

Predicting IPO Market Capitalization

Devansh Upadhyaya, Hemant Gupta, Ona Dubey, Ronit Kadakia,
Shikhar Beriwal and Yashvi Maheshwari

December 16, 2024

Contents

1	Project Idea	2
2	Motivation	2
3	Relevance	3
4	Novelty	4
5	Literature Review	5
5.1	Critical Analysis of Existing Literature	7
6	What Raw Data Did We Take? (and what they mean)	8
6.1	Cash Flow Data	8
6.2	Profit and Loss Data (From Annual Filings and Quarterly Results)	9
6.3	Balance Sheet Metrics	10
6.4	Operating Metrics	11
7	Why Did We Take This Raw Data?	12
8	From Where Did We Take the Raw Data?	13
9	How Did We Make the Raw Data into Features?	13
10	Exploratory Data Analysis	15
11	Discussion and Conclusion	18
11.1	Findings	18
11.2	Depth of Analysis	19

11.3 Strengths	21
11.4 Limitations	21
11.5 Conclusion	21
11.6 Next Steps	22
12 Policy Suggestions for IPO Market Standardization	22

1 Project Idea

The project focuses on predicting the market capitalization (market cap) for Initial Public Offerings (IPOs) using a combination of key quantitative metrics and qualitative factors. This robust predictive framework leverages data from Draft Red Herring Prospectus (DRHP) filings, incorporating detailed financial metrics such as quarterly sales, expenses, profits, and earnings per share (EPS). In addition to financial data, the project integrates influential factors such as grey market premium (GMP), assets, liabilities, etc.

To further enhance predictions, we have utilised sentiment analysis of news articles from CNBC(TV18), a prominent and trusted source for financial news. By analyzing the tone and sentiment of these articles, we aim to gauge a general perception of the company’s potential performance, adding a critical qualitative layer to the model. This approach ensures a well-rounded evaluation that captures both numerical data and market sentiment.

By applying machine learning (ML) techniques, the project predicts the market cap of IPOs, helping investors decide whether or not to subscribe. This comprehensive, data-driven tool empowers stakeholders to make informed decisions by providing actionable insights into a company’s financial health, market positioning, and broader perception.

2 Motivation

The motivation for this project stems from both personal experiences and broader market trends.

As we’re growing older—now in our 20s—the conversations around money and investments have become more frequent in our households. Our dads have started explaining the importance of concepts like compounding and how starting early with investments can make a huge difference in the long run. Naturally, this has piqued our interest in exploring the financial markets, and IPOs have emerged as one of the most exciting opportunities for young investors like us.

Take the example of IPOs like Bajaj Housing Finance—nearly everyone seemed to be applying for it. But amidst all the hype, it was challenging to figure out whether it was worth subscribing to or not. This isn't an isolated case. Since 2020, the craze for IPOs has skyrocketed, with companies from diverse sectors going public and drawing massive attention. However, this growing interest also brings its fair share of confusion and uncertainty for retail investors.

Deciding whether to invest in an IPO can feel like navigating a maze, especially when traditional methods rely heavily on historical financial metrics but fail to consider crucial qualitative aspects like industry trends or market sentiment. This gap often leaves investors, both new and experienced, at risk of misjudging the true potential of an IPO.

In today's digital age, where vast datasets and public sentiment from platforms like social media and news outlets are easily accessible, there's an opportunity to create a more data-driven and holistic framework for IPO analysis. Combining real-time public sentiment with financial data and industry trends can provide a clearer picture and help cut through the noise.

For us, this project is also deeply personal. Many of us in the group have started actively exploring investments, and we've faced the same challenges that most retail investors encounter. The lack of accessible, actionable predictions for IPOs has driven us to create a solution that doesn't just rely on numbers but also takes into account market perceptions. By combining technical innovation with practical insights, we hope to simplify decision-making for retail investors like us—and anyone else navigating the ever-growing world of IPOs.

3 Relevance

The relevance of this project lies in its ability to cater to diverse stakeholders in the financial ecosystem. For retail investors, it provides a user-friendly tool to make more informed decisions about IPO investments, reducing risks associated with traditional evaluation methods. Institutional investors can benefit from additional predictive layers, such as competitor analysis and sentiment trends, which aid in large-scale investment decisions. Financial analysts can use this project as a benchmark for incorporating non-traditional factors like social media sentiment and peer data into stock performance predictions. Moreover, for academic researchers, the project offers a unique perspective on integrating structured and unstructured data into financial modeling, paving the way for future studies in predictive analytics.

With the growing volume of IPOs worldwide and their popularity among first-time investors, this project addresses the current market’s demand for more effective evaluation methods. By bridging gaps in traditional approaches, it highlights how qualitative data sources, such as sentiment and market dynamics, can complement financial metrics to enhance predictions.

4 Novelty

The project distinguishes itself through its innovative approach to predicting IPO market capitalization, offering a unique framework that combines traditional financial analysis with modern data science techniques:

- **Multi-Source Data Integration:** Unlike conventional IPO evaluations that focus primarily on financial statements, this project synthesizes diverse data sources, including:
 - **DRHP Filings:** Extracting key financial metrics such as quarterly sales, expenses, profits, earnings per share (EPS), assets, liabilities, and risk factors.
 - **Grey Market Premium (GMP):** Incorporating GMP trends as a speculative indicator and analyzing its correlation with listing prices to assess its reliability.
 - **Sentiment Analysis:** Capturing public perception through the sentiment of news articles from CNBC(TV18), a prominent financial news source, to gauge market sentiment toward the IPO.
- **Standardization and Accessibility:** By proposing a framework for uniform data presentation across DRHP reports and centralized data repositories, the project addresses existing inefficiencies in accessing IPO-related data. These suggestions align with ongoing industry needs for better transparency and accessibility.
- **Interdisciplinary Approach:** The integration of structured financial data, unstructured qualitative data, and public sentiment sets this project apart. By balancing traditional finance with modern data science, it provides a holistic view of IPO potential, making predictions more comprehensive and actionable.
- **Timely Contribution:** With the rapid increase in IPOs and the rising interest among retail investors, this project is highly relevant. It addresses the growing demand for more reliable and accessible IPO evaluation methods, bridging the gap between speculative trends and data-backed predictions.

