

Linkedin Assignment Analysis

In this assignment, we were given 126 CSV files, the analysis of which is reported here.

Top 20 Unique Keys via Inverse Graph Ranking

Inverse Graph Ranking ranks keys based on the uniqueness of names they contain. The contribution of a name is determined by how many other keys it appears in, and the overall score for a key is the sum of these contributions.

- **Formula:**

- If a name appears in 'n' keys:

$$\text{Contribution} = 1 / n$$

- **Key Score** = Sum of contributions of all names under that key.

Table: Top five keys with highest scores

Rank	Key	score
1	Ramraj Nagar	1805.09
2	Pranjal Dubey	1797.08
3	Ravi Rajput	1781.31
4	Manoj Dewda	1303.61
5	Rohit Malviya	1286.6

Pruned Path vs. Random Walk Length

The goal of this analysis is to compare the lengths of paths in a pruned graph (the shortest path between two randomly chosen nodes) and those generated by random walks (randomly chosen paths in the graph), using statistical and visual methods.

Method:

- A random graph was created, and isolated nodes were removed to form a pruned graph.
- The shortest path length between two random nodes was calculated for the pruned graph.
- Random walks of fixed length were simulated, and their lengths were measured.
- Statistical comparisons (mean, standard deviation) and histograms were generated.

Results:

- **Pruned Path:** Shorter paths with lower variability (mean length was smaller, standard deviation lower).
- **Random Walk:** Longer and more variable paths due to randomness (mean length was larger, standard deviation higher).

Key Insights:

- **Pruned paths** are more efficient and predictable, while **random walks** exhibit greater variability and longer lengths.
- The pruned graph's structure leads to shorter, more consistent paths compared to the random nature of random walks.

Conclusion:

Pruned paths are shorter and more efficient, while random walks are longer and more unpredictable.