linux).
 - Once the download is complete, run the installer and follow the instruction prompts.
- **Python Packages**:
 - The packages are pandas, NumPy, matplotlib and seaborn.
 - Type '**pip install <package_name>**' and press Enter, replacing '**<package name>**' with the name of the package you want to install.
 - pip will download and install the packages along with any dependencies required by the package.
- **Power BI**:
 - Go to the official Microsoft Power BI website (<https://powerbi.microsoft.com/>) and download the "Power BI Desktop".

- Once the download is complete, run the downloaded executable file.

Follow the on-screen instructions to install Power BI Desktop on your computer.

Python

Python was created in the late 1980s by Guido van Rossum, a Dutch computer programmer. At the time, van Rossum was working at the National Research Institute of Mathematics and Computer Science in the Netherlands, where he was unable to find a suitable scripting language for his work. He began working on a new language which he called Python, and released the first version, Python 0.91, in 1991.

Python is a high-level, interpreted programming language that is widely used for various purposes such as web development, scientific computing, data analysis, Artificial Intelligence, and many more.

- Easy to learn python has a simple syntax and is relevant and easy to learn, making it a great language for beginners and experienced developers alike.
- Python is a high-level language which means it abstracts away many low-level details, allowing developers to focus on logic of the program without worrying about memory management and other details.
- Python has a large and active community which means there are many resources available for learning and troubleshooting.
- Python is an object-oriented language, meaning it organizes code into objects that contain data and functions that operate on that data.
- Then has a vast collection of libraries and Frameworks that make it easy to perform various tasks.
- Python can run on multiple operating systems, including Windows, macOS and Linux.
- Python Is the open-source language which means it is free to use and distribute.

Additional Packages

NumPy:

NumPy (Numerical Python) is a library for Python programming that adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. NumPy was originally created by Jim Hugunin in 2005, with contributions for several other developers.

- NumPy allows you to create and manipulate large arrays and matrices with ease.
- NumPy provides a wide range of mathematical functions to operate on arrays, including basic arithmetic operation, trigonometric functions and statistical functions.
- NumPy allows you to perform operations on an entire array at once, making it much faster than working with individual elements.
- NumPy is open-source software, which means that it is free to use and distribute, and has many contributors to its development.

Pandas:

Pandas is a software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series data.

- The key data structure in Python is Data Frame, which allows you to store and manipulate tabular data in rows of observation and columns of variables.
- Pandas is built on top of NumPy which means that it inherits many of NumPy features and capabilities,
- Pandas is free software, released under the 3 Clause BSD license.
- Pandas is designed to be fast and powerful, making it an ideal tool for data analysis and manipulation.

Matplotlib:

Matplotlib is a popular Python plotting library that provides a comprehensive set of tools for creating high-quality 2D and 3D plots, graphs and charts. It is widely used in Scientific computing, data analysis and machine learning for visualizing data and results.

- Matplotlib has a simple and intuitive API to make it easier to create plots, graphs and charts.
- Matplotlib provides a wide range of options for customizing plot appearance, including colors, fonts, labels and many more.
- Matplotlib supports a variety of plot types including line plots, pie charts, scatter plots, bar charts, histograms and many more.

Seaborn

Seaborn is a python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistics graphics.

- Seaborn is built on the top of matplotlib, which means that it inherits many of matplotlib features and capabilities.
- The Seaborn plotting function operates on Data Frames and arrays containing whole data sets, and performs the necessary same mappings and statistics to produce informative plots.
- Seaborn aims to make visualization a central part of exploring and understanding data.

4.2 Details/ Description of hardware and software used:

4.2.1 Hardware Requirements

- **RAM:** min 4GB or more
- **Processor:** i3 processor system or higher
- **Hard Disk:** Min 20 GB free space

4GB RAM:

The project requires a minimum of 4GB RAM to run efficiently. This will ensure that the system can handle the demands of the project's software and data processing. With 4GB of RAM, the system will be able to multitask and run multiple applications simultaneously without experiencing performance issues.

Having at least 4GB of RAM is required to run Jupyter and Python packages like NumPy, pandas, matplotlib and seaborn, which can consume significant memory resources, especially when working with large datasets.

Processor i3:

The project requires a minimum of an i3 processor system to run efficiently. This will ensure that the system can handle the demands of project's software and data processing. An i3 processor system provides a good balance between performance and power consumption.

