Key Financial Indicators: Unlocking Stock Performance Insights Through Explanatory Data Analysis

Radjamin Hukom

rhukom@bellarmine.edu

2/7/2025

About the Dataset

	Beta	EPS (2019)	ROE (2019)	ROA (2019)	Gross Margin (2019)	Operating Margin (2019)	Net Profit Margin (2019)	Cash Ratio (2019)	Current Ratio (2019)	Quick Ratio (2019)	...	Gross Margin (2023)	Operating Margin (2023)	Net Profit Margin (2023)	Cash Ratio (2023)	Current Ratio (2023)	Quick Ratio (2023)	A-Turnover (2023)	I-Turnover (2023)	R-Turnover (2023)	% change, Price
count	503.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	...	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000
mean	0.797630	118.382686	-0.153562	0.067533	0.424493	0.143821	0.112534	0.571659	1.511457	1.141881	...	35.789144	16.426062	13.066706	0.527319	1.404682	0.892780	0.680264	8.436862	11.903914	0.776320
std	0.584612	2523.275699	5.609127	0.079839	0.266557	0.442281	0.411438	1.262977	1.434420	1.157206	...	26.779962	17.497430	17.264114	1.114254	1.480181	1.288385	0.613431	23.766885	28.616814	1.634994
min	-0.393969	-9.604000	-104.043480	-0.363637	0.000000	-9.116667	-8.566667	0.000000	0.000000	0.000000	...	-29.881300	-93.821700	-96.559400	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.841800
25%	0.377332	1.739999	0.076486	0.024370	0.232857	0.081935	0.065771	0.083174	0.768950	0.503162	...	12.362475	8.923800	5.986450	0.055775	0.749650	0.269900	0.301950	0.000000	4.633000	0.050575
50%	0.728988	3.685285	0.143494	0.055565	0.424897	0.149117	0.117623	0.253207	1.212050	0.901789	...	35.311100	16.220200	12.056300	0.228000	1.136750	0.673150	0.540800	2.832150	6.583650	0.413550
75%	1.084444	6.253253	0.254951	0.102909	0.639748	0.220374	0.190601	0.588211	1.827400	1.415405	...	57.851425	24.378550	19.393800	0.578875	1.668925	1.056275	0.824300	5.818025	10.561750	0.977675
max	4.982928	56539.582000	20.230770	0.532833	1.000000	0.876621	0.649807	19.666666	15.330200	13.173799	...	93.116700	101.644300	101.236000	16.337400	19.430600	19.243000	4.744400	268.852500	381.935000	21.838400

8 rows × 62 columns

Ratio	Meaning	E(sign)
Beta	Indicates the stock's risk relative to the market	-
EPS	Shows profitability per share	+
ROE	Measures how efficiently equity generates profit	+
ROA	Reflects how effectively assets generate profit	+
Gross Margin	Reveals profitability after production costs	+
Operating Margin	Indicates how well operations generate profit before interest and taxes.	+
Net Profit Margin	Shows overall profitability after all expenses	+
Cash Ratio	Assesses short-term liquidity	+
Current Ratio	Measures short-term financial health	+
Quick Ratio	Evaluates liquidity without relying on inventory	+
Asset Turnover	Shows how efficiently assets generate revenue	+
Inventory Turnover	Reflects how quickly inventory is sold	+
Receivables Turnover	Measures how fast credit sales are collected	+

The dataset was found first through the usage of finsheet.io, a free tool with pay-to-use add-ins to export financial data to excel and google sheets. However, it became clear that it is not powerful enough to handle the volume of requests needed to build this dataset. The rest of the data was found through Bloomberg, a well-respected software used by financial specialists all around the world for years that the university has paid for its services and API (Application Programming Interface) with the help of a donor. The table above gives an explanation of the meaning of each financial metric recorded in the dataset and its expected sign based on theory of its correlation with the response variable, the percentage change in price of the ticker symbol's stock over the period of 2019 to 2024.

