Initial Public Offerings and Success Metrics

1. Introduction

Initial Public Offerings (IPOs) are a pivotal part for private companies that are transitioning to public companies, offering investors the chance to participate in early-stage public equity growth. These events often garner significant market attention, with many IPOs having sudden surges in value shortly after listing, while others fail to meet expectations and decrease in value rapidly. High-profile cases such as Rivian and Snowflake are great examples of the excitement and volatility surrounding IPOs. While some companies have remarkable post-IPO performance and gain significant returns, many underperform relative to the broader market benchmarks, such as the S&P 500.

This project aims to explore the outcomes of IPOs between the years 2019 and 2020 to better understand the financial and market factors that contribute to their short-term and long-term success or failure. Drawing on multiple data sources, including a Kaggle IPO dataset, web scraping from StockAnalysis.com, and supplemental financial metrics from the Yahoo Finance Python API, this analysis combines price movement data, industry classifications, and fundamental financial information to assess post-IPO performance.

Specifically, this project will aim to investigate three key research questions regarding the success of IPOs, including:

1. Are certain industries more likely to produce successful IPOs than others?
2. What is the optimal time to divest from an investment in an IPO?
3. How does the size of a company at IPO (measured by market cap) impact its post-IPO performance?

These questions are addressed in this analysis using a variety of statistical and machine learning methods, including univariate and bivariate analysis, hypothesis testing, regression, and classification modeling. The goal of this analysis is to offer data-driven insight into how investors and analysts might better assess IPO opportunities in an increasingly volatile market.

2. Data

This project uses three primary sources for data: a Kaggle dataset containing baseline IPO information, data scraped from StockAnalysis.com, and financial and industry data retrieved using the Python library Yfinance and Selenium.

2.1 IPO Baseline Data

The first set of data used in the analysis came from a Kaggle dataset titled *Company IPOs (2019 - 2021)*. This dataset provided the core IPO information needed for companies that went public between 2019 and 2021; however, I only used data for 2019-2020 and not 2021 for efficiency purposes in later scraping steps. This dataset included IPO Date, Ticker Symbol, Company Name, IPO Price, Current Price (As of December 21, 2021, when the data was collected), and Current Return (also as of December 21, 2021). This dataset served as the initial foundation for the analysis.

This data required only minimal cleaning. Columns were renamed to be easier to understand, and whitespace was also removed/ The IPO Price and Return values were converted into numerical formats by removing unnecessary dollar and percent signs. This cleaned data was then saved into the dataset ipo_df, which was the primary dataframe used to add on to the next 2 data sources.

2.2 Updated Market and Financial Data

The second dataset used in this analysis was from StockAnalysis.com, specifically their recent IPO webpage. To collect this data, I used a Selenium-based web scraping script to collect the most recent current price, current return, and market cap. I navigated through IPO pages for both 2019 and 2020 and programmed a feature to add a "Market Cap" column by triggering the site's interactive interface. This was needed to reduce the time needed to scrape this data by individually searching for each ticker symbol on the site.

This scraping was conducted using a headless Chrome browser to reduce the strain on the computer (which I'm still not confident worked). After scraping the current values from the website, the values were cleaned by removing non-numeric characters (e.g., "$", "%", "M", "B", "K") and converting the remaining results into float integers. Rows with missing or zero market cap data were filtered out, since they were either delisted or unable to be located later by Yfinance. This data was then merged with the original Kaggle dataset based on the ticker symbol to create a more comprehensive view of each company's IPO performance and characteristics.

2.3 Post-IPO Performance & Industry Data

To measure how a company performed over time, I needed to collect data from multiple points post-IPO, which was completed using the Yfinance API in combination with Selenium to retrieve time-based stock prices. Specifically, I gathered snapshots of prices at four time intervals following each IPO, including:

1. Price_1W: the price 1 week after the IPO date
2. Price_1M: the price 1 month after the IPO date
3. Price_6M: the price 6 months after the IPO date
4. Price_1Y: the price 1 year after the IPO date

Using the data collected from these 4 snapshots in time, I was able to compute values for return on investment metrics for each of the corresponding intervals above (named: Return_*Time*_pct). These return values allowed for a detailed analysis of IPO performance across multiple time horizons. With this data, I was able to locate entries with incorrect IPO price data, which was later corrected using Yfinance. In addition, I retrieved each company's industry label using Yfinance, enabling industry-level grouping for comparisons in question 1.

