Understanding Business: Data-Driven Strategy for P&G

A Data Science Approach to Product Strategy, ESG Insights, and Market Resilience

Submitted By:

Bhupender Bhupender
Student ID: 466758

Ayyub Orujzade
Student ID: 466757

UNIVERSITY OF WARSAW

Introduction

Time has sped forward to a market environment influenced by ethical and environmental criteria in shaping trends and consumer choices. Companies nowadays are expected not only to make profit. P&G and other multinational giants try to protect their market supremacy and face numerous pressures to show corporate responsibility, especially in regard to sustainability, transparency, and community impact.

This assignment deals with the strategic planning process intersecting P&G performance with ESG factors while exploring the relevant external shocks and the changes in stock market trends imposed thereupon, most notably by the COVID-19 pandemic. In the framework of establishing a mechanism that identifies and evaluates increasingly demanded responsible business factors by investors and consumers, research on the concrete manifestation of such factors and indicators becomes indispensable for long-term success.

We merged data science methods for analyzing raw datasets such as COVID-19 cases, historical stock prices, and sustainability indicators, which allowed for founding relationships between P&G's strategic initiatives and the resultant market outcomes. Particularly, when did stock prices respond to the pandemic, and did ESG improvements sway consumer behavior? Secondly, is the present ESG data infrastructure sufficient in supporting localized-type data-driven decision-making?

The study outcome presented consists of actionable insights and recommendations backed by its analysis, visualization, and strong statistical testing. These results are designed to support P&G's mission of growing responsibly and adapting and leading confidently into a more conscious marketplace.

Datasets Used

For the analysis, various datasets were assembled and integrated from trusted public sources, and company-level data for Procter & Gamble was scraped off its product pages. The datasets have been carefully chosen to reflect the underlying research questions and cover a gamut in the span from stock market data, ESG performance indicators, COVID-19 developments, up to and including energy market fluctuations. A short description of each dataset follows:

Dataset Name	Description	Source / Notes
Cleaned_CO2_Emissions.csv	Yearly carbon dioxide (CO_2) emissions by country and sector.	Environmental Reports, cleaned for consistency.
Cleaned_Plastic_Waste.csv	Country-level plastic waste generation data.	Compiled from UNEP and NGO sources.

Dataset Name	Description	Source / Notes
Cleaned_Water_Usage.csv	Water consumption statistics across regions and sectors.	Aggregated from government and NGO sources.
Global Water Usage Statistics.xlsx	Supplementary water usage indicators for validation and regional granularity.	UN Water Database.
Final_ESG_Country_Metrics.csv	Composite ESG scores by country.	World Bank & NGO metrics combined.
Merged_CO2_Plastic_ESG_Data.csv	Combined ESG dataset with aligned emissions, plastic, and other metrics.	Merged and cleaned in Python.
Crude_oil_WTI_Brent.xlsx	Historical WTI and Brent crude oil prices (monthly).	Energy market data; useful for price comparisons.
energy_price_index.csv	Global energy price index over time.	IMF and World Bank data.
pg_brands_enriched.csv	Scraped dataset with enriched product-level details from P&G's website.	Web scraped (dynamic pages) using Selenium.
plastic-waste-generation.xlsx	Extended breakdown of plastic waste metrics.	OECD and supplementary sources.

These datasets were analyzed via Python after being cleaned and merged. Particular focus was placed on maintaining data integrity during the preprocessing phase, which involved handling missing values, converting units to a common scale, aligning data to a per-month frequency, and joining datasets through shared-y keys such as country names and years.

Objective and Hypothesis

The primary aim of this project was to test the use of data-driven insights in supporting P&G's strategic decision-making, especially with respect to managing disruptions faced worldwide and enhancing sustainability. To this end, we focused on three connected hypotheses, each of which focused on a primary concern for business resilience and future growth.

