Analysis of LinkedIn job postings (2023-2024)

Introduction:

In today's dynamic job market, understanding trends in job postings can provide valuable insights for job seekers, employers and policymakers. LinkedIn as one of the largest professional networking platform, offers a rich source of data on job postings across various industries and regions. This project aims to analyze LinkedIn job postings to uncover trends, identify high demand skills, and provide actionable insights for stakeholders.

Problem statement:

This project aims to address the following key questions using data from LinkedIn job postings:

1- **What are the most frequently posted job titles on LinkedIn?**
Understanding which job titles are most commonly posted can help job seekers Identify in demand roles and tailor their job search strategies accordingly. Employers can use this information to benchmark their job postings against industry trends.

2- **What are the job titles with the highest sum of application submitted?**
Identify high demand job titles determine which job titles receive the most applications, indicating high interest and competition among job seekers. Understanding application trends analyze the factors contributing to high application rates for specific job titles, such as industry, location, and required skills.

3- **What experience levels are the most in demand in LinkedIn job postings?**
Identifying the experience levels most sought after by employers can guide individuals in their career development and help educational institutions align their programs with market needs. It also assists employers in setting realistic requirement for job openings to attract suitable candidates.

Data description:

Everyday thousands of companies and individuals turn to LinkedIn in search of talent. This dataset contains a nearly comprehensive record of **124000+ job postings** listed in 2023 and 2024 each individual posting contains dozens of valuable attributes for both postings and companies.

Files:

- **job_id:** the job id defined by LinkedIn
- **company_id:** identifier for the company associated with the job posting

- **title:** job title
- **description:** job description
- **max_salary:** maximum salary
- **med_salary:** median salary
- **min_salary:** minimum salary
- **pay_period:** pay period for salary (hourly, monthly, yearly)
- **formatted_work_type:** type of work (full time, part time, contract)
- **location:** job location
- **applies:** number of applications that have been submitted
- **original_listed_time:** original time the job was listed
- **remote_allowed**: whether job permits remote work
- **views:** number of times the job posting has been viewed
- **job_posting _url:** URL to the job posting on the platform
- **application_url**: URL where application can be submitted
- **application_type**: type of application process (offsite, onsite, etc.)
- **expiry:** expiration data or time for the job listed
- **closed_date:** time to close job listed
- **formatted_experience_level:** job experience level (entry, associate, etc.)
- **skills_desc:** description detailing required skills for job
- **listed_time:** time when the job was listed
- **posting_domain:** domain of the website with application
- **sponsored:** whether the job listed is sponsored or promoted
- **work_type:** type of work associated with the job
- **currency:** currency in which the salary is provided
- **compensation_type:** type of compensation for the job

data cleaning and processing:

for the data cleaning and processing of LinkedIn job postings, I utilized python, leveraging its powerful data manipulation libraries. Here is step by step outline of the process**:**

1. **loading the data:**
   I used the **raed_csv ()** function from the **pandas** library to load the dataset into a data frame (df).

```
>>> import pandas as pd
>>> df = pd.read_csv("linkedin job posting.csv")
```

2. **initial data exploration:**

to get an overview of the data, I used the **head (5)** function to display the first five rows.

```
>>> df.head(5)
    job_id          company_name                                      title  ... work_type currency compensation_type
0   921716   Corcoran Sawyer Smith                        Marketing Coordinator  ...  FULL_TIME      USD       BASE_SALARY
1  1829192                    NaN                Mental Health Therapist/Counselor  ...  FULL_TIME      USD       BASE_SALARY
2 10998357  The National Exemplar                     Assitant Restaurant Manager  ...  FULL_TIME      USD       BASE_SALARY
3 23221523  Abrams Fensterman, LLP  Senior Elder Law / Trusts and Estates Associat...  ...  FULL_TIME      USD       BASE_SALARY
4 35982263                    NaN                              Service Technician  ...  FULL_TIME      USD       BASE_SALARY
```

to gain more details about the dataset, such as column names, data types and the counts of rows and columns, I utilized the **info ()** function.

```
>>> df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 123849 entries, 0 to 123848
Data columns (total 28 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   job_id                      123849 non-null  int64
 1   company_name                122130 non-null  object
 2   title                       123849 non-null  object
 3   description                 123842 non-null  object
 4   max_salary                  29793 non-null   float64
 5   pay_period                  36073 non-null   object
 6   location                    123849 non-null  object
 7   company_id                  122132 non-null  float64
 8   views                       122160 non-null  float64
 9   med_salary                  6280 non-null    float64
 10  min_salary                  29793 non-null   float64
 11  formatted_work_type         123849 non-null  object
 12  applies                     23320 non-null   float64
 13  original_listed_time        123849 non-null  float64
 14  remote_allowed              15246 non-null   float64
 15  job_posting_url             123849 non-null  object
 16  application_url             87184 non-null   object
 17  application_type            123849 non-null  object
 18  expiry                      123849 non-null  float64
 19  closed_time                 1073 non-null    float64
 20  formatted_experience_level  94440 non-null   object
 21  skills_desc                 2439 non-null    object
 22  listed_time                 123849 non-null  float64
 23  posting_domain              83881 non-null   object
 24  sponsored                   123849 non-null  int64
 25  work_type                   123849 non-null  object
 26  currency                    36073 non-null   object
 27  compensation_type           36073 non-null   object
dtypes: float64(11), int64(2), object(15)
```

