ECSA PROJECT

Nolwazi Mazibuko (26195623)

Table of Contents

Introduction

In the increasingly data-centric landscape of modern industry, organizations must leverage robust analytics to maintain competitiveness and achieve sustained operational excellence. This report systematically examines the organization's operational performance, drawing from a substantial dataset encompassing customer demographics, product portfolio, sales transactions, and real-time service data. The primary aim is to provide actionable insights for upper management, facilitating informed and strategic decision-making.

The analysis begins by characterizing the customer base—a diverse group segmented by age, income, and geographic region—highlighting the potential for more targeted marketing and nuanced product differentiation strategies. The investigation extends to an evaluation of the product mix and sales trends, identifying both strengths and areas for improvement. Notably, the application of advanced statistical methodologies, including Statistical Process Control (SPC), Analysis of Variance (ANOVA), and discrete-event simulation, exposes operational bottlenecks, particularly inconsistent delivery performance in certain product categories such as KEY048, where seasonality and process variability undermine reliability.

Further, the report employs optimization modelling to scrutinize staffing and resource allocation, revealing tangible opportunities to enhance profitability and service quality through data-driven adjustments. Rather than relying on intuition or established routines, the evidence points toward a systematic approach to workforce management, aligning resources with actual operational needs.

Each section of the report is designed to build a nuanced understanding of organizational performance, progressing from descriptive analytics to more sophisticated statistical and optimization techniques. The overarching objective is to translate complex data into clear managerial implications, thereby equipping leadership with the insights necessary to refine processes, strengthen market position, and ensure the organization's long-term success through rigorous, data-informed management practices.

Part 1.2

Customer Data Analysis

The customer dataset comprises records for 5,000 individuals, each identified by a CustomerID and described via Gender, Age, Income, and City. These variables collectively enable a detailed examination of demographic and geographic patterns within the customer population. Notably, the dataset is complete—no missing values—improving the reliability and integrity of any subsequent analysis.

The data shows that the average customer is around 51.5 years old, with ages spanning from 16 up to 105. The age distribution appears pretty balanced—there's no dramatic skewing or weird outliers throwing things off. It's a solid representation across multiple age groups, from young adults to older individuals, suggesting the company connects with a broad demographic.

Customers earn an average of R80,797 annually, with incomes ranging from R5,000 up to R140,000. The distribution leans ever so slightly toward higher incomes, but not enough to claim there's a huge cluster of wealthy clients. Most customers sit in the R55,000 to R105,000 bracket, which fits the profile of working or middle-class individuals. All in all, this points to a diverse but mostly middle-income customer base, consistent with a company aiming to appeal to a broad spectrum rather than just one narrow market segment.

The customer base demonstrates notable gender diversity: 2,432 identify as female (approximately 49%), 2,350 as male (around 47%), and 218 as "Other" (4.4%). This distribution reflects a commendable level of inclusivity, indicating that marketing efforts should remain broad and not focus narrowly on traditional gender categories.

Analysing the geographic data, customers are distributed across seven cities. San Francisco leads with 780 customers, while Miami has the fewest at 647. The data points to a particularly strong market presence on the West Coast, especially in San Francisco. Meanwhile, cities like Miami and Seattle present promising opportunities for future market expansion.

Interpretation and Business Insights

Demographical Focus:

- The data reveals a customer base primarily situated in their early fifties. This demographic profile indicates a mature market segment, likely placing significant value on product reliability, convenience, and overall quality. For strategic marketing, it would be prudent to tailor messaging and offerings to resonate with these preferences, while still maintaining some outreach to younger consumers who may be entering the market.

Income Segmentation:

- From an income perspective, most customers fall within the R55,000 to R105,000 range, reflecting a predominantly middle-income audience. Consequently, pricing strategies should remain sensitive to affordability within this bracket. Nevertheless, the presence of higher-earning customers suggests an opportunity to introduce premium products or services aimed at this upper-income segment, potentially increasing overall revenue and market reach.