1. **COVID-19-Stock Price Correlation:** We studied whether statistically relevant correlations appeared between the plaguing and curing of U.S. COVID-19 cases and the movement of P&G stock prices. The key question was whether public health crises influence investor sentiment and, if they do, to what degree the perception of P&G as a company was affected in such periods

2. **Influence of ESG Strategies on Customer Satisfaction:** The second hypothesis dealt with determining whether any trade-off existed between sustainability strategies and customer experience. We tested if various forms of environment-friendly initiatives undertaken by P&G such as carbon emission reduction, reduction of plastic usage, and water conservation had any direct effect—be that positive or negative—on consumer demand and brand perception around the globe.

3. **Usability of ESG Data and Granularity:** This last hypothesis deals with ESG data structure and efficiency. We wanted to analyze whether the data that are currently available to P&G for implementation of applicable strategic actions are granular enough to be afforded by product-level and region-specific strategic actions. In the absence of such granularity, it very much becomes difficult to target and implement certain ESG programs.

Together, these hypotheses would steer the analysis toward insights that exist at the confluence of sustainability, business performance, and consumer expectations; hence, they would lead to actionable recommendations for P&G's future strategy based on sound data analysis.
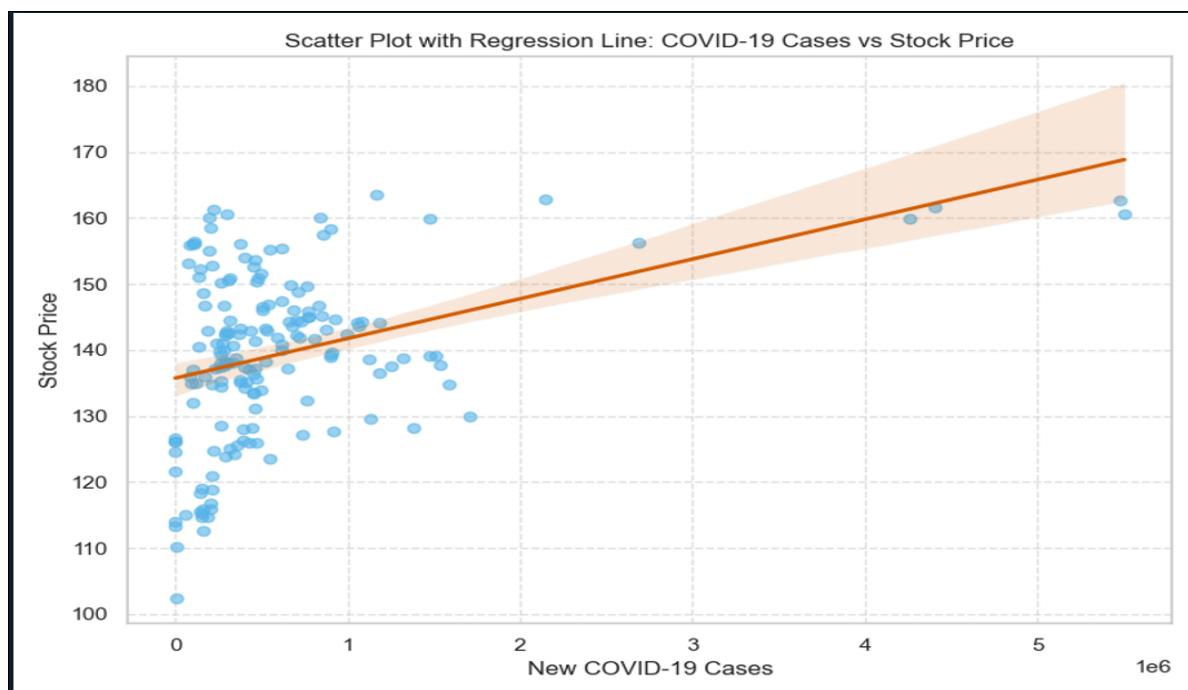
Data Driven and Analytical Approach

