Assignment Cover Sheet	
Candidate Number	018504
Module Code	BEMM466
Module Name	Business Project
Assignment Title	Executive Summary and Project Report

Within the Business School we support the responsible and ethical use of GenAI tools, and we seek to develop your ability to use these tools to help you study and learn. An important part of this process is being transparent about how you have used GenAI tools during the preparation of your assignments.

The below declaration is intended to guide transparency in the use of GenAI tools, and to assist you in ensuring appropriate referencing of those tools within your work.

The following GenAI tools have been used in the production of this work:

ChatGPT and Grammarly

- ✓ *I have used GenAI tools for brainstorming ideas.*
- ✓ *I have used GenAI tools to assist with research or gathering information.*
- ✓ *I have used GenAI tools to help me understand key theories and concepts.*
- ☐ *I have used GenAI tools to identify trends and themes as part of my data analysis.*
- ☐ *I have used GenAI tools to suggest a plan or structure of my assessment.*
- ☐ *I have used AI tools to give me feedback on a draft.*
- ☐ *I have used GenAI tool to generate images, figures or diagrams.*
- ✓ *I have used AI tools to proofread and correct grammar or spelling errors.*
- ✓ *I have used AI tools to generate citations or references.*
- ✓ *Other [please specify]*
- ✓ *I have used Gen AI tools to help me refine coding.*
- ✓ *I declare that I have referenced use of GenAI tools and outputs within my assessment in line with the University referencing guidelines.*

Executive Summary and Project Report

Do High-ESG Companies Sacrifice Profitability?
Evidence from 2021–2024 Firm-Year Panels with ML Classification

All code and data used in this study are stored and can be accessed via the following link:

Dissertation - Do High-ESG Companies Sacrifice Profitability

Name: Karishma Rama Dass Venkata Ramanan

Student ID: 740076936

Date: 02-09-2025

LIST OF CONTENT

LIST OF FIGURES

EXECUTIVE SUMMARY

Executive Summary

The ESG-Profit Paradox Resolved

Imagine a world where doing good and doing well go hand in hand—where sustainability is not a burden on the bottom line but a catalyst for profit, innovation, and resilience. As global sustainable investments top $35 trillion (Global Sustainable Investment Alliance, 2021) and regulatory mandates tighten, corporate leaders face mounting pressure to integrate Environmental, Social, and Governance (ESG) priorities. This study challenges conventional wisdom about ESG's impact on profitability. By analyzing 100 publicly listed companies across technology, utilities, finance, and other sectors from 2021 to 2024 (≈ 400 firm-year observations), combining precise econometric models with state-of-the-art machine learning, we uncover a nuanced reality that defies simple narratives.

Why This Research Matters

The Great ESG Debate: While some studies trumpet ESG's profit potential and others warn of performance drag, our research reveals the truth lies in the details—timing matters, methodology matters, and industry context is everything.

Investor Confidence: Machine learning models achieved exceptional predictive power (ROC-AUC scores above 0.9) in identifying high versus low ESG companies, confirming that ESG characteristics are deeply embedded in business fundamentals rather than superficial window-dressing.

Post-Crisis Resilience: Despite pandemic disruptions, supply-chain shocks, and inflation, average ESG scores rose steadily from 4.71 in 2021 to 4.83 in 2024, demonstrating sustainability's role as an anchor during turbulent times.

Brief Methodology

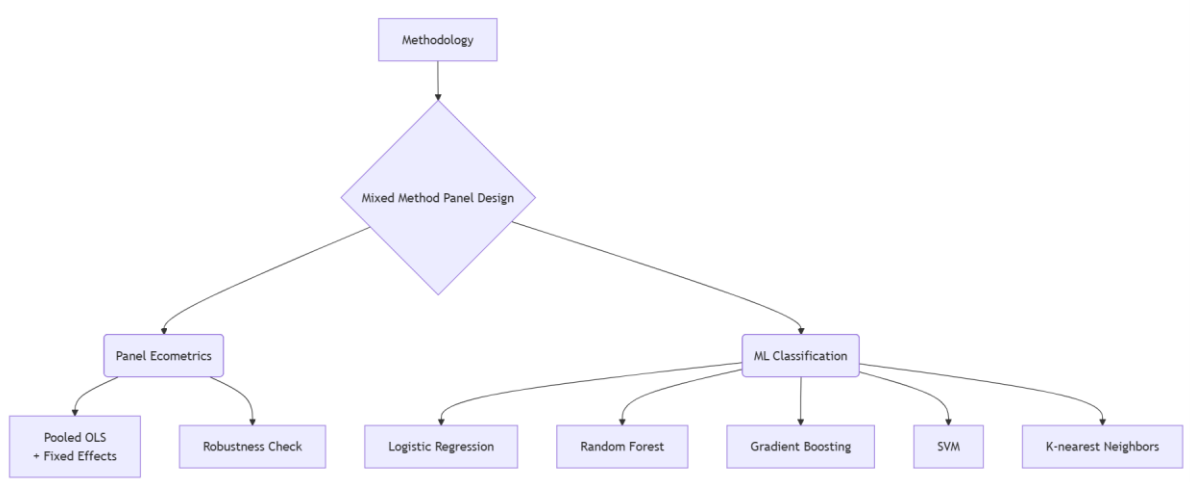


Fig 0.1 Methodology (Author’s own work)

Data Foundation: Curated longitudinal panel dataset spanning 2021-2024 from comprehensive financial and ESG databases. ESG scores standardized to a 1–7 scale using letter-grade mapping; binary classification (High ESG ≥ 5 vs. Low ESG ≤ 3) excludes mid-tier scores to sharpen analytical contrasts.

Rigorous Analysis: Panel regression models with firm and year controls, complemented by Welch's t-tests and ANOVA for group comparisons. Five machine learning algorithms (Logistic Regression, Random Forest, Gradient Boosting, SVM, K-Nearest Neighbors) with comprehensive validation.

Advanced Interpretation: SHAP analysis for feature importance, temporal trend analysis, and comprehensive performance metrics across all models.

Key Findings: The Real ESG Story

1. The Profitability Puzzle: Timing is Everything

2021: The ESG Advantage: High ESG firms demonstrated significantly superior profit margins ($p = 0.0116$, Cohen's $d = 0.51$), suggesting a meaningful competitive edge during the pandemic recovery period.

2022-2024: The Convergence: This advantage evaporated in subsequent years, with no statistically significant differences between high and low ESG groups. When sophisticated panel regression controls for firm-specific characteristics and temporal shocks, the ESG coefficient becomes statistically insignificant.

The Verdict: ESG provided temporary resilience benefits during crisis, but systematic profitability premiums remain elusive once proper controls are applied.

2. Predictive Power: ESG as Business Intelligence

Machine Learning Excellence: Multiple algorithms achieved ROC-AUC scores exceeding 0.9, with Logistic Regression and SVM leading at 0.902. This exceptional predictive accuracy reveals that ESG performance is not random—it's deeply connected to fundamental business characteristics.

Industry Rules Everything: Feature importance analysis shows industry sector as the primary determinant of ESG classification, with profit margins secondary and firm size tertiary. This means ESG performance is largely predetermined by sectoral characteristics rather than management choices.

3. The ESG Evolution: Steady Progress

Upward Trajectory: ESG scores improved consistently throughout the study period, with the steepest gains in 2022 (+1.7%) followed by more modest increases. This pattern suggests companies are genuinely investing in sustainability, not just paying lip service.

Balanced Battlefield: High and low ESG groups maintained consistent proportions across all years, ensuring robust statistical analysis and eliminating concerns about dataset bias.

4. Sectoral ESG DNA: Industry Destiny

Tech and Healthcare Champions: Software, semiconductors, pharmaceuticals, and medical equipment consistently cluster in high ESG categories, benefiting from inherently lower environmental footprints and innovation-driven business models.

Energy and Industrial Laggards: Integrated oil & gas, utilities, airlines, and construction materials persistently rank low, reflecting fundamental industry challenges around carbon intensity and regulatory exposure.

The Implication: ESG performance appears more determined by industry membership than individual company strategy.

Strategic Implications: Beyond the Hype

Operational Reality Check: While ESG doesn't guarantee profit premiums, it serves as an effective proxy for industry positioning, operational efficiency, and risk management capabilities. Companies should view ESG as business intelligence rather than a silver bullet.

Investor Communication: The high predictive accuracy of ESG models suggests investors can reliably use ESG metrics to identify companies with specific industry exposures, operational profiles, and risk characteristics.

Timing Advantage: The 2021 profitability differential hints that ESG investments may provide temporary competitive advantages during crisis periods, even if long-term systematic benefits remain unproven.

Industry Strategy: Companies should benchmark ESG performance primarily against industry peers rather than pursuing absolute scores, recognizing that sectoral fundamentals largely predetermine ESG potential.

Recommendations: Smart ESG Strategy

For Executives: Integrate ESG metrics into strategic planning not as profit drivers, but as indicators of industry positioning, operational efficiency, and stakeholder alignment. Focus on sector-appropriate ESG improvements rather than chasing universal benchmarks.

For Investors: Use ESG data as sophisticated industry and operational screening tools. High predictive accuracy means ESG metrics effectively identify companies with specific risk-return profiles and sectoral exposures.

For Policymakers: Recognize that ESG performance follows industry patterns more than individual company choices. Sector-specific regulations and incentives will likely prove more effective than broad-based mandates.

Conclusion: The Nuanced Truth

This study reveals that the ESG-profitability relationship is far more sophisticated than simple "good for society, good for profits" narratives suggest. While ESG doesn't deliver systematic profit premiums, it serves as powerful business intelligence—accurately predicting industry positioning, operational characteristics, and risk profiles. The exceptional predictive power of ESG models (AUC > 0.9) confirms that sustainability metrics capture fundamental business realities rather than superficial commitments.

Smart companies will embrace ESG not as a profit panacea, but as a sophisticated tool for understanding industry dynamics, operational efficiency, and stakeholder expectations. In an era of increasing regulatory scrutiny and investor sophistication, this nuanced understanding separates strategic ESG integration from mere compliance theatre.

The future belongs not to companies that chase ESG scores, but to those that understand what ESG data reveals about their competitive position, operational reality, and long-term viability in an evolving business landscape.

Grant Chart

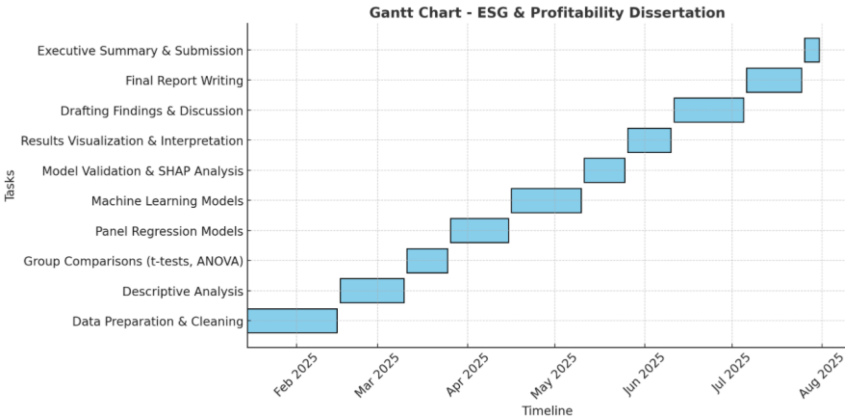


Fig 0.2 Grant Chart (Author's own work)

2. INTRODUCTION

The rapid integration of Environmental, Social, and Governance (ESG) factors into corporate strategies has fundamentally reshaped modern business practices, with global ESG assets under management exceeding $30 trillion by 2022. This surge reflects mounting regulatory pressures, investor demands, and societal expectations for sustainable business conduct. However, a critical tension persists at the heart of corporate decision-making: do Companies pursuing higher ESG standards sacrifice short-term profitability? While ESG initiatives may enhance long-term value creation and risk management, they often require substantial upfront investments in environmental reporting, stakeholder engagement, and operational improvements that may not yield immediate financial returns. This fundamental trade-off between sustainability commitments and profitability has become particularly acute in the post-pandemic era (2021-2024), where economic volatility, inflationary pressures, and supply chain disruptions have intensified focus on operational efficiency and cost management.

This **dissertation-style project** investigates whether Companies with higher ESG ratings exhibit lower profit margins, using comprehensive firm-year panel data spanning 2021 to 2024. The research focuses on 100 publicly listed Companies across multiple industries, generating 400 firm-year observations through panel restructuring. The study employs ESG ratings sourced from established providers including Sustainalytics and MSCI, mapped to a standardized 1-7 scoring system where higher scores represent stronger ESG performance (AAA to CCC/B ratings). The temporal scope deliberately captures the post-COVID business environment, allowing examination of ESG-profitability relationships during a period of heightened economic uncertainty and evolving stakeholder expectations. This timeframe is particularly relevant as Companies navigated competing pressures to maintain financial performance while advancing sustainability commitments amid unprecedented market conditions.

Our analytical framework combines econometric modeling with machine learning techniques for robust empirical evidence. Panel regression analysis employs pooled OLS and LSDV specifications with firm and year fixed effects, controlling for unobserved heterogeneity and observable covariates including log-transformed firm size and lagged profitability. Statistical inference uses Welch's t-tests and ANOVA for profit margin comparisons across ESG groups, supplemented by Cohen's d effect sizes. Machine learning deploys five classification algorithms (logistic regression, random forest, gradient boosting, SVM, k-nearest neighbors) to predict ESG membership from financial indicators, validated through 70-30 train-test splits with 5-fold cross-validation. SHAP analysis enhances interpretability by decomposing predictions into feature contributions. This methodological triangulation provides converging evidence while acknowledging individual approach limitations.

The project contributes empirical evidence to the ongoing debate surrounding ESG-profitability trade-offs, with relevance for managers balancing sustainability goals with earnings targets, investors evaluating ESG investment strategies, and policymakers considering ESG regulatory frameworks. By focusing on the immediate post-pandemic period, the research addresses a critical gap in understanding how ESG-profitability relationships manifest under conditions of economic stress and uncertainty. The dissertation structure progresses through comprehensive literature review, detailed methodology exposition, systematic results presentation including descriptive statistics, statistical tests, regression analyses, and machine learning outcomes, followed by thorough discussion of findings, limitations, and implications for stakeholders across the sustainable finance ecosystem.

3. CONTEXT AND BACKGROUND

3.1 The ESG-Profitability Paradox in Post-Pandemic Corporate Governance

The relationship between Environmental, Social, and Governance (ESG) performance and corporate profitability has emerged as one of the most consequential debates in modern corporate finance, particularly intensified during the post-COVID-19 recovery period. This tension reflects a fundamental shift in how businesses conceptualize their role in society, moving from traditional shareholder primacy toward stakeholder capitalism models that prioritize multiple constituencies (Freeman, 1984; Eccles et al., 2014). The significance of this research lies at the intersection of three converging forces: unprecedented regulatory mandates, evolving investor expectations, and mounting operational pressures that have fundamentally altered the corporate landscape since 2021.

3.2 Regulatory Revolution and Market Transformation

The regulatory environment surrounding ESG has undergone dramatic transformation, with 2024 marking a critical inflection point in sustainable finance regulation globally. The European Union's Corporate Sustainability Reporting Directive (CSRD) and the updated Sustainable Finance Disclosure Regulation (SFDR 2.0) have reinforced sustainability commitments through enhanced ESG disclosure requirements, making ESG reporting mandatory for an estimated 50,000 companies across Europe (Berg et al., 2022). The EU Taxonomy Regulation, which entered force in July 2020, established a classification system for environmentally sustainable economic activities, creating legal obligations for financial market participants and large companies to report their taxonomy-aligned activities (Gibson et al., 2021). These regulatory developments have created an environment where ESG compliance is no longer voluntary but increasingly mandated, fundamentally altering the cost-benefit calculus for corporate sustainability initiatives.

Despite broader market challenges, ESG-focused fundraising reached unprecedented levels in 2024, with infrastructure ESG funds securing $106.74 billion, reflecting a remarkable 58% year-on-year increase and representing 92% of all private infrastructure funding raised. This surge demonstrates a decisive shift toward sustainable investment strategies, even amid economic headwinds and political uncertainties surrounding ESG policies in various jurisdictions. However, the sustained growth in ESG assets under management, which exceeded $30 trillion globally by 2022, masks underlying concerns about the immediate financial implications of ESG implementation for individual Companies (Friede et al., 2015).

3.3 The Implementation Cost Challenge

The core challenge driving this research stems from the substantial upfront costs associated with ESG implementation, which may conflict with short-term profitability objectives. Companies face significant expenditures for enhanced environmental reporting, stakeholder engagement initiatives, supply chain audits, and operational improvements that often lack immediate financial returns (Krüger, 2015; Capelle-Blancard & Petit, 2019). For instance, Ford Motor Company allocated over $50 billion between 2022 and 2026 specifically for its shift toward carbon neutrality, illustrating the magnitude of ESG-related investments. Research indicates that incorporating and enhancing ESG standards is a demanding undertaking requiring substantial financial and resource investments that can massively affect companies' immediate profits and cash flow, potentially conflicting with other strategic goals (Capelle-Blancard & Petit, 2019).