The dataset consists of 503 observations with Ticker + Beta + 12 financial ratios * 5 years = 62 financial variables, providing a comprehensive overview of firm performance across multiple years. Including the response variable, % Price Change, we have 63 columns and 503 rows of stocks from S&P500. 503 ≠ 500 as some companies in the S&P500 may have more than one ticker. The mean values of key financial ratios indicate the overall industry trends, such as an average EPS of 118.38 and an ROA of 0.0675, suggesting that firms generally maintain positive, albeit modest, returns on assets. However, the standard deviations for these variables, especially EPS (2523.28) and ROE (5.61), highlight substantial variability among firms, pointing to a mix of highly profitable companies and those experiencing significant financial struggles. The presence of extreme values in these variables suggests that

some firms generate exceptionally high earnings, while others report heavy losses, contributing to the dataset's high skewness.

Examining the quartiles further supports this variability. The median EPS is only 3.69, and the 75th percentile is 6.25, yet the maximum EPS reaches an astounding 56,539.58, indicating that a few firms are driving the high mean value. Similar patterns emerge for liquidity and turnover ratios. For instance, the cash ratio ranges from 0 to 19.67, with a median of 0.25, reflecting that while most firms maintain conservative liquidity levels, a few hold unusually high cash reserves. Additionally, turnover ratios such as Inventory Turnover (I-Turnover) and Receivables Turnover (R-Turnover) exhibit large spreads, with maximum values reaching 268.85 and 381.94, respectively, suggesting that certain firms operate with highly efficient turnover cycles while others struggle with slow-moving assets.

The dataset's extreme minimum and maximum values further highlight the presence of outliers. Some firms report highly negative financial metrics, such as an ROE as low as -104.04 and a net profit margin reaching -96.56, indicating significant financial distress. On the other hand, certain firms exhibit extraordinarily high performance, with operating margins peaking at 101.64 and net profit margins at 101.24, which may be anomalies or cases of exceptionally profitable entities. The presence of such disparities suggests that data transformation techniques, such as normalization or winsorization, may be necessary to ensure a more balanced dataset suitable for predictive modeling.

