

AI in Finance Assignment 3

Yijia Zeng yzeng323@gatech.edu GTID: 903629003

Data Loading

(539318, 981)

Qualitative Features selection

Income Sheet

- **Net Income:** This is the total profit after all expenses, taxes, and interest have been deducted. Net Income is crucial for assessing overall profitability, so it is a key indicator of financial health. Companies with consistent losses may be at risk.
- **Sales:** This is the total revenue generated from goods or services sold. Sales are an essential measure of business growth and performance, indicating the company's ability to generate income.
- **EBITDA:** EBITDA reflects a company's operational profitability by excluding interest, depreciation, and admortization expenses. It is commonly used to evaluate a company's performance without the effects of financing and accounting decisions.

Balance Sheet

- **Cash:** The cash a company holds is crucial because it indicates the company's ability to handle immediate financial obligations and influences its future operational flexibility.
- **Total Assets:** The total assets of a company are essential as they show the company's resource, which is indicative of its capability for future expansion and operational efficiency.
- **Long Term Debt:** This represents the company's debt obligations due more than a year in the future. Long-Term Debt helps assess the company's leverage and long-term financial risk, showing how much the company relies on borrowed money.
- **Stockholders' Equity:** This measures the total equity owned by shareholders. It shows the company's net worth and is crucial for assessing financial stability.
- **Working Capital:** Working Capital is the difference between current assets and current liabilities. It measures a company's short-term liquidity and its ability to fund day-to-day operations.

Cash-flow Statement

- **Operating Cash Flow (OANCF):** This represents the cash generated from a company's core business operations. Operating Cash Flow is essential for assessing the sustainability of a company's operations.
- **Capital Expenditure (CAPX):** Represents investments in long-term assets such as buildings or equipment. Capital Expenditure reflects future growth potential but can also strain cash flow.
- **Net Debt Issued:** This measures the net change in debt issuance after repayments. Net Debt Issued shows the company's borrowing trends and reliance on external debt to finance its operations.
- **Net External Financing:** Net External Financing represents the amount of financing a company obtains from external sources like debt and equity. It is crucial for understanding how dependent the company is on outside capital.

Financial Ratios

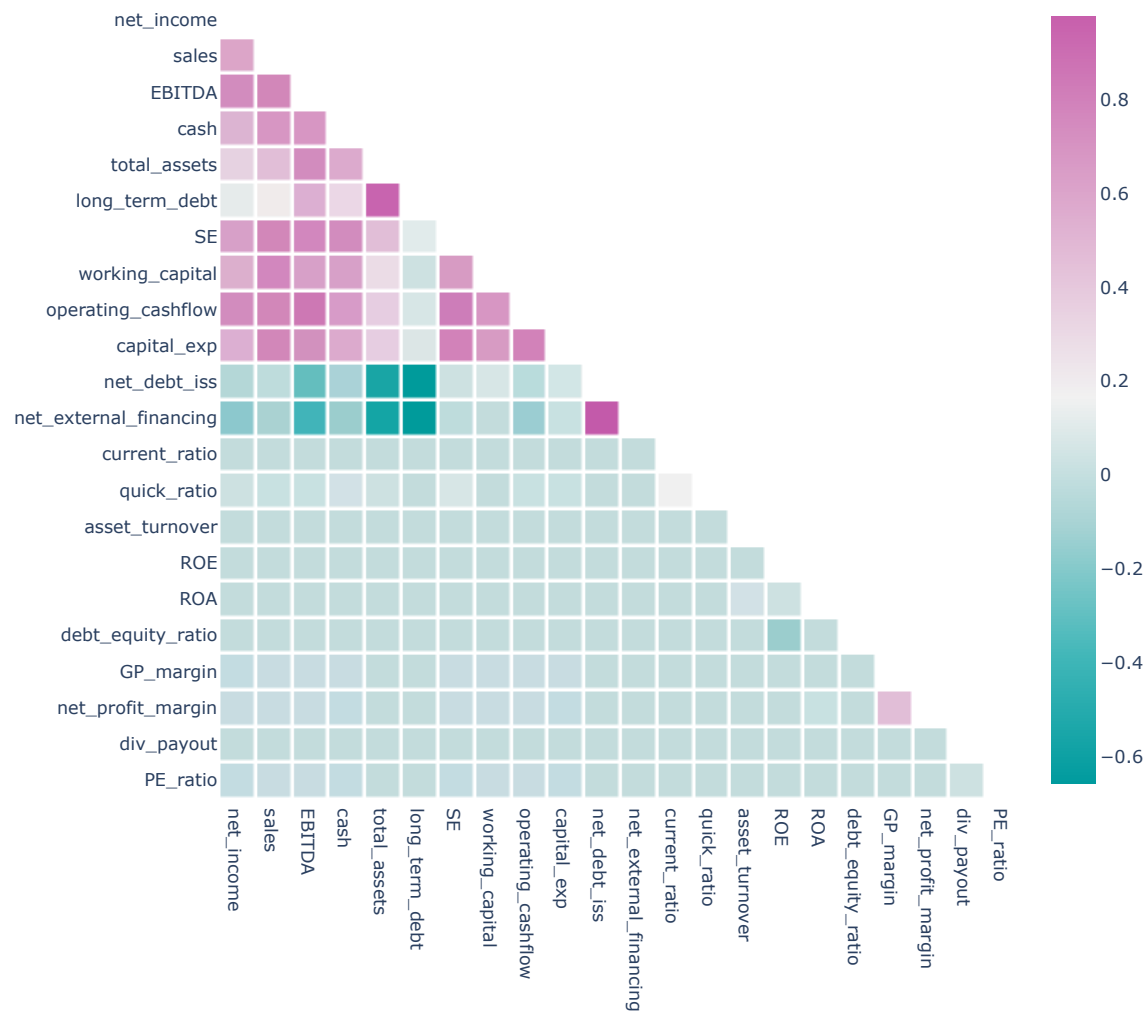
- **Current Ratio:** This is the ratio of current assets to current liabilities, which indicates a company's ability to pay off short-term obligations. A higher ratio suggests better liquidity.
- **Quick Ratio:** This is a stricter liquidity measure that excludes inventory. It provides a more conservative estimate of the company's ability to meet short-term liabilities with liquid assets.
- **Return on Assets:** This ratio shows how efficiently a company uses its assets to generate profit. ROA is important for assessing operational efficiency and overall performance.
- **Total Asset Turnover:** This indicates how efficiently a company uses its assets to generate sales. A higher turnover indicates better operational efficiency in asset utilization.
- **Return on Equity:** ROE measures how much profit a company generates with shareholders' equity. It's a key measure of profitability and management's effectiveness in using financial investments.
- **Debt-to-Equity Ratio:** This compares a company's total debt to shareholders' equity, showing the level of financial leverage and risk. A high ratio indicates heavy reliance on debt financing.
- **Gross Profit Margin:** This ratio shows the percentage of sales revenue that exceeds the cost of goods sold. Gross Profit Margin is crucial for understanding how efficiently a company manages its production costs relative to its revenue.
- **Net Profit Margin:** This ratio measures how much of a company's sales translate into actual profit after all expenses. A high margin signals strong profitability and cost control.
- **P/E Ratio:** The P/E ratio compares a company's stock price to its earnings per share. It indicates how much investors are willing to pay for each dollar of earnings, which is useful for assessing valuation.
- **Dividend Payout Ratio:** This ratio measures the percentage of net income paid out as dividends. It shows how much profit is returned to shareholders versus reinvested into the business.

