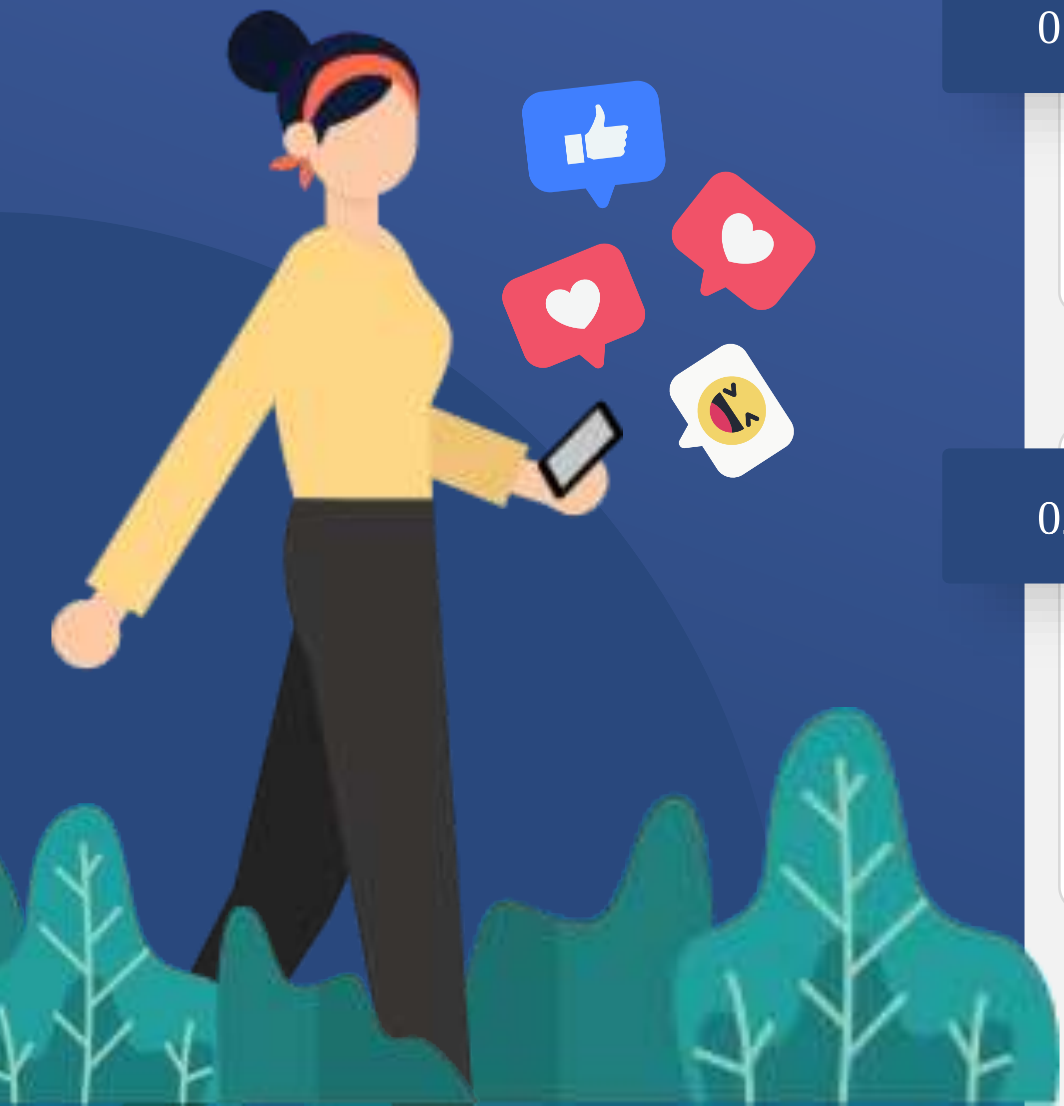
Linked In
Survey Analysis

- Rohit Dutta

Project Overview



01

What is this project about?

- A data-driven study based on a survey conducted to understand LinkedIn's role in professional networking and job search.
- Analysis of user demographics, preferences, satisfaction, challenges, and competitor usage.
- Application of **EDA, hypothesis testing, sentiment analysis, clustering, and A/B testing** to extract insights.

02

Why LinkedIn?

- LinkedIn is the **largest professional networking platform** globally (over 1B users).
- Increasing relevance for **students, job seekers, and professionals** in India and worldwide.
- Despite popularity, users often compare LinkedIn with **competitors (e.g., Naukri, Indeed, Glassdoor)**.
- Understanding **pain points and user sentiment** can help improve networking effectiveness.

Objectives of the Study

Understand User Demographics

- Who uses LinkedIn? (Students, job seekers, working professionals)
- How long have they been using LinkedIn?

Evaluate LinkedIn’s Effectiveness

- For networking, job search, skill learning.
- Satisfaction with UI, features, job recommendations.

Perform Statistical Analysis

- Hypothesis testing (UI satisfaction, networking effectiveness, job recommendations).
- Sentiment analysis on open-ended feedback.
- Clustering to segment users based on behavior.
- A/B testing to simulate product improvements.

Draw Actionable Insights

- Identify strengths, weaknesses, and opportunities for LinkedIn.
- Recommend data-backed improvements for better user experience.

