**Financial Profile Segmentation of S&P 100 Companies Using PCA and Clustering Techniques**

**Abstract**

**Introduction**:

This study aims to categorize S&P 100 companies listed on the NYSE based on their financial profiles, utilizing the comprehensive metrics provided in the corresponding fundamentals data file from Kaggle. These metrics, sourced from SEC 10K annual filings, include key financial indicators like total revenue, cost of goods sold, and various types of expenses. By applying a data-driven approach, the study seeks to reveal underlying financial patterns and trends, pivotal for a sophisticated market analysis offering insights critical for investors and market analysts.

**Methods Used**:

The analysis employed Principal Component Analysis (PCA) to reduce the complexity of the financial data and select principal components that capture the most significant variances, based on the cumulative variance they accounted for. Subsequent K-means and hierarchical clustering were applied to these principal components to segment the companies into distinct groups. The optimal number of clusters was determined using the elbow method and silhouette scores, ensuring robust and meaningful financial profiling.

**Results Highlights**:

Three distinct groups are derived using KMeans clustering. Cluster 1 includes leading corporations like Apple (AAPL), Boeing (BA), and Exxon Mobil (XOM), known for their significant market presence and high total revenues. Cluster 2 consists of major financial institutions like Bank of America (BAC) and JPMorgan Chase (JPM), with distinct financial patterns in expenses and cost of goods sold. Cluster 3, the most diverse, contains a range of companies across various industries, reflecting diverse financial strategies and performance metrics. Further analysis can be based on financial domain aspects related to features in the data.

**Conclusion**:

This study offers a comprehensive approach to segmenting S&P 100 companies based on their financial profiles. The use of PCA and clustering techniques not only simplifies the complex financial data but also reveals distinct groupings of companies with similar financial characteristics. These findings provide critical insights into the financial strategies and market positioning of these companies, invaluable for investment decisions and market analysis. Future research could extend this approach by including additional features such as stock price movements, offering a more comprehensive view of the company's financial standing and performance over time.

**References:**

<https://www.kaggle.com/datasets/dgawlik/nyse/data?select=fundamentals.csv>