Gender Balance:

- The customer base reflects an almost equal ratio of male to female individuals, indicating that marketing strategies should remain gender-neutral and centre on universal appeal. The presence of a distinct "Other" category, though smaller, signals a growing expectation for diversity and inclusion—elements that can be strategically highlighted in corporate social responsibility initiatives and brand messaging.

Geographical Focus:

- San Francisco, Los Angeles, and New York emerge as primary hubs, suggesting that allocating marketing resources toward these cities would yield the greatest impact. In contrast, Miami and Seattle, with their comparatively lower customer counts, present potential opportunities for targeted growth campaigns or the exploration of new retail locations.

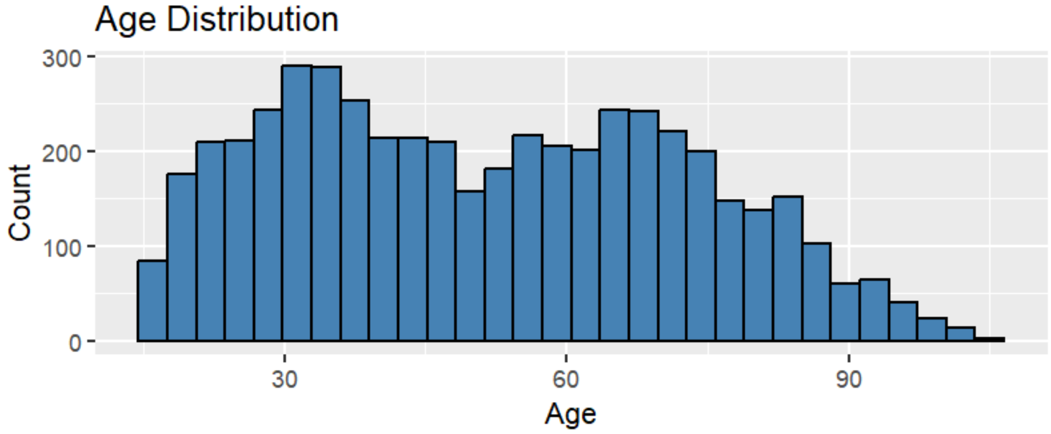
Strategic Importance

Understanding the makeup of a company’s customer base isn’t just helpful—it’s essential for effective, data-informed decision-making. When management has access to solid demographic statistics, they can pinpoint which groups to target in their marketing efforts and tailor product development accordingly. If most customers are clustered in a particular region, resources can be allocated more efficiently, ensuring efforts aren’t wasted elsewhere.

Income distribution among customers isn’t just a number; it directly shapes pricing strategies. Setting the right price points depends on knowing how much your audience can actually spend, not just guessing and hoping for the best.

A segmented approach to customer relationship management (CRM) also transforms generic messaging into meaningful engagement. Personalization becomes possible when you know who you’re talking to—making outreach more relevant and effective.

Beyond the basics, this kind of data lays the groundwork for more advanced analysis: projecting customer lifetime value, anticipating purchasing trends, and building robust market segmentation models. These activities are not optional extras—they’re critical components of any business striving for sustainable growth and a competitive edge. Without this analytical foundation, strategic planning becomes little more than guesswork.



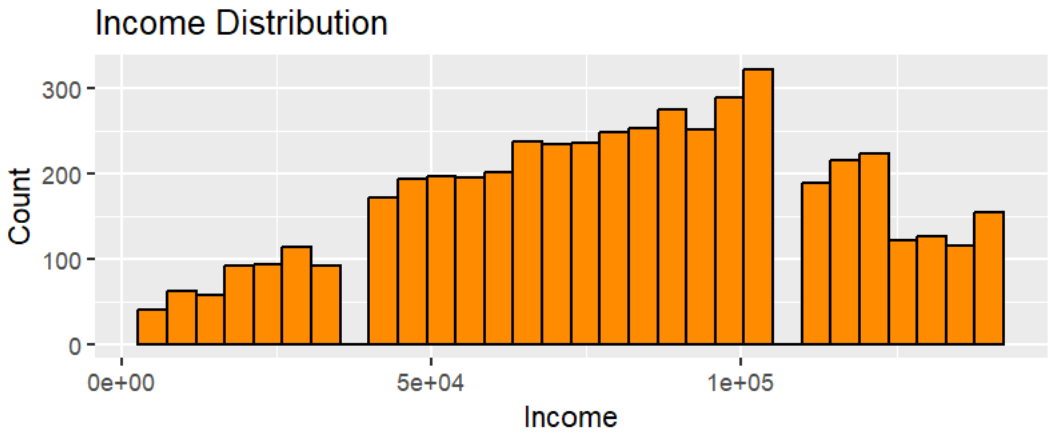
(Figure 1.1)