- **Practical Utility:** The project offers more than just theoretical insights—it delivers a potentially deployable tool that empowers investors to make informed decisions. By predicting market cap and evaluating company fundamentals, the tool simplifies the decision-making process, ensuring practical relevance for retail and institutional investors alike.

5 Literature Review

- IPO pricing is a complex process influenced by various factors. Market conditions, particularly the market P/E ratio, significantly affect both offer and list prices (Sahoo & Rajib, 2012).
- The choice of issue mechanism also impacts pricing, with book-building generally commanding higher prices than fixed-price offers (Sahoo & Rajib, 2012).
- Financial indicators like book value per share, earnings per share, and post-issue promoter group retention are important in determining IPO prices (Sahoo & Rajib, 2012). Additionally, factors such as the IPO's objectives and the company's industry type may influence pricing (Jotwani & Singh, 2012).
- Investment bankers play a crucial role in valuing companies' listing price, considering factors like growth potential and comparable firms in the industry (McCarthy, 1999).
- Interestingly, regulatory and procedural factors may contribute to the deliberate underpricing of IPOs by underwriters to maximize their expected income (Affleck-Graves & Miller, 1989).
- Financial metrics significantly influence IPO pricing, as evidenced by several studies. Net asset value, earnings per share, and price-earnings ratio positively affect IPO offer prices, while return on assets has a negative impact (Ramadhan et al., 2023).
- Pre-IPO financial performance indicators, including net asset value, return on assets, profit after tax, and return on net worth, substantially influence IPO offering prices (Yadav et al., 2023).
- One thing we would like to note is that there was significantly more amount of studies done on underpricing than there were on IPO pricing itself. In general, underpricing is a common, observable phenomenon with far-reaching implications for investors, companies, and market dynamics. It poses key

questions about market efficiency, investor behavior, and regulatory oversight, which are central to academic research in finance.

- Profitability has a significant positive effect on stock underpricing levels during IPOs, while financial leverage and firm size do not significantly impact underpricing (Evitasari & Nurhadi, 2023).
- Market sentiment and investor behavior significantly influence IPO pricing and performance. Higher investor sentiment leads to increased institutional investor bids and higher issue prices (Song & Tang, 2015).
- Positive sentiment is associated with greater underpricing, particularly for overvalued IPOs (Campbell et al., 2008).
- In hot market conditions, strong individual investor demand results in elevated IPO prices, substantial initial returns, and poor long-term performance (Derrien, 2005).
- Investor sentiment positively impacts IPO underpricing by affecting market turnover rates, with a more pronounced effect on higher-priced stocks (Chen et al., 2021).
- Underwriters play a crucial role in this process, raising issue prices based on institutional investor bids and selectively overvaluing some IPOs to maximize benefits for issuers (Song & Tang, 2015; Campbell et al., 2008).
- While information asymmetry contributes to underpricing for undervalued IPOs, better underwriter reputation leads to higher IPO valuations across the board (Campbell et al., 2008).
- Research suggests that IPO prices of comparable companies significantly influence a firm's IPO pricing and market entry decisions. Underwriters tend to select peers with higher valuations to make the issuer's shares appear conservatively priced, resulting in an upward bias in peer selection (Vismara et al., 2014; Paleari et al., 2014). This bias leads to higher underpricing and poorer long-term performance of IPOs (Paleari et al., 2014).
- The number of comparable stocks has a curvilinear relationship with IPO entry rates, partially mediated by trading volume and changes in stock returns (Chok & Qian, 2013).
- Additionally, IPOs can negatively impact stock prices of competing companies in the same industry, particularly during the listing period, due to information spillover and demand shock effects (Min, 2020).

- Factors such as operating profitability and R&D investment influence the relative competitiveness between IPO companies and their competitors, affecting competitors' share prices (Min, 2020).
- Recent studies have explored machine learning models for predicting IPO performance and underpricing.
- Artificial Neural Networks (ANN) demonstrated superior accuracy (68.11
- Random Forests outperformed eight classic machine learning algorithms in predicting initial returns on IPOs (David Quintana et al., 2017).
- XGBoost Regressor achieved the highest accuracy (91.95
- These advancements in machine learning techniques offer valuable insights for investors and financial analysts in predicting IPO performance and mitigating associated risks.
- Research on the impact of outstanding shares on IPO pricing reveals several influencing factors. S. Ramadhan et al. (2023) found that earnings per share (EPS) and net asset value positively affect IPO prices, while return on assets negatively impacts them.
- Kadek Gillang Nugraha Rianttara & Gusti Agung Krisna Lestari (2020) highlighted that the percentage of shares offered positively influences underpricing, while EPS has a negative effect on it.
- Ari yanto et al. (2020) discussed the challenges in determining IPO prices due to the absence of prior market prices, emphasizing factors like company size and underwriter reputation affecting underpricing.
- Lastly, L. Bateni & Farshid Asghari (2014) identified that the price-to-earnings (P/E) ratio significantly impacts initial offering prices on the Tehran Stock Exchange.
- Collectively, these studies underscore the complexity of factors, including outstanding shares, that influence IPO pricing dynamics.

