Project report

Project Introduction:

When people seeks a job they often search in the jobs in LinkedIn, but we noticed that when a new position becomes available, employees are often tasked with finding potential candidates who meet the job requirements, So we decided to help people find the job and in the same time help employees to increase their chances of earning bonuses, If a candidate recommended by an employee is hired, the referring employee typically receives a bonus as recognition for their contribution, employees often share job postings on LinkedIn in search of suitable matches so we want to develop AI-driven solutions that analyze these posts to identify potential matches. By automating the matching process, we aim to facilitate smoother and more effective job discovery for both job seekers and matching candidates for jobs. Moreover, in our project one of the significant advantages of the matching process is that the person who finds the job post suitable for him can send his cv to the employee who shared the post who can apply him for the job and this increases his chances to be accepted for the job.

Data Collection and Integration:

We used the following columns from the profiles table: position, posts. From the profiles dataset we have used the links of the posts to web scrape LinkedIn in order to bring the content of the posts of the profiles and we have used the title of the post to identify relevant job posts. In addition, we used the position column to extract the skills.

We planned to collect post contents and skills but we couldn't gather skills due to some constraints we will explain later on. The additional data we have collected are the posts content we have collected 3344 posts content. In order to gather the posts content, we used the scraping browser from bright data to scrape the relevant posts. We got the links for the posts from the profiles data that we already have, we used the proxy provided by bright data to scrape the posts.

In order to incorporate the data in our project we have created a new table with the following columns: title, post content, link.

The additional data we will use it for the matching between posts and profiles later in the analysis we will explain how.

The item is a post content, and the enrichment size is 3344x1 (rows x columns) the α(group) is 2.

Data Analysis:

- **Analysis Techniques:**

  Before scraping the posts, we filtered the posts using the title, posts with a title that contains one of the keywords that we defined in our notebook for example (position, job, hire…) is considered relevant, then we did the web scraping on the relevant posts.

  After scraping the posts content, we used a LLM (google Gemini) to identify which posts are job posts based on the answer of the LLM we filtered the relevant posts and we kept only posts that Gemini considered as job posts and answered on them with yes, then we have used the final job posts to identify the open positions for each post we asked Gemini what positions it contains and what are 7 skills for each position, in this stage also we did an addition filtering because Gemini sometimes returned less than 7 skills or return null as answer. Finally, with the filtered job posts we created a posts table with the following columns: link, post content, positions skills, to save the results.

Note that: positions skills are a dictionary the keys are the positions and the values are lists containing the skills for each position.

For our machine learning model, the training data are the profiles that their position is not Retired because there is no point on offering a job for a retired person and the position is not null (those with position of length less than two we consider as null).Then we extract the skills using Chat GPT on the position column for each user we have extracted 7 skills (since we assume that the current position of the user reflects is based on the skills of the user so we can extract skills using GPT).Finally, we sample from the filtered profiles a sample of size 4000.

Note: We did not use test data since the test data should be users that do not have position so we cannot extract skills for the user to use it in our machine learning algorithm and predict relevant posts for the users, so we decided to sample 115 rows from the training data to use it as a validation dataset to show our algorithm performance on it.

In addition, for tagging the user's positions as we will explain in the AI Methodologies for choosing the appropriate K in the Kmeans we have used two approaches:

(1) Since we are working with positions of users that are text we used Bert embedding that returns a vector representation of length 256 so to choose the right number of clusters we used PCA and reduced the dimensionality to 2 since it facilitates visualization and interpretation of patterns in the data, the goal of the visualization is to identify natural groupings or clusters within the dataset in **plot (2)** in the appendix as we can see there is no clear groupings or clusters and the proportion of variance that is explained by the first two principal components is very low approximately 50% thus using PCA didn't asses in choosing the value of K.

(2) The second approach: we used a bar plot representing the unique users positions in the x-axis and the number of users with each position in the y-axis and we sorted the positions count in descending order and displayed the top 100 since the size of the training set is under 4000 there is no need to cluster the data to more than 100 clusters since the average number of positions in each cluster will be less than 4 .In addition ,we saw that the number of users with positions that are in the lowest 30 are very low from 1-2 users per position thus we settled with K=70, , as we can see in **plot (1)** that represents the top 70 positions.

Furthermore, for the Kmeans that we used to cluster positions of the job posts that we will explain later on to choose the value of K we have used similar approach to what we have explained above and we concluded the following:

(1) Looking at the PCA result in **plot (3)** we don't see clear groupings or clusters thus as we can see there is no clear groupings or clusters and the proportion of variance that is explained by the first two

principal components is very low approximately 50% thus using PCA didn’t asses in choosing the value of K.