Variable Construction

(175170, 24)

Quantitative Feature Selection

Correlation Matrix



Variables Re-Grouping

Profitability

- Net Income
- EBITDA
- Sales
- Operating Cash Flow
- Gross Profit Margin
- Net Profit Margin

Capital Structure

- cash
- Total Assets
- Long-Term Debt
- Shareholders' Equity
- Net Debt Issued
- Net External Financing
- Debt-to-Equity Ratio

Liquidity

- Current Ratio

- Quick Ratio
- Working Capital

Return Ratios

- Return on Equity (ROE)
- Return on Assets (ROA)

Operating Performance

- Total Asset Turnover
- Capital Expenditure

Valuation

- P/E Ratio
- Dividend Payout

Variables Dropping

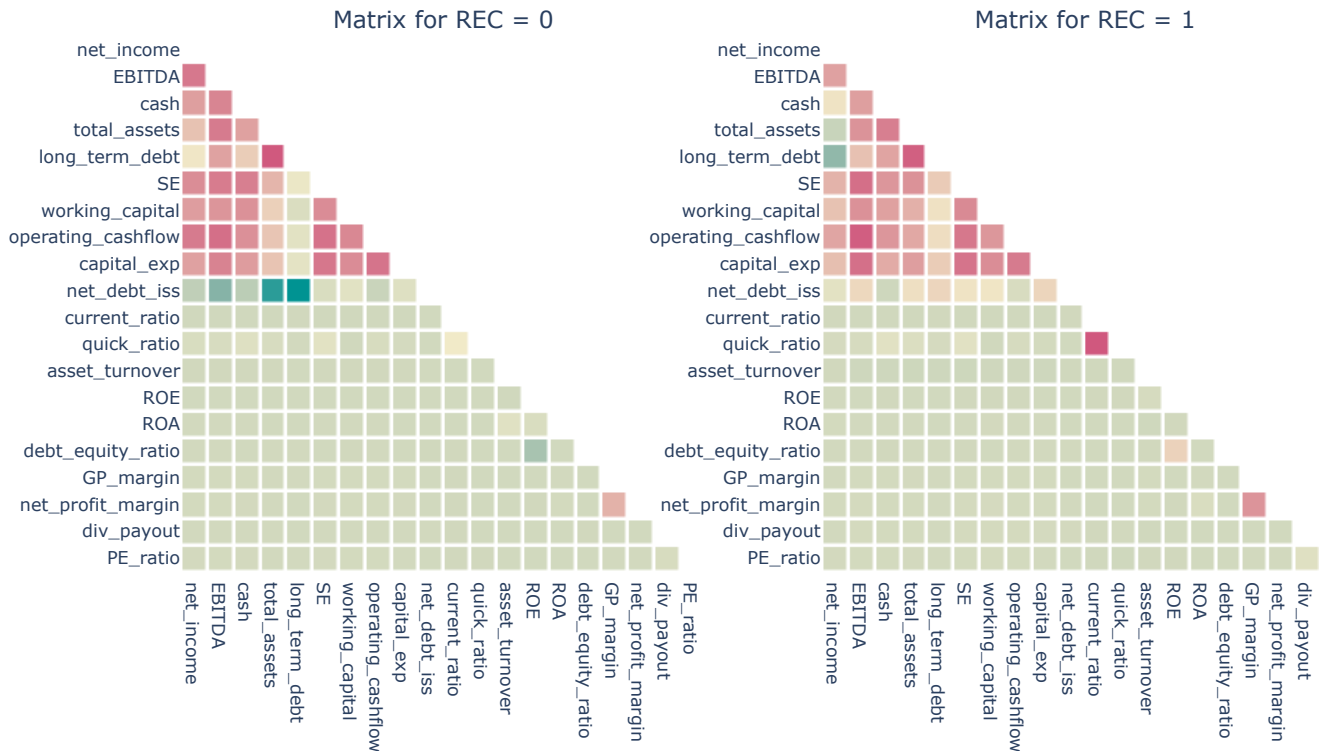
- **Sales:** Given Sales and EBITDA both reflects sales performance through operational income, and these two variables have a high correlation between each other (~ 0.8), I will drop Sales.
- **Net External Financing:** Given Net Debt Issued and Net External Financing are all closely related to a company's reliance on external sources of capital, and they have a high correlation (~ 1.0), I will drop Net External Financing to keep more focus on the Debt of the company.

(175170, 22)

Correlation Matrix for Group with different NBER recession data

Load NBER indicator

- Given NBER defined a recession as "a significant decline in economic activity that is spread across the economy and lasts more than a few months", a recession should last long enough to have a meaningful impact on economic performance. Since six months represent a significant portion of the year, I will categorize a year as recession year if it has more than or equal to six months in recession.
- If average of Yearly recession is greater than or equal to 0.5, we will consider it as NBER = 1



Observations

- Most of the correlation relationships between variables stays the same, but there are still some relationships behaves significantly different for recession year and non-recession year.
- Cash v.s. Net Income:** Correlation changes from positively correlated to not that correlated. During recessions, corporates often face declining revenues result in lower net income. Despite the fall in net income, corporates may still preserve or even increase their cash reserves due to a conservative cash management, or/and more access to external financing, or/and a cost-cutting strategies.
- Total Assets v.s. Net Income:** Correlation changes from positively correlated to not that correlated. This might due to the largely unchange total assets and a declined net income.
- Total Assets v.s. Net Debt Issued:** Correlation changes from negatively correlated to positively correlated. This might because companies increasingly rely on debt to maintain or grow their asset base when internal cash flows and equity financing become constrained. In a recession, companies may issue debt to survive, leading to a stronger relationship between assets and debt levels.
- Long Term Debt v.s. Net Debt Issued:** Correlation changes from negatively correlated to positively correlated. During non-recession period, companies are able to pay off or reduce their long-term debt bec ause of their high profitability. However, during recession, net debt issued will increase, and because of companies lack of profits, long term debt will also accumuate, leading to a positive correlation.
- Current Ratio v.s. Quick Ratio:** Correlation changes from not that related to strongly positively related. During a recession, companies focus on preserving liquidity, reducing inventory levels, and tightly managing receivables and payables. These factors cause both the current ratio and quick ratio to move in the same direction, creating a stronger positive correlation between the two. In normal conditions, inventory plays a larger role, keeping the two ratios more distinct and leading to a neutral relationship.