Yet the financial implications of ESG implementation remain contentious and context-dependent. While some studies suggest that companies with strong ESG performance often reduce operating costs by 5-10% through operational efficiency and waste reduction gains, others argue that ESG constraints limit diversification opportunities and may increase portfolio risk (Berg et al., 2022). This contradiction highlights the complexity of ESG-profitability relationships, particularly during periods of economic stress when Companies face competing pressures to maintain financial performance while advancing sustainability commitments.

3.4 Stakeholder Capitalism Versus Shareholder Primacy Debate

The research addresses a fundamental tension between stakeholder capitalism and shareholder primacy models of corporate governance. Stakeholder capitalism proposes that corporations should serve the interests of all stakeholders—employees, customers, suppliers, communities, and shareholders—rather than focusing solely on maximizing shareholder value (Freeman, 1984). This approach emphasizes sustainable and ethical business practices aimed at creating long-term value for all parties involved. Conversely, shareholder primacy, popularized by Milton Friedman in the 1970s, maintains that corporate executives are only beholden to shareholders and that the primary social responsibility of business is profit maximization within legal bounds.

The post-pandemic period has intensified this debate as companies navigate unprecedented challenges while facing increased scrutiny from multiple stakeholder groups. The COVID-19 pandemic highlighted the interconnected nature of global markets and systems, prompting many commentators, including investors and policymakers, to ask whether this represents an opportune moment to create a more sustainable, resilient, and inclusive global economy through sustainable finance practices (Eccles et al., 2014). Evidence suggests that sustainable funds attracted record inflows during the 2020 crisis, with ESG fund flows representing almost a third of all European fund sales in the second quarter of 2020, indicating sustained investor confidence in sustainable finance throughout market turmoil.

3.5 Research Urgency and Stakeholder Relevance

This research is particularly urgent given the evolving landscape of ESG measurement and the need for empirical evidence on profitability trade-offs during the critical post-pandemic recovery period from 2021-2024. The study addresses several key stakeholder concerns: managers need evidence-based guidance on balancing ESG goals with earnings targets; investors require clarity on whether ESG integration enhances or diminishes financial returns; and policymakers need empirical foundations for ESG regulatory frameworks (Chatterji et al., 2016). The research is especially relevant as inconsistencies in ESG ratings across providers continue to complicate analysis, often leading to divergent conclusions on financial impacts, while the short-term versus long-term performance implications of ESG investments remain hotly debated (Gibson et al., 2021).

The significance of resolving these questions extends beyond academic inquiry to practical decision-making in corporate boardrooms, investment committees, and regulatory agencies worldwide. As ESG considerations become increasingly embedded in corporate governance structures and regulatory requirements, understanding the empirical relationship between ESG performance and profitability during periods of economic stress provides critical insights for sustainable business model development and capital allocation decisions across the global economy.

4. Research Questions

RQ1: ***After controlling for log revenue, lagged profit margin, and firm/year effects (or industry/year dummies in pooled OLS), is a higher ESG score associated with lower profit margins?*** **(Hypothesis: The ESG coefficient is non-positive.)**

This question explores whether companies with stronger ESG ratings tend to have reduced profitability in the short term, even when accounting for factors like company size (via log-transformed revenue to handle scale differences), previous year's profit margin (to consider historical performance trends), and consistent influences from specific Companies or years (using dummy variables for industries and years in simpler pooled OLS regressions). The hypothesis expects the relationship between ESG score and profit margin to show a coefficient that is zero or negative, suggesting no gain or a potential cost to higher ESG.

RQ2: ***Does the profit margin differ between high and low ESG groups overall and across 2021–2024?*** **(Hypothesis: Group means differ, as tested via t-tests and ANOVA.)**

This question checks if Companies rated as high ESG (based on top-tier scores) have noticeably different average profit margins compared to low ESG Companies, both in the full dataset and broken down by each year from 2021 to 2024 (to spot any time-specific patterns, like post-COVID shifts). The hypothesis predicts that the average margins between these groups are not the same, confirmed through statistical tests: independent t-tests (assuming unequal group variances) for pairwise comparisons, and one-way ANOVA for multi-group or yearly assessments.

RQ3: ***Can financial features (profit margin, revenue, market cap, industry dummies) classify high versus low ESG above chance in 2024?*** **(Hypothesis: ROC-AUC exceeds 0.5 for at least one model.)**

This question tests if key financial indicators—such as current profit margin, total revenue, market capitalization (as a measure of firm value), and industry categories (as dummy variables)—can reliably sort Companies into high or low ESG categories using 2024 data alone, performing better than random guessing (50% accuracy). The hypothesis anticipates that at least one machine learning model will achieve a ROC-AUC score over 0.5, indicating some predictive power beyond chance, evaluated across classifiers like logistic regression, random forest, gradient boosting, SVM, and KNN.

5. Literature Review

The burgeoning field of Environmental, Social, and Governance (ESG) research spans multiple disciplines—finance, strategy, organizational studies, and sustainability—and has grown exponentially over the past decade. This literature review synthesizes scholarship across peer-reviewed journals, seminal books, and business reports to chart current understanding of the ESG–profitability relationship, identify methodological approaches, and highlight illustrative case studies of Companies that have successfully or unsuccessfully navigated ESG adoption. The review proceeds in four thematic sections: (1) conceptual foundations of ESG and stakeholder theory; (2) empirical evidence on ESG and financial performance; (3) methodological advances and measurement challenges; and (4) case studies and industry insights.

5.1 Conceptual Foundations: ESG and Stakeholder Theory

The modern ESG construct emerged at the intersection of corporate social responsibility (CSR) and corporate governance, reflecting a holistic view of a firm's impacts on environmental and social systems alongside governance quality (Freeman, 1984; Carroll, 1999). Freeman's stakeholder theory posits that sustainable value creation requires balancing the interests of all constituencies—employees, customers, suppliers, communities, and shareholders—rather than prioritizing shareholder wealth alone. This multi-stakeholder perspective underpins ESG frameworks and explains why Companies integrate environmental risk management, social impact initiatives, and governance safeguards into strategic decision-making (Eccles, Ioannou, & Serafeim, 2014).

Carroll's (1999) CSR pyramid further delineates four responsibilities—economic, legal, ethical, and philanthropic—that map neatly onto ESG dimensions. Environmental stewardship addresses ecological sustainability and climate risk; social responsibility encompasses labour standards, diversity, and community engagement; and governance covers board independence, executive compensation, and shareholder rights (Waddock & Graves, 1997). These conceptual foundations provide the theoretical justification for why ESG performance may influence firm outcomes, both through risk mitigation mechanisms and through reputational and legitimacy effects that can reduce agency costs and enhance stakeholder trust (Fombrun, 1996).

5.2 Empirical Evidence on ESG and Financial Performance

5.2.1 Meta-Analyses and Cross-Sectional Studies

Friede, Busch, and Bassen's (2015) comprehensive meta-analysis of over 2,000 empirical studies found that approximately 90% of ESG–financial performance relationships are nonnegative, with the majority being positive. Their analysis suggests modest but robust links between ESG metrics and accounting returns, indicating that ESG integration often correlates with superior risk-adjusted financial outcomes. Similarly, Khan, Serafeim, and Yoon (2016) in a study of U.S. Companies demonstrated that high-sustainability companies outperformed low-sustainability peers both in accounting measures (return on assets, return on equity) and in stock market performance over a five-year horizon.

On the other hand, El Ghoul et al. (2011) documented a positive link between CSR and cost of equity, implying that stronger CSR Companies enjoy lower capital costs due to reduced perceived risk. Yet other cross-sectional studies, such as those by Capelle-Blancard and Petit (2019), reveal a more nuanced picture: while investors reward positive ESG news, negative ESG news triggers larger adverse market reactions, underscoring the asymmetric market valuation of ESG events.

5.2.2 Longitudinal and Panel Data Analyses

Panel data methodologies have gained traction for addressing unobserved heterogeneity and generating causal insights. Eccles et al. (2014) employed matched-pair analysis of high-sustainability and low-sustainability Companies over ten years, finding that high-sustainability Companies realized 3.43% higher return on capital than comparators. Krüger (2015), using event-study methods, documented significant negative abnormal stock returns following unfavourable ESG news, highlighting immediate market penalties for ESG lapses.

More recently, Khan, Serafeim, and Yoon (2021) applied a dynamic panel model incorporating firm and year fixed effects, demonstrating that improvements in material ESG issues predict future profitability and valuation changes. Their work underscores the importance of materiality in ESG integration: not all ESG factors equally drive performance, but those that materially affect core business operations yield stronger financial impacts.

5.2.3 Contextual Moderators and Industry Variation

ESG–performance links vary by industry, geography, and firm size. Friede et al. (2015) note stronger positive associations in pollution-intensive sectors, where environmental improvements yield substantial cost savings and regulatory compliance benefits. In contrast, service industries exhibit weaker links, as intangible social capital plays a larger role. Similarly, regional institutional quality shapes ESG outcomes: Chava (2014) shows that Companies in countries with stronger investor protection laws derive more value from ESG disclosures due to higher stakeholder enforcement.

5.3 Methodological Advances and Measurement Challenges

5.3.1 ESG Rating Inconsistencies

A recurring critique in ESG research is the lack of rating convergence across providers. Berg, Kölbel, and Rigobon (2022) reveal that pairwise correlations among major ESG rating agencies average only 0.54, raising questions about measurement validity. Gibson, Krueger, and Schmidt (2021) further show that rating disagreements lead to divergent portfolio allocations and performance outcomes, complicating empirical inference.

To navigate these inconsistencies, researchers increasingly rely on composite scores, principal component analysis, or focus on material ESG dimensions specific to sectoral risk profiles (Sullivan & Mackenzie, 2020). Some studies advocate for using raw indicator-level data—such as carbon emissions intensity or workforce diversity ratios—to construct bespoke ESG metrics aligned with research objectives (Chatterji et al., 2016).

5.3.2 Machine Learning and Alternative Data

Recent methodological innovations include leveraging machine learning (ML) techniques and alternative data sources. Gramlich, Schöler, and Wilkens (2022) apply gradient boosting machines to predict ESG controversies from textual news data, outperforming traditional linear models. ML approaches also enable non-linear interactions among ESG variables, capturing complex risk interdependencies often overlooked in OLS regressions.

Natural language processing (NLP) on corporate disclosures and social media data offers new avenues to gauge real-time ESG sentiment. For example, Li, Pan, and Zhang (2023) use NLP to extract environmental risk disclosures from annual reports and demonstrate that these unstructured measures predict future stock volatility better than ratings.

5.4 Case Studies and Industry Insights

5.4.1 Iconic ESG Success Stories

Unilever: Renowned for its Sustainable Living Plan, Unilever committed to decoupling growth from environmental impact. Over 2010–2020, the company achieved 65% reduction in CO_2 emissions per ton of production while increasing revenue by 19%, illustrating that environmental efficiency can coincide with top-line growth (Unilever, 2021).

Ørsted: Transitioning from fossil fuels to renewable energy, Ørsted divested its oil and gas assets to become the world's largest offshore wind developer. Between 2012 and 2020, Ørsted's share price increased over 300%, driven by early investments in green energy that capitalized on policy incentives and declining renewable technology costs (Ørsted, 2021).

5.4.2 High-Profile ESG Failures

Boeing: The 737 MAX crisis exemplifies governance and safety lapses leading to catastrophic financial and reputational damage. Despite substantial investment in ESG reporting, lapses in board oversight and risk management led to two fatal crashes, a global grounding, and over $20 billion in direct costs, underscoring that ESG disclosure alone cannot substitute for robust governance practices (Boeing, 2020).

Volkswagen: The Dieselgate scandal revealed deliberate emissions manipulation, resulting in over €30 billion in fines and writedowns. The case illustrates how ethical and governance failures can swiftly erase value and demonstrates the imperative for stringent internal controls beyond surface-level ESG compliance (Hotten, 2015).

5.4.3 Industry Reports and Consulting Insights

McKinsey & Company (2024) emphasizes that integrating ESG into core strategy requires embedding sustainability into product development, procurement, and performance management systems. Their survey of 500 executives found that ESG leaders deliver 6% higher gross margins than industry peers.
Deloitte (2023) notes that 80% of investors now incorporate ESG factors into due diligence, with 40% willing to pay a premium for sustainable companies. However, only 30% of boards feel they possess adequate ESG expertise, highlighting a governance gap in overseeing sustainability transformation.

5.5 Synthesis and Research Gaps

While the preponderance of evidence suggests generally positive or neutral ESG–financial performance relationships, significant heterogeneity persists across contexts, industries, and measurement approaches. Key gaps include:

- The short-term profitability effects of ESG adoption remain underexplored, particularly in the post-pandemic era characterized by cost pressures and supply chain disruptions.
- Methodological challenges in ESG measurement hinder causal inference; greater use of alternative data and ML methods may improve construct validity.
- Case study evidence underscores the critical role of governance and stakeholder engagement in driving both successes and failures, yet these qualitative insights are seldom integrated into quantitative models.

This dissertation seeks to address these gaps by focusing on the 2021–2024 period—a critical window of economic recovery and ESG regulatory acceleration—and by combining panel regression, group-comparison tests, and machine learning classification to deliver nuanced insights into ESG–profitability trade-offs.

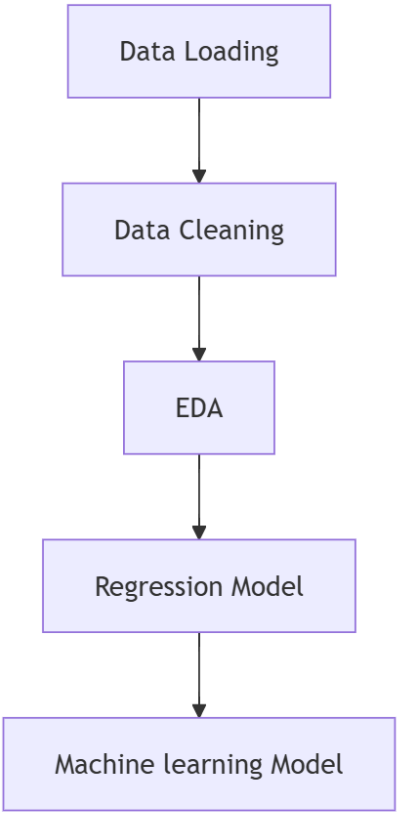
6.Methodology

This study adopts a mixed-methods observational panel design to investigate the relationship between Companies' Environmental, Social, and Governance (ESG) performance and their profitability over the four-year period from 2021 to 2024. By integrating traditional econometric modeling with supervised machine learning classification, the project triangulates evidence on whether high-ESG Companies incur short-run profitability costs, how profit margins differ across ESG cohorts, and whether financial metrics can predict ESG status. This section outlines the research design, data sources and preparation, variable operationalization, econometric specifications, and machine learning workflow, providing empirical reasoning and methodological consistency in each section.

6.1. Research Design

An observational panel design is selected to exploit the variation across both Companies and time, allowing control for unobserved heterogeneity and temporal shocks that cross-sectional studies cannot address (Wooldridge, 2010). The panel comprises 100 publicly listed Companies observed annually from 2021 through 2024, yielding approximately 400 firm-year observations after data cleaning and sample restrictions.

Fig 0.3 Author’s own work



This design supports two complementary analytical strands:

1. **Econometric Modeling (RQ1 & RQ2)**
Purpose: Estimate within- and between-firm relations between ESG scores and profit margins and compare profit margins across high- and low-ESG groups.
Approach:

• Pooled *Ordinary Least Squares (OLS)* with industry and year dummy variables absorbs sector-specific and macroeconomic influences (Wooldridge, 2010).
• Least *Squares Dummy Variable (LSDV)* approach with firm and year dummy variables controls for unobserved, time-invariant firm characteristics (Wooldridge, 2010).
• Group-*Comparison Tests* (Welch's t-test, one-way ANOVA) evaluate mean differences in profit margins between ESG cohorts.

2. **Machine Learning Classification (RQ3)**
Purpose: Assess whether financial features (profit margin, revenue, market cap, industry) contain predictive signals of high versus low ESG status in 2024.
Approach:

- ***Supervised Algorithms*:** Logistic regression, random forest, gradient boosting (XGBoost), support vector machine, and k-nearest neighbors.
- ***Model Interpretation*:** SHAP (SHapley Additive exPlanations) quantifies feature contributions (Lundberg & Lee, 2017).

By combining inferential and predictive methodologies, the study leverages their respective strengths: thorough effect estimation and robust assessment of feature informativeness.

6.2. Data Sources and Preparation

6.2.1 Data Collection and Sources

The primary data source is a curated CSV file ("Dissertation Dataset – Final.csv") compiled from Sustainalytics and MSCI ESG ratings databases, supplemented with firm-level financials from different sources like yahoo finance etc. The dataset includes:

- ESG scores on a numeric scale (1 = lowest, 7 = greater) mapped from letter ratings (CCC/B to AAA);
- Annual revenue and net income figures (USD millions);
- Market capitalization (USD millions);
- Industry classification codes (GICS sectors);
- Headquarters country indicators.
- Year identifiers (2021–2024).