(2) The second approach: We used a bar plot representing the distinct positions in the job posts as the x-axis and as the y-axis the number of posts that offers the position and we sorted them in descending order based on the frequency of each position, looking at the total number of positions which is less than 4000 we displayed only the top 100 position since more than 100 cluster will lead to a very low number of positions in some clusters less than 4 , thus from looking at the plot we saw that for the lowest 50 positions based on frequency there is less than two job posts that offers the position thus we concluded that K=50 is sufficient for our data.

- **Feature Selection:**
  To achieve our project goal, we had to select the features that significantly influence the job matching process, among various potential features we considered the "skills" and "position" as the primary attributes, for the matching process due to their direct relevance to job suitability and candidate compatibility.
  although we have considered using the location of the user as a feature but due to that almost all the job posts did not mention the location so we did not have access to the location of the open position to integrate it as a feature in the matching process, moreover most of the job posts only mentioned the job title so each feature that could not be extracted based on the job title we cannot integrate it in our data.

  Skills: This feature represents the knowledge areas of both job seekers and position requirements. It serves as a direct indicator of whether a candidate's abilities align with the position requirements.

  Position: we considered the "position" feature of the user as the job title or role being sought for the user, in addition positions with similar skills to the user's position can be considered relevant for the user. It acts as a variable that helps in narrowing down the search and match criteria to relevant fields and specializations.

AI Methodologies:

- **Tagging the users training data:**
  we wanted to tag the users in our training data based on their skills that we collect them by their position and since there are a lot of different positions in the training data, we considered positions that differs in their rank (senior , junior, …) to be the same position and after that we collected the skills based on their position using Gemini, our idea is to label the users so we get that similar users (we consider two users similar if they have similar position and skills) belongs to that same cluster and have the same label, we did that using Kmeans that is based on the embedded list of the skills (embedded with bert sentence)

  , we decided to do the clustering with K=70.
- **Clustering job posts:**
  We have 1676 job posts and each post may contain more than one offered position so each post will be associated with more than one position. First of all, to represent each position we accumulated all the 7 skills that we have

extracted in one string since each position is defined by his skills, then we used bert sentence to embed the representation.
Afterwards we decided to use Kmeans on the positions that are offered in the posts based on the embedding of the position, and we used Euclidian distance to calculate distances between the embeddings and with K= 50.
Such that all the posts that included positions that are in the same cluster will be considered as a relevant post for the whole cluster this means that when we will suggest a post on a user, we will suggest all the posts associated with the positions that are in the cluster that we predicted for the user's position.

- **Classifying user's positions:**
  We want to predict a position for users that isn't currently working based on his skills, and by using the predicted user position we will suggest for him a relevant job posts, we will explain later how we will do this suggestion.
- **We decided to use KNN to predict the user's position:**
  we accumulated the 7 skills of a user in one sentence and embedded it using bert sentence so each user is represented by their skills, then we used the KNN model with k = 7 and the Euclidian distance metric, the input is a skills representation of a user and the prediction is a position label.
- **Job posts matching:**
  For a given user we predict the position label by using KNN model that we described above and then we represented the skills for each user as an aggregation (mean) between his skills and the "similar" users, we consider two users similar if they have the same prediction label (KNN prediction). After that we used the aggregated vector for each user to get the closest centroid that we got from the Kmeans on the posts, each post in this cluster is suggested for the user.

Evaluation and Results:

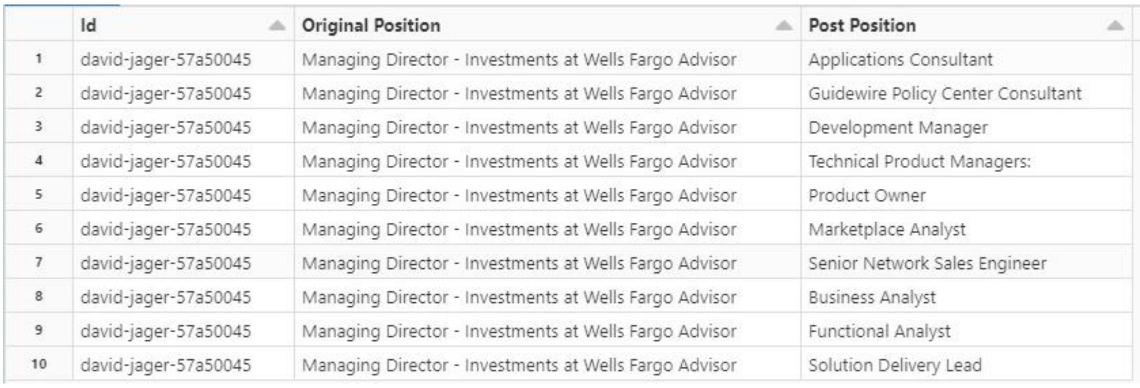
- **Evaluation Process:**
  Our project and algorithm evaluation process were designed to assess the efficiency and accuracy of our AI-driven solution in matching job seekers with relevant job posts on LinkedIn.