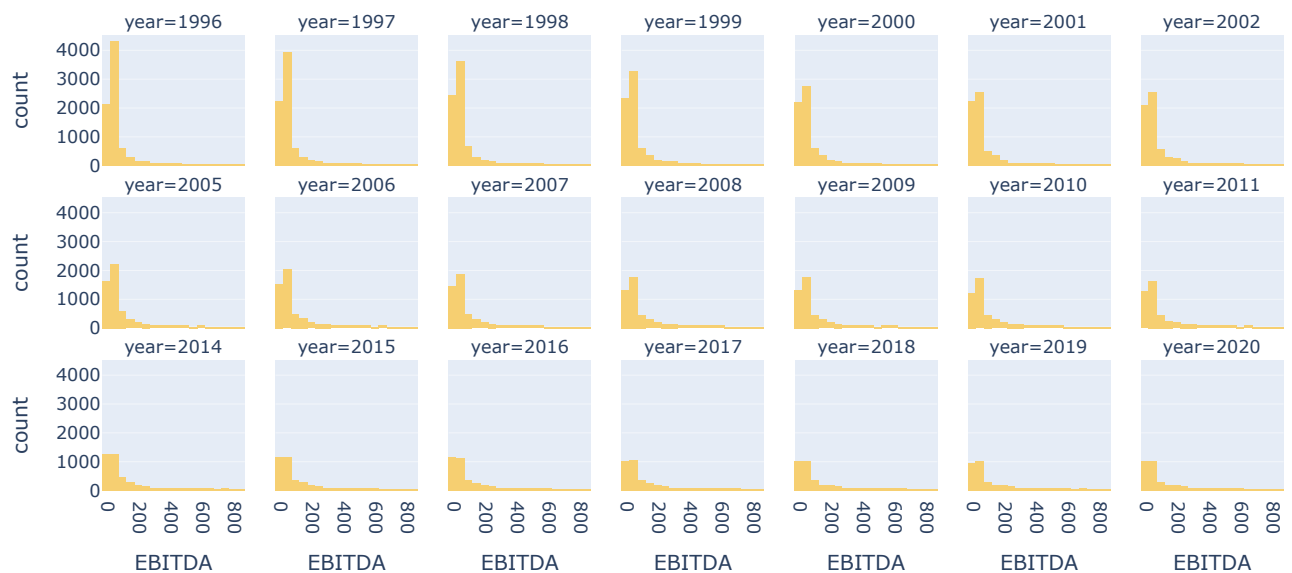
Faceted Histograms

Due to some extreme outliers and the fact that except for these outliers most of data are centered and clustered at the middle part, the following Faceted Histograms will only cover central 80% interval of the data to better reflect the pattern for the whole market.

