Joseph Muslu Adina Gitomer DS2001 12/12/2024

Company Valuation Within Same Location And Industry Visualized Large companies are an essential part of any developed society. They provide many services and products on a daily basis to millions of people around the United States. As part of the economic system, people are free to use the money they earn from their jobs however they see fit, but many times it is regarded as important to save this money or use the money they have to generate more money by investing it back into these large companies in the form of stock ownership. The investors make money in the form of dividends which are periodic payments paid out to stockholders or by capital gains: an increase in price of the selling price of the stock compared to the purchase price. But within stock investing, there are many metrics or factors to decide whether a company is a good investment. My question is whether the location of a company's headquarters has any sort of pattern on valuation, given that it is in the same industry. This question is important because it could potentially help investors decide whether to invest their money in a company, or instead stake their money against a company by shorting.

To answer this question, a dataset of the S&P 500 companies was obtained. This is a select list of some of the largest companies trading on the United States's stock exchanges that is commonly used to represent the health of the broader market. This dataset had two files. The first file was more focused on company general facts and had the company symbol, security name, GICS sector, GICS sub-industry, headquarters location, date added, CIK, and founded year. From this list, the security name was used as the company, and the GICS sector as the industry, and also used the headquarters location. I settled upon using these categories because the GICS sub-industry has over 100 distinct categories which would not have created enough overlap to represent the data. The second file was focused more on representing financials and contained some of the same information as the other file as well as the stock price, price to earnings (P/E) ratio, dividend yield, earnings per share (EPS), 52 week low stock price, 52 week high stock price, market cap, earnings before interest, taxes, depreciation, and amortization (EBITDA), price to sales (P/S) ratio, price to book (P/B) ratio, and link to SEC filings. From this list, I opted to pick the P/E ratio to represent the value of my stocks. I felt that this would be a good option as the P/E ratio is the price of the stock divided by the EPS, making it more representative than just the EPS. The EPS is useful because it is calculated by net income (earnings after expenses) or EBITDA minus expenses over the quantity of stocks, showing the income generating power of a company's individual stocks. However, the P/E ratio is even more widely used because it takes into account the price of the stock, so the P/E ratio can be compared across all stocks accurately instead of ones with similar individual share prices. With the company name, industry, location, and P/E ratio I had enough data to represent to answer my question.

Next, the data was cleaned. There were a few companies that either lacked a P/E ratio or were outlier values, and were removed. Additionally, the location data also included cities which would not have enough overlap so just state or country was used. Afterwards, the overlap between individual companies were in the same location and industry was grouped, as well as the averaged P/E ratios of all the companies in the overlap being mapped to the groups.



Figure 1: heatmap generated from python program

A heatmap was the best way to plot the data because it allows for effective comparison of P/E ratios between row and column, so trends can be observed across industries and across locations by looking at rows or columns, which makes for a streamlined viewing experience. Additionally the color scale represents the average P/E ratio in relation to the table, with the color scale centered around the mean value of the table. Blue values with lower P/E ratios represent more undervalued groups of companies, and red values being more overvalued groups of companies. An important component of this table to notice right away is that there is a fair amount of empty space. The locations depicted tend to have less companies in less populated states, or non-US located companies, with many locations not having certain industries. Also, this graph does not represent the amount of companies in each point as it is just an average for the P/E ratios of the companies, so it is possible that certain P/E ratios of locations with fewer companies skew the results in their favor in this dataset size.

Within the patterns shown in the chart, there are certain industries which tend to have higher P/E ratios, indicating being more overvalued. Information technology companies generally have rapid growth, allowing their stocks to trade at a premium. Notable examples include New York (140.30), Texas (110.88), Montana (140.30), Idaho (142.36), and Colorado (127.19).

For its location, California consistently has more overvalued companies than many other places across industries. California has the most overvalued communication service group (42.76), the third most overvalued consumer staples group (43.89), the most overvalued

healthcare group (53.05), and nearly all of California's other industry groups fell above the median of P/E ratios for the industry.

Places with undervalued companies are more difficult to spot as they are not always consistent across the board. Ireland is fairly consistent however with exception to healthcare, with a notable lowest group on the whole chart across industries for consumer discretionary (4.29) and the fifth lowest group for information technology (23.44). Ireland's industrial group (30.88) follows slightly below the industry median, and the healthcare group (34.62) slightly above the industry median. Tennessee also appears to be a place with undervalued companies, with the third lowest consumer staples group at (12.45), fourth lowest healthcare group (16.10), lowest industrial group (16.90), fifth lowest real estate group (34.16), and all other industries falling below or around the industry median.

To answer the original research question, there does appear to be patterns between location and valuation that are visualized within the heatmap. To expand upon this information and properly utilize the data, more research could be done on the individual companies that fall between appealing groups. For example, industrials in Tennessee may present a compelling buying opportunity, so individual companies could be looked at. However, another way to expand on the data would be to calculate z-scores by column (industry) and compare the z-scores for a row (location) to the other z-scores of the same row, and see if there is a z-score that stands out from among the row. This would reveal a group that is overvalued or undervalued relative to other industry groups in the same location, which could signal a clearer deviation from the norm and better buying or shorting opportunity.