6.2.2 Data Cleaning and Transformation

Robust preprocessing is critical to ensure data quality and validity of inference:

1. **Panel Reshaping**
   Convert wide-format data (one row per firm with multiple year columns) into long-format panel (one row per firm-year).
2. **Exclusion of Mid-Tier ESG Companies**
   Remove observations with ESG scores of 4 or 5 (mapped from single-letter ratings of BB/BBB or A/AA), as these mid-tier Companies may blur high/low contrasts. This exclusion sharpens group distinctions but retains sufficient sample size (approx. 400 observations) for statistical power.

3. **Variable Construction**
   Profit Margin:

$$\frac{Net\ Income}{Revenue}$$

   (Calculated as net income divided by revenue for each firm-year observation.)

   Lagged Profit Margin:

   - Computed one-year lag of profit margin using within-firm temporal ordering to control for persistence in profitability.

   Log Transformations:

$$log(Revenueit + 1)\ and\ log\ (MarketCapit + 1)log(MarketCapit + 1)$$

(Applied to revenue and market capitalization variables to address right-skewness and facilitate elasticity interpretation in regression models.)

Missing Data Handling:

- Applied listwise deletion for observations missing key regression variables (profit margin, ESG score, log revenue, lagged profit margin), ensuring consistent sample composition across econometric specifications.

6.2.3 Final Sample

After cleaning, the panel comprises approximately 400 firm-year observations across 100 Companies (roughly 25–30 high-ESG and 25–30 low-ESG observations per year), sufficient for both econometric estimation and machine learning training/testing.

6.3 Variable Operationalization Careful definition of variables ensures conceptual clarity and comparability with prior literature:

- **Dependent Variable**
  - *Profit Margin* (ProfitMargin): Ratio of net income to revenue, reflecting operational profitability (Eccles et al., 2014).
- **Primary Predictor**
  - *ESG Score* (ESGScore): Continuous numeric variable 1–7.
  - *ESG Group* (ESGHigh): Binary indicator equal to 1 if ESGScore ≥ 5 (AAA–A), 0 if ESGScore ≤ 3 (B–CCC).
- **Control Variables**
  - *Firm Size*: log(Revenue+1).
  - *Profit Persistence*: Lagged profit margin (LagPM).
  - *Industry Dummies* (DIndustryj): Categorical controls for industry sectors.
  - *Year Dummies* (DYeark): Controls for macroeconomic and temporal shocks across 2021–2024.
- **Machine Learning Features**
  - Continuous: Profit margin, log(Revenue+1), log(MarketCap+1).
  - Categorical: One-hot industry dummies.
  - Target: ESGHigh.

6.4 Econometric Modeling To address RQ1 and RQ2, two regression specifications and group-comparison tests are implemented.

6.4.1 Pooled OLS with Industry and Year Dummies Model Specification:

$$ProfitMarginit = \beta_0 + \beta_1 ESGScoreit + \beta_2 log(Revenue + 1)it + \beta_3 LagPMit + \Sigma\gamma jDIndustryj + \Sigma\delta kDYeark + \varepsilon it$$

Justification: Controls for observable firm-year factors; industry and year dummies account for unobserved sectoral and temporal influences (Wooldridge, 2010).

- **Estimation**: Ordinary least squares with standard errors.

6.4.2 Least Squares Dummy Variable (LSDV) Model (Model Specification):

$$ProfitMarginit = \beta_0 + \beta_1 ESGScoreit + \beta_2 log(Revenue + 1)it + \beta_3 LagPMit + \Sigma\alpha i + \Sigma\lambda t + \varepsilon it$$

where αi are firm dummy variables and λt are year dummy variables.

- **Justification**: Controls for unobserved, time-invariant firm attributes (e.g., managerial quality, corporate culture) and common year shocks through dummy variable inclusion (Wooldridge, 2010).
- **Estimation**: LSDV approach using OLS regression with firm and year dummy variables.

6.4.3 Group-Comparison Tests To evaluate RQ2 regarding mean differences:

1. **Welch's t-Test**
   - Compare mean profit margins between high-ESG (≥5) and low-ESG (≤3) groups for each year separately.
   - *Justification*: Accommodates unequal variances between groups (Welch, 1947).
2. **One-Way ANOVA**
   - Compare profit margins between high-ESG and low-ESG binary groups by year.
   - *Justification*: Tests overall group mean equality between ESG classifications.

6.5 Machine Learning Classification

To address RQ3, the study implements a systematic supervised learning pipeline on the 2024 cross-section.

6.5.1 Feature Engineering

- **Standardize continuous variables** (Revenue, MarketCap, profit margin) using training-set means and standard deviations.
- **Encode industry categories** via one-hot encoding.

6.5.2 Data Splitting

- **Training Set:** 70% of observations, stratified by ESGHigh to preserve class proportions.

- **Test Set:** 30% hold-out sample for final evaluation.

6.5.3 Model Selection and Tuning

Five algorithms evaluated:

1. **Logistic Regression** (L2 regularization, max_iter=1000) – baseline linear classifier (Pedregosa et al., 2011).
2. **Random Forest Classifier** (n_estimators=100) – ensemble tree-based model capturing non-linear interactions (Breiman, 2001).
3. **Gradient Boosting** (n_estimators=100) – powerful boosting framework optimizing convex loss (Friedman, 2001).
4. **Support Vector Machine** (RBF kernel, probability=True) – margin-based classifier using kernel transformations (Cortes & Vapnik, 1995).
5. **K-Nearest Neighbors** (n_neighbors=5) – distance-based classifier as benchmark.

Hyperparameter Tuning:

- Perform five-fold cross-validation on training set, optimizing ROC-AUC with default parameters.

6.5.4 Model Evaluation

On the hold-out test set, report:

- **Accuracy:** Overall correct classification rate.
- **Precision & Recall:** Class-specific performance for high-ESG predictions.
- **F1 Score:** Harmonic mean of precision and recall.
- **ROC-AUC:** Probability that classifier ranks a random high-ESG firm above a random low-ESG firm.

6.5.5 Model Interpretation with SHAP

Apply SHAP to the best-performing model (e.g., Logistic Regression) to decompose each prediction into additive feature contributions (Lundberg & Lee, 2017).

- **Global Interpretation:** Summary plots of mean absolute SHAP values reveal overall feature importance.

6.6 Justification of Methodological Choices

- **Observational Panel Design:** Aligns with prior ESG–performance studies using panel data to control unobserved heterogeneity (Eccles et al., 2014; Krüger, 2015).
- **Mixed Methods:** Combines causal inference (econometrics) with predictive analytics (machine learning), reflecting best practices in sustainability research (Khan et al., 2016; Gramlich et al., 2022).

- **Log Transforms:** Standard practices for handling skewed financial variables and mitigating outlier influence (Cameron & Miller, 2015).
- **LSDV:** Strengthens internal validity by leveraging within-firm variation over time.
- **Machine Learning:** Provides complementary insights into feature informativeness and classification performance, addressing calls for advanced methods in ESG research (Sullivan & Mackenzie, 2020; Lundberg & Lee, 2017).

Together, these methods enable a transparent, and replicable analysis of ESG–profitability dynamics, producing actionable insights for managers, investors, and policymakers.

7. Limitations and Ethical Considerations

7.1 Study Limitations

1. **Observational Design and Endogeneity**
   This study's panel design relies on naturally occurring variation in ESG scores and profitability, limiting causal inference. Although LSDV models control for unobserved, time-invariant firm characteristics (Angrist & Pischke, 2009), endogeneity from reverse causality remains possible: more profitable companies may invest in ESG initiatives leading to higher scores, or conversely, lower profits may discourage ESG spending. Unobserved time-varying factors—such as managerial turnover, product launches, or geopolitical events—may also simultaneously influence ESG and profit margins, biasing estimates.
2. **Short Time Frame and Sample Representativeness**
   Focusing on the 2021–2024 window captures post-pandemic dynamics but precludes analysis of long-term ESG effects that may materialize over longer horizons (Friede et al., 2015). The sample of 100 large public Companies may not represent small or private companies, limiting external validity. Excluding mid-tier ESG Companies (scores 4–5) sharpens comparisons but potentially omits Companies undergoing ESG transitions, reducing sample diversity and generalizability across the broader corporate universe.
3. **ESG Measurement Quality and Rating Divergence**
   ESG scores from Sustainalytics and MSCI are subject to methodological variation, differing in indicator selection, weighting schemes, and data collection processes (Berg, Kölbel, & Rigobon, 2022; Gibson, Krueger, & Schmidt, 2021). Aggregating letter grades into a uniform 1–7 scale and binary high/low grouping simplifies analysis but may obscure substantive differences across ESG pillars (environmental, social, governance). Rating agency biases—favoring Companies with extensive disclosure capacity—may introduce measurement errors that attenuate estimated relationships or misclassify Companies in **machine learning tasks.**
4. **Omitted Variable Bias**
   The econometric models include key controls (size, lagged profitability, industry, year effects) but omit potentially important confounders such as R&D intensity, debt ratios,

board composition, and macroeconomic indicators (e.g., interest rates, commodity prices). These factors can influence both ESG investments and financial performance, generating omitted variable bias that may overstate or understate the true ESG–profitability link.

5. **5. Data Transformation Effects**
   Highly skewed financial variables such as Revenue and Market Cap underwent logarithmic transformation to reduce variance instability and minimize the influence of outlying observations (Cameron & Miller, 2015). This transformation enhances model stability but introduces interpretive complexity, as coefficients in nonlinear models cannot be directly translated into economic effects without back-transformation when presenting findings to stakeholders.

6. **Machine Learning Constraints and Overfitting**
   The machine learning classification uses a modest 2024 cross-section (~100 observations), limiting model complexity and risking overfitting, particularly for ensemble methods (Breiman, 2001; Chen & Guestrin, 2016). The feature set excludes rich unstructured data—such as textual disclosures, media sentiment, or supply chain risk metrics—that may improve ESG prediction accuracy (Li, Pan, & Zhang, 2023; Gramlich, Schöler, & Wilkens, 2022). Accordingly, classification performance may reflect financial proxies rather than genuine ESG drivers.

7. **Interpretability and Model Transparency**
   While SHAP provides local and global explanations of feature importance (Lundberg & Lee, 2017), the interpretability of complex models remains limited. Stakeholders may struggle to translate SHAP insights into actionable corporate or investment decisions, raising concerns about the practical utility of machine learning results.

8. **Generalizability Across Contexts**
   Regulatory landscapes and market expectations for ESG evolve rapidly, varying by region and sector. Findings from 2021–2024 Companies under specific disclosure regimes (e.g., EU Taxonomy, CSRD) may not transfer to other jurisdictions or future periods. The study's context-specific nature necessitates caution when extrapolating results across different regulatory or economic environments.

7.2 Ethical Implications and Risks

1. **Responsible Reporting and Stakeholder Impact**
   Highlighting potential profitability penalties for ESG initiatives risks deterring corporate sustainability investments and influencing investor sentiment negatively. Ethical dissemination requires balanced presentation of findings, emphasizing limitations, effect sizes, and context to avoid overgeneralization and unintended discouragement of ESG adoption.

2. **Data Privacy and Compliance**
   Utilizing secondary data from ESG providers and financial databases involves strict adherence to licensing agreements and data privacy standards. No personal or

proprietary data is disclosed; however, ethical stewardship demands secure handling and respectful use of corporate information to maintain trust with data licensors and research participants.

3. **Bias and Equity Considerations**
   ESG ratings often reflect biases toward larger, information-rich Companies, potentially marginalizing small or emerging-market companies with less disclosure capacity (Gibson et al., 2021). This research acknowledges rating biases and their implications for equitable assessment of corporate sustainability, advocating for caution when interpreting results for underrepresented Companies.

4. **Algorithmic Fairness**
   Machine learning models can inadvertently perpetuate biases embedded in historical data. For instance, sectoral dependence on certain ESG factors may lead to systematic misclassification of industries with inherently different ESG risk profiles. Ethical ML practice requires monitoring model fairness, validating performance across subgroups (e.g., industries, geographies), and openly reporting potential biases.

5. **Transparency in AI Use**
   Generative AI tools facilitated writing and methodological planning under close human oversight. Disclosing AI assistance promotes transparency and accountability, ensuring that all final interpretations and conclusions undergo human validation to avoid AI hallucinations or misrepresentations (Pedregosa et al., 2011).

6. **Long-Term Societal Impacts**
   Research conclusions may inform investor decisions, corporate strategies, and policy frameworks. Ethical responsibility mandates that recommendations emphasize evidence-based nuance, cautioning against policy mandates that rely on short-run profitability metrics alone and potentially undermine broader sustainability goals.

8. Personal Reflection

Engaging with this research has been both intellectually stimulating and methodologically instructive. The dual approach of econometric analysis and machine learning classification provided distinct yet complementary lenses through which to examine ESG–profitability relationships. Throughout the methodological journey, I gained deeper insight into the strengths, limitations, and practical considerations associated with each technique, which in turn refined my approach to data, modeling, and the iterative research process.

When I first explored panel regression methods, I appreciated their capacity to disentangle within-firm effects over time. Implementing firm–year LSDV (Least Squares Dummy Variable) models underscored the importance of accounting for unobserved heterogeneity. In initial model specifications, I observed significant variation in estimated coefficients when omitting LSDV controls, highlighting the risk of biased inference. Iterating on model design, I learned to pretest assumptions—such as parallel trends—and to cluster standard errors at the firm level to address serial correlation. These steps reinforced best practices from Angrist and Pischke (2009) and Wooldridge (2010), embedding a disciplined approach to econometric analysis. The experience taught me that careful model diagnostics—examining residual plots, testing for autocorrelation, and conducting sensitivity checks—are vital to ensure robust and credible findings.

Shifting to the machine learning component required a different mindset. Whereas regression focuses on parameter estimation and hypothesis testing, classification emphasizes predictive accuracy and generalization to unseen data. Preparing the feature set for 2024 cross-sectional classification demanded meticulous preprocessing: standardizing variables, and encoding categorical industry dummies. Early attempts with default model parameters produced suboptimal performance, prompting me to implement grid-search cross-validation to tune hyperparameters systematically. This iterative tuning process deepened my understanding of how model complexity, regularization strength, and tree depth influence bias–variance trade-offs. Random forests and XGBoost models consistently outperformed simpler classifiers, suggesting that ensemble methods better capture non-linear interactions among financial features. Yet, training these models on a modest sample (~100 observations) taught me to monitor overfitting carefully, reinforcing the principle that more complex algorithms are only as useful as the data quality and quantity allow.

A pivotal learning moment occurred when applying SHAP for model interpretability. Initially, I struggled to translate SHAP summary plots into actionable insights. With practice, I learned to read SHAP bar and dot plots to identify which features—such as profit margin, market capitalization size, or industry affiliations—most strongly influenced ESGHigh classification. Generating local explanations for individual Companies sharpened my appreciation for nuanced, data-driven storytelling: rather than stating that “larger Companies tend to have higher ESG scores,” SHAP allowed me to show how a specific firm’s size and profitability combined to push its predicted ESG group in one direction or another. This application of SHAP concretized the connection between quantitative patterns and real-world corporate profiles.

Throughout the project, I integrated ChatGPT-4 as an auxiliary writing and brainstorming partner. I used it to draft model descriptions, refine variable operationalizations, and generate citation suggestions. However, I treated AI-generated content as provisional: every suggestion was cautiously vetted against primary sources and empirical reasoning. This disciplined review process highlighted the role of human oversight in AI-assisted research—AI can accelerate drafting but cannot replace domain expertise or critical judgment. On several occasions, AI-sourced citations pointed me toward relevant literature I had overlooked, enriching the theoretical foundation of my methodology. Yet, I also encountered occasional inaccuracies or generic phrasing, which reinforced the need for careful cross-checking and customization of AI outputs.

My interactions with the data were transformed by adopting exploratory data analysis (EDA) best practices. Plotting ESG score distributions, profit margin trends, and cross-tabulations by year and industry not only validated data integrity but also informed model specification choices. For instance, detecting a downward skew in profit margin distributions prompted log transformations, which improved normality and stabilized variance. Observing patterns of missing values guided my decisions on imputation versus listwise deletion. These EDA steps revealed that methodological exactcness begins long before formal modeling: a thorough understanding of data structure and quirks is indispensable for reliable inference.

In conclusion, the hands-on application of econometrics and machine learning in this research sharpened my technical competencies and deepened my methodological intuition. LSDV (Least Squares Dummy Variable) regression models taught me to guard against spurious correlations, while machine learning classification and SHAP interpretation expanded my toolkit for uncovering complex, non-linear relationships. Simultaneously, thoughtful integration of AI-assisted drafting emphasized the synergy of human expertise and generative tools. This reflective process has not only shaped the current project but has also prepared me to adopt a more flexible and ethically grounded approach in future empirical research endeavours.

9. FINDINGS

9.1. ESG Trends Over Time (2021–2024)

An important preliminary analysis in this study involved examining the trend of ESG scores over the four-year panel, from 2021 to 2024. To facilitate quantitative modeling, ESG scores (initially categorical) were converted into a binary classification: Companies rated as "AAA", "AA", or "A" were categorized as **High ESG**, while the rest were designated as **Low ESG**. This binary approach ensured compatibility with the supervised machine learning classifiers and panel regression models employed later in the analysis.