  Project evaluation:
  The overall evaluation was aimed to understand the impact of our solution in enhancing the job matching process. We looked at the number of successful matches made, we used LLM (gbt-3.5) to calculate the accuracy we asked him if the user have this specific position and I'm suggesting for him this position, will he fit the position or not? (Provide only yes or no answers).
  So, after that we calculated the accuracy based on the yes, no answers, for each user there are at most 5 suggested posts, if user have at least one yes, we consider that we did successful matching for this user.
  The accuracy that we got: **0.773**

  Algorithm evaluation:
  We evaluate our algorithm by calculating the f1 score for the KNN part , the f1 that we got: **0.834**
  We only evaluate the KNN part because the other machine learning algorithms we used (Kmeans) are unsupervised learning.

- **Results and Key Findings**
  1. We developed algorithms to analyze user's profiles and job posts for relevant matchings.
  2. We succeeded in matching job posts to users as we can see we got a high accuracy on the validation data.
  3. We Enhanced the overall user experience on LinkedIn by facilitating efficient job discovery and application since now in addition to jobs suggested by LinkedIn, we can suggest also relevant job posts.

- **Image of the results:**



	Id	Original Position	Post Position
1	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Applications Consultant
2	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Guidewire Policy Center Consultant
3	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Development Manager
4	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Technical Product Managers:
5	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Product Owner
6	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Marketplace Analyst
7	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Senior Network Sales Engineer
8	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Business Analyst
9	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Functional Analyst
10	david-jager-57a50045	Managing Director - Investments at Wells Fargo Advisor	Solution Delivery Lead

As we can see in the table for the given user who is a managing director our job posts matching algorithm matched for him posts that offers similar positions to his original positions which indicate that our algorithm has achieved our project goal

<u>**Note:**</u>
The table containing the full results of our matching prosses has been uploaded on the drive, link provided in the references section below.

<u>Limitations and Reflection:</u>

Data collection:

1. **Limitation:** We couldn't collect the skills from LinkedIn profiles using web-scraping because it's a private information and we had to login to be able to web-scrape data but bright data doesn't allow this.
   So, we had to generate our data, we collected the users' skills using Gemini based on the position of the user, we asked him to give us 7 skills that are required for this specific position.
   Reflection: collecting the user skills using a language model takes a lot of time, not accurate since we extracted them based on the position and the position doesn't always reflect all the skills of the user, and dealing with data that has been generated based on language model isn't easy, we had to do another preprocessing to the answers that we got from it.

2. **Limitation:** We couldn't collect the qualification for the job from the post, so we also used Gemini to collect the skills that are required for the position that is offered in job post.
   Reflection: similarly, to what we have explained above.

Computational:

1. **Limitation:** we have created a distance metric to use it in the both Kmeans models that takes as an input two skills vectors each one contains 7 embeddings for 7 skills, the distance between the two vectors is calculated as follows:
   for each element in the vector, we will calculate the cosine similarity between it and all the elements in the other vector then we will calculate the Euclidian distance between this element and the element with the highest similarity score and this distance will be added to the overall distance between the two vectors.
   But we couldn't use this metric because it will that a lot of time to run (we estimated the runtime on our data to be 12 days to finish).
   Reflection: we think if we used this metric in both Kmeans models we would get a better performance because in the case of two vectors with the same skills but not the same order we should get distance equal to 0 but that's not the case with the current representation and the used metric, we also consider to do order the skills in lexicographical order but we still get the same problem there is no guarantee that similar skills will be in the same index in the vector.
2. **Limitation**: we couldn't use all the data we have due to the computational resources' limitation so, we had to sample for the data.
   Reflection: we think if we didn't have a computational resources limitation

and trained our AI-model on all the data we got better results and performance.

Conclusions:

In conclusion, our project successfully developed an AI-driven solution to streamline the job matching process on platform LinkedIn, enhancing the job search for individuals and improving referral outcomes for employees. Despite facing challenges in data collection and computational limitations, we innovated by using language models to generate necessary data and implemented machine learning models to personalize job recommendations. Our achievements demonstrate the project's potential to significantly impact job discovery and recruitment, laying the groundwork for future advancements in AI-assisted job matching.