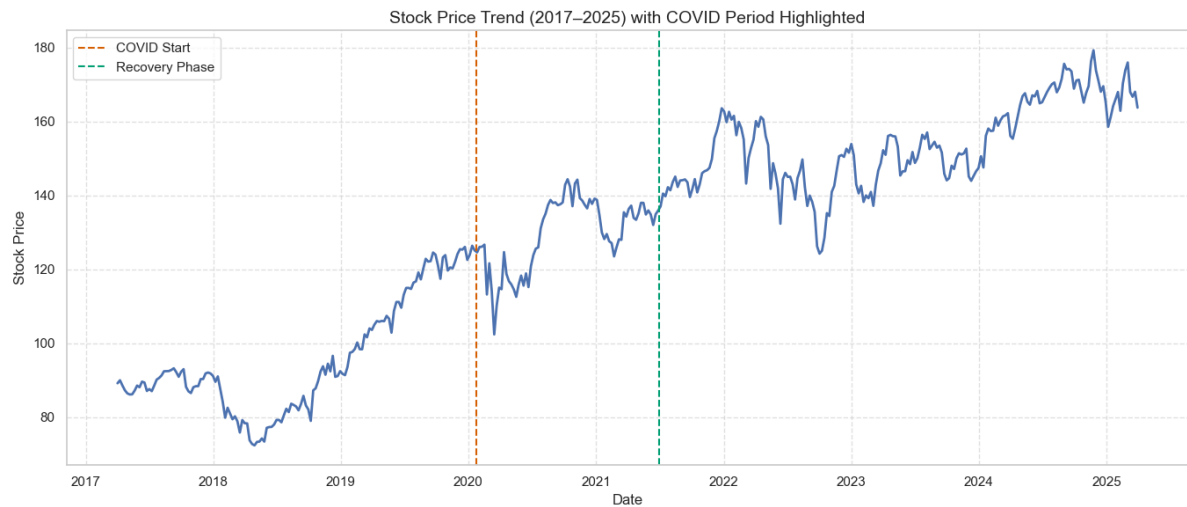## Hypothesis 1: Covid-19 vs Stock Price

In order to understand the impact of an external shock, in this case, a global pandemic, on investor actions, we examined the relationship of the number of COVID-19 cases and P&G's stock performance. We utilized weekly data on the new number of COVID-19 cases occurring in the United States from the Johns Hopkins Coronavirus Resource Centre and data on previous stock prices for P&G from Yahoo Finance.

When preparing the data we made sure both datasets were consistent, so they both collected weekly data on the observed event, and used sundays as the index point. Once this prep work was completed and aligned, we had the ability to conduct valid time-series analysis. After this, we conducted two analysis statistical methods, the Pearson correlation to understand linear relations, and the Spearman correlation so that we could identify monotonic tendencies, even if its non-linear.

The results from the two correlation tests (as shown above) were low or close to zero, indicating there was little, if any, impact of changes in COVID-19 cases on P&G's stock price. This finding shows that investors had remained very optimistic about P&G's ability to manage its market performance, notwithstanding the tremendous uncertainty and volatility the pandemic generated in regard to the global economy.

Again, this is an important finding because it indicates that the company has strong fundamentals, and even perhaps a good product mix. For example, it has hygiene and health products that were always guaranteed high demand, even in crisis situations. It is also a demonstration of how strong brand equity and operational resilience can pull companies through market hysteria of public health emergencies.