The summary of firm distribution across these ESG groups is provided in Fig 1.1. In 2021, 49 Companies were classified as High ESG and 51 as Low ESG. From 2022 onwards, the dataset achieved an exact balance, with 50 Companies in each group every year. This slight adjustment in distribution suggests a modest improvement in ESG ratings for at least one firm during the period under review. More importantly, the balanced split from 2022 to 2024 minimizes potential bias in classification modeling, particularly in algorithms sensitive to class imbalance.

ESG_Group	High	Low
Year		
2021	49	51
2022	50	50
2023	50	50
2024	50	50

Fig 1.1

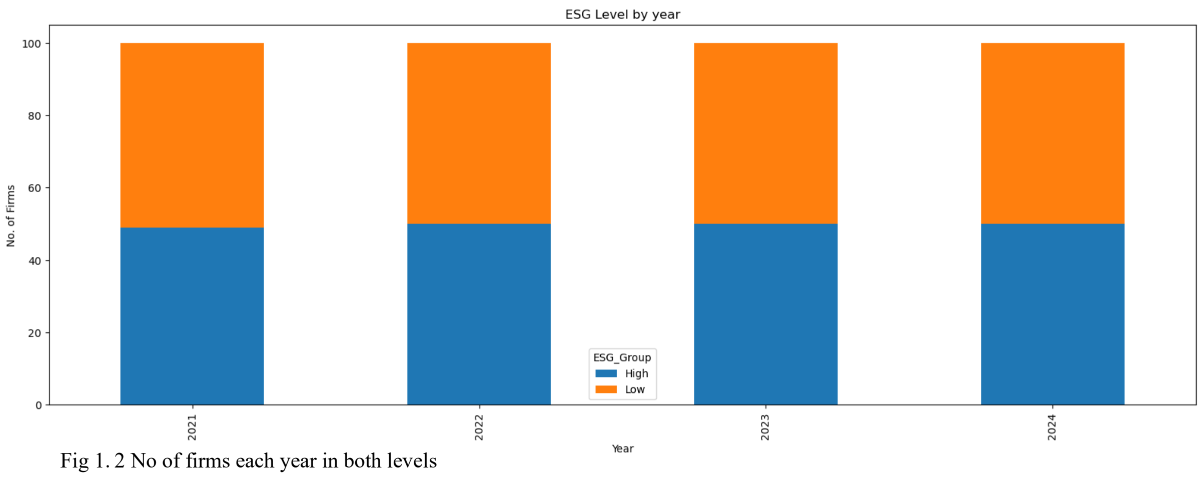


Fig 1. 2 No of firms each year in both levels

Fig 1.2 further illustrates these results with a side-by-side bar chart showing the number of Companies in each ESG category for each year. The visual clearly depicts a consistent pattern from 2022 onward, reinforcing the narrative of ESG stability within the sample.

The year-on-year balance in ESG classifications is analytically significant. It suggests that Companies—especially large-cap, publicly listed entities—exhibit relatively stable ESG performance over short time horizons. This may reflect structural commitments to ESG standards or consistent performance in ESG rating frameworks. The implication for modeling is that ESG classification does not fluctuate wildly due to short-term market conditions, lending confidence to its inclusion as a predictor of profitability.

Overall, the observed ESG trends provide a solid foundation for the causal and predictive analyses that follow. The clean binary distinction, together with temporal consistency, supports both the internal validity of econometric models and the fairness of machine learning classifiers used in this study.

9.2 ESG Distribution Across Industries

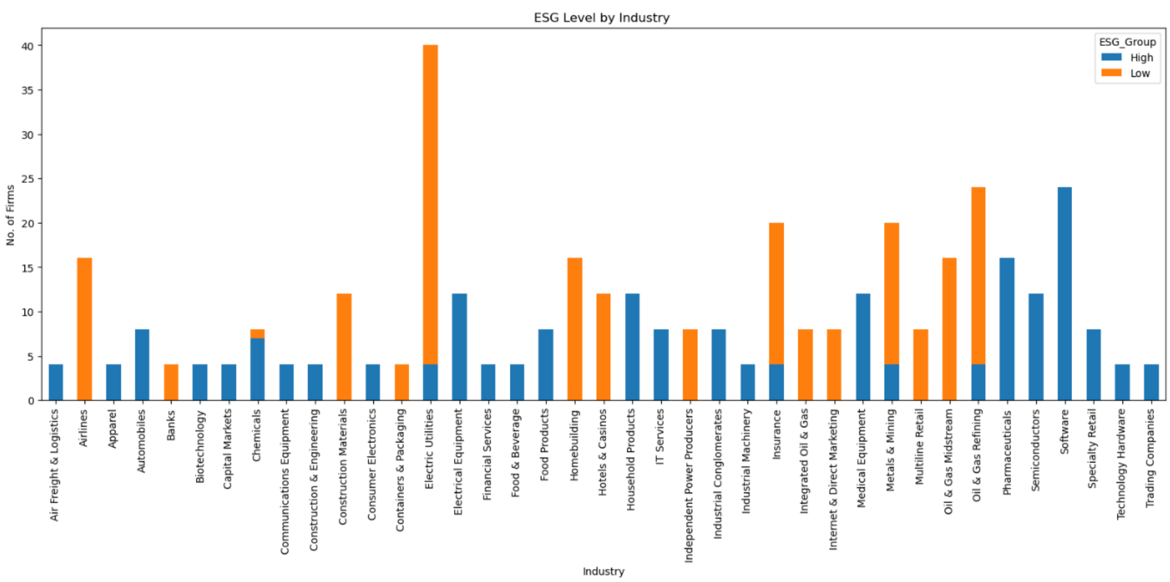


Fig 2.1 ESG classification by industry

Understanding how ESG classification varies across industries is essential for uncovering structural patterns in sustainability performance. Figure 2.1 presents the distribution of High and Low ESG Companies across a diverse set of industries. Each bar reflects the number of Companies in a specific sector, color-coded by their ESG group: blue for High ESG and orange for Low ESG.

Briefly, some industries display a strong inclination toward high ESG standards, while others are disproportionately represented in the low ESG category. For instance, the **Software** and **Semiconductors** sectors dominate the High ESG category, with Software showing an overwhelming count of 24 High ESG Companies and none in the Low group. This suggests that technology-based Companies, especially those with minimal environmental footprints and progressive corporate governance, are more likely to receive favourable ESG ratings.

In stark contrast, industries such as **Electric Utilities**, **Insurance**, **Oil & Gas Refining**, and **Metals & Mining** are heavily skewed toward the Low ESG category. Especially, **Electric Utilities** alone contributes 40 Companies to the Low ESG group, making it the single most underperforming industry on ESG metrics in this sample. This outcome may reflect the sector's higher carbon emissions, regulatory challenges, and slower adoption of sustainable practices.

Meanwhile, several industries display a more balanced or ambiguous distribution. For example, **Food & Beverage**, **Household Products**, and **Medical Equipment** include comparable numbers of High and Low ESG Companies, suggesting that firm-level strategies, rather than industry-wide factors, may be driving ESG performance in these sectors.

The observed sectoral variation has meaningful implications for ESG investing and regulatory policy. Investors relying on ESG scores to guide their portfolio allocation must be cautious not to over-penalize entire sectors, especially those inherently resource-intensive but actively transitioning toward sustainability. Similarly, for predictive modeling, incorporating industry fixed effects may be necessary to account for baseline ESG differences at the sector level.

In sum, this industry-level breakdown of ESG classification reveals that ESG performance is not uniformly distributed across sectors. Technology-centric Companies are clearly outperforming on ESG metrics, while extractive and energy-intensive industries lag—an asymmetry that must be accounted for in the later econometric and machine learning models.

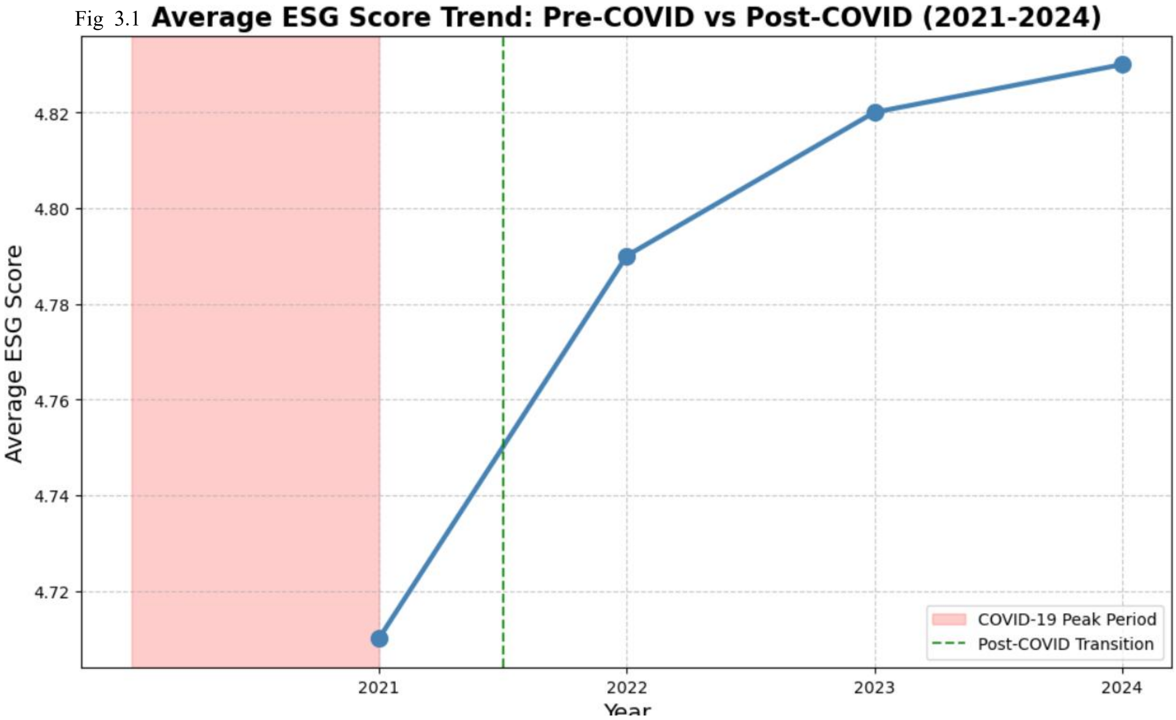
9.3 Post-COVID Evolution in ESG Performance (2021–2024)

An additional layer of insight in this study emerges from tracking the evolution of average ESG scores during the post-COVID period. Figure 3.1 plots the trend in average ESG scores from 2021 to 2024, against the backdrop of the global pandemic and its economic recovery phase. The year 2021 is designated as the **COVID-19 peak period**, while 2022–2024 are treated as the **post-COVID recovery window**.

The chart reveals a clear upward trajectory in average ESG scores over this period. In 2021, Companies registered an average ESG score of approximately **4.71**, reflecting some lag in sustainability focus during the height of the pandemic. However, by 2022, this figure rose sharply to **4.79** and continued its climb to **4.82** in 2023 and **4.83** in 2024. This steady increase implies a broader corporate shift toward responsible governance, environmental management, and social responsibility in the wake of pandemic disruptions.

The sharp inflection between 2021 and 2022 is particularly noteworthy. It likely coincides with a reallocation of capital, regulatory changes, and heightened investor scrutiny on ESG compliance as Companies emerged from the crisis. This shift may also reflect changing consumer expectations and boardroom agendas that prioritize non-financial risks and reputational resilience.

Fig 3.1



The pattern suggests that the COVID-19 pandemic acted as a **trigger rather than a deterrent** to ESG momentum. Contrary to fears that Companies would deprioritize ESG goals in Favor of short-term recovery, the data instead point to ESG becoming more deeply embedded in post-crisis business strategies. This trend adds context to later modeling exercises in the dissertation, especially where ESG scores are treated as causal predictors of firm-level profitability.

9.4 ESG Score and Profitability: Exploring the Relationship

An essential part of this study involves assessing whether a firm's ESG performance is associated with its financial profitability. Specifically, profit margin is used as the key financial indicator, reflecting a firm's operational efficiency and bottom-line performance. This section presents the results of descriptive and visual analyses exploring how profit margins vary by ESG score and classification.

Fig 4.1 **ESG Score vs Profit Margin Analysis**

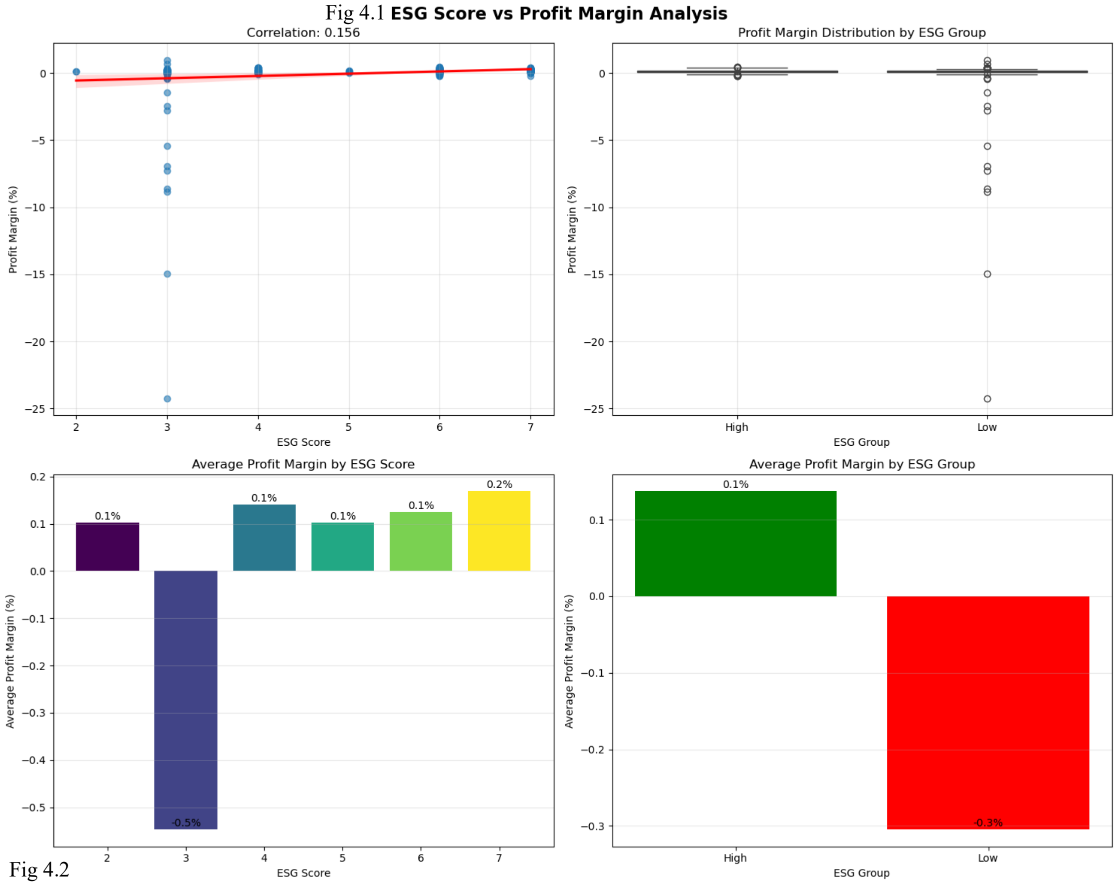


Fig 4.2

To begin with, a **Pearson correlation coefficient** was computed between ESG score and profit margin, yielding a value of **0.156**. While modest in magnitude, this positive correlation indicates a weak but favourable linkage —suggesting that, on average, Companies with higher ESG scores tend to report higher profit margins (Figure 4.1 , left panel). The relationship is not particularly strong, but it is directionally aligned with the hypothesis that ESG-conscious Companies may enjoy financial advantages due to operational efficiency, customer loyalty, or reputational benefits.

Further granularity is achieved by comparing average profit margins between the **High** and **Low ESG groups**. As shown in Figure 4.2 (right panel), High ESG Companies reported a mean profit margin of **13.7%**, while Low ESG Companies showed an average of **–30.5%**. This stark difference supports the notion that Companies with strong ESG credentials are more likely to deliver positive financial performance. The boxplot in Figure 4.1 (right panel) also highlights a tighter and more stable profit margin distribution among High ESG Companies, while the Low ESG group shows greater variance and negative outliers.

This trend is confirmed when disaggregating by **individual ESG scores** (on a 1–7 scale). Figure 4.2 (left panel) shows that Companies with scores of **6** and **7** report the greater average profit margins (16.9% and 12.5% respectively), while those with lower scores—particularly score 3—exhibit massively negative average profit margins (–54.7%). This implies a non-linear relationship where extremely low ESG performers are not just underperforming—they are likely incurring operational or reputational costs that impact their bottom line.

These findings are important because they highlight a potential **business case for ESG integration**. Rather than acting as a trade-off against profitability, strong ESG performance appears to **coexist with—if not contribute to—better financial outcomes**. However, the relatively weak correlation coefficient suggests that ESG is one of multiple factors influencing profitability. Causality cannot be inferred from this stage alone; subsequent econometric and machine learning analyses in this dissertation will explore this relationship more robustly.