Finally, to support the classification-based machine learning models, I also created a binary variable named Success_Q2 (for question 2), which classified IPOs as successful (1) if the 1-year return was greater than zero, and unsuccessful (0) if the return was negative.

Table 1: Data dictionary

Column	Type	Source	Description
IPO Date	object	Kaggle	The calendar date the company went public (Initial Public Offering).
Symbol	object	Kaggle	The stock ticker symbol representing the company.
Company Name	object	Kaggle	The full name of the company that went public.
IPO Price	float64	Kaggle	The price per share at the time of the IPO.
Current	float64	Kaggle	Stock price as of December 21, 2021.
Return	float64	Kaggle	Percent return from IPO Price as of December 21, 2021.
Market Cap	float64	StockAnalysis.com	Market capitalization in millions of dollars at the time of scraping.
Year	int64	Kaggle	The calendar year in which the IPO occurred.
Industry	object	YFinance	The industry classification for the company.
Price_1W	float64	YFinance/Selenium	Stock price 1 week after the IPO date.
Price_1M	float64	YFinance/Selenium	Stock price 1 month after the IPO date.
Price_6M	float64	YFinance/Selenium	Stock price 6 months after the IPO date.
Price_1Y	float64	YFinance/Selenium	Stock price 1 year after the IPO date.
Return_1W_pct	float64	Calculated	Percent return from IPO Price to Price_1W.
Return_1M_pct	float64	Calculated	Percent return from IPO Price to Price_1M.
Return_6M_pct	float64	Calculated	Percent return from IPO Price to Price_6M.
Return_1Y_pct	float64	Calculated	Percent return from IPO Price to Price_1Y.
Market Cap Size	object	Derived	Categorical variable: Small Cap (<300M), Mid Cap (300M–2B), Large Cap (>2B).
Success_Q2	int64	Derived	Binary indicator: 1 if Return_1Y_pct > 0, otherwise 0.

3. Analysis

3.1 Industry and IPO Success

One of the central questions guiding this analysis was whether certain industries are more likely to produce a successful IPO when compared to other industries. To explore this idea, I began by grouping all companies by their respective industries and calculated each industry's average 1-year return percentage after their IPO date (since long-term capital gains only take effect after 1 year, thus is more advantageous to look at 1-year performance as a baseline). The results revealed a wide range of performance between many sectors, with certain industries well outperforming others.

The bar graph in *Figure 1* illustrates the average return by industry. This graph only shows the top 10 highest and lowest returns industries, since there are 66 total industries reported, which becomes unreadable, but is appended as *Appendix A*. At the top of the list were Semiconductors (335.68%), Leisure (237%), and Credit Services (182.25%), while the bottom performers included Auto & Truck Dealerships (-99.8%), Beverages – Wineries & Distilleries (-93.11%), and Coking Coal (-85.75%). These extremes suggest that industry classification may play a role in predicting post-IPO performance. It is also worth noting that some companies that have gone bankrupt were removed due to a lack of information.