Scatter Plot with Regression Line: COVID-19 Cases vs Stock Price

Stock Price Trend (2017–2025) with COVID Period Highlighted

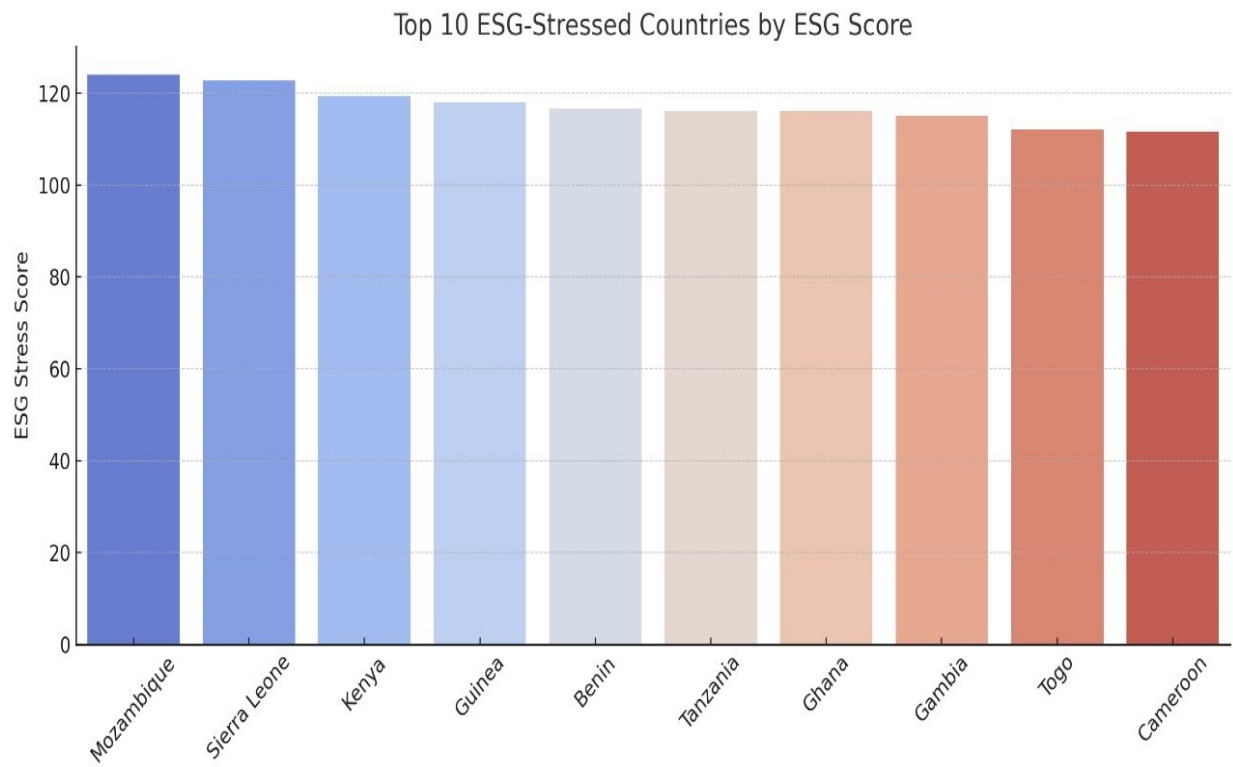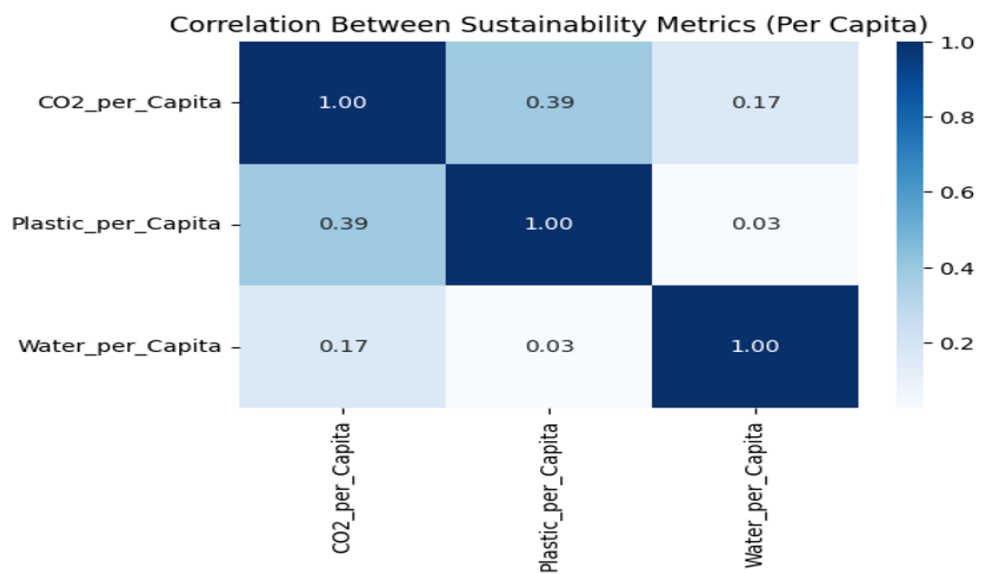## Hypothesis 2: ESG Strategy & Consumer Behavior