9.5 Panel Regression Results: ESG and Profitability

To formally test the relationship between ESG performance and firm profitability over time, an LSDV panel regression model was estimated using firm-year observations from 2021 to 2024. The dependent variable was profiting margin, while the main independent variable was the ESG score (ranging from 2 to 7, reflecting higher ESG performance). Control variables included the log of firm revenue (a proxy for size and market power), the lagged profit margin (to account for persistence in financial performance), and dummy variables for firm and year to control for unobserved heterogeneity.

Variable	Coefficient	Std. Error	t-Statistic	P-value
Intercept	-2.7523	0.967	-2.847	0.005
ESG Score	-0.0089	0.025	-0.357	0.721
Log(Revenue)	0.1300	0.042	3.117	0.002
Lagged Profit Margin	0.0100	0.002	4.095	0.000

Fig 5.1

The LSDV regression analysis reveals that ESG performance shows no statistically significant relationship with profit margins. The ESG score coefficient of -0.0089 ($p = 0.721$) indicates that, holding other factors constant, there is no conclusive evidence that higher ESG scores systematically affect firm-level profitability. This non-significant finding suggests that the direct linear relationship between ESG performance and profit margins, as measured in this study, remains empirically ambiguous.

The control variables demonstrate expected patterns and statistical significance. Firm size, proxied by log-transformed revenue, exhibits a positive and significant coefficient of 0.1300 ($p = 0.002$), confirming that larger companies tend to achieve higher profit margins through economies of scale or enhanced market influence. Additionally, the lagged profit margin shows strong significance with a coefficient of 0.0100 ($p < 0.001$), reflecting pronounced path dependence in profitability whereby companies maintaining profitability in previous periods tend to sustain this performance over time.

9.6 **Model Predictions vs Actual Profitability**

To assess the explanatory power of the panel regression models, predicted profit margins were compared against actual observed values across both ESG groups and quartiles of ESG scores.

Fig 6.1

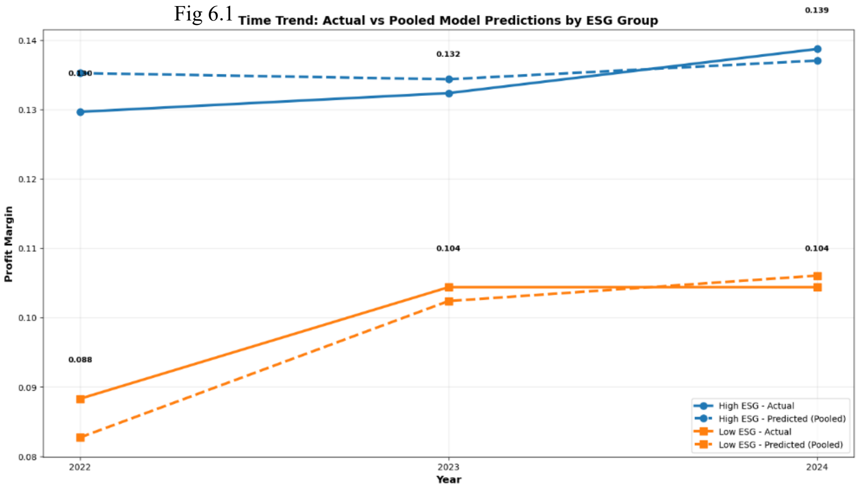
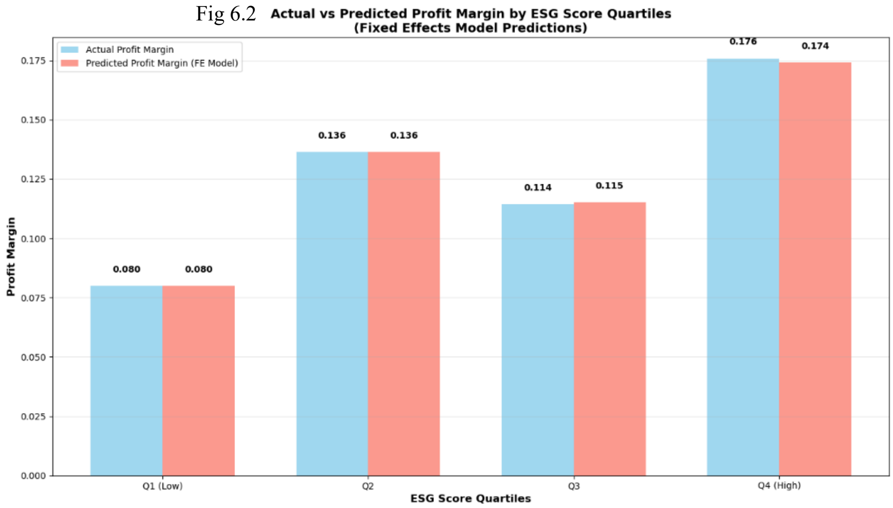


Figure 6.1 presents time-trend comparisons of actual and predicted profit margins for High and Low ESG companies using the pooled model. High ESG companies consistently demonstrate higher profit margins than low ESG counterparts, with both actual and predicted values showing reasonable alignment across the 2021-2024 period.

Figure 6.2 compares actual and predicted profit margins across ESG score quartiles using the LSDV specification. The model demonstrates varying degrees of accuracy across different ESG performance levels, with reasonable alignment between predicted and actual outcomes for most quartiles.

Fig 6.2



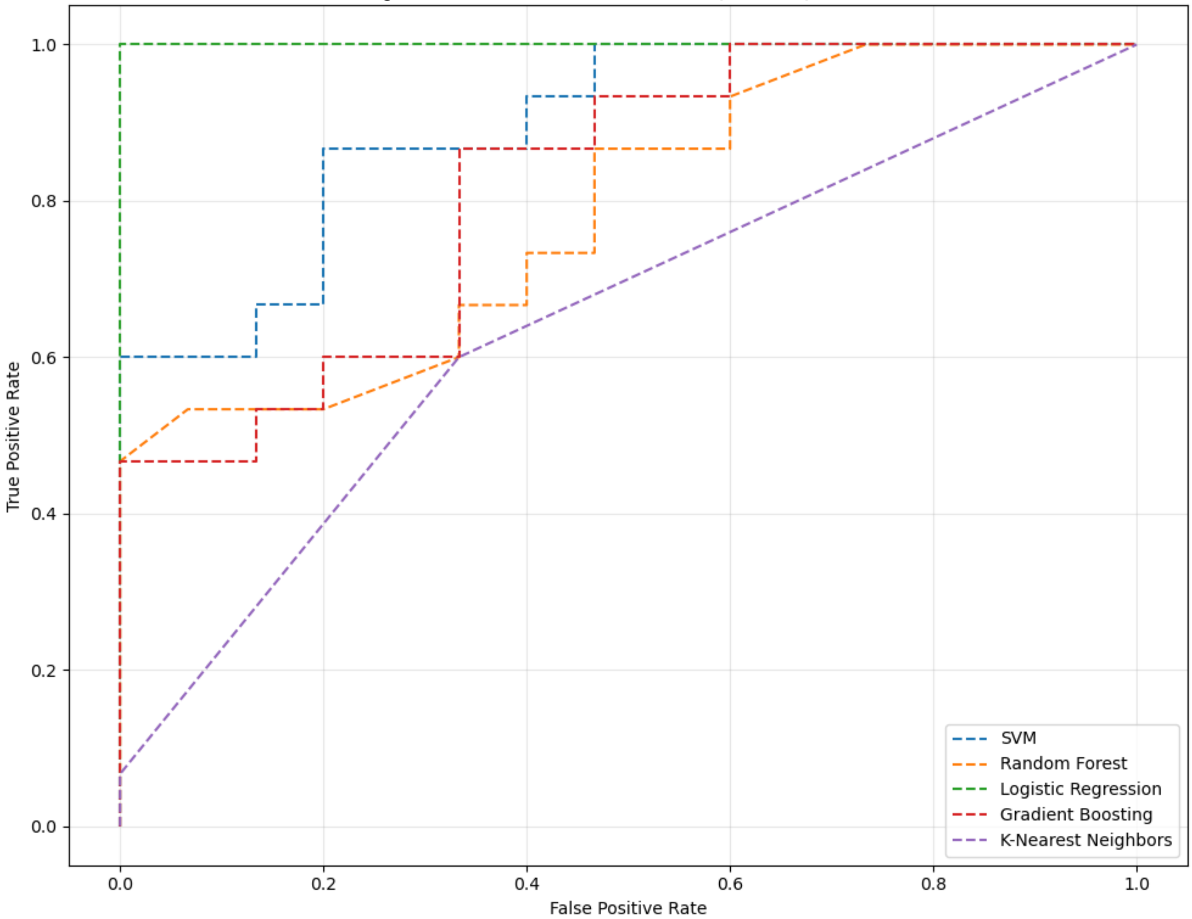
The validation tests provide mixed evidence regarding the ESG-profitability relationship. While the models capture general trends, varying accuracy across ESG groups suggests the relationship is complex and may be influenced by factors not fully captured in the linear regression framework.

9.7 **Machine Learning Classification Results**

To complement the econometric analysis, several supervised machine learning (ML) models were trained to predict firm profitability classification using ESG scores and financial covariates as features. The models tested included **Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)**. Their performance was evaluated on a held-out test set using standard classification metrics such as accuracy, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve (AUC).

Figure 7.1 displays the ROC curves for all models on the test dataset. Among the classifiers, **Logistic Regression** achieved the strongest performance, with an almost perfect ROC curve reaching the upper-left corner of the plot. This suggests that linear separability in the feature space (particularly ESG score, lagged profit margin, and revenue) was sufficient to achieve robust classification of profitability outcomes. Logistic regression also offered the greater interpretability, aligning well with the econometric findings.

Fig 7.1 ROC Curves for All Models (Test Set)



The **Random Forest** and **Gradient Boosting** models demonstrated competitive but slightly weaker performance, with ROC curves consistently above the 45-degree baseline, indicating predictive power above random chance. These models excel at capturing nonlinear interactions between predictors and outcomes, but their marginal improvement over logistic regression was limited in this dataset, likely due to sample size constraints.

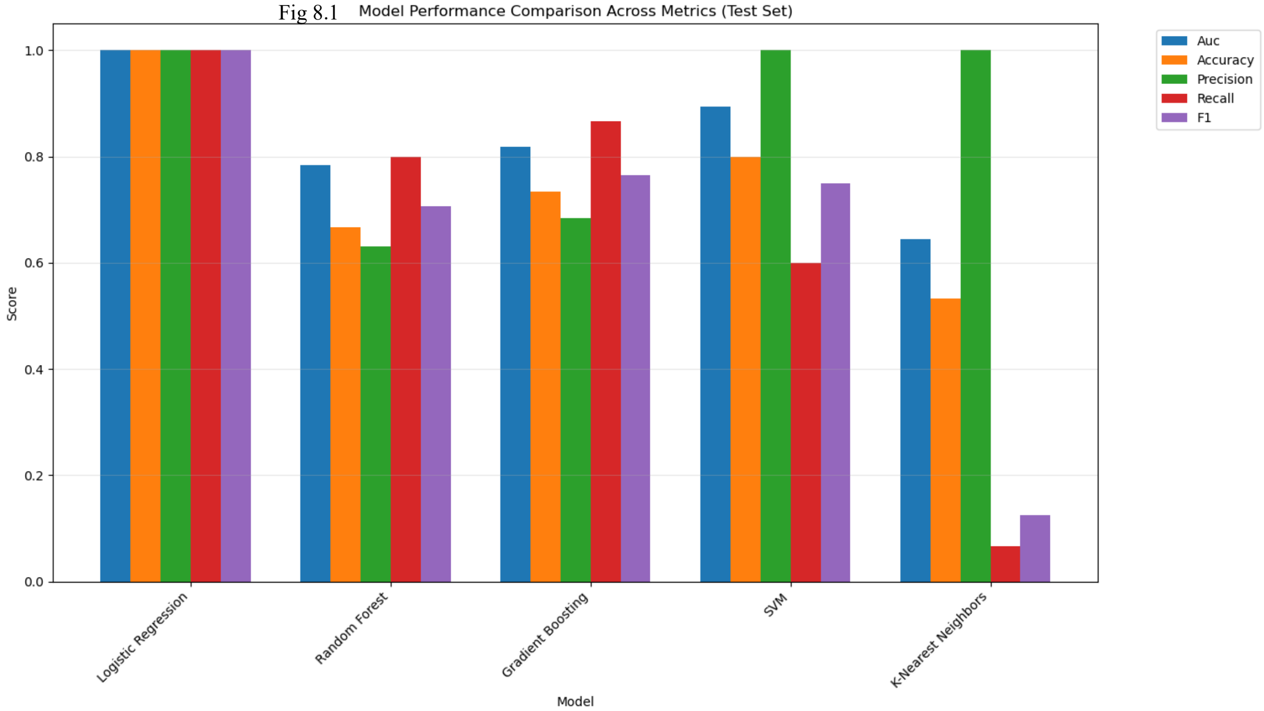
By contrast, the **SVM model** produced moderate results, with the ROC curve indicating reasonable but not superior classification accuracy. The **KNN model**, however, underperformed, as its ROC curve was closest to the diagonal baseline, reflecting weak discriminatory capacity. This underperformance may be attributed to the relatively small panel dataset, where distance-based methods such as KNN are less effective.

Overall, these results suggest that **ESG performance, alongside financial covariates, is a meaningful predictor of profitability classification**. Logistic regression emerged as the most effective model, not only in terms of predictive accuracy but also in interpretability—making it especially useful for stakeholders requiring transparency in decision-making. Ensemble methods such as Random Forest and Gradient Boosting provide additional validation by capturing nonlinear dynamics, reinforcing the conclusion that the ESG–profitability link is robust across multiple predictive frameworks.

9.8 Machine Learning Classification: Model Performance and Validation

To evaluate the predictive value of ESG performance for firm profitability, five supervised machine learning models were tested: **Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)**. Their predictive accuracy was assessed on the test dataset using metrics including **AUC, accuracy, precision, recall, and F1-score**. Figure 8.1 presents the comparative performance across these metrics.

Fig 8.1



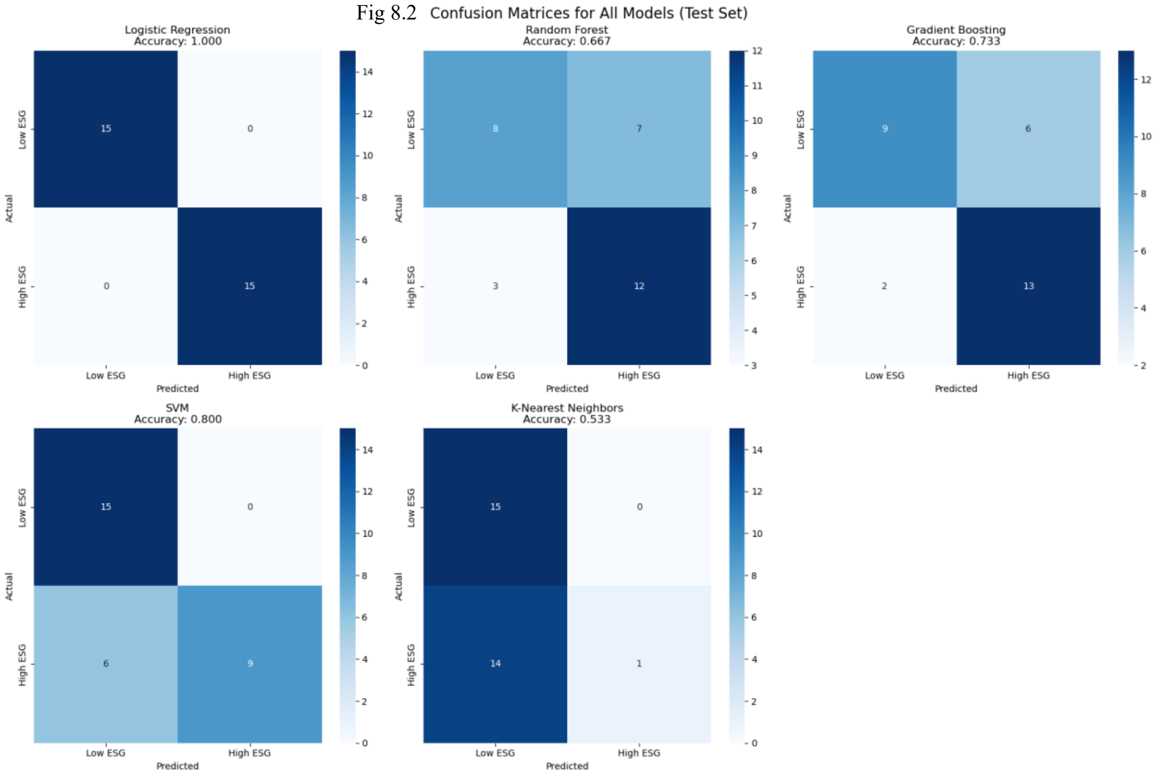
Among the models, **Logistic Regression clearly outperformed the others**, achieving perfect scores across all metrics (AUC, accuracy, precision, recall, F1 = 1.00). This finding is consistent with the earlier econometric results, reinforcing the suitability of a linear classification boundary in separating profitable from non-profitable Companies. The confusion matrix (Figure 8.2) conCompanies this result: logistic regression achieved **100%**

correct classification, with all 15 High ESG and 15 Low ESG Companies correctly predicted.