Compare with Competitors

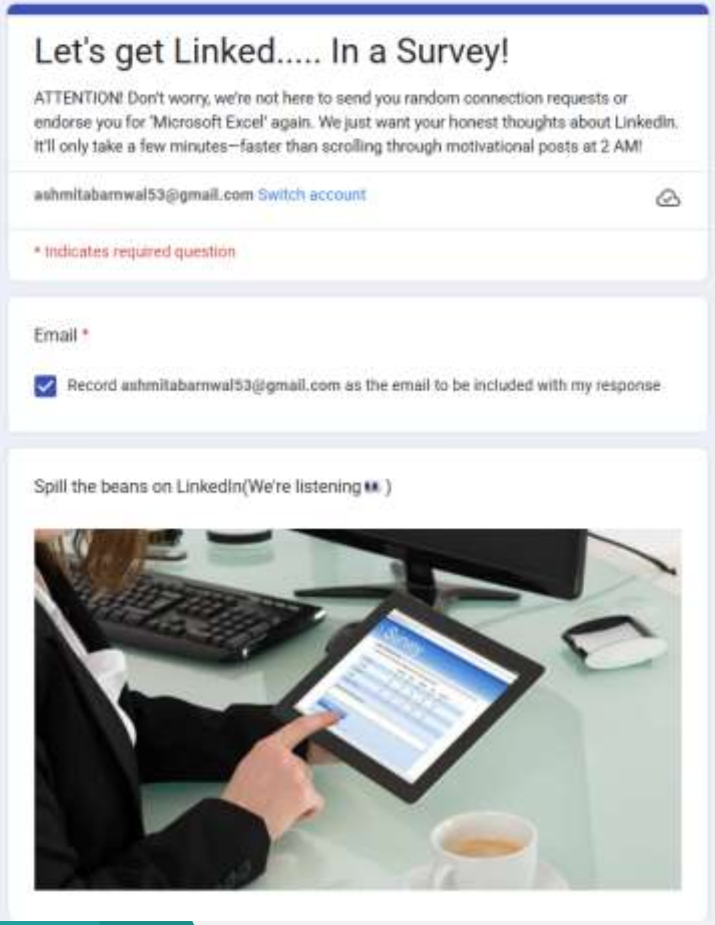
- Identify which platforms (Naukri, Indeed, Glassdoor, etc.) are also used.
- Understand where LinkedIn stands in terms of user preference.

Survey & Data Collection

Survey Tool Used

- Google Forms
- Designed with **15+ questions** covering:
  - Demographics (status, experience, etc.)
  - LinkedIn usage patterns
  - Feature satisfaction (UI, networking, job recommendations)
  - Competitor platforms used
  - Open-ended feedback

Google Form



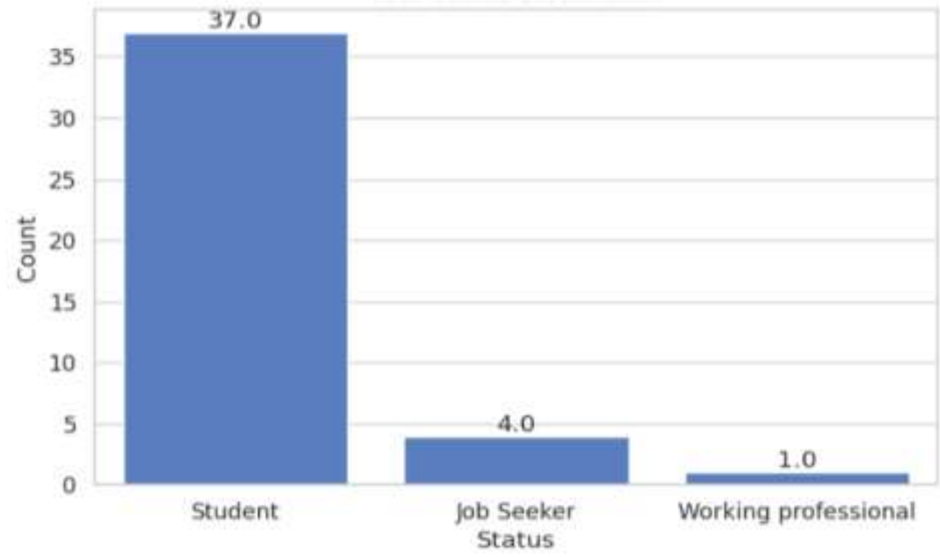
Response Collection

- Collected via peer groups, WhatsApp, LinkedIn shares, and student/professional networks
- Duration: 1 week (16-08-2025 to 23-08-2025)
- Total 42 valid responses

Response Sheet

Exploratory Data Analysis

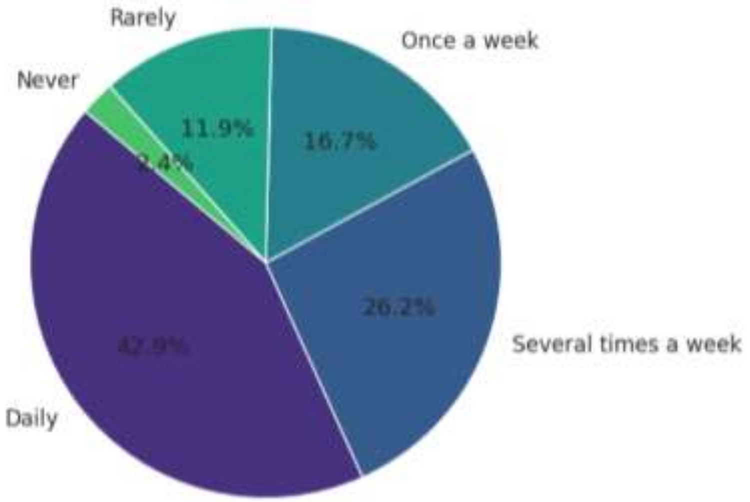
User Status Distribution



Majority of respondents are **students**, followed by **Job Seekers**.

Insight: LinkedIn is widely used by **students preparing for placements** and Job seekers for new opportunities.