The age distribution histogram provides a clear snapshot of the workforce’s demographic makeup. The majority of employees fall within the 25 to 40-year-old range, suggesting a workforce that is predominantly mid-career. This group typically brings both experience and productivity, which can be advantageous for organizational performance in the near term. At the same time, there are noticeably fewer employees at the younger and older ends of the spectrum.

After age 60, the number of employees tapers off, indicating that retirements are likely to increase in the coming years. From a management perspective, this trend signals the importance of proactive workforce planning. Organizations must anticipate potential skill gaps and begin preparing for succession well before those retirements happen.

To address these demographic shifts, it would be prudent to focus on recruiting younger employees to ensure the long-term sustainability of the organization. In tandem, enhancing mentorship and training programs could help facilitate the transfer of institutional knowledge from older to younger staff, reducing the risk of losing valuable expertise.

In summary, the analysis of age distribution supports strategic decision-making in human resource management. By planning for workforce renewal and strengthening knowledge transfer initiatives, organizations can better maintain stability and continuity as their employee demographics evolve.

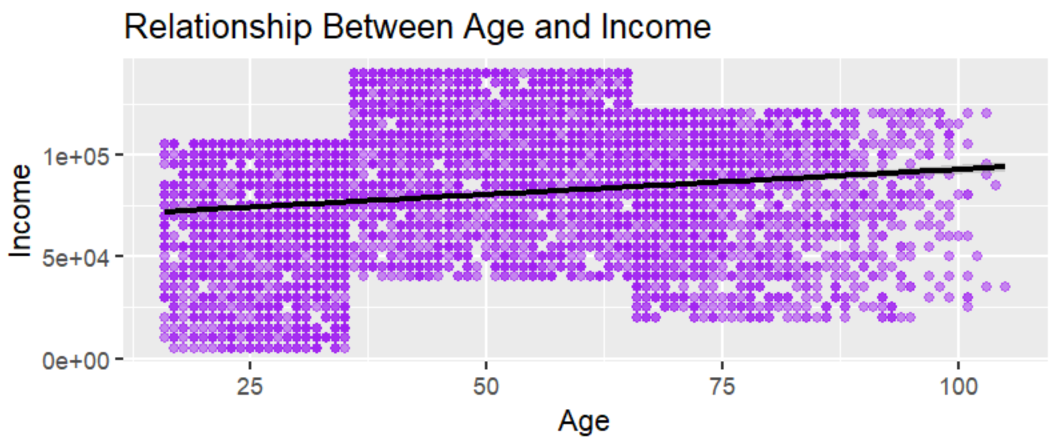


(Figure 1.2)

The income distribution histogram offers a revealing glimpse into the organization’s pay structure. Most individuals earn between $50,000 and $120,000, with a significant concentration near the $100,000 mark. This clustering indicates a large segment of the workforce occupies a middle-to-upper income tier, while relatively fewer employees fall into the lower salary brackets. The distribution is not entirely uniform; there are smaller peaks at higher income levels, likely reflecting senior roles or specialized positions.

From a management standpoint, this visualization is highly informative. It enables leaders to scrutinize whether pay distribution aligns with stated compensation policies, employee contributions, and industry standards. Recognizing income clusters can also help identify areas where salary adjustments might be necessary to maintain internal equity or retain critical talent. Furthermore, understanding the landscape of income concentrations supports more strategic decisions regarding budget planning, salary reviews, and workforce development.