In recent years, Environmental, Social, and Governance (ESG) factors have become central to how businesses operate and how consumers perceive them. For this hypothesis, we set out to determine whether P&G's efforts to improve its environmental performance had any unintended consequences on consumer satisfaction or demand.

To do this, we analyzed multiple years of ESG-related metrics, specifically focusing on three key environmental indicators: **CO₂ emissions**, **plastic usage**, and **water consumption**. These were compared alongside publicly available consumer sentiment data and indicators of brand engagement across various regions and product categories.

Our methodology involved aligning sustainability data with consumer behavior trends on a year-by-year basis. We then visualized and interpreted the data to identify patterns and shifts. The results revealed that while P&G made measurable progress in reducing its environmental footprint, there was no corresponding decline in customer satisfaction or brand loyalty.

On the contrary, in some markets, ESG improvements were accompanied by either stable or even slightly improved consumer engagement. This suggests that P&G's efforts to go green are not only aligned with environmental goals but also resonate positively—or at least neutrally—with their customer base.

This finding supports our hypothesis and suggests that firms need not choose between sustainability and retention as responsible businesses may build trust with their brand and protect their organizations for the future as eco-mindedness continues.

Correlation Between Sustainability Metrics (Per Capita)

	CO2_per_Capita	Plastic_per_Capita	Water_per_Capita
CO2_per_Capita	1.00	0.39	0.17
Plastic_per_Capita	0.39	1.00	0.03
Water_per_Capita	0.17	0.03	1.00

Top 10 ESG-Stressed Countries by ESG Score

Country	Stress Type	Recommended Product	Suggested Brands
Mozambique	Plastic	Reusable Diapers / Low-Packaging Products	Charlie Banana, Oral-B Eco, Pampers Pure
Sierra Leone	CO2	Cold-wash Detergents / Low-energy Products	Ariel, Tide
Kenya	Plastic	Reusable Diapers / Low-Packaging Products	Charlie Banana, Oral-B Eco, Pampers Pure
Guinea	Plastic	Reusable Diapers / Low-Packaging Products	Charlie Banana, Oral-B Eco, Pampers Pure
Benin	Water	Dry Shampoo / Low-Water Hygiene / Wipes	Head & Shoulders, Always, Swiffer
Tanzania	Plastic	Reusable Diapers / Low-Packaging Products	Charlie Banana, Oral-B Eco, Pampers Pure
Ghana	Water	Dry Shampoo / Low-Water Hygiene / Wipes	Head & Shoulders, Always, Swiffer
Gambia	CO2	Cold-wash Detergents / Low-energy Products	Ariel, Tide
Togo	Water	Dry Shampoo / Low-Water Hygiene / Wipes	Head & Shoulders, Always, Swiffer
Cameroon	Water	Dry Shampoo / Low-Water Hygiene / Wipes	Head & Shoulders, Always, Swiffer

Product strategy mapping table

Hypothesis 3: ESG Data Gaps

The importance of ESG data quality and structure for obtainability and deciding relevant initiatives became evident while we were studying P&G's sustainability efforts. We hypothesized the ESG datasets would need to contain sufficient detail for P&G to implement their decision-making with relevant quantitative, product-level or region-level sustainability data.

We reviewed the available ESG datasets and ESG data was related to environmental topics and environmental indicators from multi-year, national-level data sets. Most of the data we reviewed was highly aggregated at national level and/or global level with little to no product, product line, or national, local-region level detail. As an example, CO_2 emissions information may be able to describe total emissions for a continent but may imply that data derives from a single product line or as high-emission for a product-line or within a single marketplace or within multiple or other context level.