Frequency of LinkedIn Use

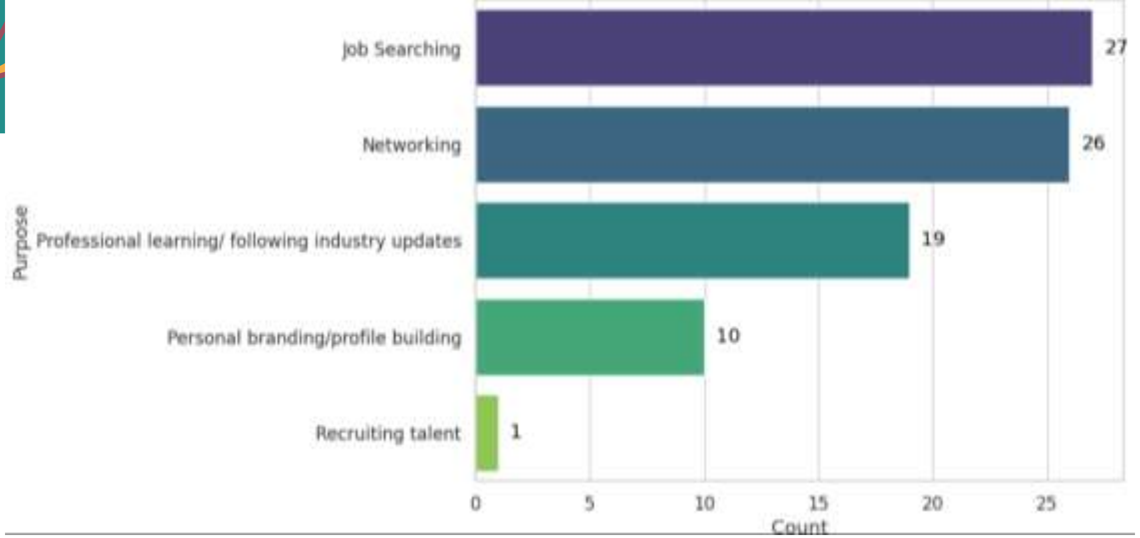


A large share of users log in **daily or several times a week**, showing **high engagement**.

Insight: LinkedIn is part of the **routine digital habits** of most users, closer to social media in frequency.

Exploratory Data Analysis

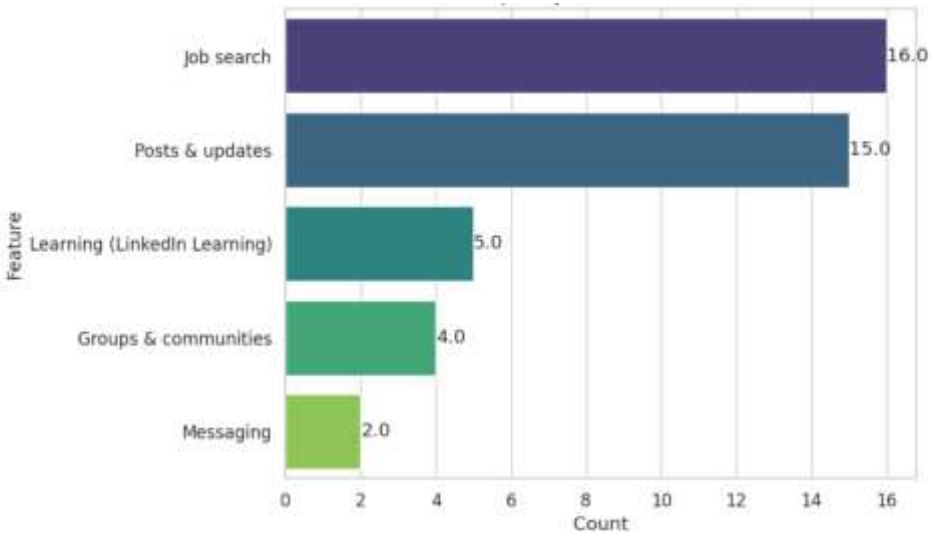
Primary Purpose of Using LinkedIn



Job search & opportunities dominate, being the top reason & **Networking & professional connections** is the second major driver.

Insight: LinkedIn is primarily seen as a **career-building tool** (jobs + networking), with less emphasis on being a **content platform**.

Most Frequently Used LinkedIn Features

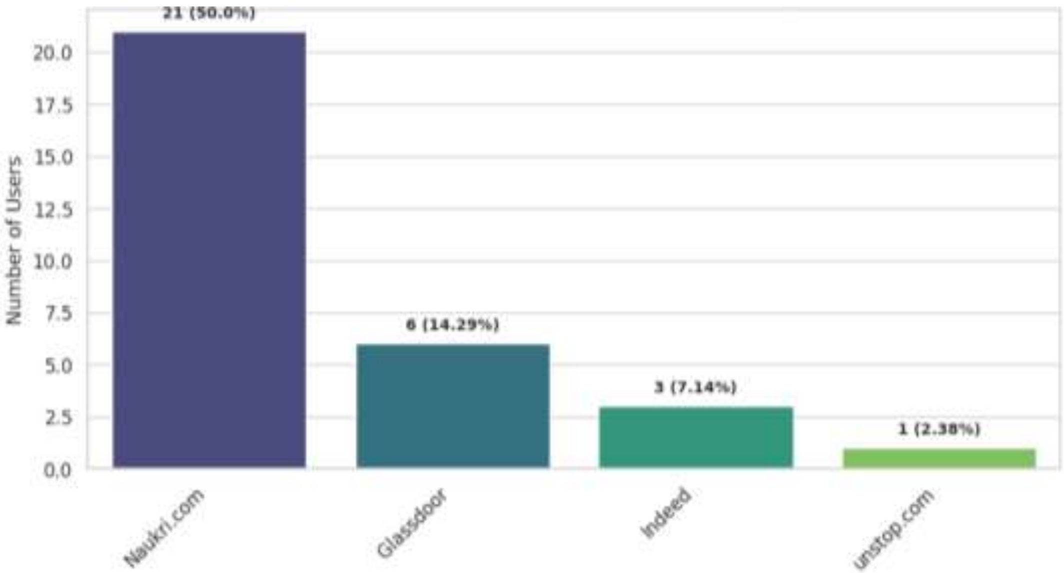


The **Job section** is the most frequently used feature. Features like **LinkedIn Learning** and **messaging** are used less.