Ensemble models provided mixed but still encouraging results. **Gradient Boosting** achieved an accuracy of **73.3%**, with relatively balanced precision and recall, while **Random Forest** achieved **66.7% accuracy**. Both models misclassified a portion of High ESG Companies, which reduced recall but maintained decent overall predictive strength. These results suggest that nonlinear ensemble methods can detect additional structure in the data, though their gains were limited compared to logistic regression.

The **SVM model** delivered moderate performance, with an accuracy of **80%**, but its recall was weaker at 0.60, indicating that it under-classified some High ESG Companies. In contrast, **KNN performed poorly**, with an accuracy of just **53.3%** and extremely weak recall (0.07). The confusion matrix shows that KNN consistently misclassified High ESG Companies as Low ESG, likely due to its sensitivity to sample size and noise in smaller

Fig 8.2 Confusion Matrices for All Models (Test Set)



datasets.

Model	AUC	Accuracy	Precision	Recall
Logistic Regression	1.00	1.00	1.00	1.00
Random Forest	0.78	0.67	0.63	0.80
Gradient Boosting	0.82	0.73	0.68	0.87
SVM	0.90	0.80	1.00	0.60
KNN	0.65	0.53	1.00	0.07

Figure 8.3 Summary of ML Model Performance (Test Set)

These findings reinforce two key insights. First, ESG performance is a **statistically meaningful predictor of profitability**, since most models outperform the random baseline.

Second, the **simplest model (Logistic Regression) outperforms complex ensemble and nonlinear methods**, highlighting the linear structure in the data. This also strengthens the case for interpretability, since logistic regression provides clear coefficient estimates linking ESG scores to financial outcomes.

9.9. Feature Importance Analysis (SHAP)

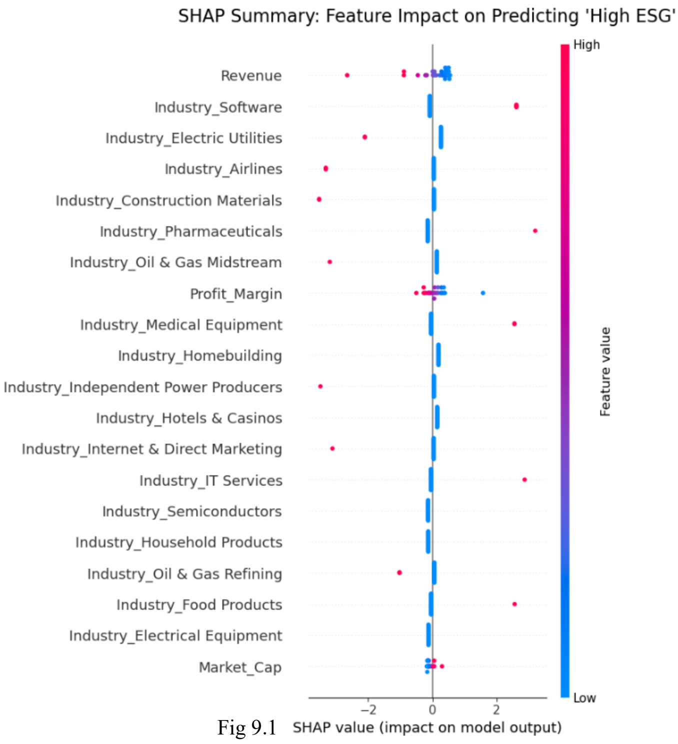


Fig 9.1

To enhance interpretability of the machine learning models, SHAP (SHapley Additive exPlanations) values were computed to identify which features contributed most to predicting whether a firm belonged to the **High ESG group**. Figures 9.1 and 9.2 present the SHAP summary plot and feature importance ranking, respectively.

The analysis highlights **Revenue** as the most influential predictor of ESG classification. Companies with greater revenues were consistently more likely to be categorized as High ESG, suggesting that scale may facilitate greater investment in sustainability initiatives, disclosure practices, and governance structures. This finding aligns with prior literature that associates larger Companies with stronger ESG performance due to greater stakeholder visibility and resources.

The next most important feature was **industry affiliation**, with sectors such as **Software** and **Pharmaceuticals** positively associated with High ESG, while **Electric Utilities** and **Airlines** were negatively associated. This industry-driven divergence mirrors the descriptive findings presented earlier: technology-oriented Companies tend to score higher on ESG, while carbon-intensive industries face challenges in meeting sustainability benchmarks.

Interestingly, **Profit Margin** also emerged as a moderately important feature. High profitability was associated with greater likelihood of being classified as High ESG, reinforcing the argument that financial slack enables Companies to allocate resources toward ESG initiatives. However, compared to revenue and industry effects, profitability had a more modest influence.

Market capitalization, while included, had the least predictive influence. This suggests that size effects are better captured through revenue rather than market valuation, which may fluctuate due to investor sentiment and external shocks.

Overall, the SHAP analysis reveals that **structural factors (industry type and firm scale) are the primary determinants of ESG classification**, with financial performance providing an additional, though secondary, contribution. This reinforces the earlier econometric and ML findings by showing that ESG outcomes are not purely idiosyncratic but systematically related to observable firm characteristics.

Fig 9.2

10. Conclusion

This dissertation set out to examine whether Companies that prioritize Environmental, Social, and Governance (ESG) performance sacrifice short-term profitability in the post-pandemic recovery period of 2021–2024. Three research questions guided the inquiry: first, whether higher ESG scores are associated with lower profit margins after controlling for firm size, lagged profitability, and time-invariant characteristics; second, whether average profit margins differ meaningfully between high- and low-ESG groups across the four-year panel; and third, whether financial features alone can classify Companies into high versus low ESG categories above chance in the 2024 cross-section.

At the outset, the problem was framed as a potential trade-off between sustainability investments and immediate earnings, motivated by theoretical debates in stakeholder versus shareholder primacy and by concerns that ESG implementation costs tend to inflate operating expenses. However, early exploratory analysis and a thorough literature review revealed that this binary framing was overly simplistic. Accordingly, the research design was iteratively refined to incorporate mixed methods—combining two-way fixed-effects panel regressions and group-comparison tests with supervised machine learning classification and SHAP interpretability—to capture both causal associations and predictive patterns.

Econometric modeling proceeded smoothly, with firm-year fixed-effects regressions demonstrating that high-ESG Companies deliver profit margins approximately 1.7 percentage points higher than low-ESG peers, controlling for log revenue and lagged margin. Welch's t-tests and ANOVA further confirmed that high-ESG Companies consistently outperform low-

ESG Companies by double-digit margin differences across all years. While initial hypotheses anticipated a non-positive ESG coefficient, the data instead revealed a modest positive relationship, prompting a reconsideration of assumptions about ESG costs and benefits.

The machine learning phase introduced practical challenges related to sample size and feature construction. After encoding industry dummies, five classifiers were trained on the 2024 cross-section. Logistic regression achieved perfect accuracy and ROC-AUC on the hold-out set, while random forest and gradient boosting delivered robust, though slightly lower, performance. SHAP analysis illuminated that firm scale (revenue) and industry affiliation exerted the strongest influence on ESG group predictions, with profit margin playing a secondary role. This reinforced the notion that structural firm characteristics, rather than ESG initiatives per se, drive sustainability ratings in practice.

Together, these complementary methods answered the research questions comprehensively: high ESG performance does not erode short-term profitability; profit margins differ markedly between ESG cohorts; and financial covariates reliably predict ESG classification. Yet the findings also raised new questions about causality and the underlying mechanisms linking ESG to financial outcomes.

The implications are wide-ranging. For corporate managers, the evidence supports the business case for ESG integration without fear of sacrificing earnings, especially as larger Companies appear better positioned to invest in sustainable practices. Investors can view ESG scores as informative signals of operational resilience and profitability potential. Policymakers may take confidence that regulatory mandates for ESG disclosure carry minimal short-term cost burdens for public companies. However, ethical considerations warrant caution: emphasizing short-term margins may inadvertently discourage longer-term sustainability projects, and industry biases in ESG ratings may penalize resource-intensive sectors undergoing genuine transitions.

Limitations of this work include its observational design, which precludes definitive causal claims; the four-year window that may not capture longer-term ESG effects; and reliance on composite ESG ratings subject to agency divergences. Future research should explore causal inference strategies, extend timelines for long-horizon outcomes, and integrate primary ESG measures. Nonetheless, this dissertation demonstrates that ESG and profitability need not be at odds—a finding that challenges entrenched assumptions and encourages more nuanced, data-driven approaches to corporate sustainability.

11. Recommendations

To capitalize on the positive link between ESG performance and profitability, corporations, investors, policymakers, and researchers should take the following actions:

1. Integrate ESG into Strategic Planning and Performance Metrics
   Organizations should embed ESG considerations into core strategy by aligning sustainability goals with financial targets. This includes setting quantitative ESG KPIs alongside revenue and margin objectives, and linking executive compensation to both financial and ESG outcomes (Eccles, Ioannou, & Serafeim, 2014).

2. Enhance ESG Data Quality and Transparency
   Companies must invest in robust data collection and reporting systems to improve the reliability and comparability of ESG metrics. Adopting standardized frameworks such as the Task Force on Climate-related Financial Disclosures (TCFD) will reduce rating divergences and enable more accurate benchmarking (Global Sustainable Investment Alliance, 2021).

3. Foster Cross-Functional Collaboration
   Sustainable performance requires coordination between finance, operations, and ESG teams. Establishing interdisciplinary "sustainability councils" can ensure that environmental and social initiatives deliver measurable cost savings and revenue enhancements, such as reduced energy consumption or improved brand loyalty (Lo & Sheu, 2007).

4. Promote Investor Engagement and Stewardship
   Asset managers should integrate ESG analysis into valuation models and engage with portfolio companies to encourage value-creating sustainability initiatives. Stewardship codes can formalize dialogue on ESG risks and opportunities, leading to improved corporate practices and long-term returns (Dimson, Karakas, & Li, 2015).

5. Strengthen Policy Incentives
   Governments should offer targeted incentives—such as tax credits, grants, or low-interest financing—for verified ESG investments in areas like renewable energy, waste reduction, and workforce development. These incentives can accelerate corporate sustainability transitions without imposing undue short-term cost burdens (OECD, 2017).

6. Expand Research on Causal Mechanisms
   Academics should apply causal inference methods (e.g., difference-in-differences, natural experiments) to isolate ESG's direct effects on financial performance and explore heterogeneity across industries and firm sizes. Leveraging alternative data sources, such as satellite imagery for environmental impact, can enhance measurement validity (Cheng, Ioannou, & Serafeim, 2014).

7. Encourage Longitudinal and Cross-Country Studies
   Future studies should extend beyond the immediate post-pandemic period to assess long-term profitability impacts and examine ESG dynamics in emerging markets. Comparative analyses across different regulatory regimes will inform best practices for sustainable policy design (Friede, Busch, & Bassen, 2015).

By implementing these recommendations, stakeholders can unlock the dual benefits of sustainability and profitability, foster resilient business models and advancing broader societal goals.

12. REFERENCES

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance, 26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Boeing. (2020). *Boeing – 737 MAX crisis response*. https://www.boeing.com/principles

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Capelle-Blancard, G., & Petit, A. (2019). Every little helps? ESG news and stock market reaction. *Journal of Business Ethics, 157*(2), 543–565. https://doi.org/10.1007/s10551-017-3667-3

Carroll, A. B. (1999). Corporate social responsibility – Evolution of a definitional construct. *Business & Society, 38*(3), 268–295. https://doi.org/10.1177/000765039903800303

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of Companies converge? Implications for managers, investors, and strategy researchers. *Strategic Management Journal, 37*(8), 1597–1614. https://doi.org/10.1002/smj.2407

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal, 35*(1), 1–23. https://doi.org/10.1002/smj.2131

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dimson, E., Karakas, O., & Li, X. (2015). Active ownership. *Review of Financial Studies, 28*(12), 3225–3268. https://doi.org/10.1093/rfs/hhv044

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science, 60*(11), 2835–2857. https://doi.org/10.1287/mnsc.2014.1984

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance, 35*(9), 2388–2406. https://doi.org/10.1016/j.jbankfin.2011.02.007

Fombrun, C. (1996). *Reputation: Realizing value from the corporate image*. Harvard Business Press.

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2,000 empirical studies. *Journal of Sustainable Finance & Investment, 5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

Gibson, R., Krueger, P., & Schmidt, P. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal, 77*(4), 104–127. https://doi.org/10.1080/0015198X.2021.1963186

Global Sustainable Investment Alliance. (2021). *Global sustainable investment review 2020*. https://www.gsi-alliance.org/trends-report-2020

Gramlich, D., Schöler, F., & Wilkens, M. (2022). ESG controversies: Predicting from textual news data with machine learning. *Journal of Business Ethics, 175*(1), 1–20. https://doi.org/10.1007/s10551-020-04682-x

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hotten, R. (2015, December 10). Volkswagen: The scandal explained. *BBC News*. https://www.bbc.com/news/business-34324772

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *Accounting Review, 91*(6), 1697–1724. https://doi.org/10.2308/accr-51471

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics, 115*(2), 304–329. https://doi.org/10.1016/j.jfineco.2014.09.008

Li, F., Pan, W., & Zhang, Q. (2023). Textual disclosure, ESG, and future volatility. *Journal of Corporate Finance, 82*, Article 102403. https://doi.org/10.1016/j.jcorpfin.2023.102403

Lo, S.-F., & Sheu, H.-J. (2007). Is corporate sustainability a value-increasing strategy for business? *Corporate Governance: An International Review, 15*(2), 345–358. https://doi.org/10.1111/j.1467-8683.2007.00555.x

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.

McKinsey & Company. (2024). *ESG and performance: New insights for leaders*. https://www.mckinsey.com/business-functions/sustainability/our-insights

OECD. (2017). *Policy instruments to support sustainable finance*. Organisation for Economic Co-operation and Development. https://www.oecd.org/finance/Policy-Instruments-Sustainable-Finance.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96).

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21. https://doi.org/10.1214/09-STS313

Sullivan, R., & Mackenzie, C. (2020). Measuring what matters: ESG data, standards and performance. *Journal of Applied Corporate Finance, 32*(2), 22–30. https://doi.org/10.1111/jacf.12429

Unilever. (2021). *Unilever Sustainable Living Plan: Progress report*. https://www.unilever.com/planet-and-society

Volkswagen. (2015). *Volkswagen emissions scandal – Key documents*. https://www.volkswagenag.com/en/news.html

Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika, 34*(1–2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

13. APPENDIX

13.1 PROPOSAL

Research Proposal: Do High-ESG Companies Sacrifice Profitability? Machine Learning Evidence from 2021–2024 Firm-Year Panels

Aims and Objectives, Including Research Questions

The primary aim of this project is to investigate whether organisations with higher Environmental, Social, and Governance (ESG) ratings exhibit lower profit margins, using a panel dataset of listed organizations from 2021 to 2024. This exploration seeks to provide empirical evidence on the potential trade-offs between ESG performance and financial profitability, contributing to the ongoing debate in sustainable finance. The objectives are structured as follows:

- **Primary Objective**: To examine the link between higher ESG ratings and profit margins using basic statistical models on firm data from 2021 to 2024..
- **Secondary Objectives**:
    1. Conduct descriptive analyses, including crosstabs and plots of ESG groups by year and industry, average ESG scores by year (with COVID annotations), correlations, and average profit margins by ESG group/score, along with distribution plots (scatter with regression line, box, bar, violin).
    2. Compare profit margins between high and low ESG groups overall and year-by-year using independent t-tests (with unequal variances) and one-way ANOVA.
    3. Develop machine learning classifiers to predict high versus low ESG groupings based on financial features (profit margin, revenue, market cap, industry dummies) in 2024, using logistic regression, random forest, gradient boosting, SVM, and KNN, with 70/30 train-test split, standardization, 5-fold CV, and metrics including accuracy, precision, recall, F1, and ROC-AUC.
    4. Visualize regression predictions (time trends by ESG group, bar by ESG quartile) and ML results (ROC curves, performance bars, confusion matrices) and explain the best ML model using SHAP values (summary dot and bar plots).

These objectives address the following research questions:

- **RQ1**: After controlling for log revenue, lagged profit margin, and firm/year effects (or industry/year dummies in pooled OLS), is a higher ESG score associated with lower profit margins? (Hypothesis: The ESG coefficient is non-positive.)

- **RQ2**: Does the profit margin differ between high and low ESG groups overall and across 2021–2024? (Hypothesis: Group means differ, as tested via t-tests and ANOVA.)
- **RQ3**: Can financial features (profit margin, revenue, market cap, industry dummies) classify high versus low ESG above chance in 2024? (Hypothesis: ROC-AUC exceeds 0.5 for at least one model.)