5.1 Critical Analysis of Existing Literature

Existing research only focuses on determining whether a given stock is overpriced or underpriced and have performed extensive statistical analysis for the same. We are working on predicting how a given stock would perform at its IPO listing, regardless of whether it is fairly priced

6 What Raw Data Did We Take? (and what they mean)

We collected comprehensive financial data from various sources to predict the performance of Indian IPOs. The following categories were considered:

6.1 Cash Flow Data

- **Cash from Operating Activity:** Cash from operating activities measures the cash generated or used by a company's core business operations. It focuses on the company's ability to produce sufficient cash flow from regular business activities such as sales, production, and service delivery. This metric excludes cash from investments or financing, making it a clear indicator of operational efficiency. Positive cash flow suggests that a company can sustain its operations without needing external funding, while negative cash flow may signal operational challenges or growth-related expenses.
- **Cash from Investing Activity:** Cash from investing activities reflects the net cash spent or gained from investments in long-term assets such as property, equipment, and acquisitions. It also includes proceeds from the sale of such assets. This metric highlights a company's growth strategies and capital expenditures. A negative value may indicate significant investments for future expansion, while a positive value could result from asset sales, signaling either successful divestments or a lack of growth-oriented investments.
- **Cash from Financing Activity:** Cash from financing activities captures the flow of cash between a company and its investors, creditors, or owners. It includes proceeds from issuing stocks or debt, as well as cash outflows such as dividends, stock buybacks, and debt repayments. This metric shows how a company funds its operations and growth. A positive value suggests new financing inflows, while a negative value indicates that the company is repaying obligations or returning capital to shareholders.
- **Net Cash Flow:** Net cash flow represents the overall change in a company's cash position over a specific period, combining cash flows from operating, investing, and financing activities. It serves as a comprehensive indicator of liquidity and financial stability. A positive net cash flow suggests that a company has increased its cash reserves, enabling potential reinvestment, debt repayment, or dividend distribution. Conversely, a negative net cash flow may require the company to raise additional capital or restructure its financial strategy.

6.2 Profit and Loss Data (From Annual Filings and Quarterly Results)

- **Sales:** Sales represent the total revenue generated from the goods or services a company sells within a specific period. It serves as a key indicator of market demand, customer base strength, and competitive positioning. Higher sales often reflect strong business performance and market acceptance, while declining sales may signal reduced demand or market challenges.
- **Expenses:** Expenses encompass all costs incurred in running a business, including operating costs, administrative expenses, and cost of goods sold. They reflect the company's ability to manage its resources efficiently. Controlling expenses while maintaining quality and growth is crucial for sustaining profitability and operational efficiency.
- **Operating Profit:** Operating profit indicates the company's profitability from its core business activities after deducting operating expenses like salaries, rent, and utilities. It excludes income from investments and non-operational activities. A higher operating profit reflects strong business operations and efficient cost management.
- **OPM % (Operating Profit Margin):** Operating Profit Margin measures the percentage of sales converted into operating profit. It highlights a company's profitability from core operations before non-operating costs and taxes. A higher OPM % indicates better cost control and pricing power, reflecting competitive and financial strength.
- **Other Income:** Other income includes revenue from non-core business activities such as interest, dividends, or asset sales. It diversifies earnings and can boost profitability, especially during operational slowdowns. However, relying too much on other income may signal weak core business performance.
- **Interest:** Interest reflects the cost of borrowing funds to finance business activities. It shows the company's financial leverage and ability to manage debt. High interest payments can strain profitability, while low or declining interest expenses suggest effective debt management or reduced borrowing needs.
- **Depreciation:** Depreciation represents the gradual reduction in value of tangible assets such as machinery, equipment, and buildings due to wear and tear or obsolescence. It impacts profit calculations and signals the company's investment in asset maintenance and capital expenditure requirements.

- **Profit Before Tax (PBT):** Profit Before Tax is the earnings figure calculated before deducting income tax expenses. It reflects a company's overall profitability from both operational and non-operational activities. Higher PBT indicates strong financial performance and the ability to cover tax obligations.
- **Tax %:** Tax % represents the proportion of profit allocated to tax payments. It indicates the effective tax burden a company faces. A lower tax percentage suggests tax optimization strategies, while a higher rate may reflect limited tax-saving measures or a high taxable income base.
- **Net Profit:** Net profit is the company's final earnings after deducting all expenses, including operating costs, interest, taxes, and depreciation. It reflects the overall profitability and financial health of the business. Sustained net profits are critical for growth, investment, and shareholder returns.
- **EPS in Rs (Earnings Per Share):** Earnings Per Share measures the company's profitability on a per-share basis, indicating how much profit is attributable to each outstanding share of common stock. It is a key metric for assessing shareholder returns and is often used by investors to evaluate a company's financial performance.
- **Dividend Payout %:** Dividend Payout % shows the proportion of net profit distributed to shareholders as dividends. It reflects the company's dividend policy and commitment to returning profits to investors. A higher payout ratio indicates a focus on rewarding shareholders, while a lower ratio suggests reinvestment in business growth.

6.3 Balance Sheet Metrics

- **Equity Capital:** Equity capital represents the funds contributed by shareholders through the issuance of shares. It forms the foundation of a company's capital structure, signifying ownership interest. A strong equity base indicates financial stability and a lower reliance on debt, reducing overall financial risk.
- **Reserves:** Reserves consist of retained earnings and other surplus funds set aside for future use, such as expansion, debt repayment, or contingencies. They serve as a financial buffer, enhancing the company's ability to withstand economic fluctuations and invest in growth opportunities.
- **Borrowings:** Borrowings reflect the company's debt obligations from loans, bonds, or other financial instruments. They indicate the company's reliance

on external financing. Higher borrowings can signal growth-driven investments but also increase financial risk through interest obligations and repayment pressures.

- **Other Liabilities:** Other liabilities encompass financial obligations not directly related to core business operations, such as deferred taxes, provisions, or outstanding payables. They reflect the company's broader financial responsibilities and potential claims against its assets.
- **Total Liabilities:** Total liabilities represent the cumulative financial obligations a company owes to creditors, lenders, and suppliers. This metric provides a comprehensive view of the company's debt burden and its ability to meet long-term and short-term obligations.
- **Fixed Assets:** Fixed assets include long-term tangible assets like buildings, machinery, and equipment used in business operations. These assets indicate the company's investment in infrastructure and its capacity to generate future revenue through operational productivity.
- **CWIP (Capital Work in Progress):** CWIP refers to assets under construction that are not yet operational but are expected to become productive in the future. It highlights the company's ongoing investment in expansion projects, signaling growth potential and future revenue generation.
- **Investments:** Investments include financial assets such as stocks, bonds, and mutual funds acquired for income generation or strategic purposes. They reflect the company's diversification strategy and its ability to generate returns beyond core business operations.
- **Other Assets:** Other assets cover miscellaneous items such as patents, trademarks, goodwill, and prepaid expenses. These assets support business operations indirectly and contribute to the company's valuation and competitive advantage.
- **Total Assets:** Total assets represent the combined value of everything the company owns, including fixed assets, investments, and current assets like cash, inventory, and receivables. This metric reflects the company's overall resource base and its ability to generate future economic benefits.