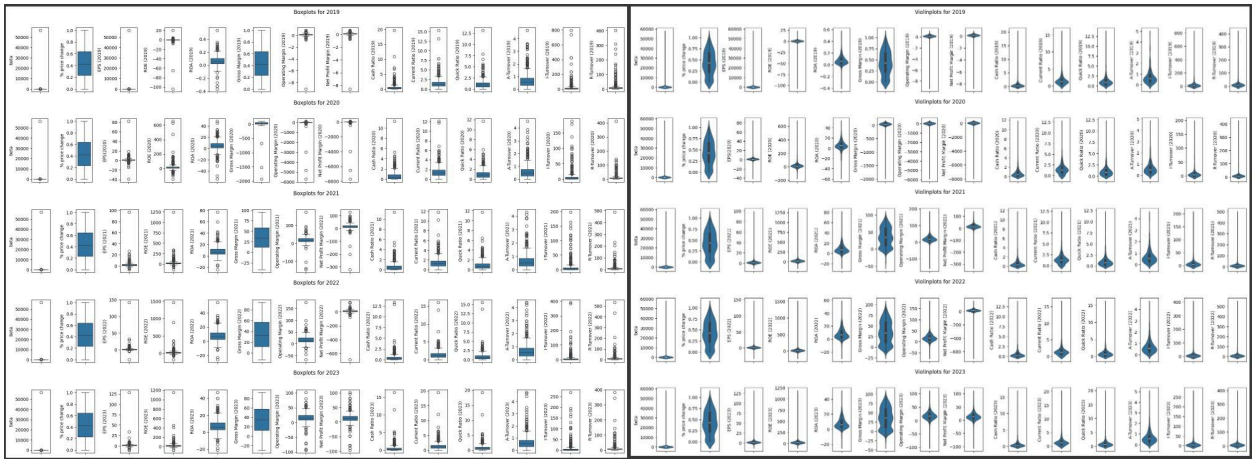
On the Data Values

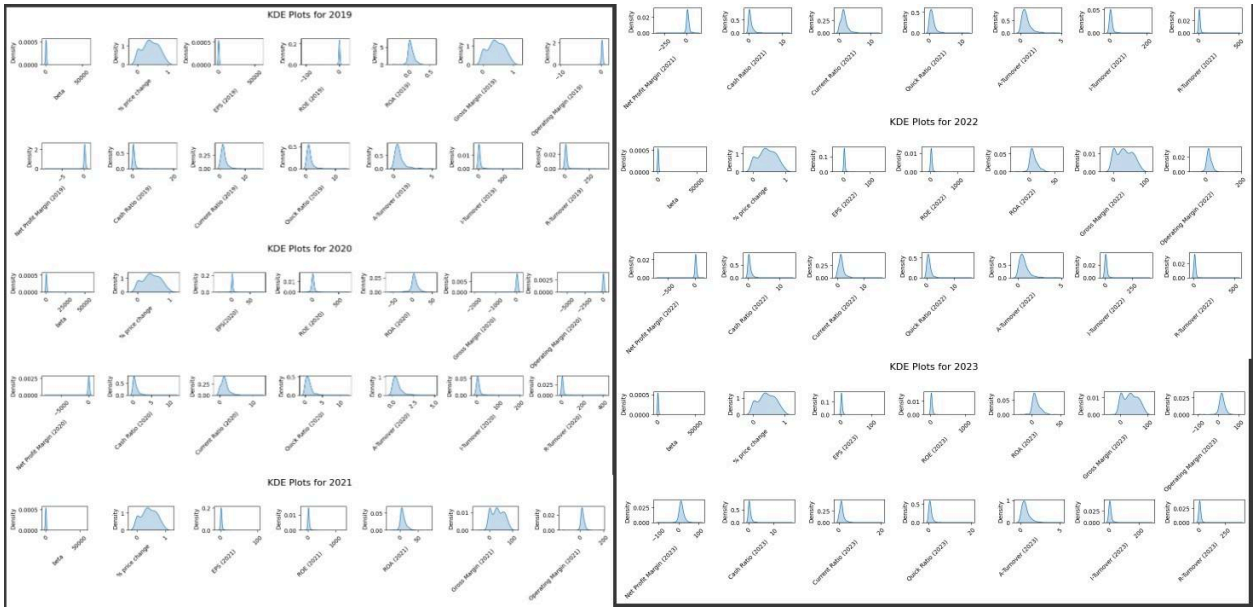
	Year	Column	Column Type	Outlier Count	Outlier Percentage
0	2019	EPS (2019)	ratio	38	7.54%
1	2019	ROE (2019)	ratio	67	13.29%
2	2019	ROA (2019)	ratio	29	5.75%
3	2019	Gross Margin (2019)	ratio	0	0.00%
4	2019	Operating Margin (2019)	ratio	29	5.75%
5	2019	Net Profit Margin (2019)	ratio	28	5.56%
6	2019	Cash Ratio (2019)	ratio	38	7.54%
7	2019	Current Ratio (2019)	ratio	37	7.34%
8	2019	Quick Ratio (2019)	ratio	35	6.94%
9	2019	A-Turnover (2019)	ratio	37	7.34%
10	2019	I-Turnover (2019)	ratio	60	11.90%
11	2019	R-Turnover (2019)	ratio	57	11.31%
12	2020	EPS(2020)	ratio	50	9.92%
13	2020	ROE (2020)	ratio	56	11.11%
14	2020	ROA (2020)	ratio	44	8.73%
15	2020	Gross Margin (2020)	ratio	3	0.60%
16	2020	Operating Margin (2020)	ratio	37	7.34%
17	2020	Net Profit Margin (2020)	ratio	47	9.33%
18	2020	Cash Ratio (2020)	ratio	27	5.36%
19	2020	Current Ratio (2020)	ratio	29	5.75%
20	2020	Quick Ratio (2020)	ratio	29	5.75%
21	2020	A-Turnover (2020)	ratio	29	5.75%
22	2020	I-Turnover (2020)	ratio	46	9.13%
23	2020	R-Turnover (2020)	ratio	48	9.52%
24	2021	EPS (2021)	ratio	30	5.95%
25	2021	ROE (2021)	ratio	43	8.53%
26	2021	ROA (2021)	ratio	24	4.76%
27	2021	Gross Margin (2021)	ratio	0	0.00%
28	2021	Operating Margin (2021)	ratio	22	4.37%