Insight: Usage is **goal-driven (jobs + visibility)** rather than **engagement-driven (content + messaging)**.

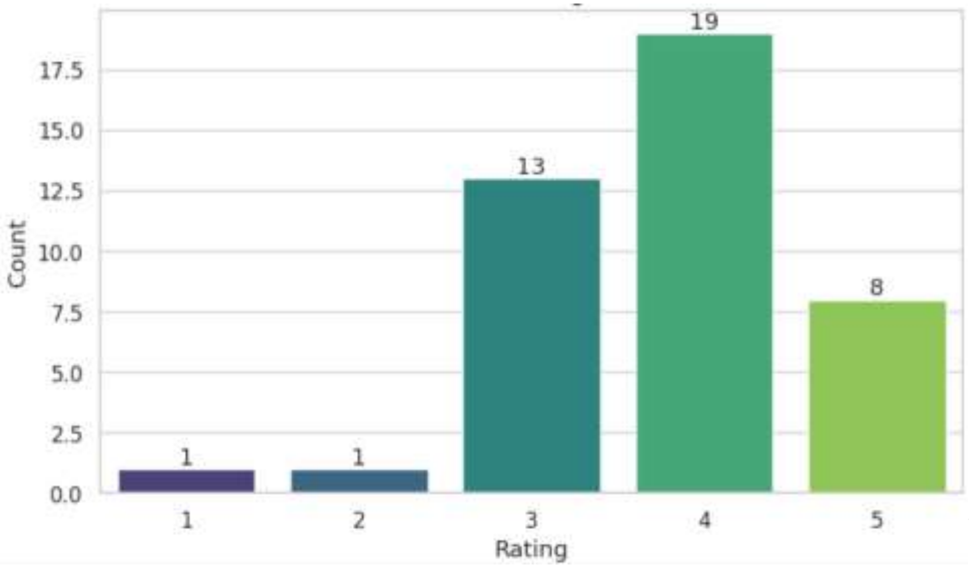
Exploratory Data Analysis

Top Competitors



Insight: 50% of the respondents use Naukri.com which shows people rely on other sites too for job searching.

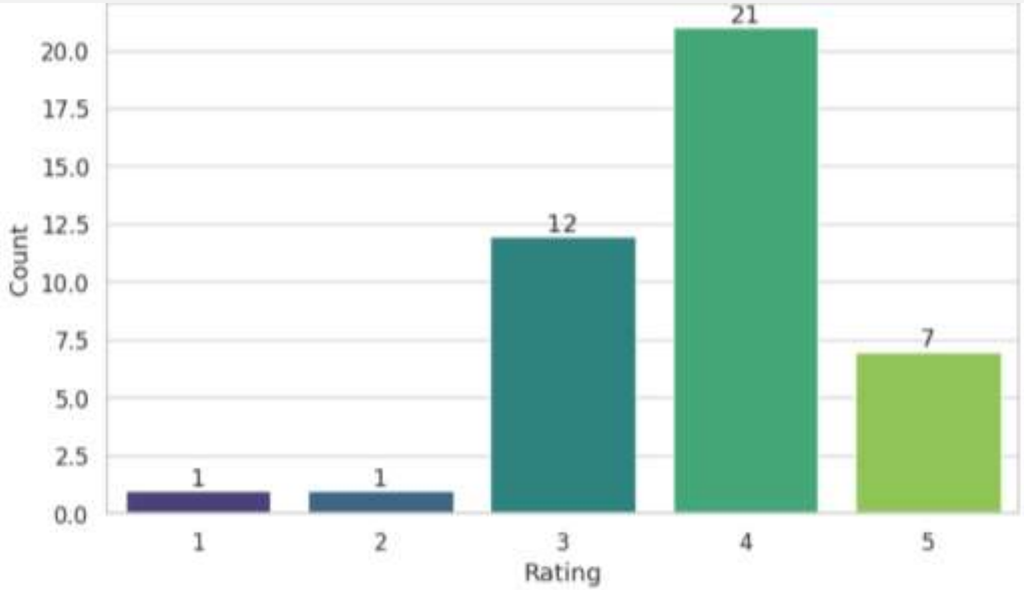
UI Satisfaction



Insight: Users generally find the interface intuitive.

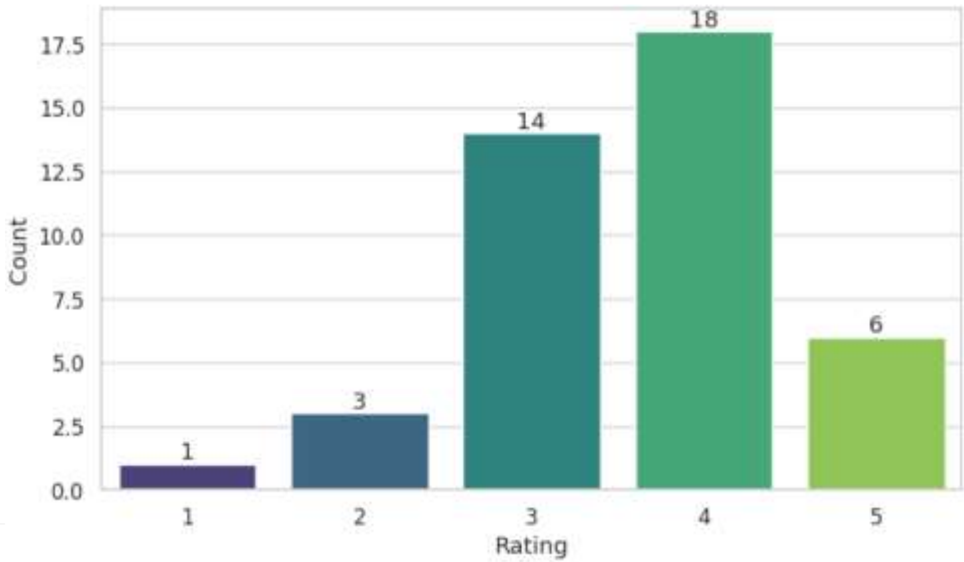
Exploratory Data Analysis

Networking Effectiveness



Insight: Users finds neutral building professional network over Linkedin.