To remove of any duplicates records I used drop_duplicates () function

```
>>> df.drop_duplicates()
           job_id                  company_name                                      title  ... work_type currency compensation_type
0          921716          Corcoran Sawyer Smith                          Marketing Coordinator  ...  FULL_TIME      USD       BASE_SALARY
1         1829192                           NaN              Mental Health Therapist/Counselor  ...  FULL_TIME      USD       BASE_SALARY
2        10998357       The National Exemplar                   Assitant Restaurant Manager  ...  FULL_TIME      USD       BASE_SALARY
3        23221523       Abrams Fensterman, LLP  Senior Elder Law / Trusts and Estates Associat...  ...  FULL_TIME      USD       BASE_SALARY
4        35982263                           NaN                            Service Technician  ...  FULL_TIME      USD       BASE_SALARY
...           ...                           ...                                           ...  ...        ...      ...               ...
123844  3906267117                Lozano Smith              Title IX/Investigations Attorney  ...  FULL_TIME      USD       BASE_SALARY
123845  3906267126                   Pinterest    Staff Software Engineer, ML Serving Platform  ...  FULL_TIME      NaN               NaN
123846  3906267131                EPS Learning           Account Executive, Oregon/Washington  ...  FULL_TIME      NaN               NaN
123847  3906267195  Trelleborg Applied Technologies                   Business Development Manager  ...  FULL_TIME      NaN               NaN
123848  3906267224                   Solugenix              Marketing Social Media Specialist  ...  FULL_TIME      USD       BASE_SALARY

[123849 rows x 28 columns]
```

3. **Identifying missing values:**
   I employed the **isnull (). sum ()** function to determine the number of missing values in each column.

```
>>> df.isnull().sum()
job_id                        0
company_name               1719
title                         0
description                   7
max_salary                94056
pay_period                87776
location                      0
company_id                 1717
views                      1689
med_salary               117569
min_salary                94056
formatted_work_type           0
applies                  100529
original_listed_time          0
remote_allowed           108603
job_posting_url               0
application_url           36665
application_type              0
expiry                        0
closed_time              122776
formatted_experience_level 29409
skills_desc              121410
listed_time                   0
posting_domain            39968
sponsored                     0
work_type                     0
currency                  87776
compensation_type         87776
dtype: int64
```

4. **Handling missing values:**
   In dealing with missing values, there are several options: dropping the rows, replacing the missing values or leaving them as is. For this project, I choose to leave the missing values intact. This decision was made to avoid inadvertently affecting other columns or compromising the credibility of the analysis.

```
>>> # leaving the missing values without making any chaning in the data frame
```

This approach ensures that the data remains as true to the original source as possible, maintaining the integrity of the analysis while allowing for comprehensive exploration and insights.

Analysis and Modeling:

In this section, I performed various analysis to answer the key questions of the project. I utilized python for data manipulation and analysis and power bi for visualization. Here are the steps and method used:

1. **Analyzing the most posted job titles:**
   To identify the most frequently posted job titles, I used the value_counts () function along with to_frame () to convert the result into a data frame.

```
>>> #finding the top5 jobs title posting in linkedin
>>> top_jobs = df[['title']].value_counts().to_frame()
>>> top_jobs.head(5)
                                count
title
Sales Manager                     673
Customer Service Representative   373
Project Manager                   354
Administrative Assistant          254
Senior Accountant                 238
```