6.4 Operating Metrics

- **Debtor Days:** Debtor days measure the average number of days a company takes to collect payments from its customers after a sale. It indicates credit

terms offered and the efficiency of accounts receivable management. A lower debtor days figure suggests timely collections and strong cash flow management, while a higher figure may signal potential cash flow issues or lenient credit policies.

- **Inventory Days:** Inventory days represent the average number of days a company holds inventory before selling it. This metric assesses inventory management efficiency, indicating how quickly goods are turned into sales. A lower value points to efficient stock management and reduced holding costs, while a higher value may suggest overstocking or slow-moving products.
- **Days Payable:** Days payable measures the average time a company takes to pay its suppliers after receiving goods or services. It reflects supplier payment terms and cash management strategies. A higher days payable figure can improve cash flow but may risk supplier relationships if payment delays become excessive.
- **Cash Conversion Cycle (CCC):** The cash conversion cycle evaluates how quickly a company converts its investments in inventory and receivables into cash from sales, minus the time taken to pay suppliers. A shorter CCC indicates efficient cash flow management and operational efficiency, while a longer cycle may highlight liquidity challenges or operational inefficiencies.
- **Working Capital Days:** Working capital days track how long a company's net working capital (current assets minus current liabilities) is tied up in day-to-day operations. It reflects the company's short-term liquidity and operational efficiency. A lower value suggests effective use of working capital, while a higher value indicates potential liquidity constraints.
- **ROCE % (Return on Capital Employed):** ROCE % measures the profitability and efficiency of a company's capital investments. It indicates how well the company generates profits from its capital employed, including equity and debt. A higher ROCE % reflects strong financial performance and efficient resource utilization, making the company attractive to investors.

7 Why Did We Take This Raw Data?

We selected these metrics because they provide a comprehensive overview of a company's financial health, operational efficiency, and potential for future growth. Historical financial performance is crucial for assessing whether a company is likely to succeed post-IPO. Financial indicators such as cash flow, profit and loss metrics, and balance sheet components reveal the company's operational efficiency,

profitability, and asset-liability management. Metrics like revenue, expenses, and net profit highlight historical financial performance and profitability, which are crucial indicators of future success. Operating metrics like debtor days and ROCE % offer insights into operational efficiency and financial health. We also included qualitative data from CNBC TV18 articles, providing sentiment analysis and contextual market information to capture market perceptions that might influence IPO success.

8 From Where Did We Take the Raw Data?

We sourced the financial data from:

- Screener <https://www.screener.in/>
- Moneycontrol <https://www.moneycontrol.com/>
- CNBC TV18 <https://www.cnbctv18.com/>
- DRHP Reports of the IPO candidate companies.

9 How Did We Make the Raw Data into Features?

To transform raw data into actionable features, we carefully selected metrics that most effectively represent the company's financial performance. From the balance sheet, we focused on core indicators such as equity capital, reserves, and total liabilities, which reflect the company's capital structure and financial obligations. In the cash flow statement, we included cash from operating, investing, and financing activities to evaluate liquidity and cash management. From the profit and loss statement, we extracted revenue, expenses, net profit, and EPS in Rs, focusing on profitability and operational efficiency. Similarly, quarterly results data helped capture recent business performance trends. To enrich the dataset with qualitative insights, we applied the OpenAI API on CNBC TV18 articles, generating sentiment scores, price targets, and investment recommendations. To ensure consistency and eliminate scale-induced biases, we normalized key features as percentages of the IPO price, making them comparable across companies. We also integrated historical data by combining previous year and quarter results with current metrics, enabling a dynamic evaluation of financial performance trends.

From Balance Sheet

- Equity Capital
- Reserves
- Borrowings
- Total Liabilities
- Fixed Assets
- Total Assets

From Cash Flows

- Cash from Operating Activity
- Cash from Investing Activity
- Cash from Financing Activity
- Net Cash Flow

From Profit and Loss

- Revenue
- Expenses
- Net Profit
- EPS in Rs

From Quarterly Results

- Revenue
- Expenses
- Net Profit
- EPS in Rs

Additionally, we scraped CNBC TV18 for articles on each company, extracting relevant content such as overall sentiment. We further enhanced our dataset by combining metrics from the relevant previous year or quarter for each stock to track historical performance. To analyze qualitative data, we used the OpenAI API on CNBC TV18 content, generating the following features:

- **Sentiment Score:** A value from 0 to 1, where 0 indicates extremely negative sentiment, 0.5 is neutral, and 1 is extremely positive.
- **Price Target:** The average target price from the articles or 'N/A' if not mentioned.
- **Recommendation:** A one-word recommendation, either 'Subscribe' or 'Don't Subscribe,' based on sentiment and price targets.

Finally, we normalized relevant features as percentages of the IPO price to reduce scale-induced biases.