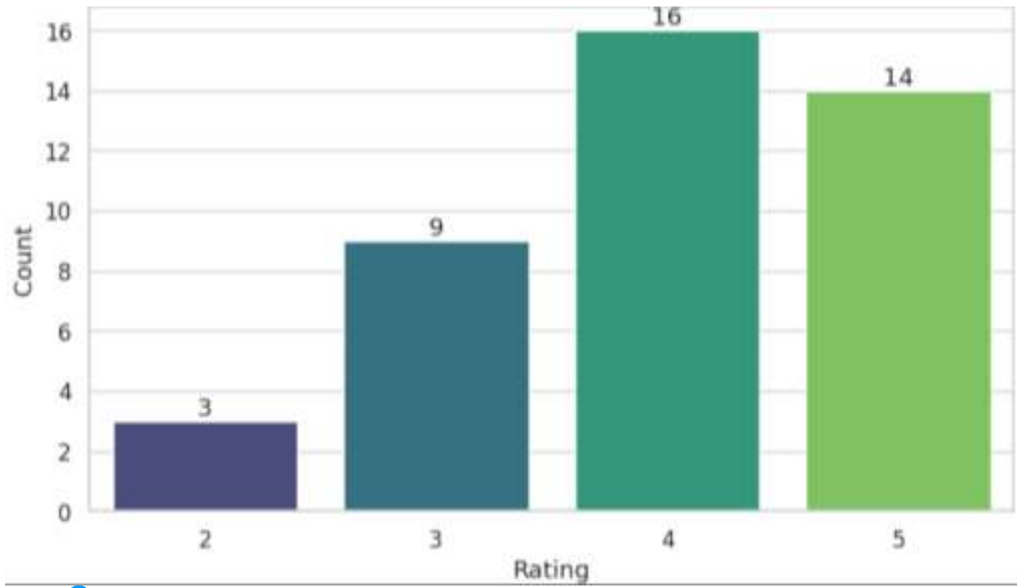
Job Recommendation Usefulness



Insight: A polarized distribution (some 1–3, some 4–5) suggests LinkedIn's job recommendations may work very well for some industries but not others.

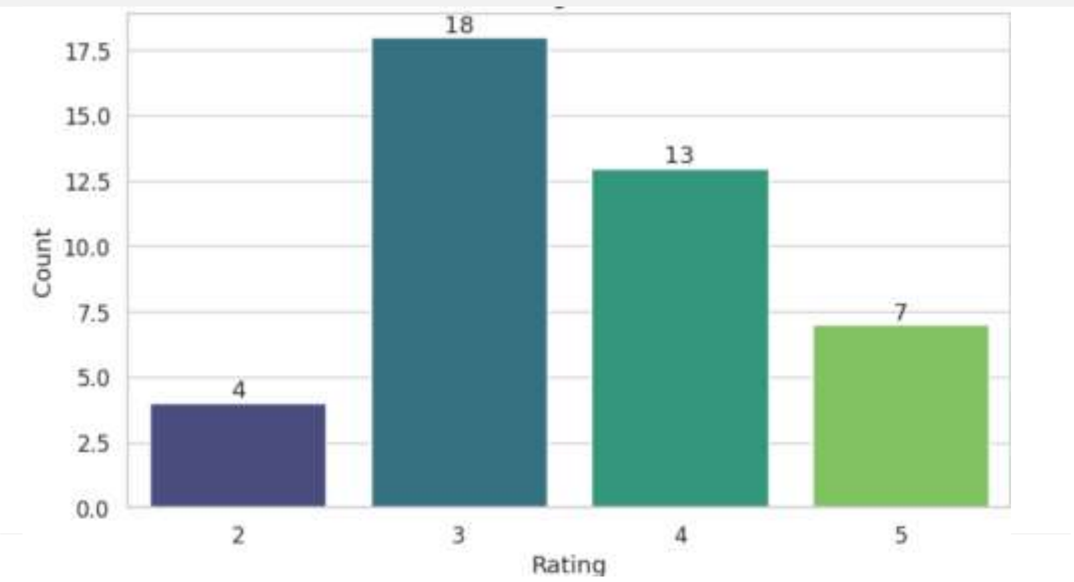
Exploratory Data Analysis

Recommend Rating Distribution



Insight: Maximum respondents are promoting Linkedin to others

Trust Rating Distribution



Insight: Users seems to have trust issues with the shared informations, posts, articles over the platform .

Hypothesis Testing

Hypothesis 1 — Usage Frequency vs UI Satisfaction

Aim:
To evaluate whether the frequency of LinkedIn usage has an impact on users' satisfaction with its interface.

Hypotheses:
Null Hypothesis (H0): UI satisfaction is the same across all usage frequency groups.
Alternative Hypothesis (H1): UI satisfaction differs depending on usage frequency.

Kruskal-Wallis Formula

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

Where:
N = total number of observations
k= number of groups
Ri = sum of ranks for group iii
ni= number of observations in group iii

Test Used:
Since UI satisfaction is measured on an ordinal scale (ratings) and the independent variable (frequency of usage) is categorical with multiple groups, the **Kruskal–Wallis test** was selected. This is a **non-parametric test** that compares the median ranks of more than two groups. It is preferred over ANOVA when the assumption of normality is not guaranteed.