References:

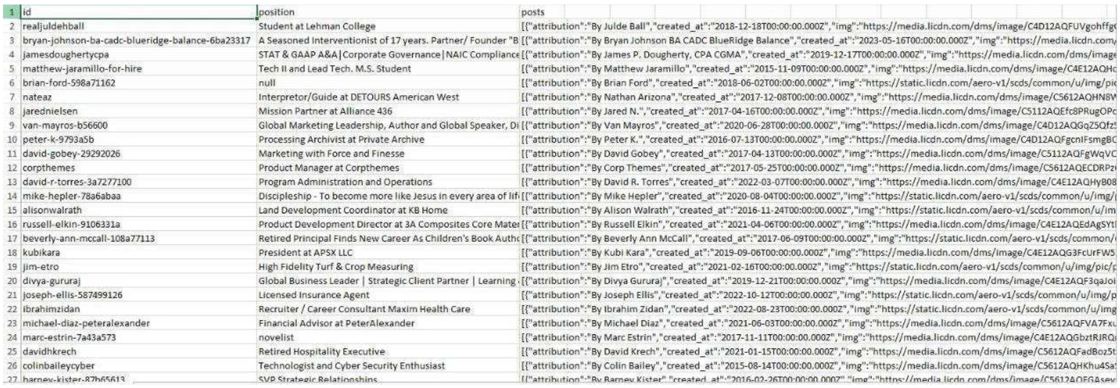
- link to our project in GitHub: **https://github.com/samarsamara/LinkedIn-Job-Matching**

- link to our collected data: **https://drive.google.com/drive/folders/13OAMDhdAgQ-oBwqrtC9uEFlfRiCb5KRx?usp=drive_link**

APPENDIX:

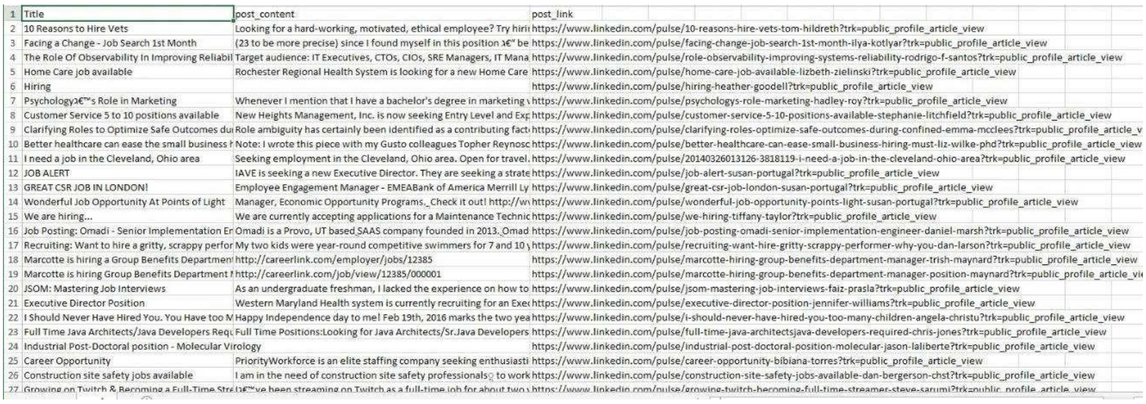
- **Data Collection and Integration relevant images and datasets:**

Users table:



1	id	position	posts
2	realjuldehball	Student at Lehman College	[{"attribution":"By Julde Ball","created_at":"2018-12-18T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4D12AQFUVgohffg
3	bryan-johnson-ba-cadc-blueridge-balance-6ba23317	A Seasoned Interventionist of 17 years. Partner/ Founder "B	[{"attribution":"By Bryan Johnson BA CADC BlueRidge Balance","created_at":"2023-05-16T00:00:00.000Z","img":"https://media.licdn.com/
4	jamesdoughertycpa	STAT & GAAP A&A|Corporate Governance|NAIC Compliance	[{"attribution":"By James P. Dougherty, CPA CGMA","created_at":"2019-12-17T00:00:00.000Z","img":"https://media.licdn.com/dms/image
5	matthew-jaramillo-for-hire	Tech II and Lead Tech. M.S. Student	[{"attribution":"By Matthew Jaramillo","created_at":"2015-11-09T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQHc
6	brian-ford-598a71162	null	[{"attribution":"By Brian Ford","created_at":"2018-06-02T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/img/pic
7	nateaz	Interpretor/Guide at DETOURS American West	[{"attribution":"By Nathan Arizona","created_at":"2017-12-08T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQHN8W
8	jarednielsen	Mission Partner at Alliance 436	[{"attribution":"By Jared N.","created_at":"2017-04-16T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5112AQEfc8PRugOPc
9	van-mayros-b56600	Global Marketing Leadership, Author and Global Speaker, Di	[{"attribution":"By Van Mayros","created_at":"2020-06-28T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4D12AQGqZ5Qfz5
10	peter-k-9793a5b	Processing Archivist at Private Archive	[{"attribution":"By Peter K.","created_at":"2016-07-13T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4D12AQFgcnIFsmgBC
11	david-gobey-29292026	Marketing with Force and Finesse	[{"attribution":"By David Gobey","created_at":"2017-04-13T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5112AQFgWqVC
12	corpthemes	Product Manager at Corpthemes	[{"attribution":"By Corp Themes","created_at":"2017-05-25T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQECDRPz
13	david-r-torres-3a7277100	Program Administration and Operations	[{"attribution":"By David R. Torres","created_at":"2022-03-07T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQHyB08
14	mike-hepler-78a6abaa	Discipleship - To become more like Jesus in every area of lif	[{"attribution":"By Mike Hepler","created_at":"2020-08-04T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/img/
15	alisonwalrath	Land Development Coordinator at KB Home	[{"attribution":"By Alison Walrath","created_at":"2016-11-24T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/im
16	russell-elkin-9106331a	Product Development Director at 3A Composites Core Mater	[{"attribution":"By Russell Elkin","created_at":"2021-04-06T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQEdAgSYt
17	beverly-ann-mccall-108a77113	Retired Principal Finds New Career As Children's Book Auth	[{"attribution":"By Beverly Ann McCall","created_at":"2017-06-09T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/
18	kubikara	President at APSX LLC	[{"attribution":"By Kubi Kara","created_at":"2019-09-06T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQG3FcUrFW5
19	jim-etro	High Fidelity Turf & Crop Measuring	[{"attribution":"By Jim Etro","created_at":"2021-02-16T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/img/pic/p
20	divya-gururaj	Global Business Leader | Strategic Client Partner | Learning	[{"attribution":"By Divya Gururaj","created_at":"2019-12-21T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQF3qaJoi
21	joseph-ellis-587499126	Licensed Insurance Agent	[{"attribution":"By Joseph Ellis","created_at":"2022-10-12T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/img/p
22	ibrahimzidan	Recruiter / Career Consultant Maxim Health Care	[{"attribution":"By Ibrahim Zidan","created_at":"2022-08-23T00:00:00.000Z","img":"https://static.licdn.com/aero-v1/scds/common/u/img
23	michael-diaz-peteralexander	Financial Advisor at PeterAlexander	[{"attribution":"By Michael Diaz","created_at":"2021-06-03T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQFVA7Fxu
24	marc-estrin-7a43a573	novelist	[{"attribution":"By Marc Estrin","created_at":"2017-11-11T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C4E12AQGbztRJRQ
25	davidhkrech	Retired Hospitality Executive	[{"attribution":"By David Krech","created_at":"2021-01-15T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQFadBoz5t
26	colinbaileycyber	Technologist and Cyber Security Enthusiast	[{"attribution":"By Colin Bailey","created_at":"2015-08-14T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQHKhu4Sa
27	barney-kister-87b65613	SVP Strategic Relationships	[{"attribution":"By Barney Kister","created_at":"2016-02-26T00:00:00.000Z","img":"https://media.licdn.com/dms/image/C5612AQEGAsev

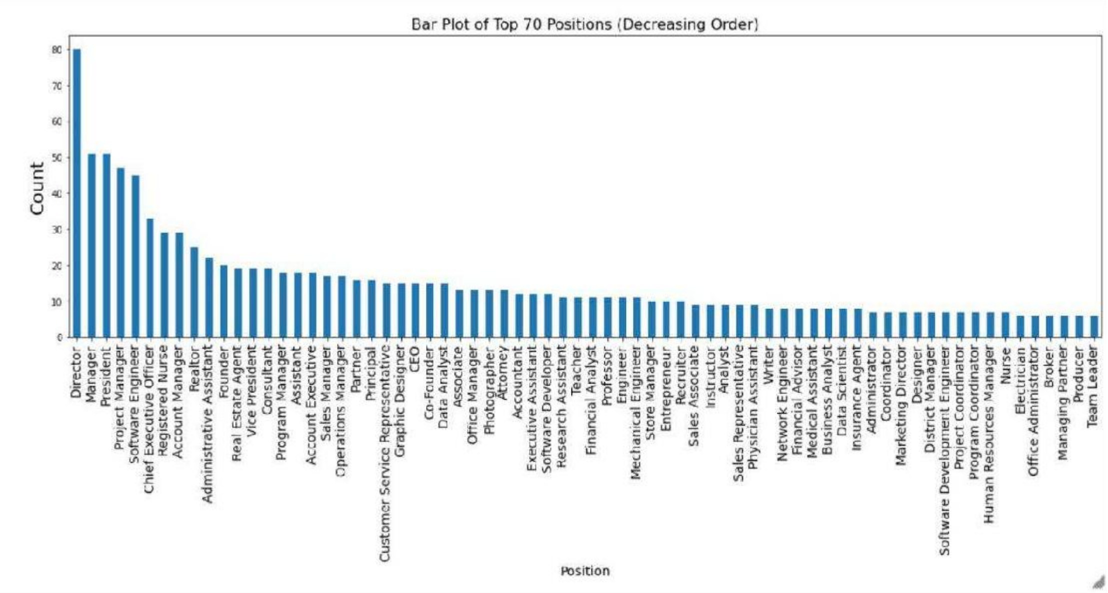
Posts table:



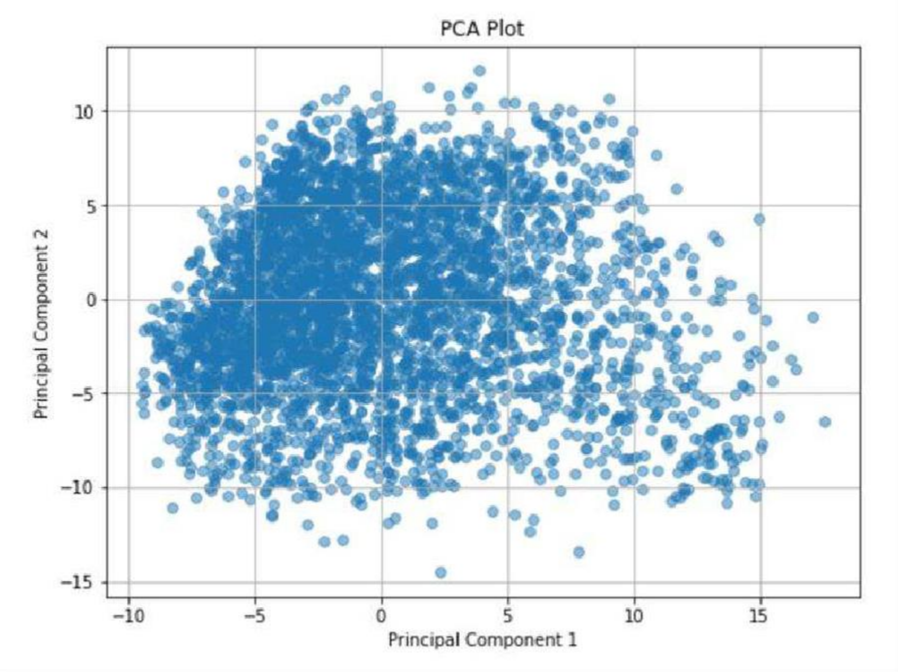
1	Title	post_content	post_link
2	10 Reasons to Hire Vets	Looking for a hard-working, motivated, ethical employee? Try hiri	https://www.linkedin.com/pulse/10-reasons-hire-vets-tom-hildreth?trk=public_profile_article_view
3	Facing a Change - Job Search 1st Month	(23 to be more precise) since I found myself in this position â€" be	https://www.linkedin.com/pulse/facing-change-job-search-1st-month-ilya-kotlyar?trk=public_profile_article_view
4	The Role Of Observability In Improving Reliabil	Target audience: IT Executives, CTOs, CIOs, SRE Managers, IT Mana	https://www.linkedin.com/pulse/role-observability-improving-systems-reliability-rodrigo-f-santos?trk=public_profile_article_view
5	Home Care job available	Rochester Regional Health System is looking for a new Home Care	https://www.linkedin.com/pulse/home-care-job-available-lizbeth-zielinski?trk=public_profile_article_view
6	Hiring		https://www.linkedin.com/pulse/hiring-heather-goodell?trk=public_profile_article_view
7	Psychologyâ€™s Role in Marketing	Whenever I mention that I have a bachelor's degree in marketing	https://www.linkedin.com/pulse/psychologys-role-marketing-hadley-roy?trk=public_profile_article_view
8	Customer Service 5 to 10 positions available	New Heights Management, Inc. is now seeking Entry Level and Exp	https://www.linkedin.com/pulse/customer-service-5-10-positions-available-stephanie-litchfield?trk=public_profile_article_view
9	Clarifying Roles to Optimize Safe Outcomes du	Role ambiguity has certainly been identified as a contributing fact	https://www.linkedin.com/pulse/clarifying-roles-optimize-safe-outcomes-during-confined-emma-mcclees?trk=public_profile_article_v
10	Better healthcare can ease the small business	Note: I wrote this piece with my Gusto colleagues Topher Reynosc	https://www.linkedin.com/pulse/better-healthcare-can-ease-small-business-hiring-must-liz-wilke-phd?trk=public_profile_article_view
11	I need a job in the Cleveland, Ohio area	Seeking employment in the Cleveland, Ohio area. Open for travel.	https://www.linkedin.com/pulse/20140326013126-3818119-i-need-a-job-in-the-cleveland-ohio-area?trk=public_profile_article_view
12	JOB ALERT	IAVE is seeking a new Executive Director. They are seeking a strate	https://www.linkedin.com/pulse/job-alert-susan-portugal?trk=public_profile_article_view
13	GREAT CSR JOB IN LONDON!	Employee Engagement Manager - EMEABank of America Merrill Ly	https://www.linkedin.com/pulse/great-csr-job-london-susan-portugal?trk=public_profile_article_view
14	Wonderful Job Opportunity At Points of Light	Manager, Economic Opportunity Programs._Check it out! http://w	https://www.linkedin.com/pulse/wonderful-job-opportunity-points-light-susan-portugal?trk=public_profile_article_view