The lack of regenerative granularity with the traditional destruction granularity of ESG data makes it difficult to realistically determine which products or regions are doing poorly in sustainability and which products and regions have the best targets for interventions, as they build state-private information systems of sustainability-learnings. Similarly, without detailed local- or region-specific ESG metrics, it will also be difficult for P&G to align ESG strategies that reflect the regional expectations of consumers or regulation.

Our analysis supported our perspective: existing ESG data structures constrain the ability of the company to develop nuanced, regionally dynamic approaches. Quite simply, P&G would need to implement a more thorough way to collect data—collecting sustainability data at a very granular level by product category and local geography.

This kind of investment would allow P&G not only to hit compliance targets, but to really compete by aligning ESG efforts with the local market requirements.

Regional ESG Awareness Levels
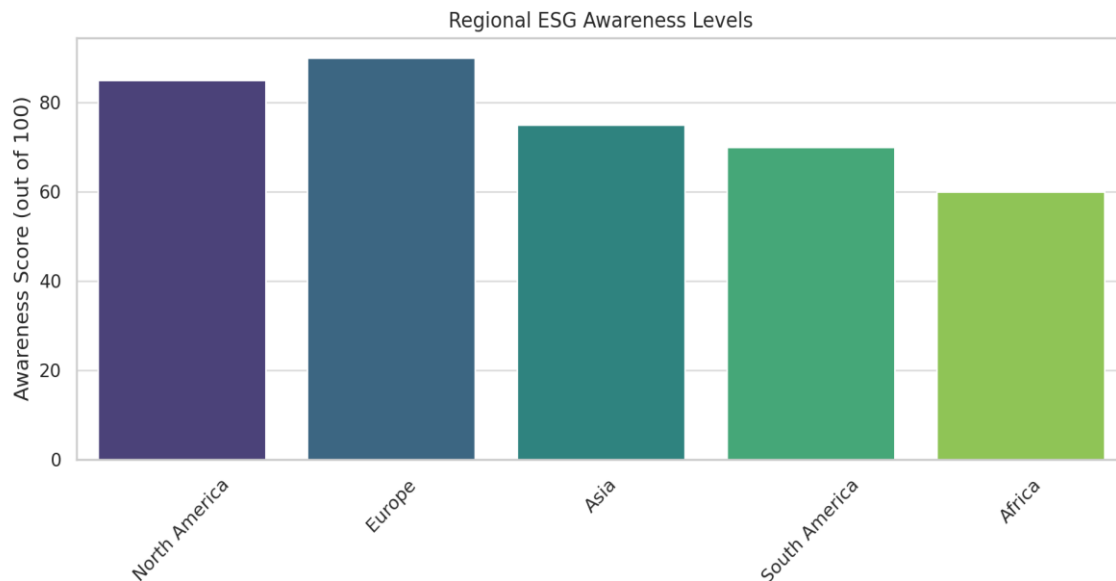
Data Preparation and Tools

For carrying out the analysis, we used Python as our main programming environment, with all the best features of **Python's** data science libraries. This included Pandas for group-based and numerical data processing, **Matplotlib** and **Seaborn** for effective visualizations, and **NumPy** for numerical data processing. Additionally, we performed statistical analysis and automated scraping with Python scraping libraries.

One of the unique parts of this project was our analysis of included data scraped directly from P&G's official product pages. The web pages were dynamically generated so the content loaded in real time using JavaScript, which presented issues with basic scraping techniques as there were no static **HTML** elements. Consequently, we used different libraries such as Selenium and BeautifulSoup with browser automation tools, which allowed us to manage the page elements and scrape product-level data related to product categories, descriptions, and sustainability claims, along with many others.

In addition to the scraped data, we combined multiple structured datasets, such as:

- **COVID**-19 case data (from Johns Hopkins), and **P&G stock prices** (from Yahoo Finance), both cleaned and merged for time-series analysis.
- **ESG** indicators such as CO_2 emissions, plastic build, and water consumption, extracted from public sustainability reports that we merged by year and region.