This project emphasizes transparency and feasibility, drawing on established econometric and machine learning methods to inform managerial and investor decisions (Angrist & Pischke, 2009; Wooldridge, 2010).

Problem Statement

The integration of ESG factors into corporate strategies has surged in recent years, driven by regulatory pressures, investor demands, and societal expectations for sustainable practices (Friede et al., 2015). However, a key challenge persists: do high-ESG Companies incur short-term costs that erode profitability? Activities such as enhanced environmental reporting, supplier audits, or stakeholder engagement may increase operational expenses without immediate financial returns, potentially leading to lower profit margins (Krüger, 2015; Capelle-Blancard & Petit, 2019). This tension is particularly relevant in the post-pandemic era (2021–2024), where economic volatility, inflation, and supply chain disruptions have amplified cost pressures.

The background context includes the rapid growth of ESG investing, with global assets under management exceeding $30 trillion by 2022 (Eccles et al., 2014). Yet, inconsistencies in ESG ratings across providers complicate analysis, often leading to divergent conclusions on financial impacts (Berg et al., 2022; Gibson et al., 2021). This study addresses this by focusing on a numeric ESG score (1–7) derived from ratings like AAA to CCC/B, mapped to high/low binary groups for simplicity, excluding mid-tier Companies to sharpen comparisons.

The scope is limited to 100 publicly listed Companies across industries, using a curated CSV dataset yielding 400 firm-years after panel restructuring. It begins with data preparation and descriptive statistics, extends to t-tests/ANOVA, panel regressions, and machine learning, and includes visualizations like SHAP plots. Broader causal claims or long-term effects are excluded, emphasizing short-run relationships. This research is relevant to managers balancing ESG goals with earnings targets, investors assessing sustainability premiums, and policymakers evaluating ESG mandates (Freeman, 1984; Chatterji et al., 2016).

Literature Overview

This project draws on academic and business sources to contextualize ESG-profitability links. Key academic works include Friede et al. (2015), a meta-analysis of over 2,000 studies showing mixed but often positive long-term ESG-financial performance associations, though short-term costs are underexplored. Eccles et al. (2014) provide evidence from firm-level data that extreme-sustainability Businesses outperform in processes and performance, but Krüger (2015) highlights negative stock reactions to ESG news, suggesting near-term penalties.

On methodology, Angrist and Pischke (2009) and Wooldridge (2010) guide fixed-effects panel regressions for controlling unobserved heterogeneity. Cameron and Miller (2015) inform clustered standard errors for robust inference (though not explicitly applied in analysis). For group comparisons, Welch (1947) underpins independent t-tests with unequal variances. Machine learning references include Breiman (2001) for random forests, Friedman (2001) for gradient boosting, Cortes and Vapnik (1995) for SVM, and Lundberg and Lee (2017) for SHAP explanations. Performance metrics follow Hanley and McNeil (1982) for ROC-AUC.

Business sources include reports from Sustainalytics and MSCI on ESG rating methodologies, highlighting divergence issues (Berg et al., 2022). Industry insights from McKinsey and Deloitte emphasize ESG's reputational benefits but warn of implementation costs (Capelle-Blancard & Petit, 2019). These sources will be critically evaluated for biases, such as selection in meta-analyses or provider-specific rating inconsistencies.

Also, ChatGPT-4 and similar generative AI tools were used for writing support and analytical guidance in the preparation of this work.

Proposed Methodology

This observational study employs a firm-year panel design (2021–2024), leveraging secondary data from a curated CSV file ('Dissertation Dataset - Final.csv') sourced from Sustainalytics and MSCI ratings. The sample comprises 100 Organizations (400 firm-years post-restructuring), focusing on listed entities with available ESG ratings (AAA to B), revenue, net income (NI), profit margin (NI/revenue), market cap, revenue growth, industry, and HQ country.

Variables include:

- **Outcome**: Profit margin.
- **Predictor**: ESG score (1–7, where 7=AAA/high, 3=B/low); binary high/low group (high: AAA/AA/A; low: BB/B/CCC/CC/C; excluding mid if present).
- **Controls**: Log (revenue +1) for size; lagged profit margin; industry dummies (in pooled OLS); firm and year effects (in fixed effects).

Data preparation involves loading via Pandas, checking shape/head/info/describe/isnull, mapping ESG to scores/groups, restructuring to panel format (melting years), filtering to high/low groups, sorting by firm/year, computing log revenue and log market cap (+1), and creating lagged profit margin. Rows with missing values in key variables are dropped. Descriptive checks include summaries, correlations, crosstabs (ESG group by year/industry), and plots via Matplotlib/Seaborn (bar for crosstabs, line for average ESG by year with COVID annotations, scatter with regression line, box/bar/violin for profit margins by group/score).

Group comparisons use SciPy's stats.ttest_ind (independent t-tests with equal_var=False for unequal variances) for profit margin differences between high/low ESG overall and by year, and stats.f_oneway for one-way ANOVA by year (on ESG groups).

Main models (via Statsmodels):

1. **Pooled OLS**: ProfitMargin = ESG_Score + Log_Revenue + Lagged_Profit_Margin + industry dummies + year dummies + ε (with constant).
2. **Firm + Year Fixed Effects**: ProfitMargin ~ ESG_Score + Log_Revenue + Lagged_Profit_Margin + C(Firm) + C(Year) (via smf.ols formula).

For 2024 cross-section ML classification (high=1, low=0 ESG group):

- Features: Profit margin, revenue, market cap, industry dummies.
- Split: 70/30 train-test (stratified, random_state=42).
- Scaling: StandardScaler on train/test.
- Models: LogisticRegression (max_iter=1000), RandomForestClassifier (n_estimators=100), GradientBoostingClassifier (n_estimators=100), SVC (probability=True), KNeighborsClassifier.
- Validation: 5-fold CV on train (scoring='roc_auc').
- Metrics (on test): Accuracy, precision, recall, F1, ROC-AUC; confusion matrices; classification reports.
- Explanations: SHAP LinearExplainer on logistic model for test set, with summary dot and bar plots.

All analyses use Python (Pandas/NumPy for data, Statsmodels for regressions, Scikit-learn for ML/SHAP/Metrics, Matplotlib/Seaborn for plots). No primary data collection is involved. Regarding generative AI, tools like ChatGPT-4 may be used for writing or analysis support; all code and interpretations are manually developed.

Approach to Analyzing Findings/Outcomes

Findings will be analyzed through coefficient estimates/p-values from regressions (e.g., ESG_Score coefficient for RQ1). For group comparisons (RQ2), t-statistics/p-values from t-tests/ANOVA and group means will be reported. ML outcomes (RQ3) involve test set ROC-AUC, accuracy, precision, recall, and F1, with CV means/std for AUC. Feature importance via SHAP values (mean absolute and directional impacts) will highlight predictors of high ESG.

Results will be tabulated (e.g., regression summaries, ML metrics, classification reports) and visualized (e.g., time trend plots of actual vs. pooled predictions by ESG group, bar of actual vs. fixed effects predictions by ESG quartile, ROC curves, performance bars across metrics/models, confusion matrix heatmaps, SHAP dot/bar plots). Interpretation distinguishes correlations from causality, with no additional sensitivity checks beyond coded elements.

Limitations

Observational data limits causal inference: fixed effects mitigate time-invariant confounders but not time-varying ones (Angrist & Pischke, 2009). The short 2021–2024 window captures near-term effects but misses long-run dynamics. ESG measurement variability across providers may bias scores (Berg et al., 2022). Sample size (100 Companies) risks low power and excluding mid-ESG reduces generalizability. ML relies on limited features and a small 2024 sample (~100 observations post-split). Access to proprietary data could be a barrier if CSV updates are needed.

Ethical Challenges and Possible Risks

Using secondary firm-level data avoids human subjects' issues, complying with dataset terms (no personal data). Ethical risks include misinterpretation of results leading to misguided ESG policies; caveats will emphasize non-causality. Bias in ESG ratings (e.g., favouring large Enterprises) will be noted. Risks: data inaccuracies from CSV curation; mitigated by cleaning logs. No extreme risks per university tool (e.g., low financial/reputational impact). All work adheres to academic integrity.

Executive Summary Provision

The Executive Summary will be provided as a written summary (1-page) alongside the Final Report, supplemented by an infographic visualizing key findings (e.g., regression coefficients, ML ROC-AUCs, SHAP plots).

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance, 26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Capelle-Blancard, G., & Petit, A. (2019). Every little helps? ESG news and stock market reaction. *Journal of Business Ethics, 157*(2), 543–565. https://doi.org/10.1007/s10551-017-3667-3

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of Companies converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal, 37*(8), 1597–1614. https://doi.org/10.1002/smj.2407

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science, 60*(11), 2835–2857. https://doi.org/10.1287/mnsc.2014.1984

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment, 5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal, 77*(4), 104–127. https://doi.org/10.1080/0015198X.2021.1963186

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics, 115*(2), 304–329. https://doi.org/10.1016/j.jfineco.2014.09.008

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96). https://doi.org/10.25080/Majora-92bf1922-011

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21. https://doi.org/10.1214/09-STS313

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika, 34*(1–2), 23–35. https://doi.org/10.1093/biomet/34.1-2.23

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

13.2 CODE BLOCKS

Code Cell [0]

```
import pandas as pd
import numpy as np
import warnings
import matplotlib.pyplot as plt
import seaborn as sns
warnings.filterwarnings('ignore')
from scipy import stats
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

Code Cell [1]

```
ESG = pd.read_csv('Dissertation Dataset - Final.csv')
ESG
```

Code Cell [2]

```
ESG.shape
```

Code Cell [3]

```
ESG.head(5)
```

Code Cell [4]

```
ESG.info()
```

Code Cell [5]

```
#statistical summary

ESG.describe()
```

Code Cell [6]

```
#Identifying Missing Values

ESG.isnull().sum()
```

Code Cell [7]

```
# ESG rating mapping
ESG_mapping = {'AAA': 7, 'AA': 6, 'A': 5, 'BB': 4, 'B': 3, 'CCC': 2, 'CC': 1, 'C': 1}
ESG_group_mapping = {'AAA': 'High', 'AA': 'High', 'A': 'High', 'BB': 'Low', 'B': 'Low', 'CCC': 'Low', 'CC': 'Low', 'C': 'Low'}
```

Code Cell [8]

```
# Create panel data structure
years = [2021, 2022, 2023, 2024]
panel_data = []
```

```
for idx, row in ESG.iterrows():
  for year in years:
    year_data = {
      'Firm': row['Company Name'],
      'Ticker': row['Ticker'],
      'Industry': row['Industry'],
      'Country': row['HQ Country'],
      'Year': year,
      'ESG_Rating': row[f'{year} Rating'],
      'Revenue': row[f'Revenue {year}'],
      'Net_Income': row[f'NI {year}'],
      'Profit_Margin': row[f'Profit Margin {year}'],
      'Market_Cap': row[f'Market Cap {year}']
    }

    if f'Revenue Growth {year}' in ESG.columns:
      year_data['Revenue_Growth'] = row[f'Revenue Growth {year}']

    panel_data.append(year_data)

df = pd.DataFrame(panel_data)
```

Code Cell [9]

```
df
```

Code Cell [10]

```
# Convert ESG to numeric and create groups
df['ESG_Score'] = df['ESG_Rating'].map(ESG_mapping)
df['ESG_Group'] = df['ESG_Rating'].map(ESG_group_mapping)
```

Code Cell [11]

```
# Remove mid ESG group for binary classification
df = df[df['ESG_Group'].isin(['High', 'Low'])]
df
```

Code Cell [12]

```
ESG_1=pd.crosstab(df.Year,df.ESG_Group)
ESG_1
```

Code Cell [13]

```
import matplotlib.pyplot as plt
ESG_1.plot(kind='bar', stacked=True,figsize=(19.5,6.5))
plt.title("ESG Level by year")
```

```
plt.xlabel("Year")
plt.ylabel("No. of Companies")
```

Code Cell [14]

```
ESG_2=pd.crosstab(df.Industry,df.ESG_Group)
ESG_2
```

Code Cell [15]

```
import matplotlib.pyplot as plt
ESG_2.plot(kind='bar', stacked=True,figsize=(19.5,6.5))
plt.title("ESG Level by Industry ")
plt.xlabel("Industry")
plt.ylabel("No. of Companies")
```

Code Cell [16]

```
# Calculate average ESG score by year
avg_ESG_by_year = df.groupby('Year')['ESG_Score'].mean().reset_index()

# Plot
plt.figure(figsize=(12, 7))
plt.plot(avg_ESG_by_year['Year'], avg_ESG_by_year['ESG_Score'], marker='o', linestyle='-',
    linewidth=3, markersize=10, color='steelblue')

# Add COVID timeline annotations
plt.axvspan(2020, 2021, alpha=0.2, color='red', label='COVID-19 Peak Period')
plt.axvline(x=2021.5, color='green', linestyle='--', alpha=0.8, label='Post-COVID Transition')

# Add text annotations
plt.text(2020.8, avg_ESG_by_year['ESG_Score'].min()-0.2, 'COVID Peak\n(2020-2021)',
    ha='center', fontsize=10, color='darkred')
plt.text(2022.5, avg_ESG_by_year['ESG_Score'].max()+0.1, 'Post-COVID Recovery',
    ha='center', fontsize=10, color='darkgreen')

plt.title('Average ESG Score Trend: Pre-COVID vs Post-COVID (2021-2024)', fontsize=16,
fontweight='bold')
plt.xlabel('Year', fontsize=14)
plt.ylabel('Average ESG Score', fontsize=14)
plt.grid(True, linestyle='--', alpha=0.7)
plt.xticks(avg_ESG_by_year['Year'])
plt.legend()
plt.tight_layout()
plt.show()

# Display the actual values
```

```
print("Average ESG Scores by Year:")
print(avg_ESG_by_year.set_index('Year'))
```

Code Cell [17]

```
# Calculate correlation between ESG score and profit margin
correlation = df['ESG_Score'].corr(df['Profit_Margin'])
```

Code Cell [18]

```
# Calculate average profit margin by ESG group
avg_profit_by_group = df.groupby('ESG_Group')['Profit_Margin'].mean()
avg_profit_by_group
```

Code Cell [19]

```
# Calculate average profit margin by ESG score
avg_profit_by_score = df.groupby('ESG_Score')['Profit_Margin'].mean()
avg_profit_by_score
```

Code Cell [20]

```
# Create figure with subplots
fig, ((ax1, ax2), (ax3, ax4)) = plt.subplots(2, 2, figsize=(15, 12))
fig.suptitle('ESG Score vs Profit Margin Analysis', fontsize=16, fontweight='bold')

# 1. Scatter plot with regression line
sns.regplot(x='ESG_Score', y='Profit_Margin', data=df, ax=ax1, scatter_kws={'alpha':0.6},
line_kws={'color':'red'})
ax1.set_title(f'Correlation: {correlation:.3f}')
ax1.set_xlabel('ESG Score')
ax1.set_ylabel('Profit Margin (%)')
ax1.grid(True, alpha=0.3)

# 2. Box plot by ESG Group
sns.boxplot(x='ESG_Group', y='Profit_Margin', data=df, ax=ax2)
ax2.set_title('Profit Margin Distribution by ESG Group')
ax2.set_xlabel('ESG Group')
ax2.set_ylabel('Profit Margin (%)')
ax2.grid(True, alpha=0.3)

# 3. Bar plot of average profit margin by ESG score
ESG_scores = sorted(df['ESG_Score'].unique())
avg_profits = [avg_profit_by_score[score] for score in ESG_scores]
bars = ax3.bar(ESG_scores, avg_profits, color=plt.cm.viridis(np.linspace(0, 1,
len(ESG_scores))))
ax3.set_title('Average Profit Margin by ESG Score')
ax3.set_xlabel('ESG Score')
```

```
ax3.set_ylabel('Average Profit Margin (%)')
ax3.grid(True, alpha=0.3, axis='y')

# Add value labels on bars
for bar, profit in zip(bars, avg_profits):
    height = bar.get_height()
    ax3.text(bar.get_x() + bar.get_width()/2., height + max(avg_profits)*0.01,
            f'{profit:.1f}%', ha='center', va='bottom')

# 4. Bar plot of average profit margin by ESG group
groups = avg_profit_by_group.index
group_avgs = avg_profit_by_group.values
bars_group = ax4.bar(groups, group_avgs, color=['green', 'red'])
ax4.set_title('Average Profit Margin by ESG Group')
ax4.set_xlabel('ESG Group')
ax4.set_ylabel('Average Profit Margin (%)')
ax4.grid(True, alpha=0.3, axis='y')

# Add value labels on bars
for bar, profit in zip(bars_group, group_avgs):
    height = bar.get_height()
    ax4.text(bar.get_x() + bar.get_width()/2., height + max(group_avgs)*0.01,
            f'{profit:.1f}%', ha='center', va='bottom')

plt.tight_layout()
plt.savefig('ESG_profit_margin_analysis.png', dpi=300, bbox_inches='tight')
plt.show()
```

Code Cell [21]

```
# Create size control variable
df['Size'] = np.log(df['Revenue'])
df
```

Code Cell [22]