15	We are hiring...	We are currently accepting applications for a Maintenance Technic	https://www.linkedin.com/pulse/we-hiring-tiffany-taylor?trk=public_profile_article_view
16	Job Posting: Omadi - Senior Implementation Er	Omadi is a Provo, UT based SAAS company founded in 2013. Omad	https://www.linkedin.com/pulse/job-posting-omadi-senior-implementation-engineer-daniel-marsh?trk=public_profile_article_view
17	Recruiting: Want to hire a gritty, scrappy perfor	My two kids were year-round competitive swimmers for 7 and 10 y	https://www.linkedin.com/pulse/recruiting-want-hire-gritty-scrappy-performer-why-you-dan-larson?trk=public_profile_article_view
18	Marcotte is hiring a Group Benefits Departmen	http://careerlink.com/employer/jobs/12385	https://www.linkedin.com/pulse/marcotte-hiring-group-benefits-department-manager-trish-maynard?trk=public_profile_article_view
19	Marcotte is hiring Group Benefits Department	http://careerlink.com/job/view/12385/000001	https://www.linkedin.com/pulse/marcotte-hiring-group-benefits-department-manager-position-maynard?trk=public_profile_article_vie
20	JSOM: Mastering Job Interviews	As an undergraduate freshman, I lacked the experience on how to	https://www.linkedin.com/pulse/jsom-mastering-job-interviews-faiz-prasla?trk=public_profile_article_view
21	Executive Director Position	Western Maryland Health system is currently recruiting for an Exe	https://www.linkedin.com/pulse/executive-director-position-jennifer-williams?trk=public_profile_article_view
22	I Should Never Have Hired You. You Have too M	Happy Independence day to me! Feb 19th, 2016 marks the two yea	https://www.linkedin.com/pulse/i-should-never-have-hired-you-too-many-children-angela-christu?trk=public_profile_article_view
23	Full Time Java Architects/Java Developers Req	Full Time Positions:Looking for Java Architects/Sr.Java Developers	https://www.linkedin.com/pulse/full-time-java-architectsjava-developers-required-chris-jones?trk=public_profile_article_view
24	Industrial Post-Doctoral position - Molecular Virology		https://www.linkedin.com/pulse/industrial-post-doctoral-position-molecular-jason-laliberte?trk=public_profile_article_view
25	Career Opportunity	PriorityWorkforce is an elite staffing company seeking enthusiasti	https://www.linkedin.com/pulse/career-opportunity-bibiana-torres?trk=public_profile_article_view
26	Construction site safety jobs available	I am in the need of construction site safety professionals to work	https://www.linkedin.com/pulse/construction-site-safety-jobs-available-dan-bergerson-chst?trk=public_profile_article_view
27	Growing on Twitch & Becoming a Full-Time Stre	I†™ve been streaming on Twitch as a full-time job for about two	https://www.linkedin.com/pulse/growing-twitch-becoming-full-time-streamer-steve-sarumi?trk=public_profile_article_view