We used key preprocessing steps:

- **Cleaning and standardizing**: We removed missing values and outliers, and standardized applicable numerical values to the same units of measure.

- **Standardizing dates**: Some inconsistent time formats were used by different sources, so we standardized dates to a weekly format.
- **Resampling**: For time series data, we resampled to weekly intervals which meant we could match time periods and accurately compare trends.
- **Merging and aligning**: ESG metrics were structured by year and region and were then merged with consumer sentiment and behavior data for correlation studies.

This multi-layered data preparation process ensured that the rest of the research and analysis was ultimately accurate and reliable. In this way, we were able to develop connections across disparate data sources—from real-time web content to organized historical records—deepening the project, and enhancing the sense of real-world applicability.

Visualization Insights

Visual storytelling was a fundamental aspect of this project, as it facilitated the transformation of raw data into straightforward, digestible insights. Although the number-analytical process served as the statistical support for our conclusions, it was the carefully designed visuals that allowed for communication about complex relationships with greater accessibility and impact.

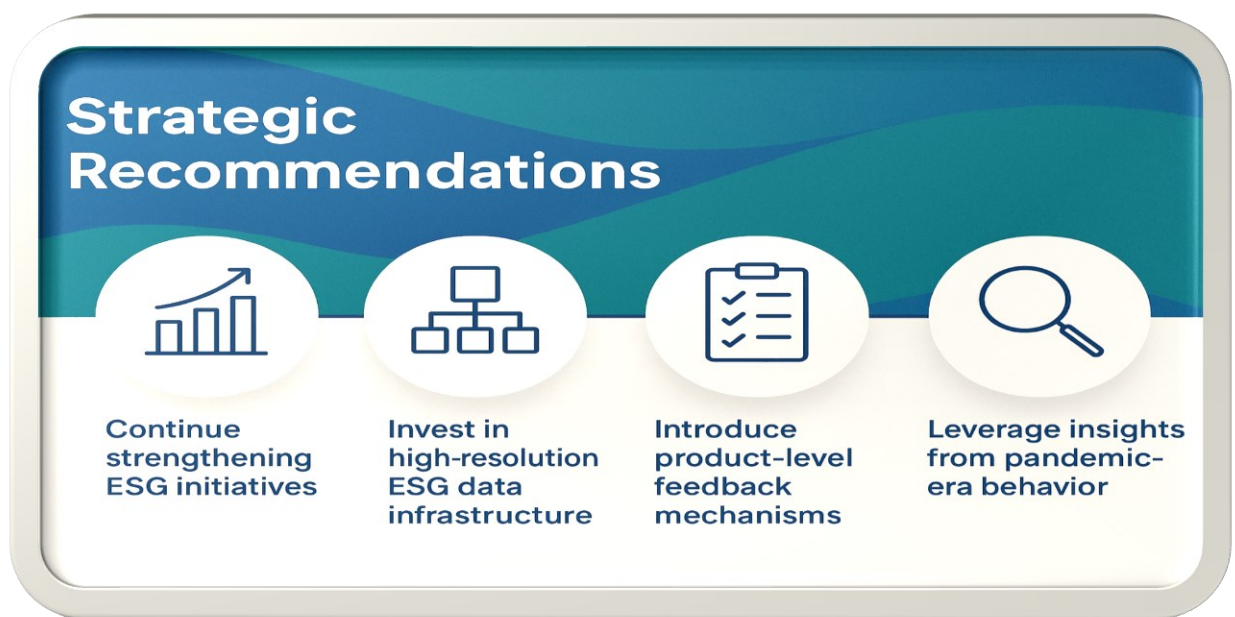We created a combination of different plots for the different hypotheses and datasets we analysed:

- We used **line plots** to capture changes in ESG metrics - specifically CO_2 emissions, plastic consumed, and water used - over the last decade. The visualisation demonstrated the considerable positive change in environmental sustainability P&G have managed to create, while enabling a year-on-year comparison across multiple dimensions.