20GB Hard Disk memory:

The project requires a minimum of 20GB of free space on the hard disk to store the project file data set, dashboard and install packages of Python. This will ensure that there is enough storage for the project requirements.

4.2.2 Software Requirements:

- **Dataset:** file_name.csv
- **IDE:** Jupyter
- **Programming Language:** Python
 - **Libraries:** Pandas, NumPy, Matplotlib, and Seaborn.
- **Data Visualization Tools:** Power BI

Dataset in .csv format

The dataset is the collection of data that we are working with in Python. In this case the dataset is assumed to be a CSV (comma separated values) file, which is a common format for tabular data.

The CSV file contains rows and columns, where each row represents a single observation or record and each column represents a variable of features. The CSV file will be easily imported into Python using the pandas library.

IDE: Jupyter

A **Jupyter** is a popular open-source Integrated Development Environment (IDE) that provides a web-based interactive environment for working with Python and other programming languages. It's widely used in data science, data analytics, scientific computing and education.

- **Notebooks:** Jupyter's core feature is the notebook, a web-based document that allows users to create and share documents that contain live code, visualization and navigate text.
- **Package Management:** Jupyter provides easy package management, allowing users to install and manage Python packages and libraries with ease.
- **Collaborative:** Jupiter's real time collaboration features enables teams to work together on data analysis and visualization projects, improving communication and reducing errors.

LinkedIn Job Posting Analysis

- **Interactive data analysis:** Jupyter's interactive environment and support for NumPy, pandas, matplotlib and seaborn enables users to perform interactive data analysis and visualization.

Python:

In this project, Python is the chosen programming language for data analysis, manipulation, and visualization.

Python is developed by Guido van Rossum. Guido van Rossum Started implementing Python in 1989, Python has become one of the most popular programming languages in the world.

- **High level:** Python is a high level programming language, meaning it abstract away many low-level details, allowing developers to focus on the logic of the problem.
- **Interpreted:** Python code is interpreted line by line, rather than being compiled all at once.
- **Easy to learn:** Python has simple syntax and is relatively easy to learn, making it a great language for beginners.
- **Cross platform:** Python can run on multiple operating systems including Linux, Windows and MacOS.
- **Open source:** Python is open-source language, which means it is free to use, modify and distribute.

Libraries: Pandas, NumPy, Matplotlib, and Seaborn.

1. **Pandas** provides data structures and functions for efficient data manipulation and analyzing
2. **NumPy** offers support for large, multi-dimensional arrays and matrices, and is the foundation of most scientific computing in Python.

LinkedIn Job Posting Analysis

3. **Matplotlib** is a popular plotting library for creating static, animated and interactive visualizations.
4. **Seaborn is a visualization** library based on matplotlib that provides a high-level interface for creating informative and attractive statistical graphics.

Power BI:

Power BI is a business analytics service by Microsoft that enables users to create interactive visualization and business intelligence reports. It's used for:

- Data visualization
- Business analytics
- Reporting and dashboard

It provides a wide range of data visualization tools including:

- **Charts:** column chart, bar chart line chart pie chart and more to help users understand and compare data.
- **Tables:** Interactive tables that allow users to sort, filter, drill down into data.
- **Dashboard:** customizable dashboard that allows users to combine multiple visualizations and KPIS into a single view.
- **Report** interactive report that enables users to create and share insights with others.

5. System Evaluation

5.1 Dashboard

Dashboards are visual representations of data that provide a snapshot of key metrics and insights. They are designed to be easily understood and consumed, often used to track performance, identify trends, and make data-driven decisions.

Power BI is a business intelligence (BI) tool developed by Microsoft. It allows users to connect to various data sources, transform and clean data, and create interactive visualizations and dashboards.

- **Power BI Service:** While Power BI Desktop is a standalone application, you can also sign up for a Power BI Service account to share your reports and collaborate with others.
- **Online Learning:** Microsoft offers a variety of online resources and tutorials to help you learn how to use Power BI. Check out the Power BI documentation and training materials on the Microsoft website.
- **Dashboard creation:** Design custom dashboards to present key metrics and insights in a visually appealing way.
- **Collaboration and sharing:** Share dashboards with colleagues and stakeholders, enabling data-driven decision-making across the organization.
- **Community:** There's a thriving Power BI community where you can ask questions, share tips, and learn from others. Visit the Power BI forums or join online communities to connect with fellow users.