10 Exploratory Data Analysis

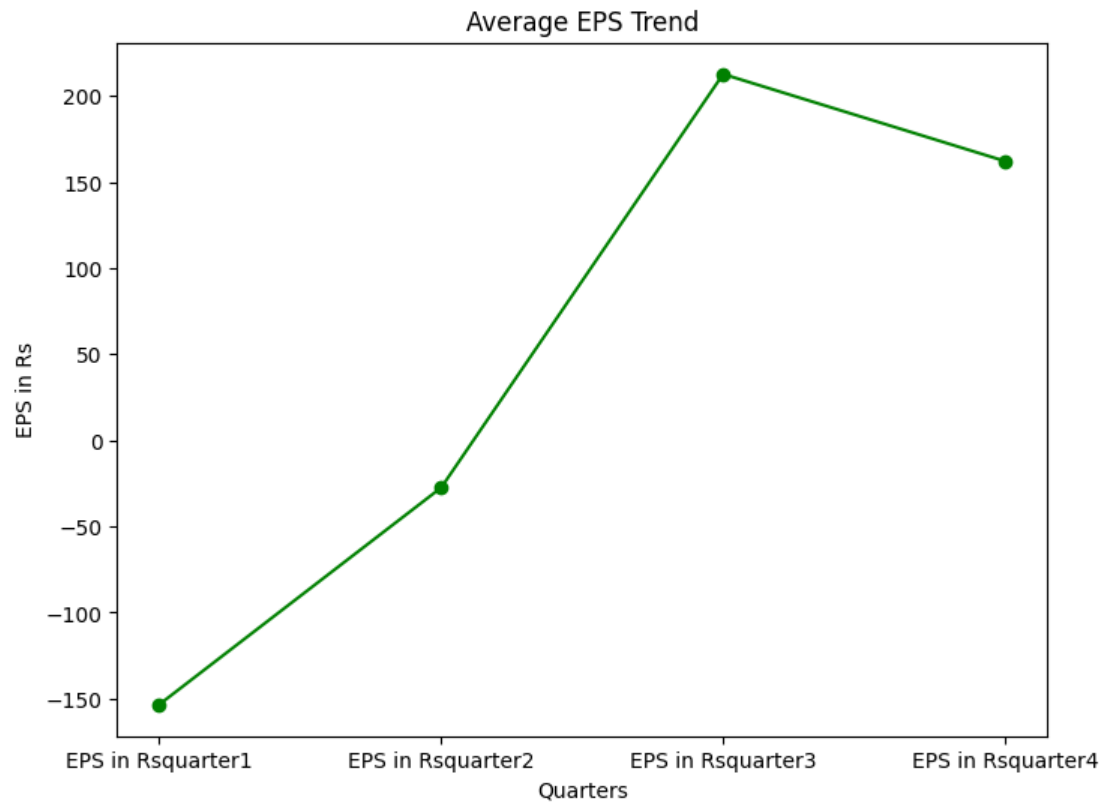


Figure 1: Average EPS Trend

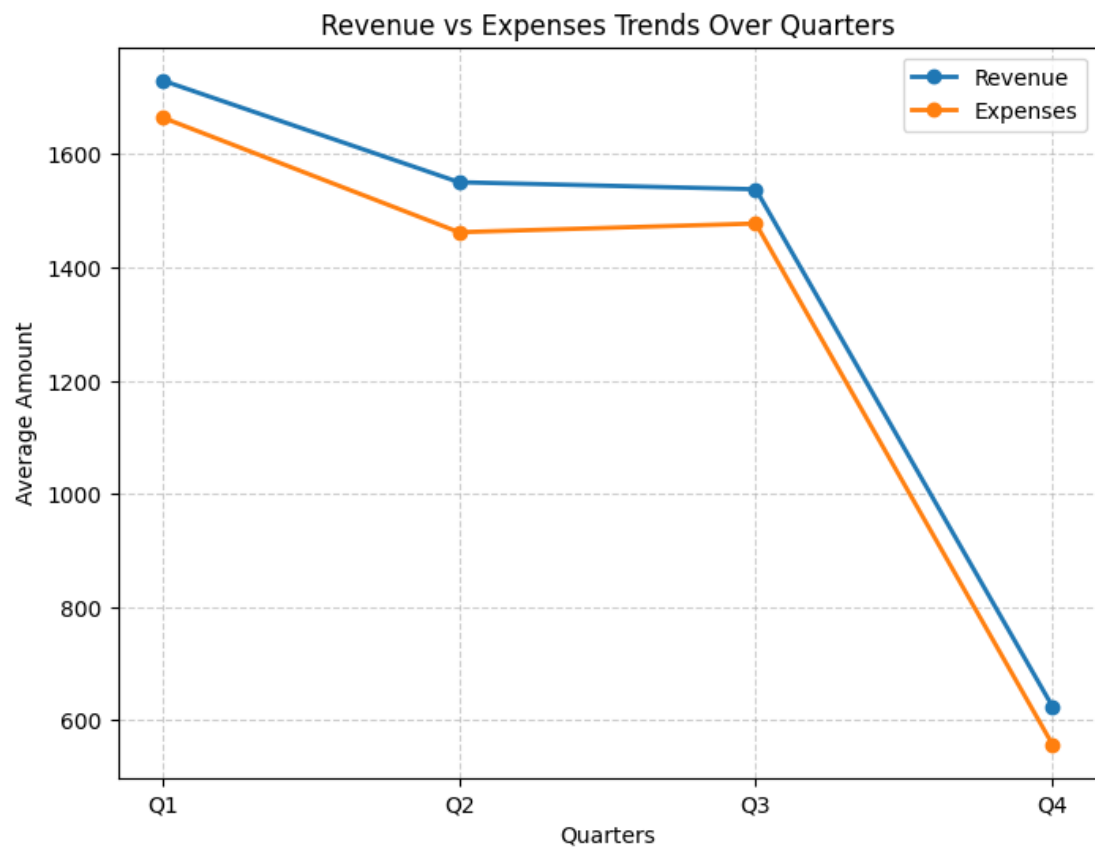
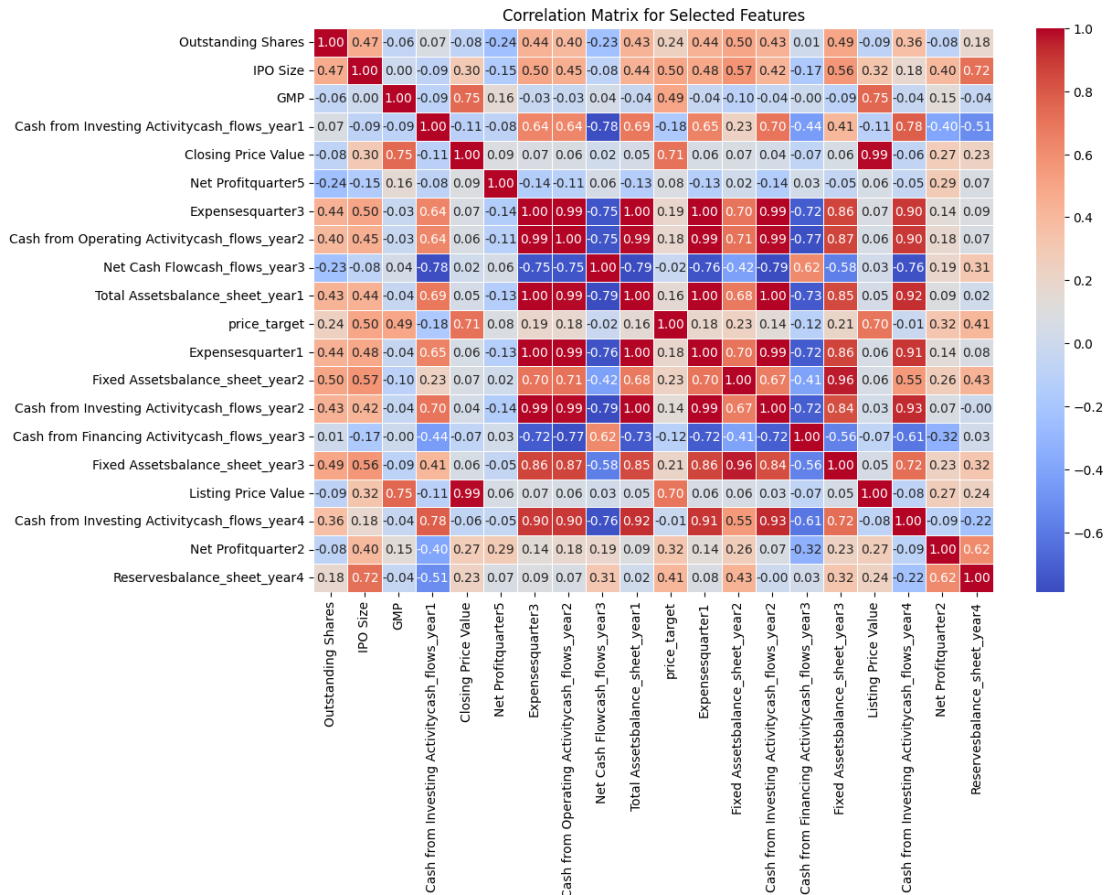