- **Data analysis:**

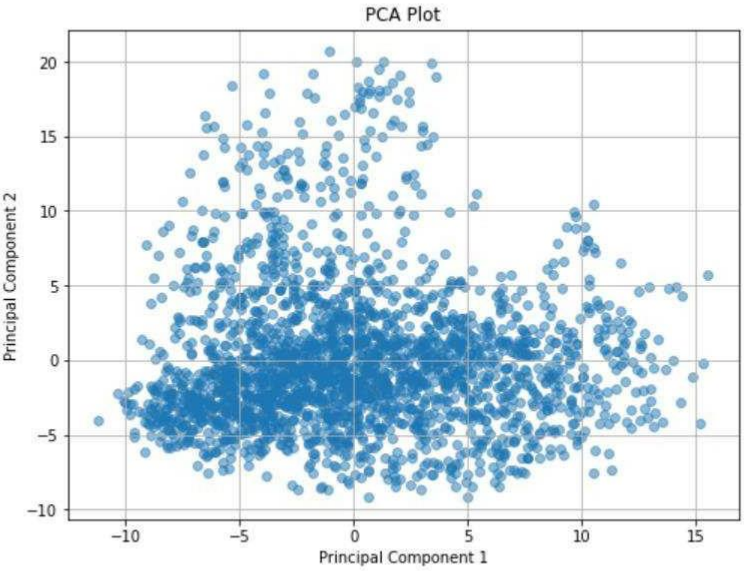
(1) **User′s positions Kmeans analysis for user's position Kmeans:**



(2) **Pca for user′s positions:**

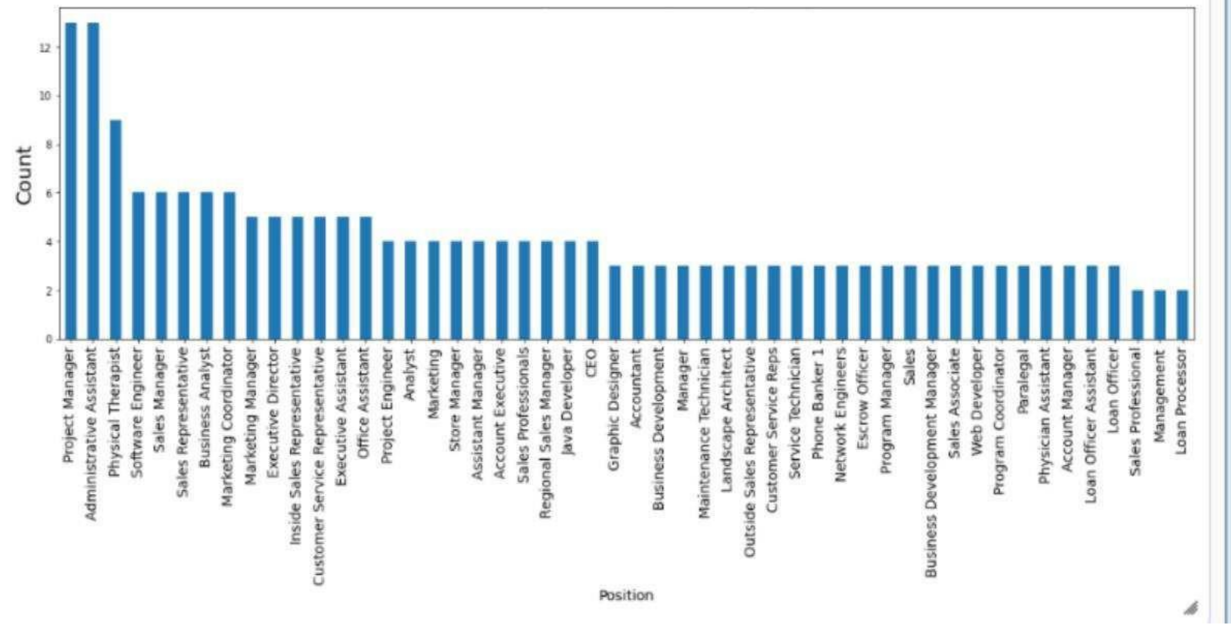


(3) **PCA for positions of the job posts:**



(4) **Bar plot of the positions in the job posts:**

Bar plot of the top 50 positions (Descending order)


Count
Project Manager
Administrative Assistant
Physical Therapist
Software Engineer
Sales Manager
Sales Representative
Business Analyst
Marketing Coordinator
Marketing Manager
Executive Director
Inside Sales Representative
Customer Service Representative
Executive Assistant
Office Assistant
Project Engineer
Analyst
Marketing
Store Manager
Assistant Manager
Account Executive
Sales Professionals
Regional Sales Manager
Java Developer
CEO
Graphic Designer
Accountant
Business Development
Manager
Maintenance Technician
Landscape Architect
Outside Sales Representative
Customer Service Reps
Service Technician
Phone Banker 1
Network Engineers
Escrow Officer
Program Manager
Sales
Business Development Manager
Sales Associate
Web Developer
Program Coordinator
Paralegal
Physician Assistant
Account Manager
Loan Officer Assistant
Loan Officer
Sales Professional
Management
Loan Processor
Position