Results:
Kruskal–Wallis H-statistic: **5.548**
p-value: **0.2356**

Inference:
Since the p-value (0.2356) is greater than the significance level (α = 0.05), we **fail to reject H0**. This indicates that **UI satisfaction does not significantly differ across users with different LinkedIn usage frequencies**.

Interpretation:
This suggests that whether someone uses LinkedIn daily, weekly, or rarely, their satisfaction with the platform's interface is not significantly different. LinkedIn's UI seems to provide a consistent experience irrespective of engagement frequency.

Hypothesis Testing

Hypothesis 2 — User Status vs Networking Effectiveness

Aim:
To examine whether professional status (e.g., student, employed, freelancer) influences how effective users find LinkedIn for networking.

Hypotheses:
Null Hypothesis (H0): Networking effectiveness ratings are the same across all status groups.
Alternative Hypothesis (H1): Networking effectiveness differs across status groups.

Kruskal-Wallis Formula

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

Where:
N = total number of observations
k= number of groups
Ri = sum of ranks for group iii
ni= number of observations in group iii

Test Used:
The independent variable (Status) is categorical, and the dependent variable (Networking Effectiveness) is ordinal. Since there are more than two groups, and normality cannot be assumed, the **Kruskal–Wallis test** is used.

Results:
Kruskal–Wallis H-statistic: **1.388**
p-value: **0.4995**

Inference:
With a p-value greater than 0.05, we **fail to reject H0**. This implies that **user status does not significantly affect perceptions of LinkedIn's networking effectiveness**.

Interpretation:
This suggests that whether someone uses LinkedIn daily, weekly, or rarely, their satisfaction with the platform's interface is not significantly different. LinkedIn's UI seems to provide a consistent experience irrespective of engagement frequency.

Hypothesis Testing

Hypothesis 3 — Favorite Feature vs Job Recommendation Usefulness

Aim:
To determine whether user's most frequently used LinkedIn feature impacts how useful they perceive the job recommendation system.

Hypotheses:
Null Hypothesis (H0): Job recommendation usefulness ratings do not differ by favorite feature.
Alternative Hypothesis (H1): Job recommendation usefulness differs across favorite features.

Kruskal-Wallis Formula

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

Where:
N = total number of observations
k= number of groups
Ri = sum of ranks for group iii
ni= number of observations in group iii

Test Used:
The independent variable (Favorite Feature) is categorical with multiple groups, and the dependent variable (Job Recommendation Usefulness) is ordinal. The Kruskal–Wallis test is appropriate here.

Results:
Kruskal–Wallis H-statistic: **1.834**
p-value: **0.7662**

Inference:
Since the p-value is much greater than 0.05, we **fail to reject H0**. This means **users' favorite feature does not significantly impact how they perceive the job recommendation system**.

Interpretation:
Even if a user spends most of their time on job posts, networking, or learning resources, their satisfaction with LinkedIn's job recommendations remains similar. This suggests that perceptions of job recommendation quality are independent of general usage patterns on the platform.

Hypothesis Testing

Hypothesis 4 — Years of LinkedIn Use vs Likelihood to Recommend

Aim:

To test if longer LinkedIn usage duration increases users' likelihood to recommend LinkedIn to others.

Hypotheses:

Null Hypothesis (H0): Years of LinkedIn use have no effect on recommendation likelihood.

Alternative Hypothesis (H1): Years of LinkedIn use positively correlates with recommendation likelihood.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

ρ = Spearman's rank correlation coefficient

d_i = difference between the two ranks of each observation

n = number of observations

Test Used:

Here, both variables are ordinal/numerical. To check correlation between two such variables, the **Spearman Rank Correlation** test is appropriate.

Results:

Spearman correlation coefficient (ρ): **0.168**

p-value: **0.2866**

Inference:

The correlation is weak and statistically insignificant ($p > 0.05$). We **fail to reject H0**, meaning there is **no significant relationship between years of LinkedIn use and likelihood to recommend**.

Interpretation:

Users who have been on LinkedIn longer are not necessarily more likely to recommend it compared to newer users. This suggests that **time on platform does not build advocacy**, implying LinkedIn must actively improve features to drive recommendations.

Hypothesis Testing

Hypothesis 5 — Overall Sentiment vs User Status

Aim:
To check whether overall perception (sentiment score) of LinkedIn differs based on professional status.

Hypotheses:
Null Hypothesis (H0): Overall sentiment is the same across all status groups.
Alternative Hypothesis (H1): Sentiment differs across different statuses.

Kruskal-Wallis Formula

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

Where:
N = total number of observations
k= number of groups
Ri = sum of ranks for group iii
ni= number of observations in group iii

Test Used:
Since sentiment scores are continuous (derived from text or ratings) and user status is categorical, the **Kruskal–Wallis test** was chosen.

Results:

Kruskal–Wallis H-statistic: **0.693**
p-value: **0.7073**

Inference:

With $p > 0.05$, we **fail to reject H0**. There is **no significant difference in sentiment across user status groups**.

Interpretation:

This indicates that students, professionals, and freelancers hold broadly similar perceptions of LinkedIn. Sentiment does not strongly vary by user category, which implies that overall attitudes toward LinkedIn are consistent across demographics.

Sentiment Analysis

Improvements Feedback – Sentiment Analysis

Phrase Frequency: The most common improvement suggestions revolved around *"more relevant"*, *"entry level"*, and *"separate section"*. This shows users want:

- More relevant recommendations (likely job posts/content).
- Entry-level opportunities to be highlighted better.
- Better structured sections within the platform.