Figure 2: This shows that on average, companies that apply for an IPO are profitable



11 Discussion and Conclusion

We have used Random Forest Regression to predict Market Capitalization using 72 features derived from 194 companies' historical financial data, financial news sentiment analysis, and IPO details such as GMP (Grey Market Premium), IPO Size, and Outstanding Shares. The dataset includes a combination of quarterly financial metrics (e.g., revenue, expenses, net profit), cash flow indicators, and balance sheet data spanning multiple years.

The model was designed to identify the most critical factors driving Market Capitalization, providing data-driven insights for investors, analysts, and underwriters. We chose market capitalization as our target variable as it is a direct indicator of the company's valuation which can be benchmarked against other companies' market valuations. Stock prices alone can't compare company valuations because they don't account for differences in the number of shares outstanding, hence are not an indicator of the company's worth. P/E ratios cannot be used to compare companies as our set of companies do not belong to the same sector, hence P/E cannot be used with a fixed multiple to account for the company's value making market capitalization a more accurate measure.

To handle data inconsistencies and missing values, we:

- Dropped features with more than 50% missing data.
- Applied KNN Imputation to fill in remaining gaps, ensuring minimal information loss.

The Random Forest Regression model was chosen for its ability to handle large datasets with non linearity, and its ability to rank feature importance effectively. Our analysis revealed that features such as Outstanding Shares, IPO Size, and GMP were the most influential predictors of Market Capitalization (refer to the Feature Importance Graph).

11.1 Findings

The top features driving the model predictions (Market Cap) are as follows:

- **Outstanding Shares:** This is the most critical feature, as it directly influences the market capitalization, which is calculated as Outstanding Shares \times Price.
- **IPO Size:** Larger IPO sizes are often linked to higher market interest and valuations, directly impacting Market Cap.