Faceted Histogram for net_income



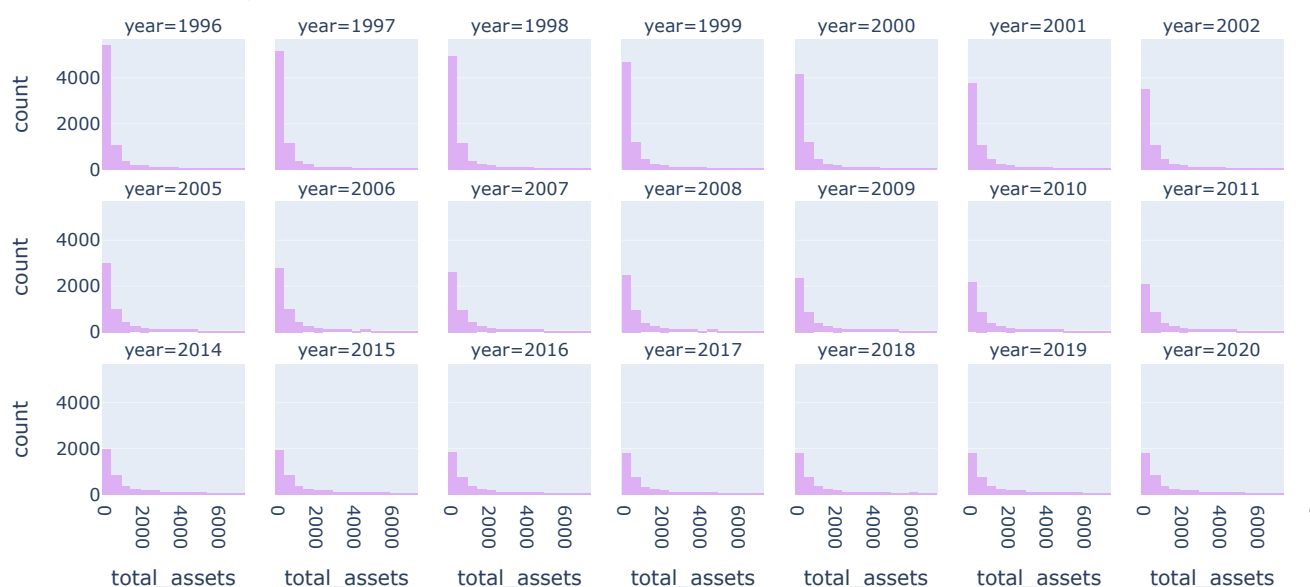
Faceted Histogram for EBITDA



Faceted Histogram for cash



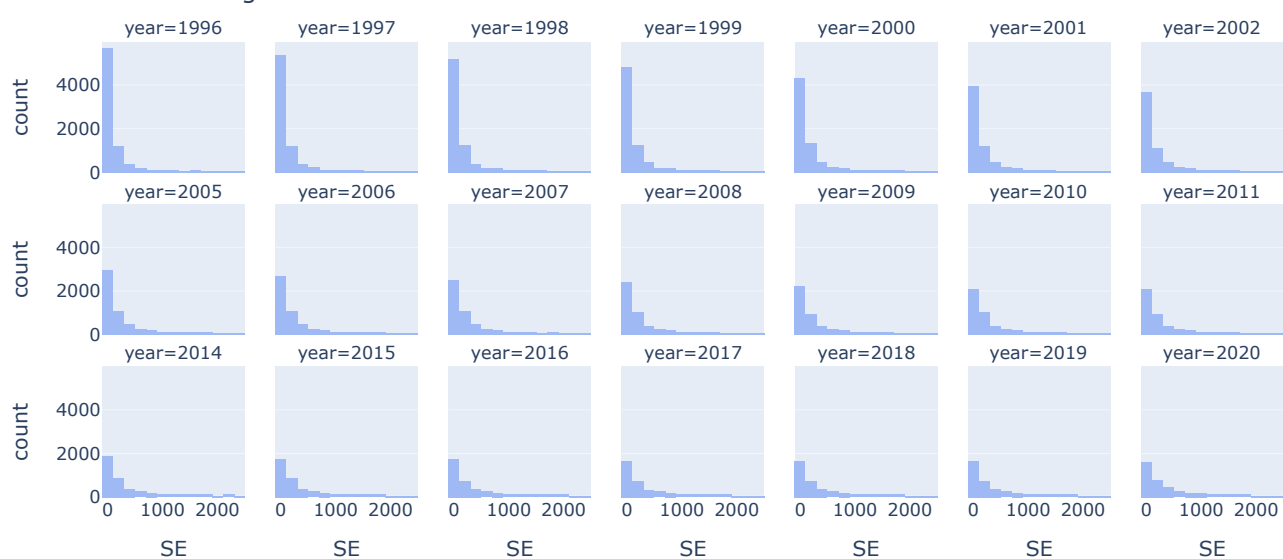
Faceted Histogram for total_assets



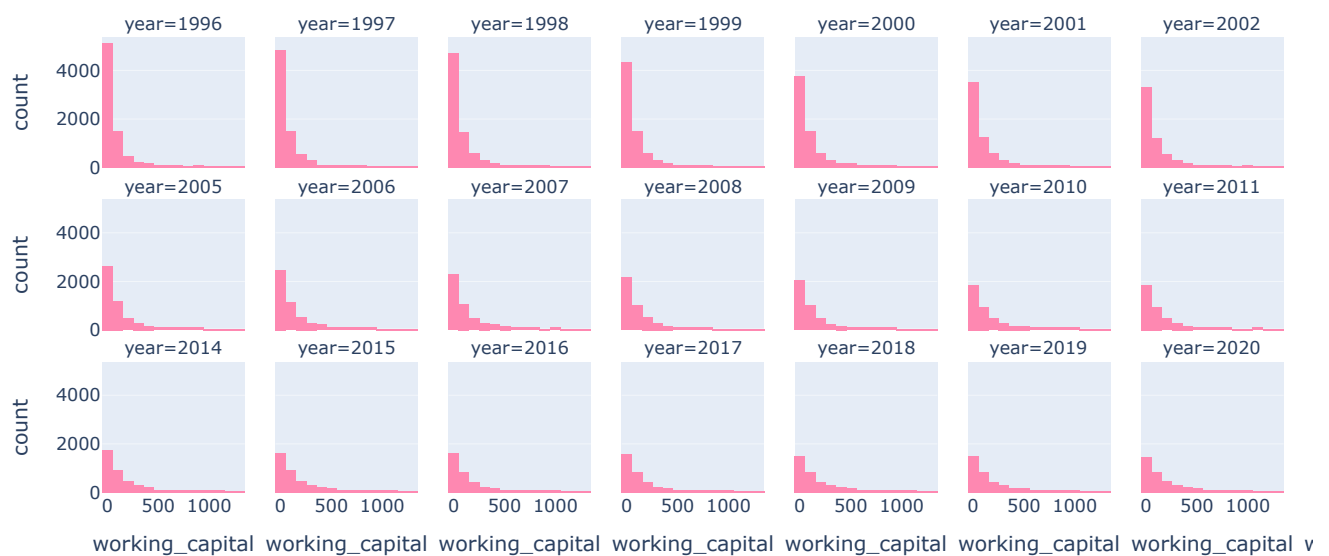
Faceted Histogram for long_term_debt



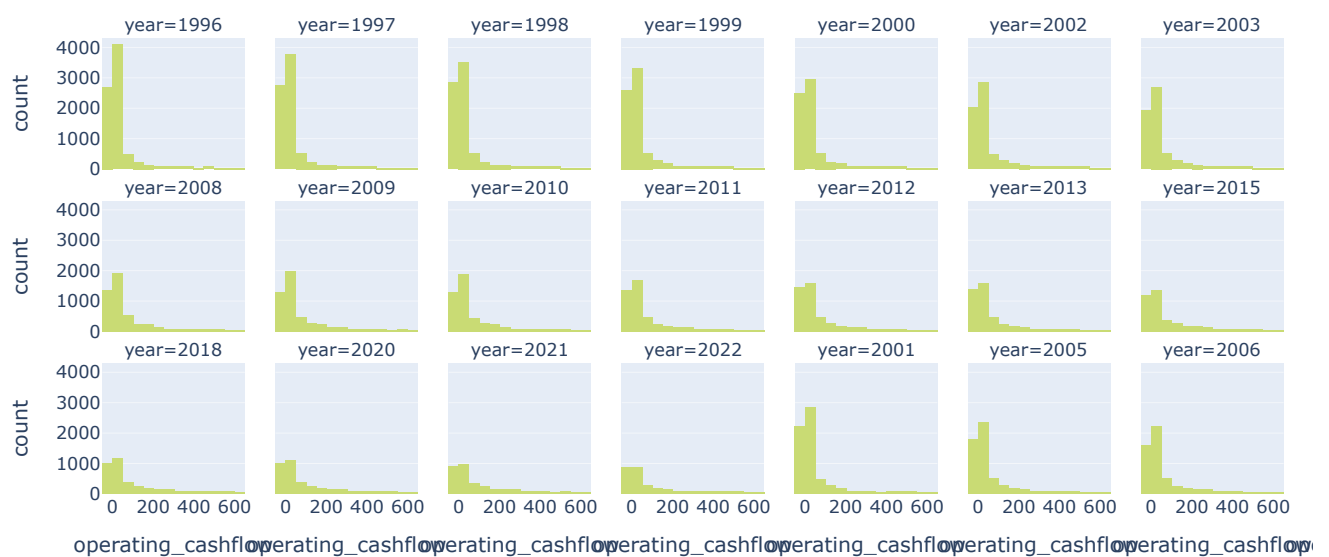
Faceted Histogram for SE



Faceted Histogram for working_capital



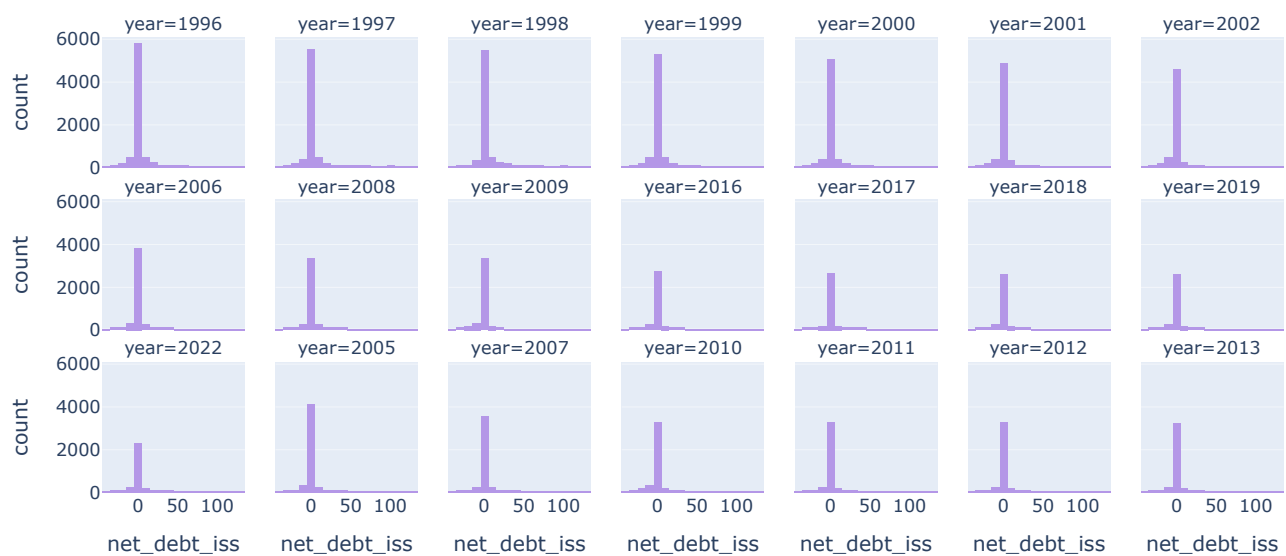
Faceted Histogram for operating_cashflow