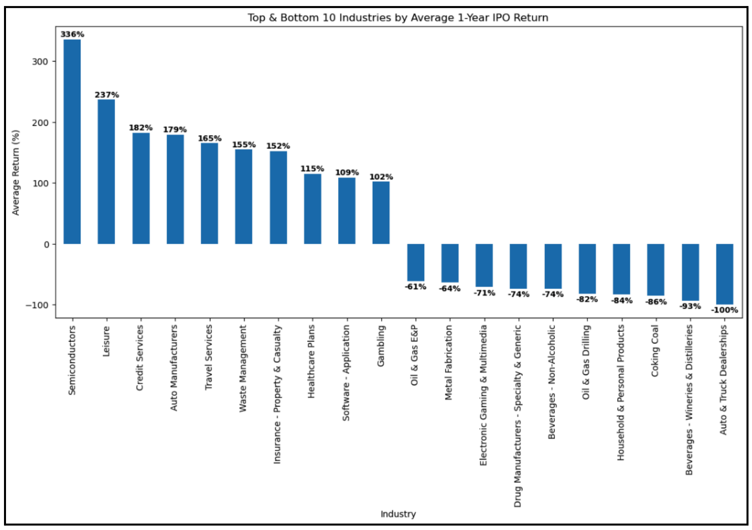


Figure 1 - Industry Return Bar Graph

Additionally, I created a boxplot, *Figure 2*, which was used to show the return distributions for the 20 industries presented in Figure 1. This was created to show the different levels of variance between different industries.

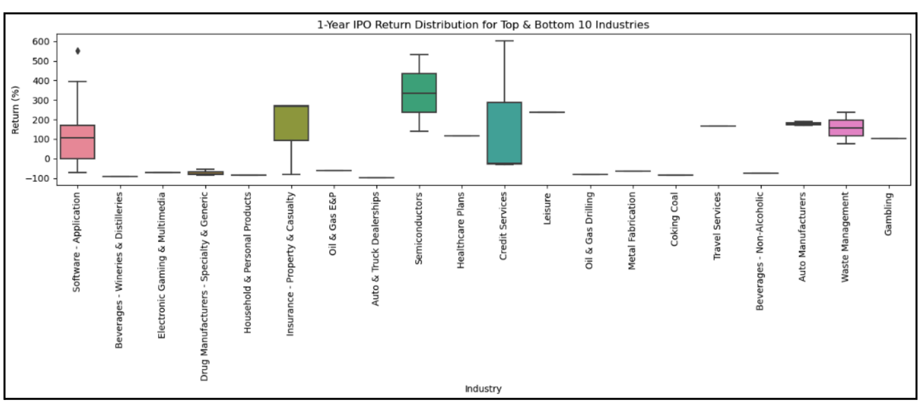


Figure 2 - Industry Return Box Plots

To statistically validate the findings from Figures 1 and 2, I performed a one-way ANOVA test. This test is set to evaluate whether the mean return differs significantly across different groups (industries). The test produced an F-statistic (similar to a T-statistic) of 3.5053 and a p-value of 1e-06, which is far below the alpha level of 0.5. This leads to the conclusion that we can reject the null hypothesis and assert that industries differ in average IPO return in a significant way.

Additionally, I decided to analyze how industry classification and market cap together help predict an IPO's success using logistic regression. This was done by defining a variable for success (1-year return > 0), and one-hot encoding features for both industries and market cap. This model achieved an accuracy score of 98.9%, indicating that combining industry and market cap size is highly predictive of IPO success. This is shown in the confusion matrix (*Figure 3),* which has a nearly perfect classification between IPO classifications.

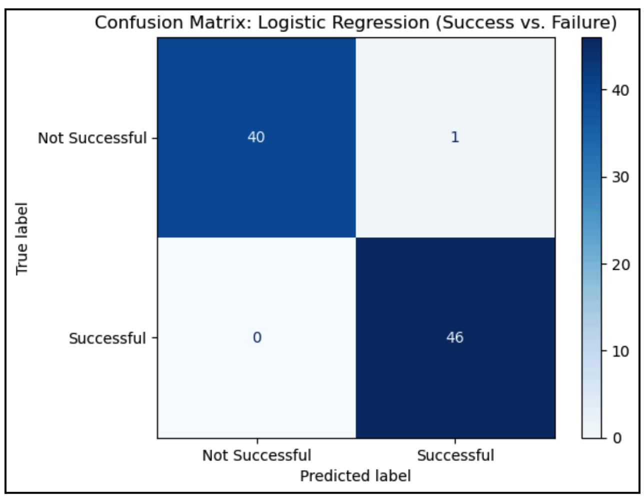


Figure 3 - IPO Success Confusion Matrix

3.2 IPO Performance Over Time and Optimal Divestment Timing

To explore optimal divestment timing following an IPO, I decided to examine stock performance across various times post-IPO, including 1 week, 1 month, 6 months, and 1 year. The goal of this was to see how returns evolved over time and whether there was an optimal time in this range to divest, or if early success could help predict long-term success.