29	2021	Net Profit Margin (2021)	ratio	29	5.75%
30	2021	Cash Ratio (2021)	ratio	36	7.14%
31	2021	Current Ratio (2021)	ratio	28	5.56%
32	2021	Quick Ratio (2021)	ratio	27	5.36%
33	2021	A-Turnover (2021)	ratio	31	6.15%
34	2021	I-Turnover (2021)	ratio	47	9.33%
35	2021	R-Turnover (2021)	ratio	48	9.52%
36	2022	EPS (2022)	ratio	36	7.14%
37	2022	ROE (2022)	ratio	39	7.74%
38	2022	ROA (2022)	ratio	25	4.96%
39	2022	Gross Margin (2022)	ratio	0	0.00%
40	2022	Operating Margin (2022)	ratio	22	4.37%
41	2022	Net Profit Margin (2022)	ratio	30	5.95%
42	2022	Cash Ratio (2022)	ratio	44	8.73%
43	2022	Current Ratio (2022)	ratio	32	6.35%
44	2022	Quick Ratio (2022)	ratio	25	4.96%
45	2022	A-Turnover (2022)	ratio	33	6.55%
46	2022	I-Turnover (2022)	ratio	54	10.71%
47	2022	R-Turnover (2022)	ratio	43	8.53%
48	2023	EPS (2023)	ratio	46	9.13%
49	2023	ROE (2023)	ratio	35	6.94%
50	2023	ROA (2023)	ratio	26	5.16%
51	2023	Gross Margin (2023)	ratio	0	0.00%
52	2023	Operating Margin (2023)	ratio	23	4.56%
53	2023	Net Profit Margin (2023)	ratio	35	6.94%
54	2023	Cash Ratio (2023)	ratio	41	8.13%
55	2023	Current Ratio (2023)	ratio	41	8.13%
56	2023	Quick Ratio (2023)	ratio	30	5.95%
57	2023	A-Turnover (2023)	ratio	34	6.75%
58	2023	I-Turnover (2023)	ratio	57	11.31%
59	2023	R-Turnover (2023)	ratio	46	9.13%