5.1.1 Job seeker

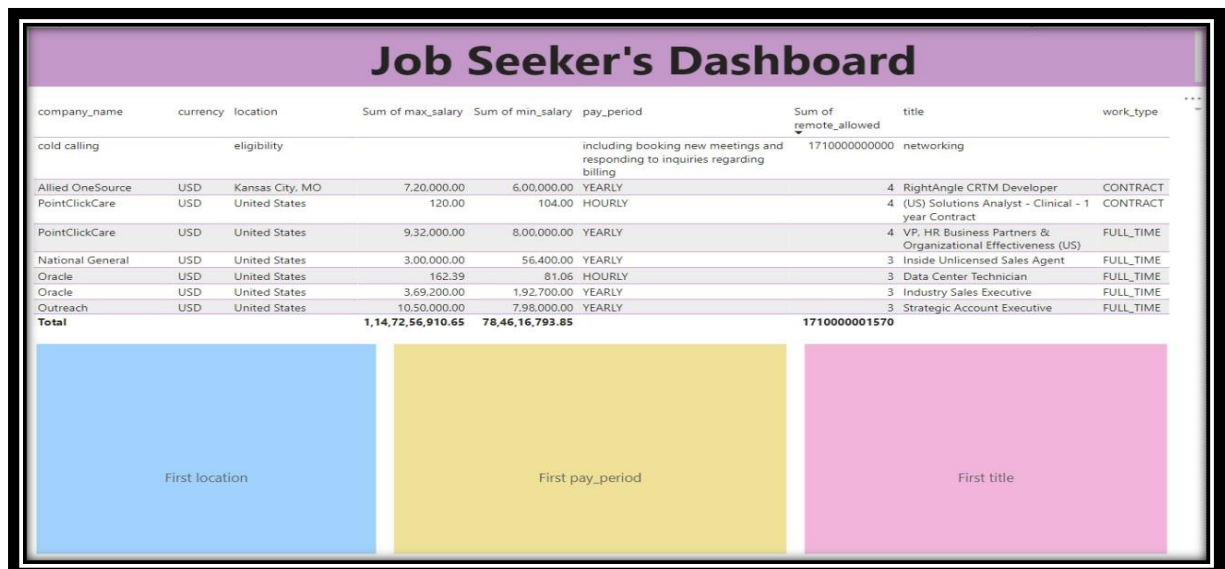
- **Job Postings:** The dashboard lists 7 job titles, suggesting a focus on various roles within the company.

LinkedIn Job Posting Analysis

- **Applications:** There are a significant number of applications for the "Senior Scientist" and "Software Engineer" roles, indicating high demand for these positions.
- **Views:** The dashboard displays a high number of views (1710000000000), potentially signifying strong online presence and job postings reaching a wide audience.

Key Insights:

- **High Demand Roles:** "Senior Scientist" and "Software Engineer" are in high demand, attracting a considerable number of applicants. The recruiting team might need to expedite the hiring process for these roles.
- **Geographic Focus:** San Diego, CA, stands out as the location with the most applications, suggesting a successful recruitin strategy or a strong existing presence in the area. The team could explore further opportunities or initiatives in San Diego.
- **Effective Marketing:** The high number of views indicates strong online visibility and effective marketing efforts, leading to widespread reach. This could be attributed to well-structured job postings or targeted advertising campaigns.



