LinkedIn Data Analysis

20.04.2025

—

Arpita Tripathi

202410101200022

su-24023@sitare.org

Objective

The objective of this project is to clean and analyze LinkedIn data of 126 students by treating each file name as a node and the content within each file as its adjacent nodes.

Data Cleaning

1. Removed emojis, special characters, and capitalized all row values for consistency.

2. Standardized column names, removed duplicates, and ensured file name uniqueness.

3. Converted `.xlsx` files to `.csv` format with uniform naming.

4. Fixed parsing issues, verified required columns, cleaned irrelevant or oversized data and rename all files .

Making of Adjacency List and Operation

1. Treated file names as nodes and file data as adjacent nodes, generating new files from adjacent nodes that were not already present as keys in the adjacency list(Folder).
2. Total numbers of files (key in adjacency list) are '28474'.
3. Find multiple random paths between two students, prune the paths, and calculate the "count," "min," "max," "avg" lengths, as well as the length of each random path or prune path.

Key Insight

1. Top Ten Company:-

 1. HCLTech(1200), Sitare University(1155), Google(673), Amazon(560), Dakshana   .   .
.Foundation(478), Microsoft, Tata Consultancy Services, ZingHR, Selfemployed, Freelance

2. Students have Most Connections:-

1.Rohit Malviya(4670), Ravi Rajput(4120), Manoj Dewda(3940) and many more.
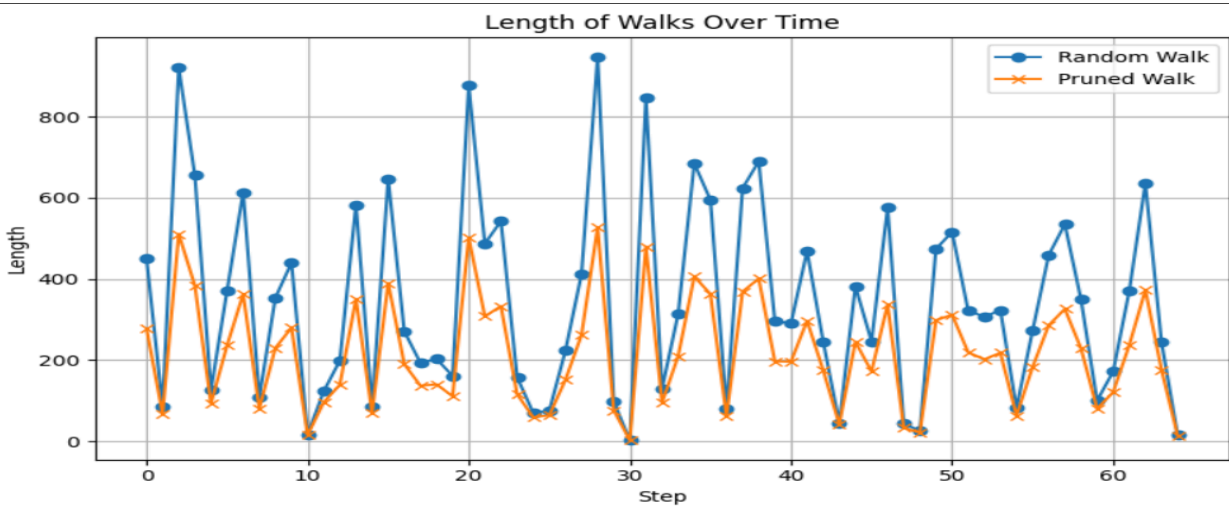
3. Most Common Last Name:-
1. Kumar and Singh dominate, reflecting common surnames in the region.
2. Names like Idabu, Dulcey, and Raji indicate cultural diversity

4. Company Size Categories:-

1.Solo and Small enterprises comprise most companies, suggesting a preference for
.   startups or independent ventures.

2.Large corporations are fewer but significant.

Comparison of Random Walk and Prune Path

Conclusion

The data reveals Sitare students building powerful professional networks (4,000+ connections), with surprising global reach beyond India. Dominant tech/freelance career trends and efficient connection paths showcase our institution's strong industry and international linkages.