The histogram in *Figure 4* shows that the distribution of 1-year trends is highly skewed to the right. This is expected, since a company could only decrease in value by at most 100%, while it could, in theory, infinitely increase in value. While the average return was 41.7%, many companies experienced losses, which was evident from the median return being just 5.2%. Some outliers had exceptionally high returns, with some companies obtaining an over 600% return. The histogram makes clear that while IPOs can deliver strong gains, they can also carry substantial downside risks.

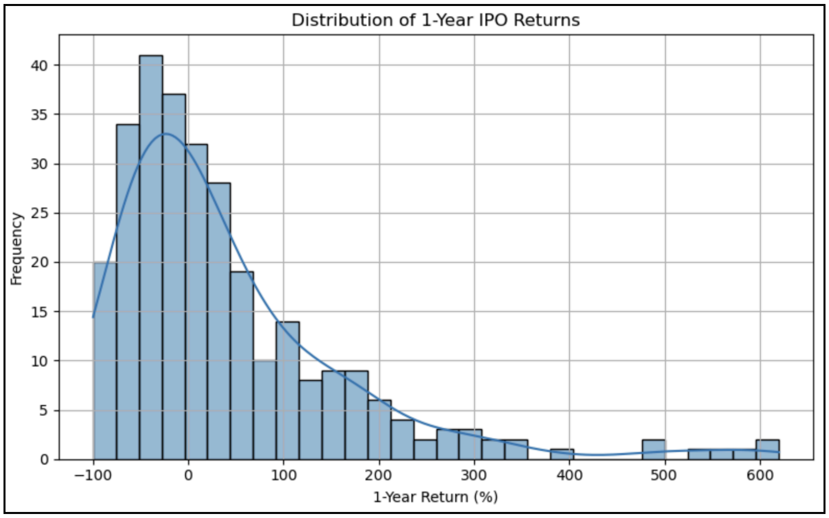


Figure 4 - Distribution of 1-year Returns

To better understand the change in return over time, I calculated the overall average return for each interval (*Figure 5*). From the graph, I was able to see that on average, returns tend to steadily increase from the 1-week mark to the 1-year mark. With an average 1-week return of around 26% and an average 1-year return of around 42%. However, returns in the first month were relatively flat, if not slightly declined, suggesting that immediate post-IPO momentum may stall before longer-term trends start to emerge.

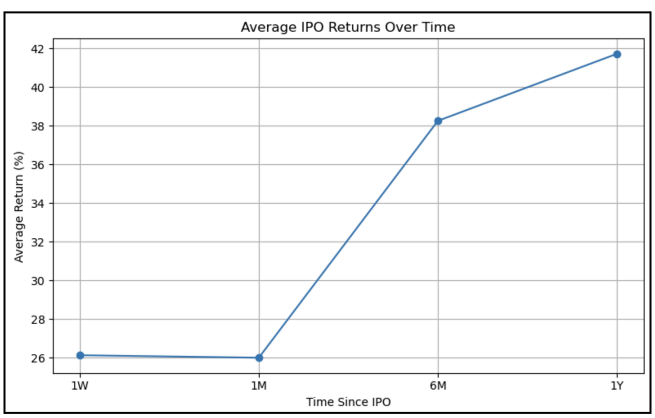


Figure 5 - Overall Average IPO Return Over Time

To account for company size in addition to return over time, I grouped returns into 3 market cap categories (small cap, mid cap, and large cap), and then proceeded to plot their performance over time (*Figure 6).* Across all periods of time, large-cap IPO consistently outperformed smaller firms. Additionally, all 3 types increased over a 6-month period, however, small-cap stocks take a sharp decline at the 1-year mark, while large and mid-cap stocks continue to grow. This suggests that market cap could play a critical role in post-IPO performance, which will be analyzed more in depth in question 3.