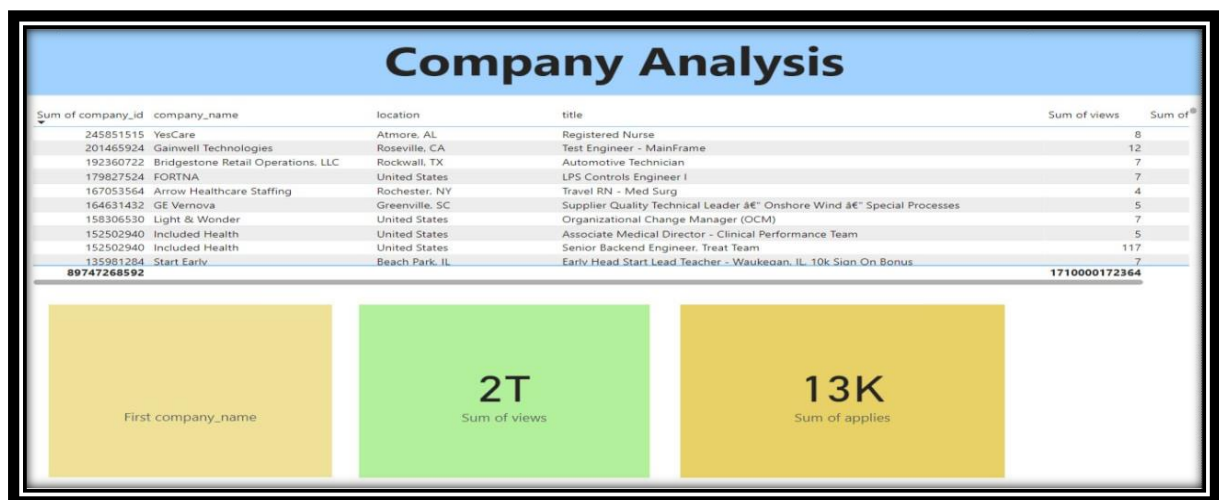
LinkedIn Job Posting Analysis

5.1.2 Company

- Sum of Job IDs: This number represents the total number of job postings.
- Sum of Company IDs: This number reflects the total number of companies involved in the recruiting process.
- Sum of Applies: This number indicates the total number of applications received for all job postings.
- Sum of Views: This number represents the total number of times job postings were viewed by candidates.

Key Insights:

- Most Popular Job Titles: The chart "Sum of applies by title" shows the job titles with the highest number of applications. This indicates which positions are attracting the most interest from candidates.
- Top Recruiting Locations: The chart "Count of company_name by location and company_name" highlights the geographic locations with the most active recruiting. This provides insights into where the company is focused on finding talent.
- Company Activity: The table listing company ID, company name, and calls reveals the level of engagement from each company in the recruiting process.



LinkedIn Job Posting Analysis

5.1.3 Recruiters:

- **Sum of company_id:** This likely represents the number of companies the recruiting team is working with.
- **Sum of job_id:** This likely represents the number of job postings the team is managing.
- **Sum of views:** This likely represents the number of times job postings were viewed.
- **Sum of applies:** This likely represents the total number of applications received.

Key Insights:

- The team is working with a large number of companies and has a high volume of job postings.
- The team's efforts are focused primarily on San Diego, CA, followed by Denver, CO and Los Angeles, CA.



6. User Operational Manual

A User Operational Manual for a LinkedIn job posting analysis project typically outlines the steps and guidelines for users to effectively analyze and extract insights from LinkedIn job postings.

The manual would likely cover the following aspects:

1. **Job Posting Analysis Objectives:** Clearly define the objectives of the analysis, such as identifying job trends, understanding job requirements, or analyzing company hiring strategies.
2. **Data Analysis:** Describe the methods and techniques for analyzing the job posting data, such as identifying keywords, analyzing job descriptions, and visualizing data trends.
3. **Insight Generation:** Provide guidelines on how to generate insights from the analyzed data, including identifying job trends, understanding job requirements, and analyzing company hiring strategies. Identify job trends by analyzing the frequency of job postings over time. Understand job requirements by analyzing the skills and qualifications mentioned in job descriptions. Analyze company hiring strategies by examining the types of jobs they post and the locations they operate in.

6.1 List of Abbreviations

Abbreviations:

CSV- Comma separated file

EDA- Exploratory Data Analysis

IDE- Integrated Development Environment

BI- Business Intelligence

6.2 References

- **Dataset Source**

<https://www.kaggle.com/datasets/arshkon/linkedin-job-postings>

- **Data Analysis**

<https://www.geeksforgeeks.org/what-is-data-analytics/>

- **EDA**

<https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/>

- **Book**

- **Title:** Informatics Practices with Python
- **Author:** Preeti Aror
- **Publisher:** Sultan chand

6.3 Conclusion

The LinkedIn job posting analysis project aimed to identify job trends, understand job requirements, and analyze company hiring strategies in the operations industry. Through a systematic approach, we collected, cleaned, and analyzed a large dataset of job postings from LinkedIn, and generated insights that can inform business decisions and talent acquisition strategies. The LinkedIn job posting analysis project provided valuable insights into the operations industry, highlighting job trends, job requirements, and company hiring strategies. The project demonstrated the potential of job posting analysis to inform business decisions and talent acquisition strategies, and contributed to the development of a systematic approach to analyzing job posting data. Future research should aim to build upon these findings, addressing the limitations and exploring new directions to further advance the field of job posting analysis.