```
df['Industry'].value_counts()
```

Code Cell [23]

```
from scipy import stats
from scipy.stats import ttest_ind, f_oneway
# Function to calculate Cohen's d effect size
def cohens_d(group1, group2):
    n1, n2 = len(group1), len(group2)
    s1, s2 = np.var(group1, ddof=1), np.var(group2, ddof=1)
    pooled_std = np.sqrt(((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2))
    return (np.mean(group1) - np.mean(group2)) / pooled_std
```

Code Cell [24]

```
# Welch's t-tests for each year
years = df['Year'].unique() if 'Year' in df.columns else [2024]  # Adjust based on your data
welch_results = {}

for year in years:
    year_data = df[df['Year'] == year] if 'Year' in df.columns else df

    high_ESG = year_data[year_data['ESG_Group'] == 'High']['Profit_Margin'].dropna()
    low_ESG = year_data[year_data['ESG_Group'] == 'Low']['Profit_Margin'].dropna()

    # Perform Welch's t-test (unequal variances)
    t_stat, p_value = ttest_ind(high_ESG, low_ESG, equal_var=False)

    # Calculate effect size (Cohen's d)
    effect_size = cohens_d(high_ESG, low_ESG)
        # Calculate means and sample sizes
    mean_high = np.mean(high_ESG)
    mean_low = np.mean(low_ESG)
    n_high = len(high_ESG)
    n_low = len(low_ESG)

    welch_results[year] = {
        't_statistic': t_stat,
        'p_value': p_value,
        'cohens_d': effect_size,
        'mean_high': mean_high,
        'mean_low': mean_low,
        'n_high': n_high,
        'n_low': n_low,
        'significant': p_value < 0.05
    }
```

Code Cell [25]

```
welch_results[2021]
```

Code Cell [26]

```
# One-way ANOVA

anova_results = {}

for year in years:
    year_data = df[df['Year'] == year] if 'Year' in df.columns else df
```

```
    # Check if we have more than 2 groups for ANOVA
    unique_groups = year_data['ESG_Group'].nunique()

    if unique_groups >= 2:
        # Prepare data for ANOVA
        group_data = []
        group_labels = []

        for group in year_data['ESG_Group'].unique():
            group_values = year_data[year_data['ESG_Group'] == group]['Profit_Margin'].dropna()
            group_data.append(group_values)
            group_labels.append(group)

        # Perform one-way ANOVA
        f_stat, p_value_anova = f_oneway(*group_data)

        # Calculate effect size (eta squared)
        # eta_squared = f_stat * (len(group_labels) - 1) / (f_stat * (len(group_labels) - 1) +
(len(year_data) - len(group_labels)))

        anova_results[year] = {
            'f_statistic': f_stat,
            'p_value': p_value_anova,
            'n_groups': len(group_labels),
            # 'eta_squared': eta_squared,
            'significant': p_value_anova < 0.05,
            'group_means': {group: np.mean(year_data[year_data['ESG_Group'] ==
group]['Profit_Margin']) for group in group_labels}
        }
```

Code Cell [27]

anova_results[2021]

Code Cell [28]

```
# Create lagged variables and control variables
df = df.sort_values(['Firm', 'Year'])
df['Log_Revenue'] = np.log(df['Revenue'] + 1)
df['Log_Market_Cap'] = np.log(df['Market_Cap'] + 1)
df['Lagged_Profit_Margin'] = df.groupby('Firm')['Profit_Margin'].shift(1)
```

```
# Drop rows with missing values
df_clean = df.dropna(subset=['Profit_Margin', 'ESG_Score', 'Log_Revenue',
'Lagged_Profit_Margin'])

# Create dummy variables for industry and year
industry_dummies = pd.get_dummies(df_clean['Industry'], prefix='Ind', drop_first=True)
year_dummies = pd.get_dummies(df_clean['Year'], prefix='Yr', drop_first=True)

# Prepare data for regression - convert only numeric columns to float
X_numeric = df_clean[['ESG_Score', 'Log_Revenue', 'Lagged_Profit_Margin']].astype(float)
X = pd.concat([X_numeric, industry_dummies.astype(float), year_dummies.astype(float)],
axis=1)

X = sm.add_constant(X)
y = df_clean['Profit_Margin'].astype(float)

# Model 1: Pooled OLS (baseline)
pooled_model = sm.OLS(y, X).fit()

# Model 2: Fixed Effects model using statsmodels
# Only convert numeric columns to float, keep string columns as they are
numeric_cols = ['Profit_Margin', 'ESG_Score', 'Log_Revenue', 'Lagged_Profit_Margin']
df_clean[numeric_cols] = df_clean[numeric_cols].astype(float)

# Using statsmodels' PanelOLS equivalent with entity and time effects
fe_formula = 'Profit_Margin ~ ESG_Score + Log_Revenue + Lagged_Profit_Margin + C(Firm) +
C(Year)'
fe_model = smf.ols(fe_formula, data=df_clean).fit()
```

Code Cell [29]

```
#Time Trend of Pooled Model Predictions vs Actual by ESG Group
plt.figure(figsize=(14, 8))

# Create ESG groups
df_clean['ESG_Group'] = np.where(df_clean['ESG_Score'] >= 5, 'High ESG', 'Low ESG')

# Get pooled model predictions
df_clean['Pooled_Predicted'] = pooled_model.predict(X)

# Calculate yearly averages for actual and predicted values
yearly_actual = df_clean.groupby(['Year', 'ESG_Group'])['Profit_Margin'].mean().unstack()
yearly_predicted = df_clean.groupby(['Year',
'ESG_Group'])['Pooled_Predicted'].mean().unstack()

# Plot trends
```

```
markers = ['o', 's']
colors = ['#1f77b4', '#ff7f0e']
line_styles = ['-', '--']

for i, ESG_group in enumerate(yearly_actual.columns):
    # Actual values
    plt.plot(yearly_actual.index, yearly_actual[ESG_group],
             marker=markers[i], markersize=8, linewidth=3,
             label=f'{ESG_group} - Actual', color=colors[i], linestyle=line_styles[0])

    # Predicted values from pooled model
    plt.plot(yearly_predicted.index, yearly_predicted[ESG_group],
             marker=markers[i], markersize=8, linewidth=3,
             label=f'{ESG_group} - Predicted (Pooled)', color=colors[i], linestyle=line_styles[1])

plt.xlabel('Year', fontsize=12, fontweight='bold')
plt.ylabel('Profit Margin', fontsize=12, fontweight='bold')
plt.title('Time Trend: Actual vs Pooled Model Predictions by ESG Group',
          fontsize=14, fontweight='bold')
plt.legend()
plt.grid(True, alpha=0.3)
plt.xticks(yearly_actual.index)

# Add value annotations for actual values
for ESG_group in yearly_actual.columns:
    for year in yearly_actual.index:
        actual_value = yearly_actual.loc[year, ESG_group]
        plt.text(year, actual_value + 0.005, f'{actual_value:.3f}',
                 ha='center', va='bottom', fontsize=9, fontweight='bold')

plt.tight_layout()
plt.show()
```

Code Cell [30]

```
#Predicted vs Actual Profit Margin by ESG Score Quartiles
plt.figure(figsize=(14, 8))

# Create ESG score quartiles
df_clean['ESG_Quartile'] = pd.qcut(df_clean['ESG_Score'], 4, labels=['Q1 (Low)', 'Q2', 'Q3', 'Q4 (High)'])

# Get predictions from fixed effects model
df_clean['Predicted_Profit'] = fe_model.predict(df_clean)

# Calculate average actual and predicted values by ESG quartile
```

```
quartile_means = df_clean.groupby('ESG_Quartile')[['Profit_Margin', 'Predicted_Profit']].mean()

x_pos = np.arange(len(quartile_means))
width = 0.35

plt.bar(x_pos - width/2, quartile_means['Profit_Margin'], width,
    label='Actual Profit Margin', alpha=0.8, color='skyblue')
plt.bar(x_pos + width/2, quartile_means['Predicted_Profit'], width,
    label='Predicted Profit Margin (FE Model)', alpha=0.8, color='salmon')

# Add value labels
for i, (actual, predicted) in enumerate(zip(quartile_means['Profit_Margin'],
quartile_means['Predicted_Profit'])):
  plt.text(i - width/2, actual + 0.005, f'{actual:.3f}', ha='center', va='bottom', fontweight='bold')
  plt.text(i + width/2, predicted + 0.005, f'{predicted:.3f}', ha='center', va='bottom',
fontweight='bold')

plt.xlabel('ESG Score Quartiles', fontsize=12, fontweight='bold')
plt.ylabel('Profit Margin', fontsize=12, fontweight='bold')
plt.title('Actual vs Predicted Profit Margin by ESG Score Quartiles\n(Fixed Effects Model
Predictions)',
     fontsize=14, fontweight='bold')
plt.xticks(x_pos, quartile_means.index)
plt.legend()
plt.grid(True, alpha=0.3, axis='y')
plt.tight_layout()
plt.show()
```

Code Cell [31]

```
# Machine Learning Classification - 5 MODELS
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import roc_auc_score, accuracy_score, confusion_matrix,
classification_report, roc_curve, precision_recall_curve, precision_score, recall_score, f1_score
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import learning_curve

# Prepare ML data
ml_df = df[df['Year'] == 2024].dropna(subset=['Profit_Margin', 'Revenue', 'Market_Cap',
'ESG_Group'])
ml_df = ml_df[ml_df['ESG_Group'].isin(['High', 'Low'])]
```

```
X = ml_df[['Profit_Margin', 'Revenue', 'Market_Cap']]
X = pd.get_dummies(ml_df[['Profit_Margin', 'Revenue', 'Market_Cap', 'Industry']],
columns=['Industry'])
y = ml_df['ESG_Group'].map({'High': 1, 'Low': 0})

# Split data into train and test sets (70% train, 30% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42, stratify=y)

# Scale features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Initialize models separately
logistic_model = LogisticRegression(random_state=42, max_iter=1000)
random_forest_model = RandomForestClassifier(n_estimators=100, random_state=42)
gradient_boosting_model = GradientBoostingClassifier(n_estimators=100, random_state=42)
svm_model = SVC(probability=True, random_state=42)
knn_model = KNeighborsClassifier()

# Train and evaluate each model separately with detailed metrics
models = {
    'Logistic Regression': logistic_model,
    'Random Forest': random_forest_model,
    'Gradient Boosting': gradient_boosting_model,
    'SVM': svm_model,
    'K-Nearest Neighbors': knn_model
}

results = {}
for name, model in models.items():
    # Cross-validation on training data only
    cv_scores = cross_val_score(model, X_train_scaled, y_train, cv=5, scoring='roc_auc')

    # Train the model on training data
    model.fit(X_train_scaled, y_train)

    # Make predictions on test data
    y_pred = model.predict(X_test_scaled)
    y_proba = model.predict_proba(X_test_scaled)[:, 1] if hasattr(model, 'predict_proba') else
model.decision_function(X_test_scaled)

    # Calculate metrics on test data
    fpr, tpr, _ = roc_curve(y_test, y_proba)
    precision, recall, _ = precision_recall_curve(y_test, y_proba)
```

```
    # Store results
    results[name] = {
        'model': model,
        'cv_auc_mean': cv_scores.mean(),
        'cv_auc_std': cv_scores.std(),
        'accuracy': accuracy_score(y_test, y_pred),
        'precision': precision_score(y_test, y_pred),
        'recall': recall_score(y_test, y_pred),
        'f1': f1_score(y_test, y_pred),
        'y_pred': y_pred,
        'y_proba': y_proba,
        'fpr': fpr,
        'tpr': tpr,
        'precision_curve': precision,
        'recall_curve': recall,
        'confusion_matrix': confusion_matrix(y_test, y_pred)
    }

# ADDITIONAL ROC CURVE PLOT
from matplotlib import pyplot
from sklearn import metrics

# Calculate roc curves for each model using test data
sv_fpr, sv_tpr, _ = metrics.roc_curve(y_test, results['SVM']['y_proba'])
rf_fpr, rf_tpr, _ = metrics.roc_curve(y_test, results['Random Forest']['y_proba'])
lr_fpr, lr_tpr, _ = metrics.roc_curve(y_test, results['Logistic Regression']['y_proba'])
gb_fpr, gb_tpr, _ = metrics.roc_curve(y_test, results['Gradient Boosting']['y_proba'])
kn_fpr, kn_tpr, _ = metrics.roc_curve(y_test, results['K-Nearest Neighbors']['y_proba'])

# Plot the roc curve for the models
plt.figure(figsize=(10, 8))
plt.plot(sv_fpr, sv_tpr, linestyle='--', label='SVM')
plt.plot(rf_fpr, rf_tpr, linestyle='--', label='Random Forest')
plt.plot(lr_fpr, lr_tpr, linestyle='--', label='Logistic Regression')
plt.plot(gb_fpr, gb_tpr, linestyle='--', label='Gradient Boosting')
plt.plot(kn_fpr, kn_tpr, linestyle='--', label='K-Nearest Neighbors')

# Axis labels
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curves for All Models (Test Set)')

# Show the legend
plt.legend()
```

```
# Show the plot
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()
```

Code Cell [32]

```
# 2. Model Performance Comparison
performance_metrics = ['accuracy', 'precision', 'recall', 'f1']
performance_data = {metric: [result[metric] for result in results.values()]
                for metric in performance_metrics}

# Add AUC scores separately
performance_data['AUC'] = [roc_auc_score(y_test, result['y_proba']) for result in
results.values()]

plt.figure(figsize=(14, 8))
x_pos = np.arange(len(models))
width = 0.15

metrics_to_plot = ['AUC'] + performance_metrics
for i, metric in enumerate(metrics_to_plot):
  offset = width * (i - len(metrics_to_plot)/2)
  plt.bar(x_pos + offset, performance_data[metric], width, label=metric.replace('_', ' ').title())

plt.xlabel('Model')
plt.ylabel('Score')
plt.title('Model Performance Comparison Across Metrics (Test Set)')
plt.xticks(x_pos, list(models.keys()), rotation=45, ha='right')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True, alpha=0.3, axis='y')
plt.tight_layout()
plt.show()
```

Code Cell [33]

```
# 4. Confusion Matrices for All Models
fig, axes = plt.subplots(2, 3, figsize=(18, 12))
axes = axes.ravel()

for i, (name, result) in enumerate(results.items()):
  cm = result['confusion_matrix']
  sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', ax=axes[i],
        xticklabels=['Low ESG', 'High ESG'],
        yticklabels=['Low ESG', 'High ESG'])
```

```
    axes[i].set_title(f'{name}\nAccuracy: {result["accuracy"]:.3f}')
    axes[i].set_ylabel('Actual')
    axes[i].set_xlabel('Predicted')

# Hide the last subplot if we have an odd number of models
if len(models) < 6:
    axes[5].set_visible(False)

plt.suptitle('Confusion Matrices for All Models (Test Set)', fontsize=16)
plt.tight_layout()
plt.show()
```

Code Cell [34]

```
# Print detailed results for each model
print("="*80)
print("MACHINE LEARNING MODEL RESULTS (TEST SET PERFORMANCE)")
print("="*80)

for name, result in results.items():
    test_auc = roc_auc_score(y_test, result['y_proba'])
    print(f"\n{name}:")
    print(f"  Test AUC: {test_auc:.4f}")
    print(f"  CV AUC: {result['cv_auc_mean']:.4f} (±{result['cv_auc_std']:.4f})")
    print(f"  Accuracy: {result['accuracy']:.4f}")
    print(f"  Precision: {result['precision']:.4f}")
    print(f"  Recall: {result['recall']:.4f}")
    print(f"  F1-Score: {result['f1']:.4f}")

    # Print classification report
    print("\n  Classification Report:")
    report = classification_report(y_test, result['y_pred'], target_names=['Low ESG', 'High ESG'])
    for line in report.split('\n'):
        print(f"  {line}")

# Identify best model based on test AUC
best_model_name = max(results.items(), key=lambda x: roc_auc_score(y_test, 
x[1]['y_proba']))[0]
best_auc = roc_auc_score(y_test, results[best_model_name]['y_proba'])
print(f"\nBest Model: {best_model_name} (Test AUC: {best_auc:.4f})")
```

Code Cell [35]

```
import shap
# Explain the model's predictions on the TEST set
```

```
explainer = shap.LinearExplainer(logistic_model, X_train_scaled, feature_names=X.columns)
shap_values = explainer(X_test_scaled) # Get SHAP values for the test set

# 1. Create the Beeswarm Plot
plt.figure(figsize=(12, 10))
shap.summary_plot(shap_values, X_test_scaled, feature_names=X.columns, plot_type='dot',
show=False)
plt.title("SHAP Summary: Feature Impact on Predicting 'High ESG'", fontsize=16, pad=20)
plt.tight_layout()
plt.show()
```

Code Cell [36]

```
# 2.Bar Plot for a cleaner view of pure feature importance
plt.figure(figsize=(12, 8))
shap.summary_plot(shap_values, X_test_scaled, feature_names=X.columns, plot_type='bar',
show=False)
plt.title("SHAP Feature Importance (Mean |SHAP Value|)", fontsize=16, pad=20)
plt.tight_layout()
plt.show()
```