Sentiment Distribution:

- **Neutral (30 responses)** → Majority of suggestions were written in a neutral, factual tone (users simply pointing out missing features).
- **Positive (10 responses)** → Some users acknowledged useful aspects but suggested small enhancements.
- **Negative (2 responses)** → Very few strongly dissatisfied responses, suggesting LinkedIn is generally well-received but has improvement areas.

Inference: Most improvement feedback is constructive, not emotionally charged. Users are **not angry**, but they want *practical enhancements* for usability and relevance.

Sentiment Analysis

Overall Impression – Sentiment Analysis

Top Positive Phrases:

“good” (19 mentions), *“work”*, *“great”*, *“job”*, *“networking”*, *“helpful”*, *“knowledge”*.

Indicates users see LinkedIn as a **positive professional networking tool**.

Neutral/Mixed Phrases:

“professional” (4 neutral mentions), *“platform”*, *“people”*.

These terms often appear in factual descriptions without strong emotion.

Negative Mentions (rare):

Found in contexts with *“connect”*, *“gain”*, or *“people”* → Some users felt networking quality is inconsistent.

Inference: Overall impressions are **heavily positive**, especially around networking and career benefits. A few mixed responses highlight areas where user experience in connections could be better.

ashutoshkr2154@gmail.com

LinkedIn is a valuable professional platform that connects people, shares opportunities, and fosters growth, though it could benefit from stronger measures against spam and irrelevant content.

goraipriyanka96@gmail.com

Good for networking and job finding

contact.piyushkumarbharti@gmail.com

it's alright nothing good nothing bad

Clustering Objective & Methodology

Why Clustering?

Objective: Group LinkedIn users based on their experience ratings.

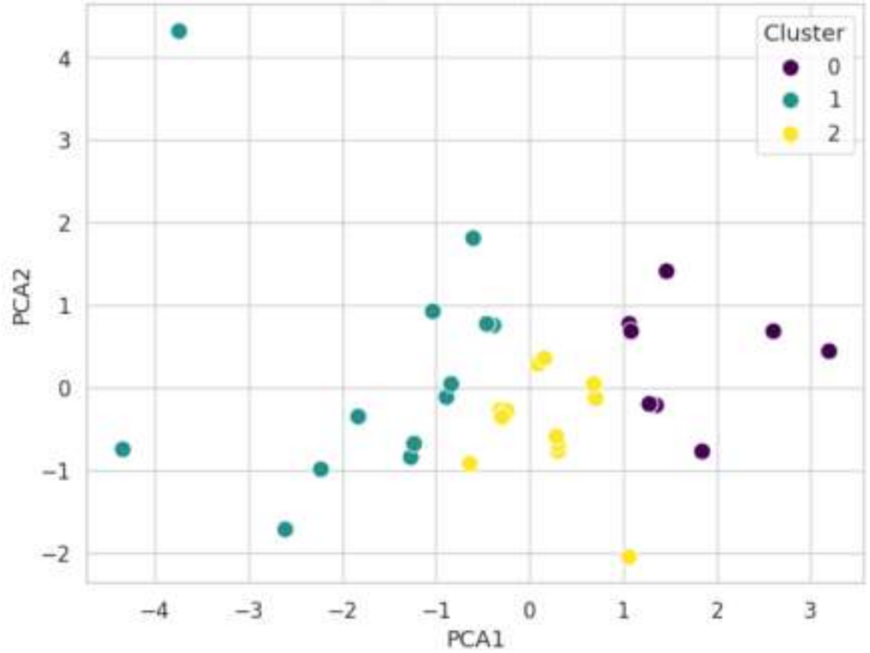
Why?

- Identify **user personas** (happy users, dissatisfied users, neutral users).
- Helps LinkedIn **target improvements** for specific groups.
- Supports **personalized feature development** (e.g., better job recs for dissatisfied cluster).

Features Used for Clustering:

- UI Satisfaction
- Networking Effectiveness
- Job Recommendation Usefulness
- Recommend to Others
- Trust

K-Means Clustering of Linkedin Users



Silhouette Scores
k=2 → 0.237
k=3 → 0.152
k=4 → 0.177
k=5 → 0.246
Chose **k=3** (interpretable and reasonably distinct).

