**Assignment 3: Code Modifications and Analytical Decisions**

**Student: FADZAI BONDAYI 48770140**

**Unit: COMP6200**

**Centrality Extension (Task 1)**

Modifications:

Added betweenness centrality alongside degree and eigenvector centrality. Calculated metrics on both the largest connected graph and a top-100 subgraph to improve performance.

Rationale:

Betweenness reveals strategic bridge directors who can influence across disconnected company clusters. A degree shows how many boards a director is connected to. Eigenvector reveals influence based on connectedness to other high-ranking individuals.

**Code Repair (Task 2)**

Fixes:

1. Boolean Mapping

Changed lambda x: x == 't' to. map({'t': True, 'f': False}) for clarity.The previous line was not readable and may silently break if values aren't exactly 't' or 'f'. So the fixed line improves clarity and safety by using an explicit dictionary. Future readers will understand exactly how values are transformed.

2. Chained Assignment Warning

Added copy() when manipulating comp\_raw to avoid SettingWithCopyWarning

3. Repeated Groupby

Grouped once and aggregated age, compensation, and gender together.

4. Log Compensation

Replaced. map(math.log10) with np.log10() for performance and clarity.

In the original code, the data was grouped three separate times using. groupby('NAME') for each of the age, compensation, and gender columns. This was inefficient and harder to read. I replaced this with a single. groupby() operation is stored in a variable called grouped, and then all aggregations at once. This makes the code cleaner and more performant, especially for larger datasets.

Additionally, the original code was used. map(math.log10) to compute the log of compensation, which operates element-by-element and isn't the standard practice with pandas. I replaced this with numpy.log10() is applied directly to the column, which is vectorized and much more efficient. This also avoids errors when processing large numeric columns. I also added a .copy() of the raw DataFrame to prevent Pandas’ SettingWithCopyWarning during name standardization.

Finally, I included a. head() call at the end to show the result, ensuring the transformed data is visible for validation. These changes align with best practices in data science coding: write clear, efficient, and maintainable code.

**Feature Exploration (Task 3)**

Feature:software\_background

17% of directors had a software background. Their average compensation was slightly higher. Visualized this trend with bar charts.

We analyzed the software\_background feature to determine whether directors with a background in software or technology differ from others in terms of representation and compensation.

Only 17.4% of directors have a software background, meaning technical directors are a minority on company boards.Despite being fewer in number software-background directors earn slightly higher average compensation of 683205 compared to those without that are earning 640620

This suggests that while directors with technical backgrounds are underrepresented, their expertise is valued in compensation decisions. In the context of this project, such directors may play an important role in digital transformation, innovation, or tech-driven acquisition strategy making them potentially influential figures for a venture capital firm targeting high-growth or tech-based companies.

**Complementary Dataset (Task 4)**

Dataset: SEC EDGAR 10-K Financial Filings

Why: Combines financial capacity with network influence.

Use: Rank companies by both network centrality and revenue/assets.

Complementary dataset URL: [https://www.sec.gov/edgar.shtml](https://www.google.com/url?q=https%3A%2F%2Fwww.sec.gov%2Fedgar.shtml)

We would collect revenue and profit data for companies in the network, and combine it with their centrality scores. This would help us answer:

Are the most connected companies also the wealthiest? Which companies have both influence and financial capability to acquire others? Are there under-the-radar companies with high financials, low centrality, worth attention?

Using this, we could provide the Venture Capital fund with a ranked list of acquisition candidates, backed by both network influence and financial strength.

**Visualisation (Task 5a)**

Built horizontal bar charts for eigenvector centrality. Focused on directors only. Ensured results were easy to interpret for non-technical stakeholders.

The chart above shows the top 10 directors ranked by eigenvector centrality a measure of how connected they are to other influential nodes in the corporate network. These individuals act as power brokers and are ideal targets for strategic engagement by the VC fund.

Their influence makes them likely to facilitate deals, build trust across boards, or unlock access to key acquisition targets.

**Ethics Reflection (Task 5c)**

Reflected on the ethical implications of targeting individuals using network metrics. Raised concerns around privacy, bias, and transparency.

This project uses network analysis to identify powerful individuals who may influence major corporate decisions. While this is valuable for venture capitalists, it raises important ethical concerns:

Targeting Individuals

Identifying and approaching directors based solely on network position may pressure them into acting against their judgment or ethics.

Privacy & Consent

Even though data is public, repurposing it for influence strategies could be seen as manipulative or invasive.

Bias

The network may over-represent certain demographics e.g. older white men and reinforce systemic inequalities in board access and influence.

Transparency

Stakeholders affected by decisions employees, communities have no visibility into how influence-based decisions are made.

Conclusion

While data science can support smart business decisions, it must be balanced with accountability, fairness, and transparency. Ethical review and stakeholder inclusion are vital.