LinkedIn Class Network Analysis Report

Objective

- To clean and unify raw LinkedIn connection data from multiple Excel files.
- To build an adjacency list and graph representation of the class network.
- To assess the employment status distribution among students.
- To generate useful insights and suggest future actions.

Process Overview

Data Cleaning:

- Unified Excel files with inconsistent headers.
- Converted to CSV and standardized column names.
- Removed incomplete entries and formatted names.

Graph Construction:

- Created an adjacency list from mutual connections.
- Ensured node consistency by adding missing references.
- Stored as JSON for easy reuse.

Analysis and Algorithms:

- Computed degree of each node (connection count).
- Performed greedy algorithm to find a dominating set.
- Used Python (matplotlib) to visualize data distribution.

Key Insights

Top 10 Most Connected Individuals

- Bar chart showed most connected people.
- These nodes can be potential peer mentors or community leads.

Employment Status Distribution

- Pie chart showed:

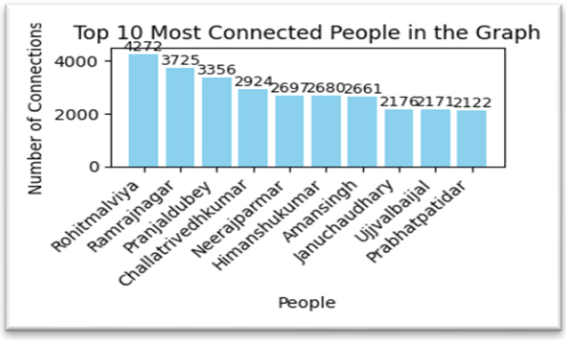
- Employed: 60.2% and Unemployed: 39.8%
- Helps identify job placement rate and target support.

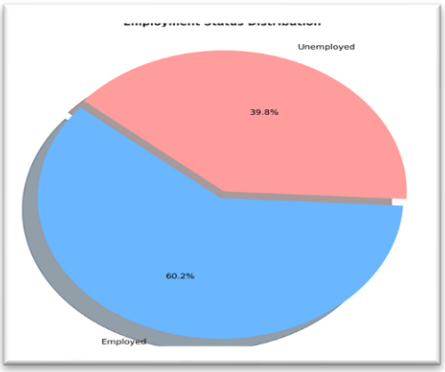
Connection Strength Ranges

- Horizontal bar chart categorized students:
  - 0–500: Low connection count
  - 1000–2000: Strong presence on LinkedIn
  - 3000+: Highly active networkers

Dominating Set Analysis

- Found minimum nodes needed to influence entire graph.
- Key to creating impactful campaigns or group initiatives.





Recommendations

- Organize "LinkedIn Growth" sessions for low-connectivity students.
- Assign mentors based on dominating set or top connectors.
- Leverage top influencers to boost class branding and reach.

Tech Stack Used

- **Languages**: Python
- **Libraries**: pandas, json, matplotlib, doctest
- **Tools**: VS Code, Excel, Jupyter Notebook
- **Concepts**: Graph Theory, Data Cleaning, Statistical Analysis.

Prepared By : Pragati Chauhan