	Column Name	Column Type	Count of Outliers	% of Outliers
0	Ticker	nominal	0	0.00%
1	beta	ratio	38	7.55%
2	% price change	ratio	0	0.00%

The dataset contains a significant number of outliers, particularly in financial ratios such as EPS, ROE, ROA, and turnover metrics, with outlier percentages exceeding 10% in some cases. Notably, gross margin remains relatively stable across all years with minimal outliers. High outlier counts in key profitability and efficiency ratios suggest substantial variability in company performance, which could pose challenges for using Multiple Linear Regression (MLR). Outliers can distort coefficient estimates, reduce model accuracy, and increase heteroscedasticity. Despite the presence of extreme values, it has been determined that outliers should be retained due to the inherent characteristics of financial data. Unlike in other domains where outliers may result from data entry errors or measurement inconsistencies, financial data often exhibits genuine extreme values driven by market conditions, firm-specific events, or economic cycles. For example, companies experiencing rapid growth, financial distress, or industry disruptions may naturally produce extreme EPS, ROE, or margin figures. Removing these outliers could eliminate crucial insights about firms that significantly outperform or underperform their peers, leading to a misrepresentation of real-world financial dynamics.

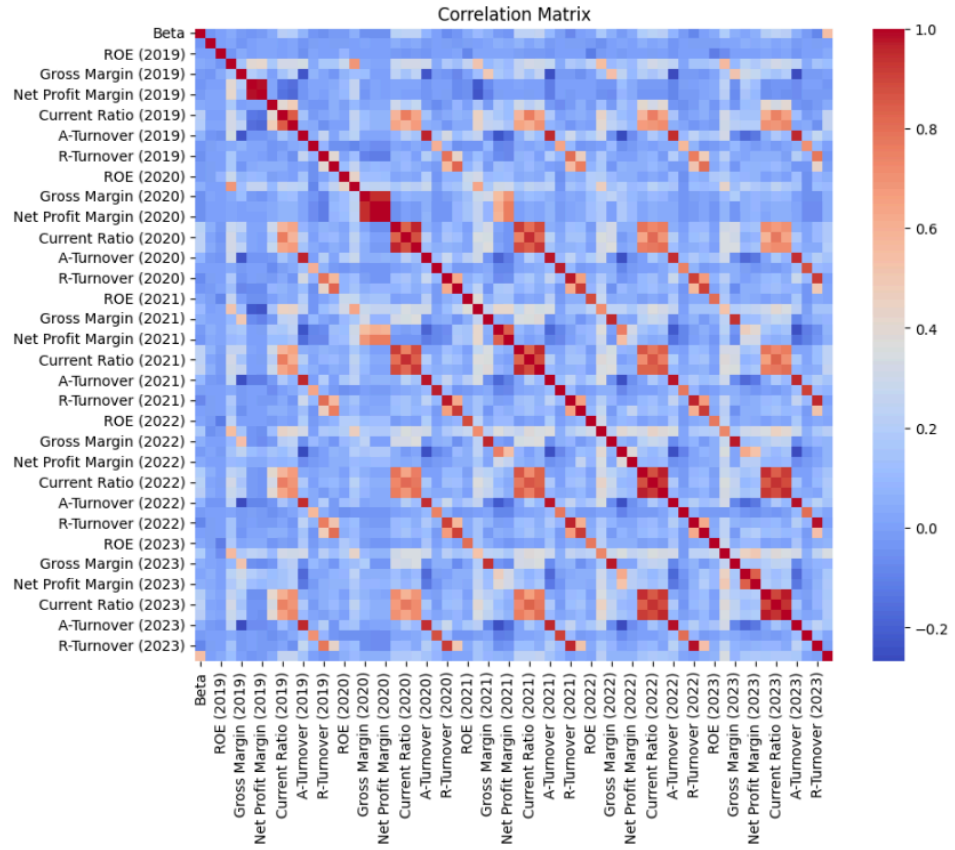
Distribution Analysis

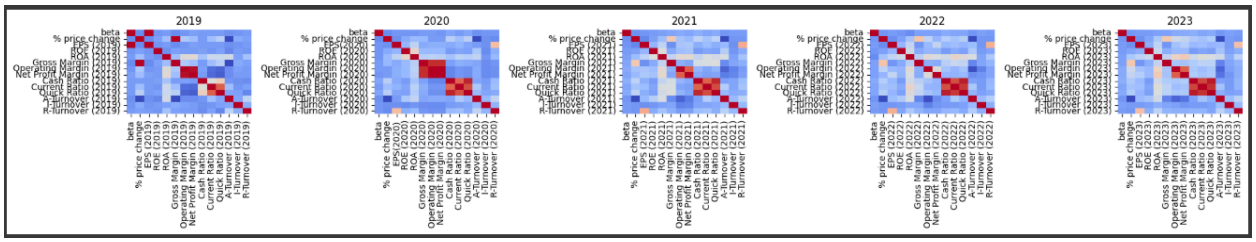




The dataset exhibits significant skewness across multiple financial ratios, which can negatively impact the assumptions of normality required for Multiple Linear Regression (MLR). Notably, EPS and ROE display extreme skewness across all years, with values exceeding ±10 in several instances, suggesting the presence of outliers that could distort regression results. Additionally, net profit and operating margins show substantial negative skewness, indicating that most firms have relatively small or negative margins, while a few extreme positive values pull the mean downward. Liquidity ratios, such as the Cash Ratio, Current Ratio, and Quick Ratio, also exhibit considerable positive skewness, meaning some firms have abnormally high liquidity compared to the rest of the sample. Similarly, turnover ratios, including Inventory Turnover (I-Turnover) and Receivables Turnover (R-Turnover), tend to have high positive skewness, reflecting a few firms with exceptionally high turnover levels.