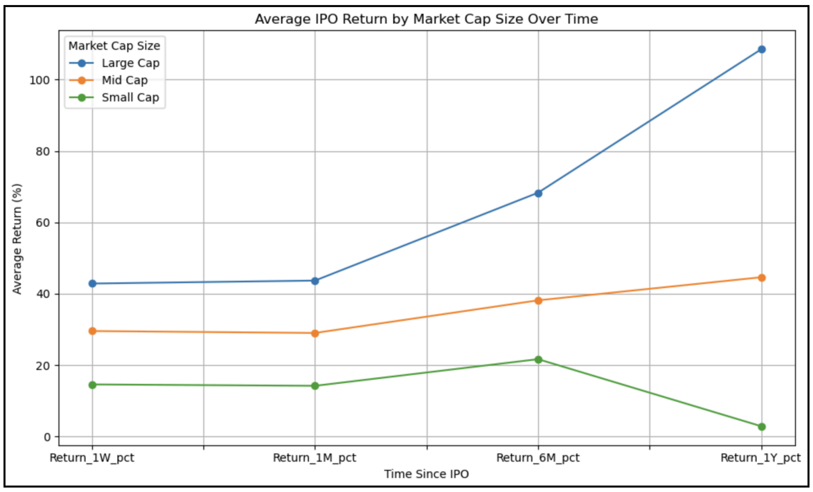


Figure 6 - IPO Return by Market Cap Over Time

To determine whether different intervals were statistically distinct, I performed a paired T-test comparing 6-month and 1-year returns. The P-value of 0.517 greatly exceeded the alpha level of 0.05, indicating the improvement was not statistically significant. However, when comparing the 1-month and 1-year returns, the resulting P-value is 0.022, which is less than the alpha level of 0.05. These results suggest that returns begin to plateau after 6 months, but have substantial growth before that, indicating that waiting till at least the 6-month point to divest is beneficial.

Finally, I constructed a decision tree (*Figure 7*) trained on the 1-week and 1-month returns, which achieved an accuracy score of 83% in predicting whether a stock would be successful (i.e., have a positive 6-month return). This model suggests that it may be possible to identify successful companies early using simple return-based indicators.

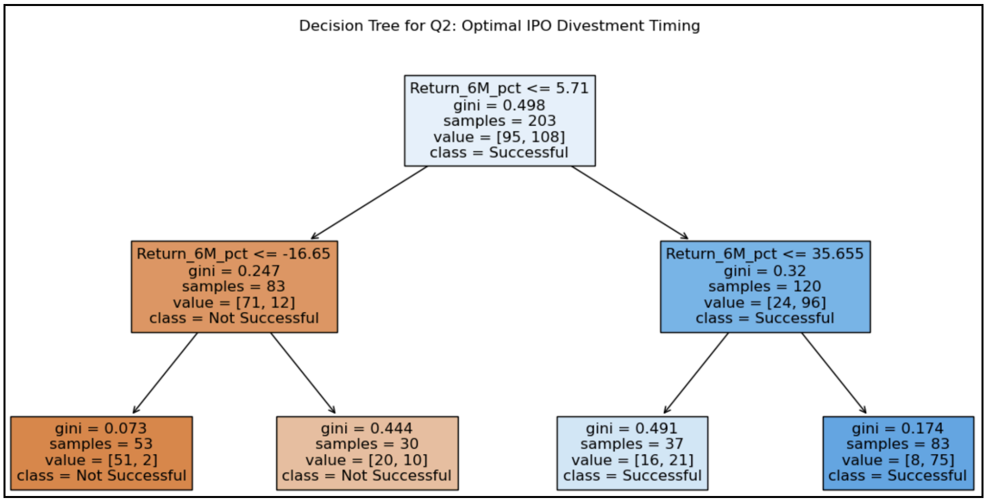


Figure 7 - Decision Tree

3.3 Market Cap and Post-IPO Performance

To investigate the impact of company size on IPO performance, I first compared IPO returns across market cap groups using both bar and box plots. *Figure 8* shows that small-cap companies made up the largest portion of IPOs, but *Figure 9* revealed that large-cap companies achieved higher median and maximum returns.