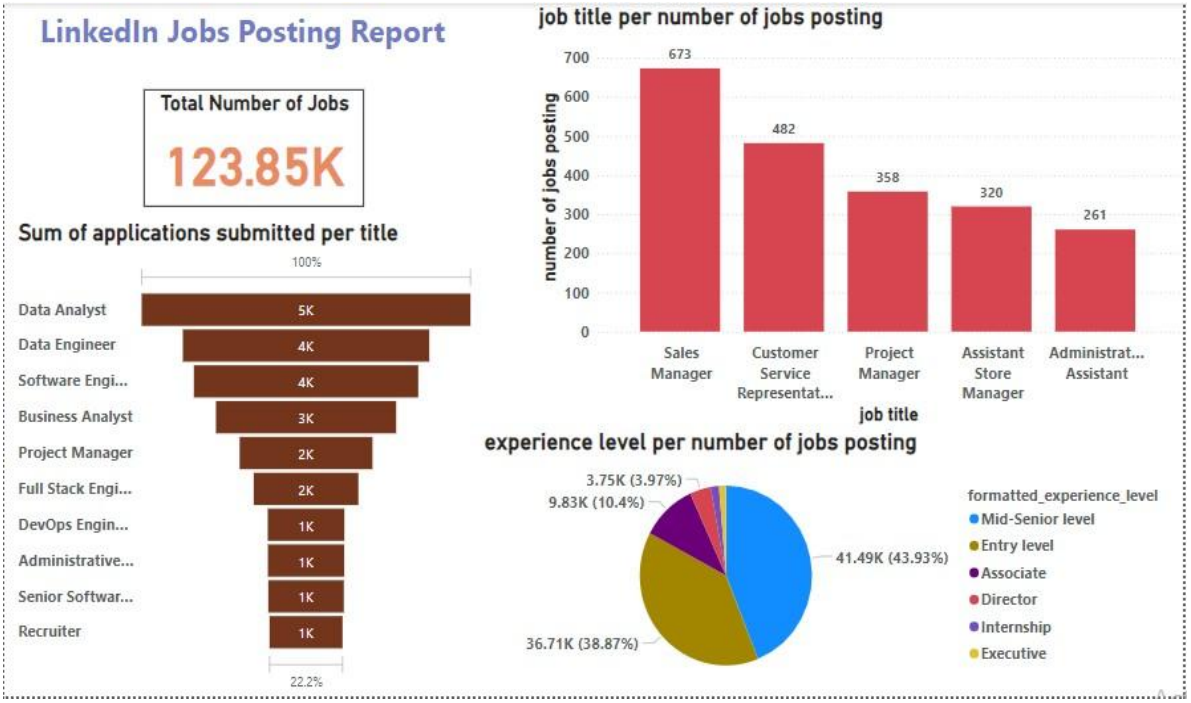
2. **Analyzing job titles with the highest sum of applications submitted:**
   To find the job titles with the highest sum of applications submitted, I used the groupby () function to group the data by the job title. Then I applied the sum () function to aggregate the applications and sort_values () to order them in descending order.

```
>>> # finding the job title with high number of applications submitted
>>> num_applies = df[['title','applies']]
>>> num_applies = num_applies.groupby(['title'],as_index = False).sum()
>>> num_applies = num_applies.sort_values(by = "applies", ascending = False)
>>> num_applies.head(5)
                  title  applies
15743       Data Analyst   5163.0
15898      Data Engineer   3867.0
61529  Software Engineer   3496.0
7974    Business Analyst   2768.0
47243    Project Manager   2083.0
```

3. **Analyzing the most in-demand experience levels:**
   To determine the most demanded experience levels in job postings, I used the value_counts () function again with to_frame ().

```
>>> #finding the experience level per number of jobs
>>> exp_level = df["formatted_experience_level"].value_counts().to_frame()
>>> exp_level
                           count
formatted_experience_level
Mid-Senior level           41489
Entry level                36708
Associate                   9826
Director                    3746
Internship                  1449
Executive                   1222
```

4. **Visualization with power bi:**
   After performing the analysis in python, I used power bi to create visualization that effectively communicate the results. Power bi's interactive and dynamic visualization capabilities helped in presenting the data insights clearly and engagingly.



Results:

The analysis of LinkedIn job postings revealed the following key insights:

1. **Most posted job titles:**
   - Sales manager: 673 postings

- Customer service representative: 482 postings
- Project manager: 358 postings
- Assistant store manager: 320 postings
- Administrative assistant: 261 postings

2. **Job titles with highest application submitted:**
    - Data analyst: 5K applications
    - Data engineer: 4K applications
    - Software engineer: 4K applications
    - Business analyst: 3K applications
    - Project manager: 2K applications

3. **Experience level demand:**
    - Mid-senior level: 41.49K postings (43.93%)
    - Entry level: 36.71K postings (38.87%)
    - Associate: 9.83K postings (10.4%)

Conclusion:

The analysis of LinkedIn job postings provided valuable insights into the current job market dynamics. Key finding include:

- **High demand for the management and customer- facing roles:**
  Sales manager and customer service representative are among the most frequently posted job titles, including a strong demand for professional in these areas.
- **Significant interest in technical and analytical positions:**
  Roles such as data analyst, data engineer and software engineer attract the highest number of applications, reflecting a competitive landscape for these positions.
- **Experience level preference:**
  Mid- senior level positions are the most sought-after, suggesting that employers are looking for experienced professionals to fill critical roles.

These insights can help job seekers tailor their job search strategies and career development plans to align with market demands. Employers can use this information to optimize their recruitment efforts and attract the right talent. Overall the analysis underscores the important of technical skills and experience in current job market, providing actionable intelligence for all stakeholders involved.

Appendix:

Here is link for the dataset: https://www.kaggle.com/datasets/arshkon/linkedin-job-postings.

Data Analyst: Saeed Yassin

Phone: 0096899379304

Address: Muscat- Oman