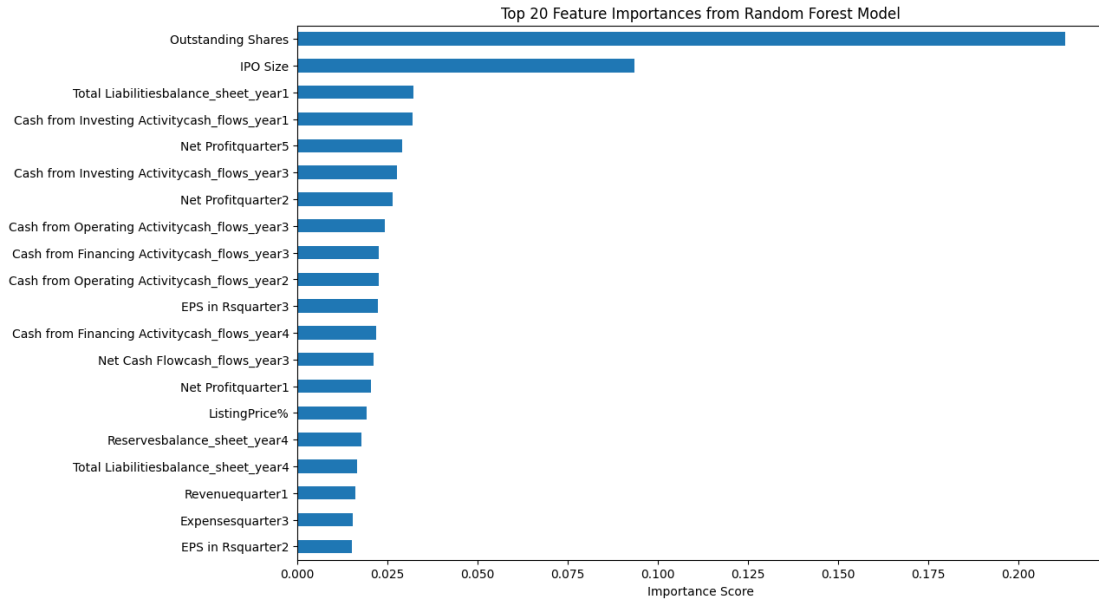


Figure 3: Top Features

- **GMP (Grey Market Premium):** Reflects investor demand and the expected listing price premium, which contributes significantly to post-IPO Market Cap.
- **Cash from Investing Activity (Year 1):** Indicates the firm's capital investment patterns, cash rich companies can invest more, signalling growth potential and long-term value.
- **Closing Price Value:** The stock's closing price plays a direct role in determining Market Cap.
- Other relevant features, such as Net Profit (Quarter 5), Expenses (Quarter 3), and Operating/Investing Cash Flows, emphasize the role of financial performance and operational efficiency in influencing market value.

11.2 Depth of Analysis

The analysis highlights the importance of financial metrics (profits, expenses) and operational cash flows in predicting Market Cap. A few key insights include:

- **Outstanding Shares and Closing Price Value:** These are fundamental factors in Market Cap computation. Hence, their high importance aligns with expectations.

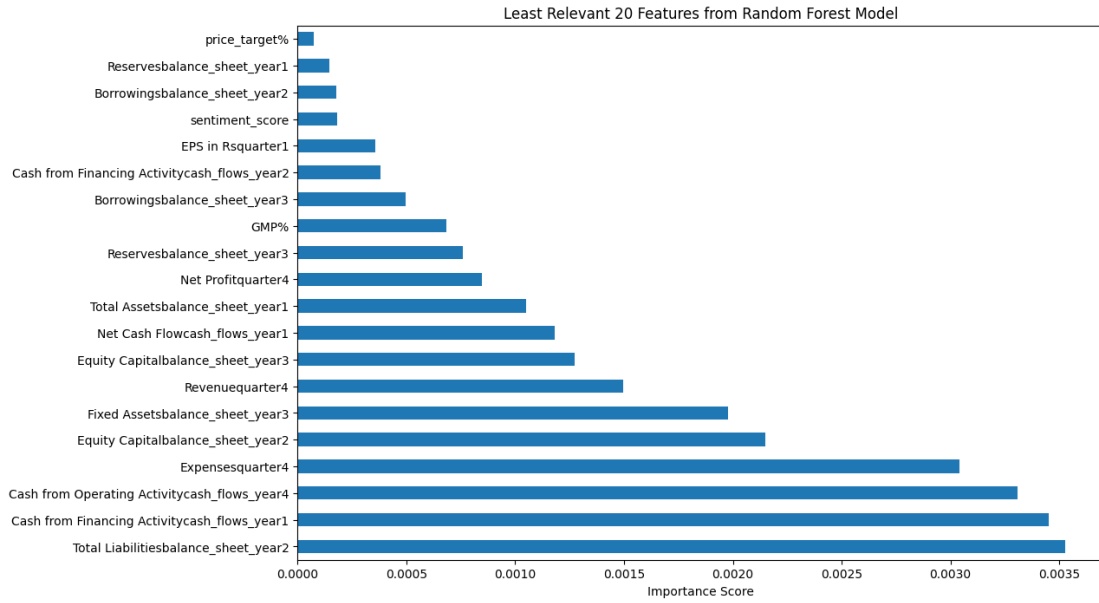


Figure 4: Some of the features we expected to be pivotal but didn't: Sentiment analysis , GMP (Grey Market Premium)

- **IPO Metrics (IPO Size and GMP):** These metrics reveal investor confidence and demand during the IPO stage, which significantly correlates with market performance post-listing.
- **Investing Cash Flow:** High importance suggests that firms investing in growth opportunities tend to achieve better market valuations.
- **Operating Cash Flow:** Operational efficiency and liquidity management strongly influence investor perception.
- **Expenses and Profits:** These metrics reflect the company's efficiency in managing costs and achieving profitability, which is critical for long-term valuations.

The results are consistent with financial theory, where profitability, cash flow, and share-related metrics strongly influence company valuations.

11.3 Strengths

- **Robust Cleaning and Imputation:** Missing values were systematically handled using KNN imputation, ensuring minimal data loss.
- **Feature Importance Analysis:** The use of Random Forest regression allowed identification of the most influential features in predicting Market Cap.
- **Comprehensive Dataset:** The dataset includes quarterly financial metrics, cash flows, IPO-related data, and sentiment-based metrics like GMP and price targets, making it information-dense and ideal for predictive modeling.
- **Choice of Model:** Random Forest regression is effective in handling large datasets, missing data, outliers, and noisy features.

11.4 Limitations

- **Insufficient Data:** Dropping features with $> 50\%$ missing values may have caused loss of potentially valuable information. Limited data points may restrict the model's generalizability and robustness.
- **No External Factors:** External market conditions (e.g., economic indicators, sector trends) are not included in the dataset, which may impact the market cap and hence the performance of the stock at IPO.
- **Overfitting Risk:** While Random Forest is less prone to overfitting, the performance may still vary due to small dataset size.

11.5 Conclusion

The analysis successfully identified key drivers of Market Cap, such as Outstanding Shares, IPO Size, and Cash Flows, highlighting their critical role in determining market performance. While the model demonstrates promising insights, limitations such as data sparsity and potential feature overlap restrict its scalability. Future improvements can include incorporating external market factors and expanding the dataset to improve model robustness.

We also found these to be our evaluation metrics for the Random Forest regression model we ran:

- **R^2 :** 0.75

- **Adjusted R^2 :** 0.60
- **RMSLE:** 0.61

11.6 Next Steps

- Collect more data, including economic indicators and sector-specific metrics.
- Perform feature selection or dimensionality reduction to handle correlated features.
- Explore advanced models (e.g., XGBoost, LightGBM) for improved predictions.

