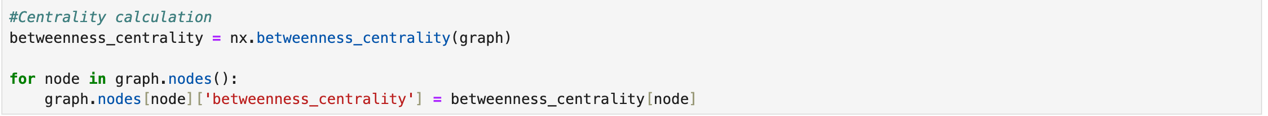
Tasks 1

Betweenness centrality identifies nodes in a network that act as bridges. In the relationship between directors and companies, directors with higher betweenness connect different clusters of companies, while companies with higher betweenness may become key links between influential directors. This helps to discover key intermediaries and potential influencers in corporate governance.



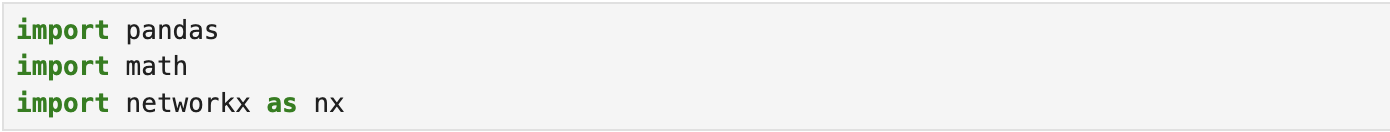
Betweenness Centrality:

* Directors: Strategic connectors between corporate communities.
* Companies: Gateways that facilitate interaction across the network.

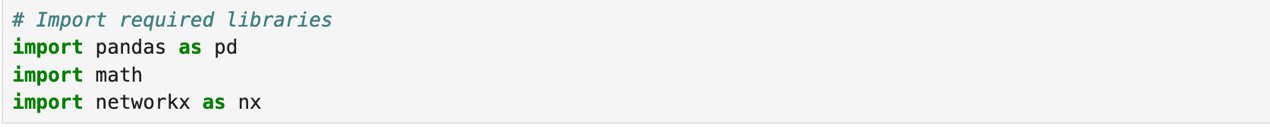
Tasks 2

1. Add Clear Comments

Original:



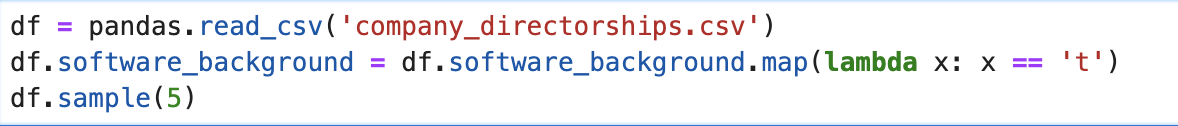
Improved:



The original code lacks documentation, and pandas is renamed according to the pd standard. This can improve the clarity of the pandas library and make it easier for people who use the code to understand it. Adding comments at the beginning can reduce the reading and comprehension time, especially for staff who need to check the code, so that they code more easily.

1. Clarify Boolean mapping

Original:

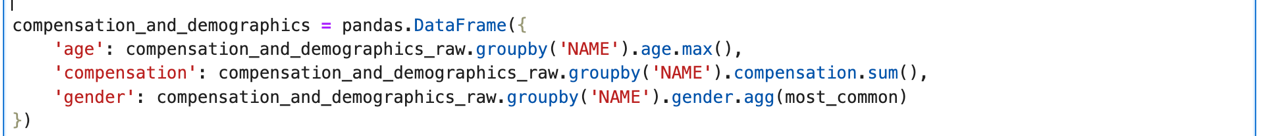


Improved:



It is not obvious in the code what values are in this column and how to interpret them. By using standardized input code (strip().lower()), the accuracy of the code can be improved, and the intention can be made clear to the reader.

1. Avoid repetition in groupby for clean code

Original:

Improved:

图片包含 公司名称

AI 生成的内容可能不正确。

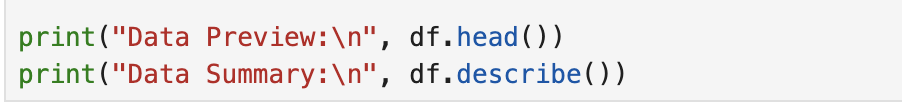
Repeated calls to groupby are inefficient and reduce readability. Assigning values to variables can improve performance and make the logic clearer.

1. Delete the useless data

Original:



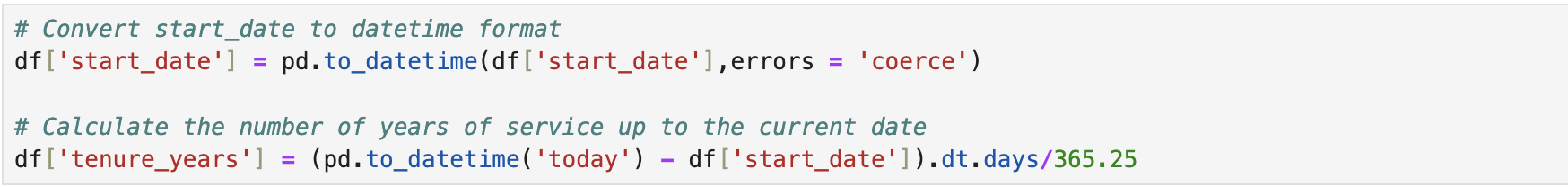
Improved:



Because: df.sample() will display randomly, which is not conducive to debugging and understanding the data structure. df.head() and df.describe() is more stable and more helpful for understanding variable distribution and missing values

Tasks 3

Unused feature is start\_date

Selection this field is used to calculate director tenure. It can helps analyze the correlation between experience and compensation or gender representation. Calculate the number of years of tenure and check its correlation with compensation. 

Tests 4

Dataset Source and Overview

* Dataset Name: Indian Startup Funding Dataset
* Source: Kaggle - Startup Investments in India
* Key Fields: StartupName, CityLocation, IndustryVertical, InvestorsName, InvestmentType, AmountInUSD, Date

The original project focused on company directorships and personal networks. However, it lacked context around market dynamics and investment behavior. By adding this dataset of startups, we can gain insight into which industries are seeing the most entrepreneurial activity, and which are attracting funding.  
While there is no direct ID association, the data has been linked by city. We compared the “city location” in the startup dataset with the supervisor company location to analyze regional entrepreneurial activity. We created a visualization of the top startups using seaborn and matplotlib.

Tests 5

(a):The data visualization effect has been improved. The seaborne bar chart is used to compare the number of directors by gender, which simply and clearly shows the number of male and female directors. The second picture is the best entrepreneurial city, ranking the top ten cities, using different colors to distinguish, making the chart clearer图表, 条形图

AI 生成的内容可能不正确。

(b): Implementing Tests 4 and 5 using complementary data: Complementary datasets and refined analysis: Using the startup dataset from Test 4, we performed city-level aggregation to identify regions with the highest startup activity. We cross-referenced this data with the directorship data to find geographic overlap. We explored visual associations that could be achieved without direct joins. 图表

AI 生成的内容可能不正确。

Conclusion，Combining directorship and financing data provides a broader picture of firm influence and entrepreneurial dynamics across regions, enriching the structure and context of the analysis.