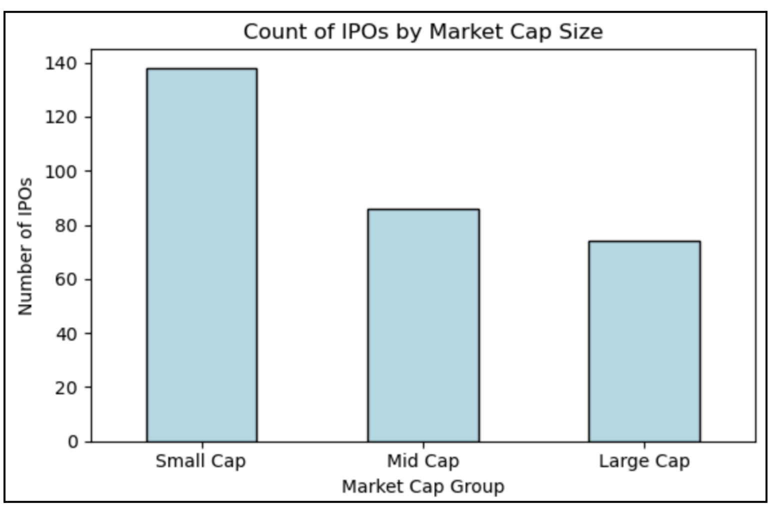


Figure 8 - IPO Count by Market Cap

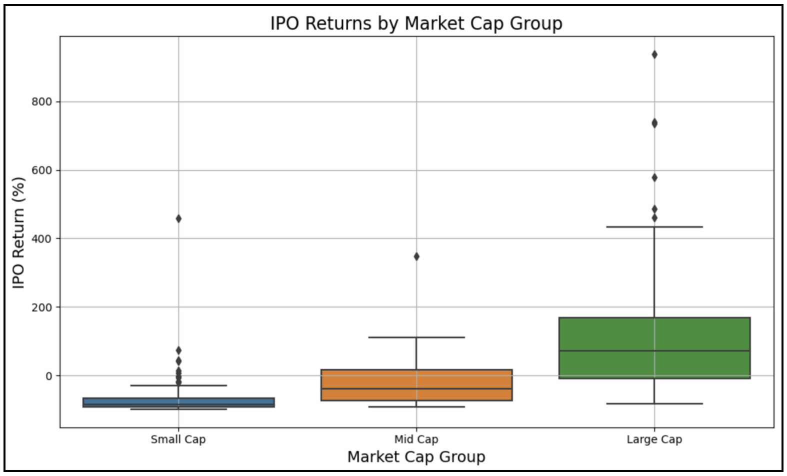


Figure 9 - IPO Return by Market Cap

A linear regression (*Figure 10)* was conducted next to quantify the relationship between market cap and IPO return, however, the R-squared value was only 0.11, which suggests a weak linear relationship. To validate the difference in performance further, I conducted a t-test comparing IPO returns between small-cap and large-cap firms, with the results showing a statistically significant difference with a P value less than 0.001. This reinforces the idea that larger-cap firms tend to perform better compared to other firms after going public.