Correlation Analysis





The correlation analysis from 2019 to 2023 highlights Gross Margin as the most consistently significant factor, showing strong positive correlations with percentage price changes across all years. Other financial metrics, such as Operating Margin, Net Profit Margin, Quick Ratio, and Cash Ratio, exhibit occasional significance but with weaker and less consistent correlations. Interestingly, EPS, ROE, and ROA generally have low or negligible correlations, suggesting they may not be strong predictors of price changes in this dataset. The presence of multicollinearity, particularly with highly correlated variables like Gross Margin, could affect the reliability of multiple linear regression (MLR) by inflating variance and reducing the precision of coefficient estimates. This

suggests that while MLR can still provide insights, alternative techniques such as principal component regression (PCR) or ridge regression may help mitigate multicollinearity and improve model robustness.

Attempt at using MLR

The multiple linear regression (MLR) results for the years 2019 to 2023 reveal significant variations in the factors influencing percentage price changes. In 2019, the model demonstrated an exceptionally high R-squared value of 0.996, indicating that nearly all variations in price change were explained by the independent variables. Gross Margin (2019) was the most significant predictor with a coefficient of 0.9760 ($p < 0.001$), while other financial metrics, such as EPS, ROE, and ROA, showed little to no statistical significance. However, the large condition number (5.4e+04) suggests potential multicollinearity issues.

In contrast, the MLR model for 2020 exhibited a sharp decline in explanatory power, with an R-squared value of 0.259. Gross Margin (2020) remained a significant predictor ($p < 0.001$), along with ROA ($p = 0.002$), Current Ratio ($p = 0.014$), Quick Ratio ($p = 0.017$), and A-Turnover ($p < 0.001$). Notably, EPS had a marginal impact ($p = 0.062$), and several variables, including ROE and Cash Ratio, were not statistically significant. The Durbin-Watson statistic of 1.982 suggests minimal autocorrelation.

For 2021, the R-squared value improved to 0.348, indicating a moderate explanatory power. Gross Margin (2021) was again highly significant ($p < 0.001$), along with Current Ratio ($p = 0.003$) and A-Turnover ($p < 0.001$). However, EPS ($p = 0.083$) and Quick Ratio ($p = 0.082$) were only marginally significant. The condition number (511) indicates a lower risk of multicollinearity compared to 2019.

The 2022 regression results continued the trend of moderate explanatory power, with an R-squared value of 0.383. Gross Margin (2022) remained a dominant predictor ($p < 0.001$), while Operating Margin ($p = 0.007$), Current Ratio ($p = 0.013$), and A-Turnover ($p < 0.001$) also had significant effects. EPS, ROE, and ROA did not show significant relationships with price change. The Durbin-Watson statistic (1.915) suggests little autocorrelation.

By 2023, the R-squared value was 0.366, similar to the previous year. Gross Margin (2023) ($p < 0.001$), Operating Margin ($p = 0.035$), Current Ratio ($p = 0.021$), and A-Turnover ($p < 0.001$) were significant predictors. EPS, ROE, and other financial ratios showed weak or no significance. Overall, the results indicate that Gross Margin consistently played a critical role in explaining price changes across all years, while the significance of other financial metrics varied. The declining explanatory power post-2019 suggests changing market dynamics, where additional unaccounted factors may have influenced price movements.

Conclusion

The dataset presents a range of challenges for developing a robust prediction system, primarily due to the significant skewness and presence of outliers in key financial ratios such as EPS, ROE, and margins. Columns like EPS, ROE, and operating margins exhibit extreme skewness, with values far exceeding ±10, suggesting that outliers could distort model accuracy. Additionally, liquidity ratios and turnover metrics show substantial positive skewness, reflecting a few firms with unusually high values. These outliers and skewed distributions could violate the assumptions of normality in regression models, potentially leading to inaccurate predictions. However, given the nature of financial data, removing these outliers could obscure meaningful insights, so alternative approaches such as robust regression or data transformation may be needed to address these issues. In conclusion, while the dataset contains challenges like skewness and outliers, these characteristics are inherent in financial data, and with careful handling, they can be managed to produce a more accurate and realistic predictive model.

Works Cited

Bloomberg. (n.d.). *Company profiles*. Bloomberg. Retrieved from https://www.bloomberg.com

Finsheet. (n.d.). *Financial data analysis tools*. Finsheet. Retrieved from https://www.finsheet.io