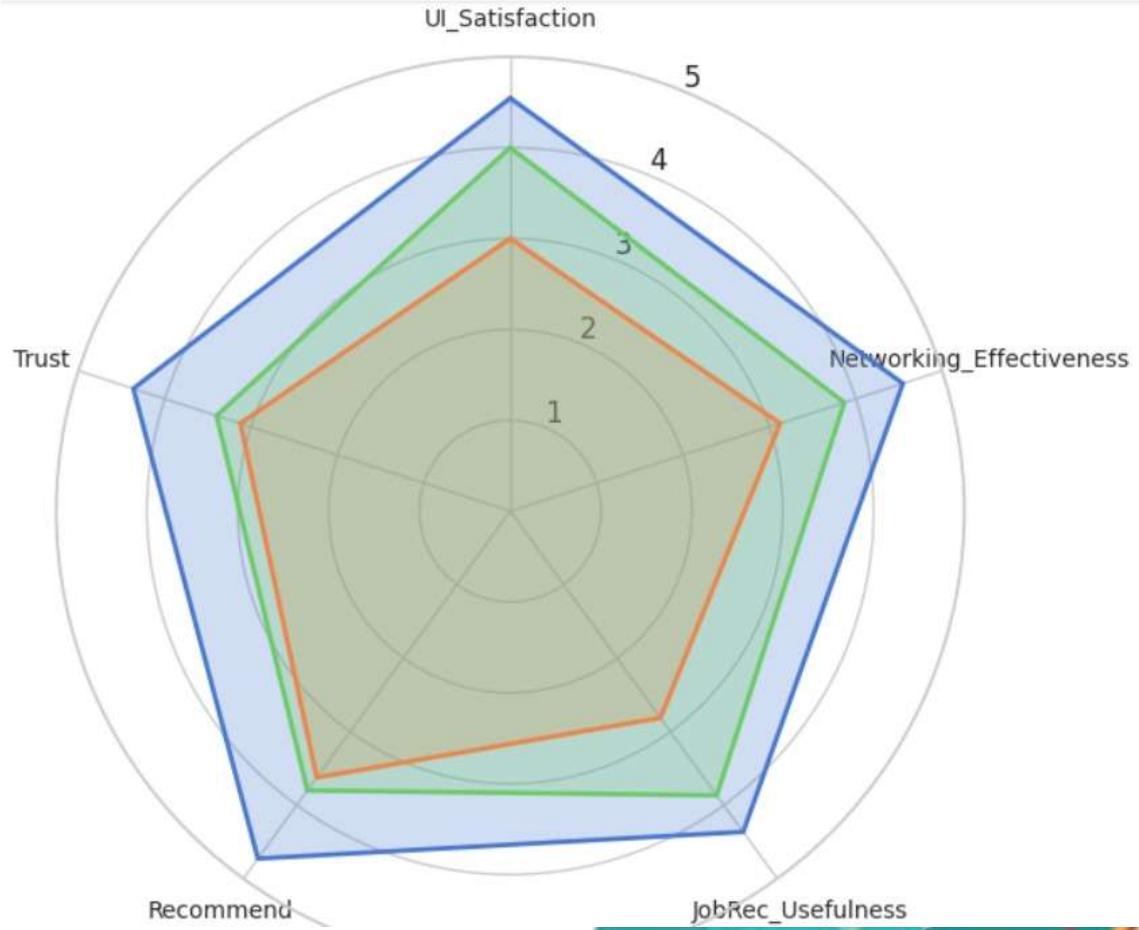
Insights & Implications

Cluster 0 (Happy Users)
High satisfaction, trust, and recommendation likelihood.
Need **retention strategies** to keep engagement high.
Cluster 1 (Dissatisfied Users)
Struggle most with **job recommendations** and **trust**.
Must prioritize **improving job-matching algorithms** and **networking support**.
Cluster 2 (Neutral/Moderates)
Decent satisfaction but **less enthusiasm**.
Potential to convert into **promoters** by focusing on **trust-building**.

Key Takeaway

Clustering revealed **3 LinkedIn user personas**.
Action Plan:

- **Fix job recommendations** (biggest weakness).
- **Boost trust features** for neutral/dissatisfied users.
- **Continue engaging happy users**.



A/B Test Setup

Objective: Compare LinkedIn-only users vs. Competitor users

Groups Defined:
Group A: LinkedIn Only (users not using any competitor)
Group B: CompetitorUsers (users also using competitors)

Metrics Tested (ratings scale: 1–5):
UI Satisfaction
Networking Effectiveness
Job Recommendation Usefulness
Trust
Recommend (Likelihood to Recommend)

Test Used: Independent **t-test** (unequal variances)

Significance Level (α): 0.05

A/B Test Results:

Metric	Mean (LinkedIn Only)	Mean (Competitors)	P-value	Significant
UI Satisfaction	3.64	3.81	0.53	No
Networking Effectivenes s	3.91	3.71	0.46	No
JobRec Usefulness	3.55	3.61	0.85	No
Trust	3.45	3.58	0.68	No
Recommend	4.18	3.90	0.34	No

Insights & Takeaways

- **No statistically significant differences** found between LinkedIn-only users and Competitor users across all 5 metrics.

- **Mean scores are similar** across groups → user satisfaction and trust levels are **comparable**, regardless of competitor usage.

- **LinkedIn-only users** show slightly higher “Recommend” scores (4.18 vs 3.90), but **not statistically significant**.

- **Interpretation**:
  - LinkedIn competes on **equal footing** with other platforms in terms of satisfaction.
  - Users who try competitors do not rate LinkedIn significantly worse (or better).
  - Future strategy: LinkedIn can focus on **differentiating features** to convert competitor users, since current satisfaction is similar.

Challenges Faced

Data-Related Challenges

- **Missing Values**: Some survey responses were incomplete (e.g., ratings not filled). Required **imputation or exclusion**, which may have reduced the usable sample size.
- **Imbalanced Groups**: More users were in the *Competitor* category compared to *LinkedIn-only*, which can affect **statistical power** and interpretation of A/B tests.
- **Self-Reported Bias**: Responses are **subjective** and may not fully reflect actual behavior (e.g., satisfaction may differ from actual usage).
- **Non-Normality**: Rating data is **ordinal** (1–5 scale) and may not meet assumptions of normality, requiring **non-parametric tests** (like Kruskal-Wallis).

Methodological Challenges

- **Choosing the Right Test**: For different hypotheses, had to carefully decide between **ANOVA, Kruskal-Wallis, Chi-square, and t-tests** depending on distribution and variable type.
- **Multiple Hypotheses Problem**: Testing several metrics increases the chance of **Type I errors (false positives)**. Needed careful interpretation of p-values.
- **Limited Sample Size**: Small subgroup sizes (e.g., LinkedIn-only users in some cases) reduced the reliability of statistical inference.

Interpretation Challenges

- **No Significant Differences**: In A/B testing, all results were statistically insignificant, making it harder to draw strong conclusions.
- **Practical vs. Statistical Significance**: Even if differences were not statistically significant, small **practical differences in means** could still be relevant for business decisions.
- **Generalizability**: Results are based on a specific dataset; user behavior might differ in real-world LinkedIn global user base.

Appendix

All code, data cleaning, and analysis can be accessed at:

Linkedin_Survey_analysis.ipynb

Github Repository

Thank You!