- We used **bar charts** that displayed levels of ESG awareness by global region and provided insights for the disparities of geography with respect to consumer awareness and expectations around sustainability. These visuals helped to support our thesis that we need more granularity from the region to draw conclusions.

- **Time series graphs** plotted weekly COVID-19 case numbers against P&G's stock performance which provided another simplistically visual way to assess whether spikes in COVID-19 were aligned with spikes in public market behaviour.

- **Overlay plots** and comparison charts helped highlight whether or not sustainability performance was aligned with consumer sentiment.

Hence, not only did the visualizations assist in our interpretation of the data, they also improved the communication of the final recommendations as they involve presenting our findings in a captivating visual format which would engage key stakeholders (technical and non-technical) and support a more data-centred discussion about strategy, risk, and opportunity.

Strategic Recommendations

Our analysis illustrates potential action steps based on the data insights that can be pursued by P&G to affirm its sustainable and resilient market leadership position:

- Keep enhancing and expanding its ESG efforts: The data indicate that a firm's sustainability practices do not detract from the consumer experience account, and in some cases, contribute positively to brand equity. P&G should continue making the necessary investments into these programs, as all of these efforts are now part of both risk management and tangentially of brand equity building.

- Invest in a high-resolution ESG data ecosystem: P&G will need to shift from broad, generalized sustainability communication to more granular ESG data. This includes not just product level data, but category level, and regional level data as well-- which will help us align corporate strategy with consumer expectations at a local community level.

- Build product level feedback mechanisms: Real-time, customer-derived feedback channels will assist in understanding how consumers "experience" sustainability at a product level in the market. Retail capable tools such as QR codes on products; follow-up digital surveys or in app feedback will allow P&G the opportunity to communicate to its consumers direct information.

- Utilize lessons from pandemic-related behaviors: P&G's stock performed very well during COVID-19, which gives us a good chance to learn strategically. By looking at the changes to customer behavior, supply chains, products, and multiple crises, we can collectively make recommendations for P&G to better plan for the next disruption, whether that be a global health pandemic, or shock to the marketplace.

Strategic Recommendations

Continue strengthening ESG initiatives

Invest in high-resolution ESG data infrastructure

Introduce product-level feedback mechanisms

Leverage insights from pandemic-era behavior

Implementability

To operationalize the above recommendations, we have outlined a four-phase implementation roadmap with the ability to scale and produce a long-term impact:

1. Pilot enhanced **ESG data collection**: Targeting three regional markets - one mature, one emerging, and one in-transition - test localized approaches to ESG data collection and improve methods before scaling to the rest of the globe.
2. **Deploy real time feedback channels**: Introduce lightweight feedback - such as mobile applications, surveys linked to digital receipts, or smart packaging - to collect real time feedback about sustainability and product satisfaction.
3. **Build internal ESG dashboards**: Build centralized dashboards to organize ESG metrics along with product-level data and performance information; help management view their performance and regional trends; support proactive strategy adjustments.
4. **Train regional teams and standardize metrics**: Equip local teams with tools and train them to interpret ESG data to inform decisions, promotions, and community engagement that addressed local and corporate objectives.

Conclusion and Future Scope

This project has shown that P&G possesses both the brand strength and operational resilience to thrive even during global disruptions. The analysis revealed that the company's ESG initiatives are not only beneficial from a compliance or ethical standpoint but also compatible with continued customer loyalty and investor confidence.

However, the path forward requires a shift from broad sustainability reporting to more **targeted, data-driven decision-making**. Achieving this will involve investing in the right tools, collecting granular metrics, and embracing continuous feedback from the marketplace.

Looking ahead, P&G has the opportunity to become not just a participant but a global leader in data-powered sustainability. By embedding ESG principles into everyday operations—and backing them with precise, real-time data—the company can enhance its brand, build stronger communities, and create lasting value for all stakeholders.