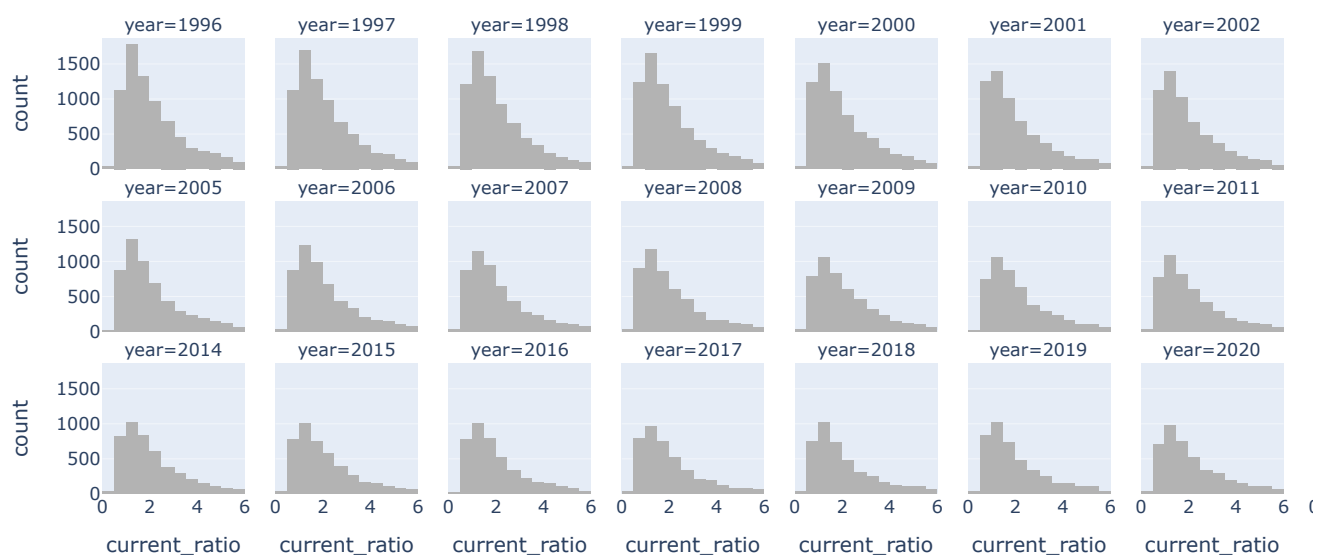
Faceted Histogram for capital_exp



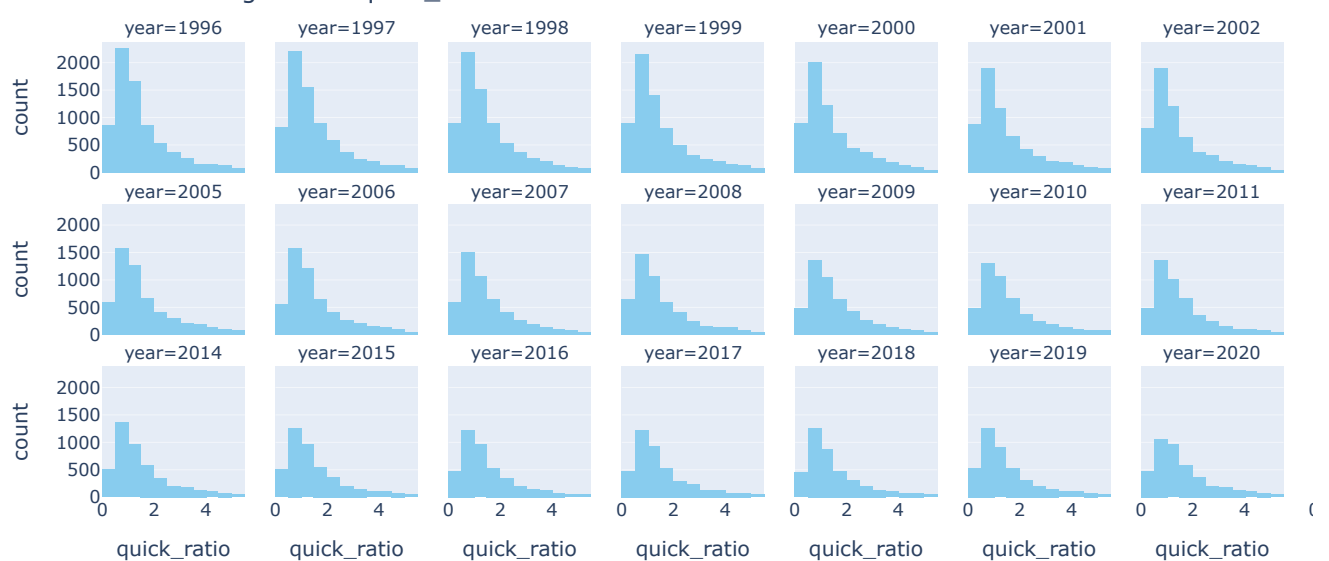
Faceted Histogram for net_debt_iss



Faceted Histogram for current_ratio



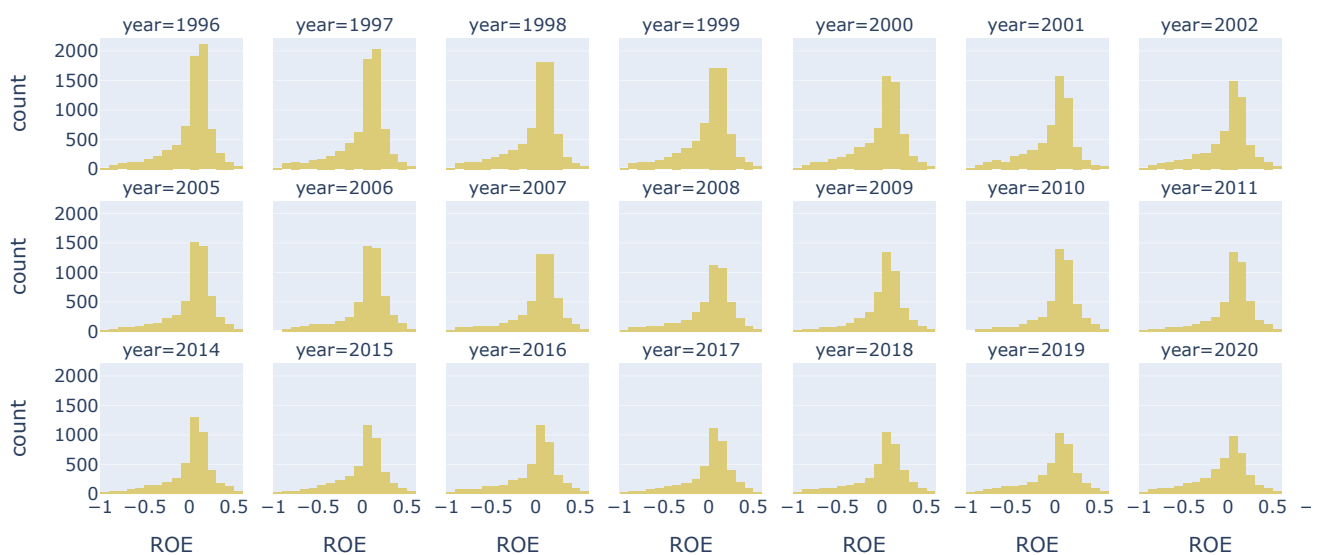
Faceted Histogram for quick_ratio



Faceted Histogram for asset_turnover



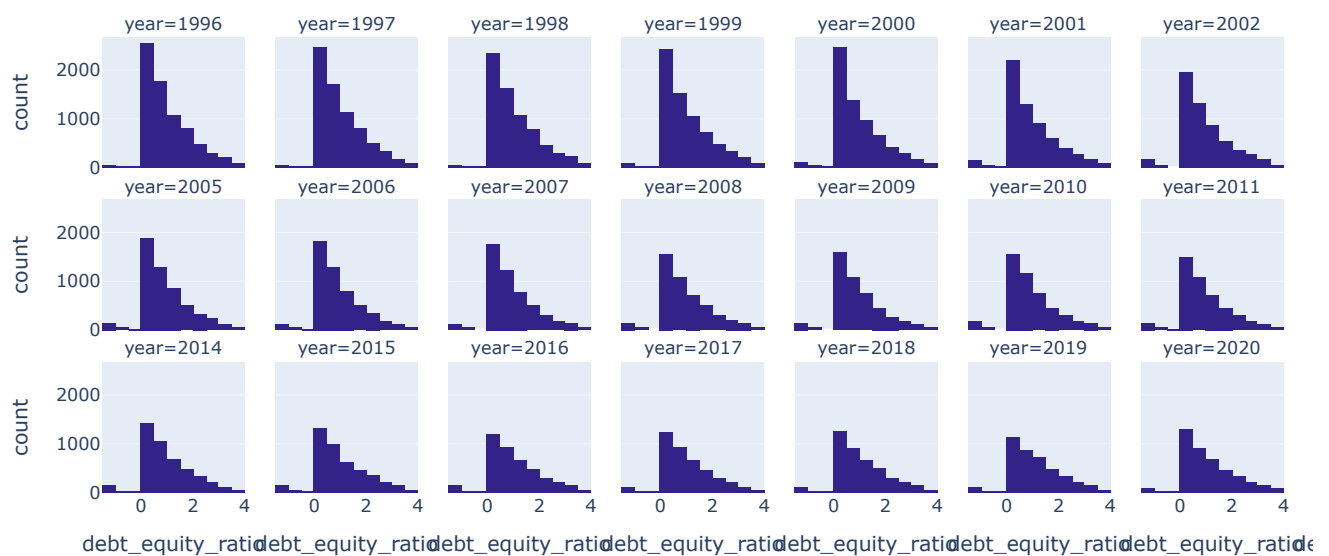
Faceted Histogram for ROE



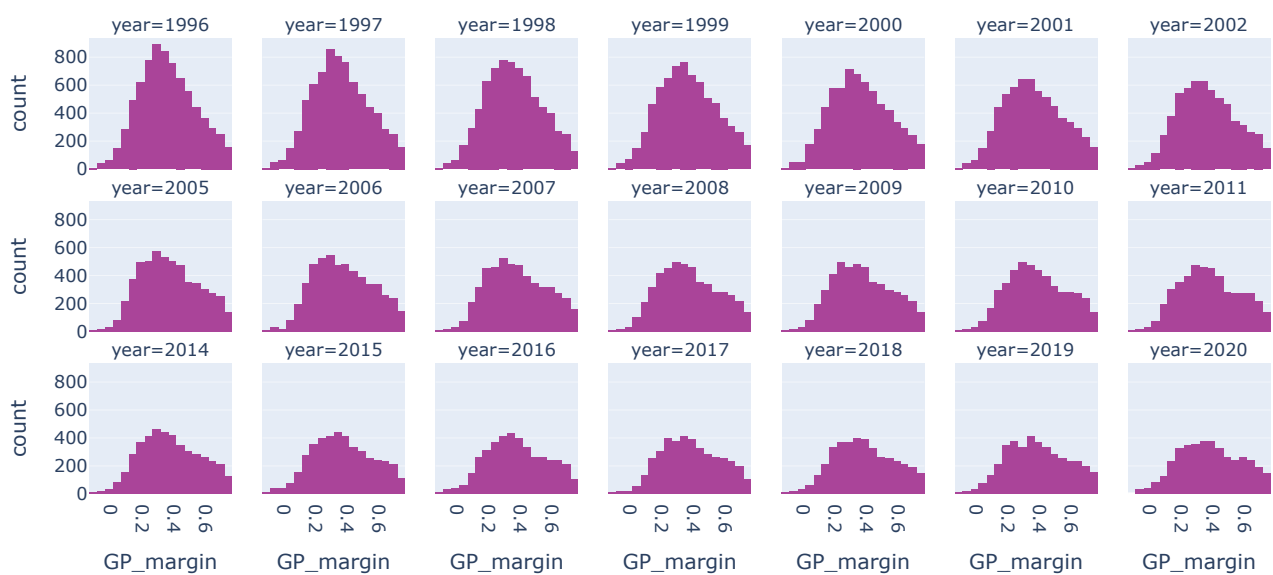
Faceted Histogram for ROA



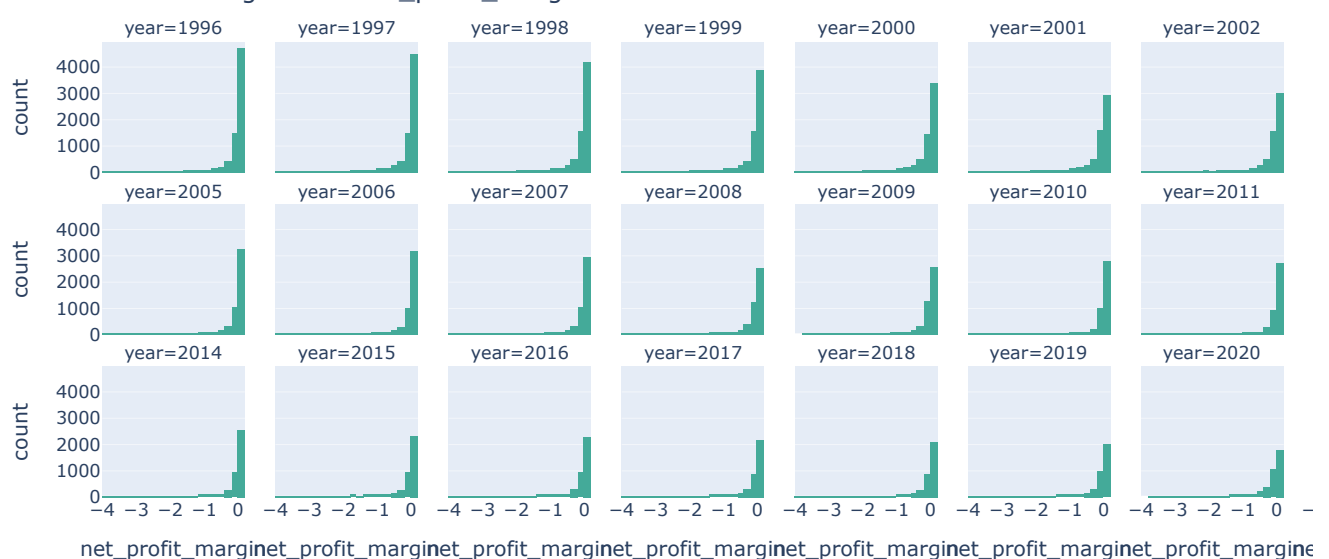
Faceted Histogram for debt_equity_ratio



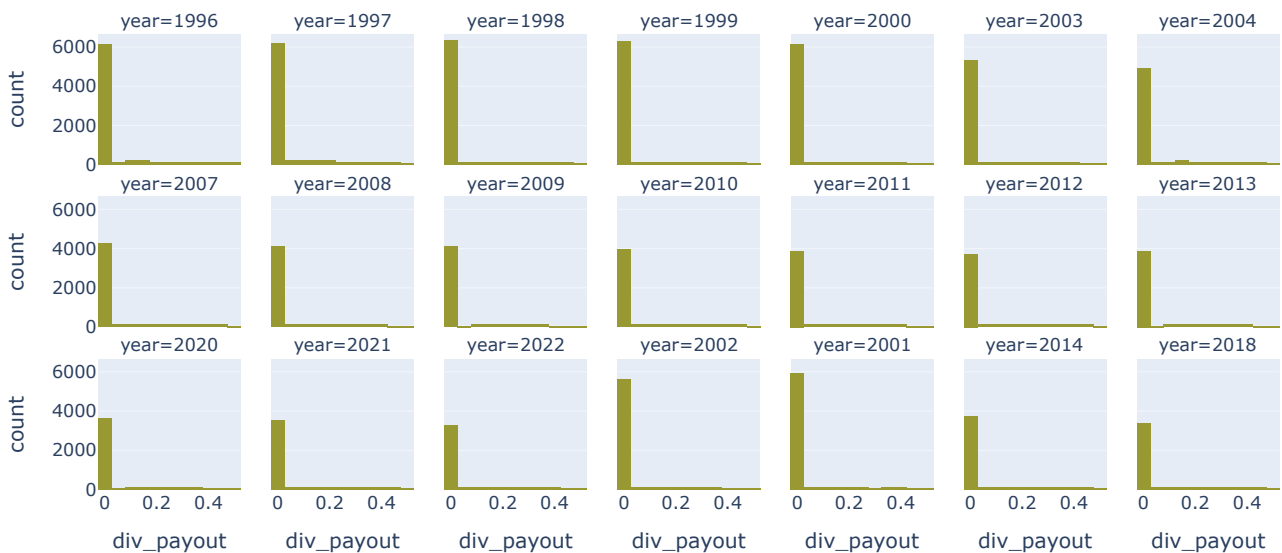
Faceted Histogram for GP_margin



Faceted Histogram for net_profit_margin



Faceted Histogram for div_payout



Faceted Histogram for PE_ratio



Histogram Analysis

- Net Income:** The distributions of net income are right-skewed from 1996 to 2022 with the most corporates stay at the center part with a net income of -20 to 20. As time pass, it seems that the total number of corporates decrease and the net income distribution slightly spread out, with higher percentage of corporates having net income above 20 or lower than -20. The right-skewed distribution indicates that a large proportion of companies have net income near zero, with fewer companies showing high profits. Over time, the total number of corporates decreases, likely due to market consolidation or companies exiting the market. As the distribution spreads out, more companies have net income either above 20 or below -20, indicating growing disparities in profitability. This suggests that while some firms become more profitable, others experience larger losses, reflecting increasing polarization in corporate financial performance.
- EBITDA:** The distribution of EBITDA are also right skewed with most data located around 0, showing most of companies had a break-even EBITDA, but a few performed better with a positive EBITDA. As time goes, the number of corporates decrease, and the distribution is more right-skewed with fewer in center, indicating the corporates survive in the market had a better EBITDA, and those that can barely break-even probably experienced bankrupts or absorbed by other companies.
- Cash, Total Assets, Long Term Debt, Stockholders' Equity, Working Capital, and Operating Cashflow, Capital Expenditure** mainly follows a same right-skewed distribution shape, with most data centered around 0, and a same changing direction as EBITA and Net Income with a large decrease in center part and more spread out to positive part. This similarities are probably due to their high correlations with each other.

- **Net Debt Issued:** The distributions of net debt issued are symmetric with most corporates stay at the center part with a net debt issued of -5 to 5 and a slight right skewed. Such right-skewed distribution pattern continues while having a decreased center part. The right skew in the distribution suggests that a smaller number of firms are issuing significantly more debt compared to those repaying, leading to more outliers on the higher end. Over time, the center of the distribution decreases, meaning fewer companies have minimal debt issuance or repayment. This could reflect increased reliance on debt by certain firms or greater variability in financing strategies.