12 Policy Suggestions for IPO Market Standardization

1. Standardization for Retail Firms:

- Establish uniform guidelines for how retail firms present IPO-related information, ensuring consistency and transparency. This would include presenting key metrics like GMP (Grey Market Premium), listing price projections, and company fundamentals in a comparable format.
- Require retail firms to clearly disclose the basis of GMP calculations and any assumptions involved to avoid misleading retail investors.

2. Correlation Between GMP and Listing Price:

- EConduct and mandate regular audits to assess the correlation between GMP and actual listing prices. If GMP is found to be systematically misleading, introduce regulations to curb its misuse.
- Develop an investor education campaign on understanding GMP, its limitations, and its role as a speculative indicator rather than a definitive predictor.

3. Uniform DRHP Reports:

- Implement standardized formatting for Draft Red Herring Prospectus (DRHP) reports to make key financial metrics—such as quarterly sales, profits, EPS, assets, and liabilities—easily accessible and comparable across companies.

- Require DRHP filings to include simplified summary sections for retail investors, highlighting key takeaways in layman-friendly language.

4. Centralized Data Repository:

- Create a centralized platform, either government-run or managed by private entities under regulation, where all IPO-related data, including DRHPs, GMP trends, industry comparisons etc. is made available.
- Ensure that this platform provides real-time updates and accurate links, addressing issues like broken URLs and scattered information across multiple websites.
- Include API access for developers to integrate IPO data into analytical tools, making it easier for both institutional and retail investors to analyze IPOs effectively.
- Require the centralized platform to include archived data of previous IPOs, such as historical GMP trends, listing prices, and performance metrics. This would enable better analysis and research for future IPO predictions.

5. Technology-Driven Data Submission for Companies:

- Develop a user-friendly backend platform for companies to upload DRHP data and other IPO-related information directly in a standardized format. This platform should automate the formatting, validation, and submission process, reducing manual errors and inefficiencies.
- Encourage tech-enabled innovations to streamline the preparation and submission of DRHP filings, ensuring smaller firms with limited resources can comply without excessive technical barriers.

References

- [1] Patnaik, D., & Others. (2023). Stock price forecasting using machine learning techniques: A systematic review. *2023 International Conference on Data Science & Informatics (ICDSI)*.
- [2] Siwach, P., & Others. (2024). Pricing and performance of initial public offerings: Directions for future research. *Qualitative Research in Financial Markets*.
- [3] Sonsare, P., & Others. (2023). A comparative analysis of machine learning algorithms for IPO underperformance prediction. *Journal of Advanced Applied Scientific Research*.

- [4] Ramadhan, S., & Others. (2023). Financial performance factors influencing initial public offering (IPO) of companies in Indonesia. *Global Research Review in Business and Economics*.
- [5] Rianttara, I. K. G. N., & Lestari, G. A. K. (2020). Pengaruh persentase saham ditawarkan, profitabilitas, dan earning per share terhadap underpricing pada perusahaan IPO di Bursa Efek Indonesia tahun 2017-2018. *Journal Research of Accounting (JARAC)*.
- [6] Yanto, A., & Others. (2020). Analysis of factors which affect company shares underpricing level at initial public offering (IPO) on the Indonesia Stock Exchange period 2013-2018. *International Journal of Economics and Management Studies*.
- [7] Bateni, L., & Asghari, F. (2014). Study of factors affecting the initial public offering (IPO) price of the shares on the Tehran Stock Exchange. *CURRENT: Jurnal Kajian Akuntansi dan Bisnis Terkini*.
- [8] Song, S. (2015). Investor sentiment, underwriters' behavior, and IPO pricing: Empirical analysis from off-line institutional investors' bids. *Journal of Financial Economics*.
- [9] Campbell, C., & Others. (2008). Market sentiment, IPO underpricing, and valuation. *Journal of Financial Economics*.
- [10] Derrien, F. (2005). IPO pricing in "hot" market conditions: Who leaves money on the table? *Journal of Finance*.
- [11] Chen, B., & Others. (2021). The impact of investor sentiment on IPO underpricing. *Proceedings of the 2021 International Conference on Public Relations and Social Sciences (ICPRSS 2021)*.
- [12] Yadav, A., & Others. (2023). Pre-IPO financial performance and offer price estimation: Evidence from India. *Journal of Risk and Financial Management*.
- [13] Vito, J., & Soenarno, V. (2023). Pengaruh debt to equity ratio, intellectual capital, earnings per share, return on equity, return on asset terhadap nilai perusahaan yang melakukan IPO. *Prosiding Working Papers Series in Management*.
- [14] Evitasari, R., & Nurhadi, N. (2023). Pengaruh financial leverage, profitabilitas, dan firm size terhadap tingkat underpricing saham IPO di Bursa Efek Indonesia (BEI). *CURRENT: Jurnal Kajian Akuntansi dan Bisnis Terkini*.

- [15] Gupta, C. P., & Suri, S. (2017). Valuation and pricing of Indian IPOs. *Journal of Financial Economics*.
- [16] Özyeşil, M. (2024). Macro-economic determinants of the IPO waves: An assessment of G-7 countries. *F1000Research*.
- [17] Min, J. (2020). Information spillover and demand shock effect of the IPOs on the stock price of competitors: Evidence from the Korean stock market. *Journal of Financial Economics*.
- [18] Vismara, S., & Others. (2014). Changes in underwriters' selection of comparable firms pre- and post-IPO: Same bank, same company, different peers. *Journal of Corporate Finance*.
- [19] Paleari, S., & Others. (2014). How do underwriters select peers when valuing IPOs? *Journal of Financial Economics*.
- [20] Chok, J., & Qian, J. (2013). Comparable stocks, boundedly rational stock markets, and IPO entry rates. *PLOS ONE*.
- [21] Abhyuday, A., & Others. (2024). A survey on machine learning techniques for stock market price prediction. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*.
- [22] Quintana, D., Sáez, Y., & Isasi, P. (2017). Random forest prediction of IPO underpricing. *Journal of Financial Markets*.