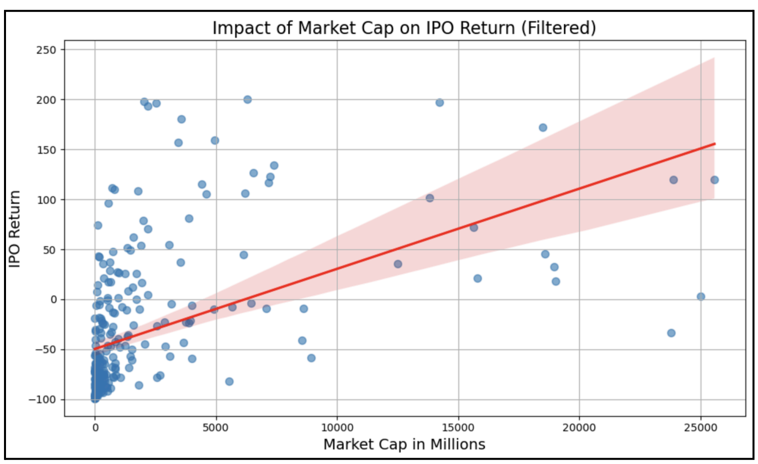


Figure 10 - IPO Linear Regression

Finally, I tested whether market cap could predict 1-year IPO returns using a linear regression model (*Figure 11*). However, the model's R-squared value was just 0.288, confirming that while there is a significant difference, market cap alone is not a strong enough predictor of individual

IPO performance. These findings suggest that although large-cap companies on average outperform smaller ones, market cap alone is insufficient for forecasting post-ipo success, and other variables likely play an important role.

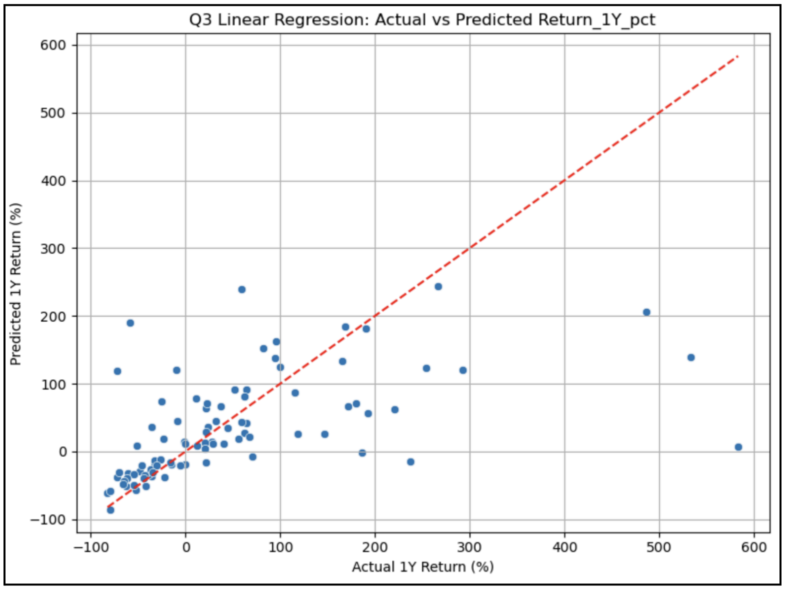


Figure 11 - Actual vs. Predicted 1-year Return

4. Conclusion

In this project, I explored the post-IPO performance of companies that went public between 2019 and 2020, focusing on industry, timing, and company size impacts on return. Based on the three research questions proposed, the following conclusions were reached:

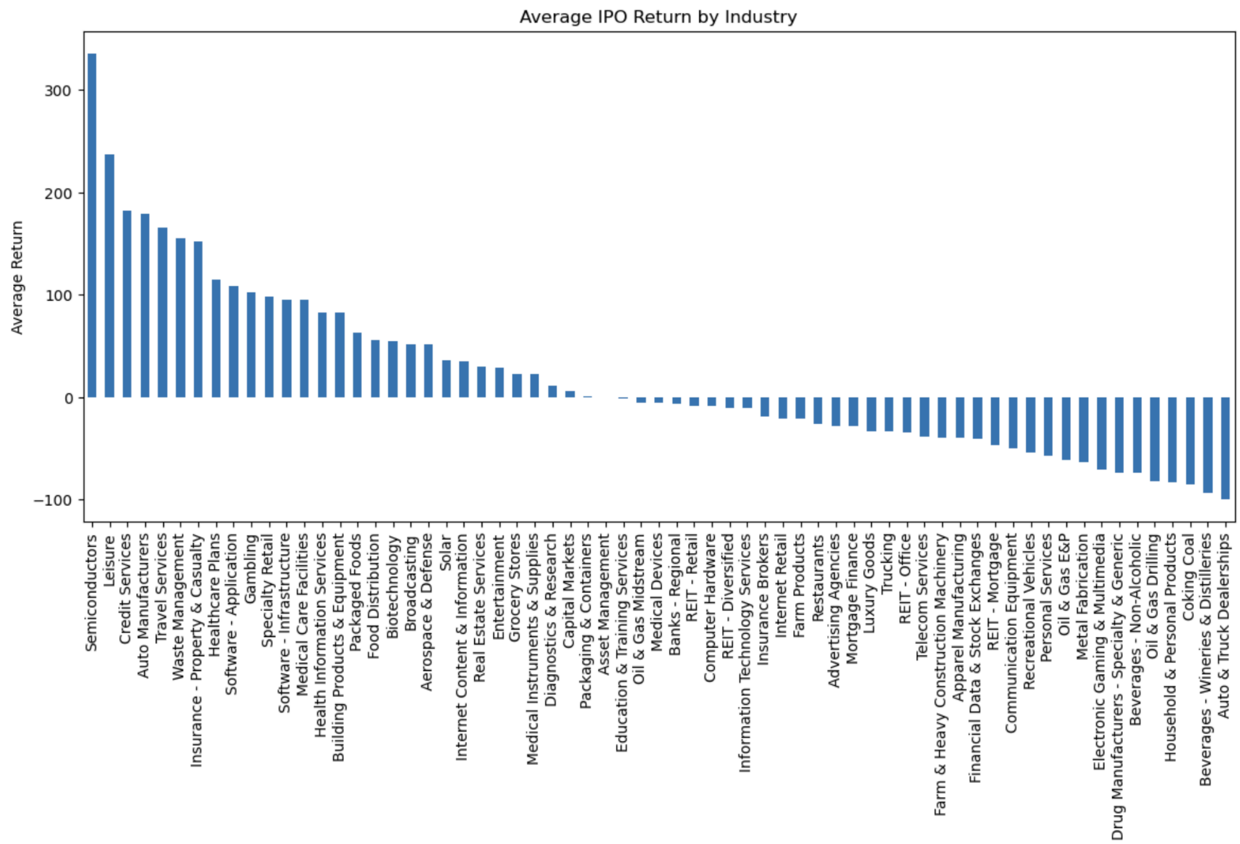
1. **Are certain industries more likely to produce successful IPOs than others?**
   I found that some industries consistently outperform others. For example, the Semiconductors, Leisure, and Credit Services industries showed the highest average 1-year returns, while industries like Auto & Truck Dealerships and Beverages - Wineries & Distilleries performed poorly. Additionally, the ANOVA test confirmed that industries differed significantly in IPO performance.
2. **Is there an optimal time to divest from an IPO?**
   In general, IPO returns increase over time, with average returns rising from 1 week to 1 year post-IPO. However, a paired t-test showed no significant difference between 6-month and 1-year returns, suggesting that gains may plateau after this period. A

decision tree also highlighted 6-month performance as a key predictor for whether an IPO would ultimately be successful, indicating that this may be a critical point for divestment.

3. **How does the size of a company (measured by market cap) impact its post-IPO performance?**
   Large-cap companies had notably higher average returns, while small caps made up the majority of IPOs but performed worse on average. A t-test confirmed a significant difference between small and large-cap returns. Still, linear regression models showed weak correlation, suggesting that market cap alone cannot reliably predict individual IPO performance.

This analysis had some limitations, including the reliance on publicly scraped data and the use of only certain data points (e.g., 1-year return only) rather than a consistent daily view. Future work could explore additional predictors such as revenue metrics, employee numbers, or look into more years of data to get a more complete picture. Overall, this project provides a helpful insight into the metrics that predict IPO performance and offers a baseline for investors to look at when evaluating IPO opportunities.

Appendix



Appendix A - All Industry Returns