- **Current Ratio:** The distributions of Current Ratio are right-skewed with most data centered between 0.5 to 1.5. As time passed the right-skewed pattern is maintained but become more spread out with less proportion of data located between 0.5 to 1.5 and more proportion of data located above 1.5. The right-skewed distribution of the Current Ratio, with most data initially centered between 0.5 to 1.5, suggests that the majority of companies had just enough liquid assets to cover their short-term liabilities. Over time, while the right-skew remains, the distribution has spread out, indicating a shift toward stronger liquidity. This could signal improved financial health or more conservative liquidity strategies over time.
- **Quick Ratio:** The distributions of Quick Ratio and the change of such distributions looks similar with the distribution of the Current Ratio, but with a obvious portion of data points located below 0.5. This distribution indicates that majority of firms could comfortably cover their short-term liabilities without relying on inventory. Over time, the distribution spreads out, with fewer companies falling within the 0.5 to 1.5 range and more showing Quick Ratios above 1.5, suggesting improved liquidity for a portion of firms. However, a significant proportion of companies also exhibit Quick Ratios below 0.5, highlighting liquidity challenges for some firms. The Quick Ratio and Current Ratio distributions are similar because both measure liquidity, with the Quick Ratio being a stricter measure since it excludes inventory.
- **Asset Turnover:** The distributions of Asset Turnover changed from nearly uniformly distributed between 0 to 2 into a more right skewed pattern as time passed, with more data located below 0.5. In the earlier period, companies showed a more diverse range of efficiency in asset utilization, with many firms having higher asset turnover ratios. Over time, the shift towards more data clustering below 0.5 indicates a growing number of firms are generating less revenue per unit of assets, reflecting lower operational efficiency.
- **ROE & ROA:** The distributions of ROE and ROA are all obviously left-skewed with most data centered around 0 to 0.2. As time pass, the distributions become more spread out with less data located around 0 to 0.2, and more proportion of data located below 0. This distribution indicates that most companies have moderate profitability. As time passes, the distribution spreads out, with fewer companies maintaining ROE and ROA in this range and more companies showing negative values. This shift suggests an increasing proportion of companies are experiencing losses or declining profitability over time, showing growing polarization in financial performance.
- **Debt Equity Ratio:** The distributions of Debt Equity Ratio are right skewed with most data located between 0 and 0.5 and having a consistent small tails on the left for ratios between -1 to -1.5. As time pass the distribution spread out, with less data located at the 0 to 0.5. Such distribution indicates that many firms rely more on equity than debt. The small left tail represents companies with negative equity, possibly due to significant losses or accumulated debt. As time progresses, the distribution spreads out, meaning more companies adopt higher debt levels, leading to fewer firms in the 0 to 0.5 range. This reflects a shift where some companies take on more leverage.
- **GP Margin:** The distribution of GP margin are mainly symmetric with most data located around 0.3. As time passes, such distribution becomes more spreadout with larger proportion of data located above 0.4. The distribution of Gross Profit margin being mainly symmetric around 0.3 indicates that, for most companies, 30% of sales contribute to covering operational and overhead costs, reflecting consistent profitability in earlier periods. As time passes, the distribution spreads out, with a larger proportion of companies having GP margins above 0.4. This shift suggests that more firms are improving their production efficiency or pricing power, allowing them to retain a higher percentage of sales as gross profit.
- **Net Profit Margin:** The distributions of Net Profit Margin are significantly left-skewed with most data located between 0 to 0.2 and remaining data are all below 0. As time passes, distribution is more spread out, with less companies having net income margin between 0 to 0.2, making the negative proportion become larger. This shift suggests that as time passes, more companies are struggling with profitability, leading to an increasing number of firms experiencing losses.
- **Dividend Payout:** The distributions of Dividend Payout are right skewed with more than 90% of data located between -0.025 to 0.025. Such pattern maintains as time passes with a decrease in the amount of data located between -0.025 to 0.025. The skewness suggests that the majority of firms either pay minimal dividends or reinvest their earnings and a small number of companies pay significantly higher dividends compared to the majority.
- **PE ratio** The distributions of PE ratio are bimodal with few data located between 0 and 5, while having a left-skewed distribution below 0 and a right skewed distribution above 5. Most data located either between 0 and -5 or between 5 to 15. As time passes, the shape of such distributions maintained with a overall decrease across all kind of PE ratio. The gap between two clusters suggests that companies are either undervalued or moderately valued, but very few companies have P/E ratios close to 0, meaning few companies have very low or near zero earnings. The concentration of data between 0 and -5 reflects companies that are just slightly unprofitable, suggesting potential financial struggles, while more extreme negative P/E ratios might be rarer. The right

skewed distribution on the right suggests that most companies are in a mature phase where their earnings are reasonably priced by the market.

Feature Scaling

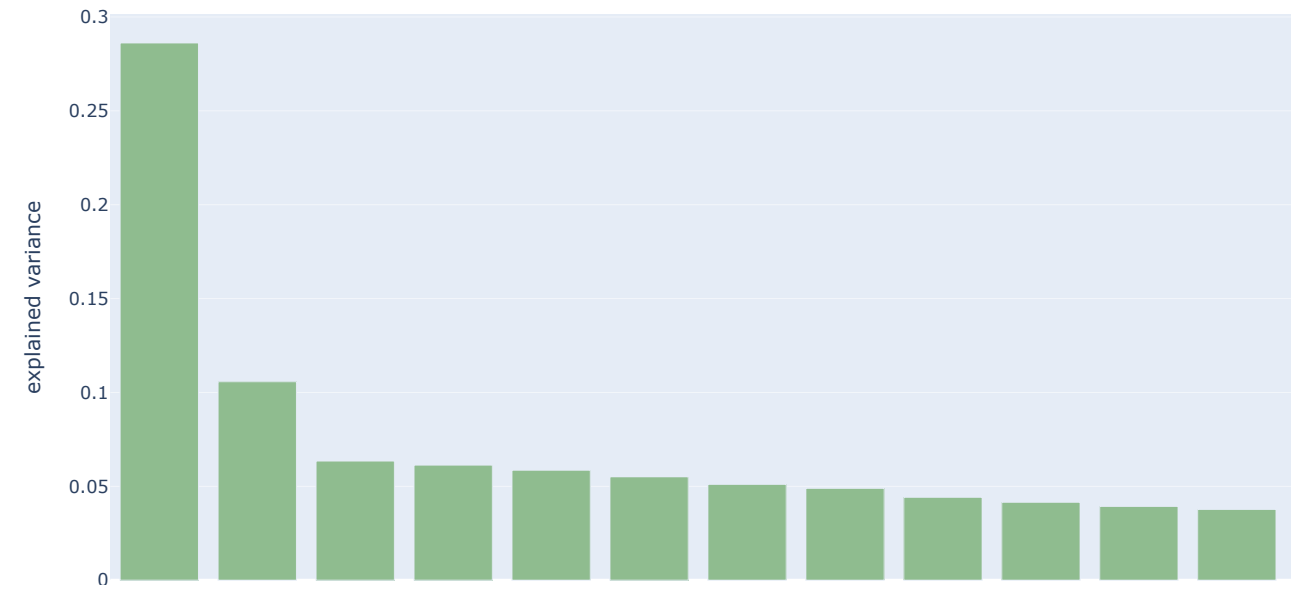
Since the Feature Scaling is performed to prepare for the following PCA analysis, we will use normalization (StandardScaler) to maintain the variance of each variables, so that the PCA result which indicates the percentage of variance each variables contribute to the total variance will make sense.

Principal Component Analysis

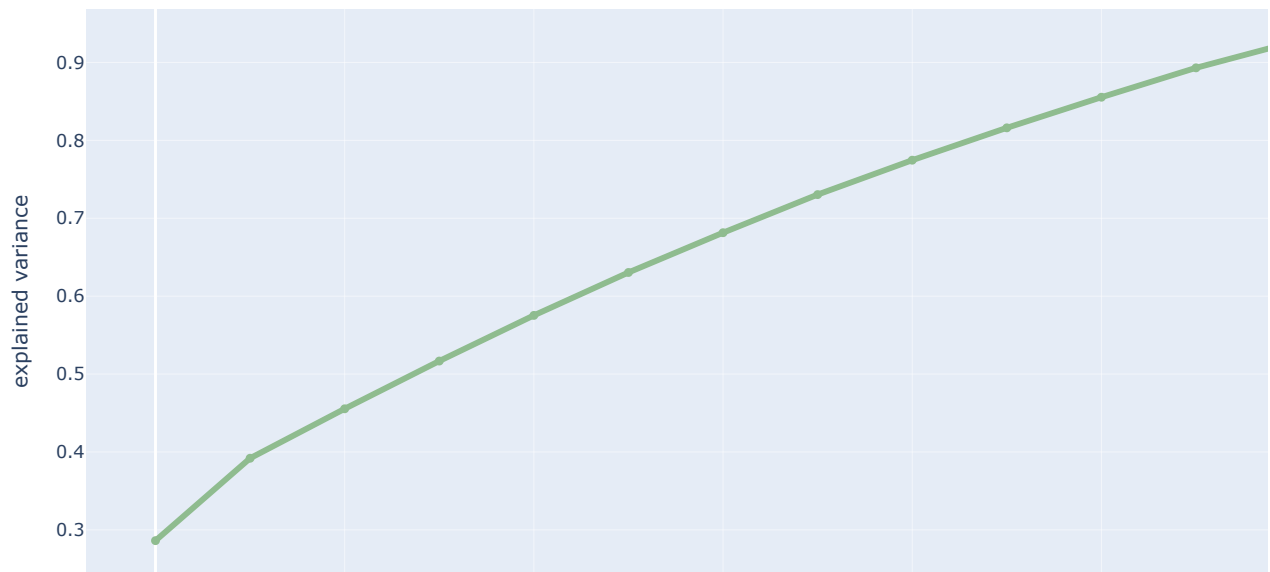
Number of components cover 90% variance is 13 (Precise Coverage: 92.54 %)
This is the array of how much each of them cover:
[0.28618151 0.10573623 0.06348596 0.06131397 0.0585829 0.05508251
0.05105293 0.04892539 0.04416151 0.04150996 0.03930828 0.03770627
0.03233681]

```
{'EBITDA',  
 'GP_margin',  
 'PE_ratio',  
 'ROA',  
 'ROE',  
 'asset_turnover',  
 'current_ratio',  
 'div_payout',  
 'long_term_debt',  
 'net_debt_iss',  
 'quick_ratio'}
```

Explained Variance for PCA



Cumulative Explained Variance for PCA



Conclusion

- According to the number of pca model components, 13 principal components are enough to cover 90% of the variance in the data.
- Moreover, after finding the most important features for each components, it seems that **EBITDA, GP Margin, PE ratio, ROE, ROA, Asset Turnover, Curret Ratio, Dividend Pauout, Long Term Debt, Net Debt Issued, Quick Ratio** are the most important features based on the result of this PCA analysis.