In sum, analysing this histogram is instrumental for evidence-based human resources and financial management. It highlights patterns that can influence employee motivation, organizational fairness, and the overall effectiveness of compensation strategies.



(Figure 1.3)

The scatter plot presents the association between age and income, with each data point corresponding to an individual's income at a particular age. While the trend line does slope gently upward—implying that, on average, income tends to rise with age—the relationship is not especially strong. This suggests that, although greater experience and seniority are often linked to higher earnings, a variety of other factors (such as occupation, education, and career trajectory) also play significant roles in determining income.

Notably, the considerable vertical spread of data points within most age groups highlights substantial variability in income levels. In other words, not every older employee earns more than their younger counterparts, and some younger individuals already surpass older peers in earnings. This complexity reflects the multifaceted nature of compensation.

From a management or industrial engineering perspective, these findings are relevant for compensation planning, performance assessment, and workforce strategy. The data enables organizations to evaluate whether income progression with age is both equitable and consistent across departments, while also identifying potential disparities that could influence employee morale or retention. Ultimately, this type of analysis supports data-driven decisions in designing salary structures, planning career paths, and ensuring pay equity within organizations.

Products Data Analysis

The dataset contains both categorical and numerical variables. Specifically, categorical variables are Product ID, Category, and Description, while Selling Price and Markup fall under numerical data. Looking at the numbers, the mean selling price is R 4,493.59, but there's substantial variation—standard deviation clocks in at R 6,503.77. Prices go from as low as R 350.45 up to R 19,725.18, which clearly points to a broad range, from budget items to high-end products. The distribution isn't symmetrical, either; a positive skewness of 1.43 shows that most products are on the lower end, with a handful of much pricier outliers.

As for markup, the average sits at 20.46%, with a standard deviation of 6.07%. The range goes from 10.13% up to 29.84%. So, there's some fluctuation, but overall, markup percentages remain relatively consistent across categories. This pattern suggests that pricing strategies are generally controlled, though there's still room to optimize margins for select products.

Each of the six product categories contains exactly 10 items, ensuring a balanced foundation for comparing profitability and pricing structures across the board. This even distribution makes it much easier to benchmark performance and plan strategically, since no single category throws off the analysis.

Findings of the results

The analysis highlights notable aspects of the company’s product offerings and pricing methods. The broad range of selling prices reflects a diverse product portfolio, appealing to a spectrum of market segments. This assortment, which includes both lower- and higher-priced items, suggests the company aims to attract customers with varying purchasing power and preferences.

Price Dispersion and Product Mix

- A significant standard deviation in pricing, though, points to possible misalignments between certain products and their intended audiences. This could indicate opportunities for the company to refine its pricing strategies, ensuring that each product’s price accurately mirrors its value proposition and the expectations of its target demographic.

Consistency in Markup

- Regarding markup consistency, the data indicates a relatively uniform approach. While a standardized markup structure can streamline internal processes, it may overlook the potential to adjust margins for products with stronger demand or greater perceived value. To maximize profitability, the company could benefit from further analysing price elasticity and benchmarking against competitors, allowing for more nuanced pricing adjustments that respond to both market dynamics and consumer behaviour.

Category Breakdown and Strategic Focus

- With equal representation across all product types, any observed differences in performance are more likely tied to pricing strategies and market demand, rather than sampling inconsistencies. This setup enables clearer, more accurate comparisons of profitability and operational efficiency between product groups.

Managerial Implications

From an industrial engineering standpoint, this dataset offers substantial opportunities to enhance operational performance, refine pricing strategies, and optimize profitability. Several points warrant attention from senior management:

1. Product Portfolio Management:

The significant variance in selling prices indicates the importance of distinguishing between high- and low-margin products. Management should leverage this data to focus investments on products that yield higher profitability or align most closely with overarching strategic objectives.

2. Profitability Optimization:

Products that exhibit low markup percentages merit closer examination for potential inefficiencies in cost structure or pricing methods. Industrial engineers can play a critical role by conducting detailed analyses of production, procurement, and distribution processes to identify areas where costs can be reduced.

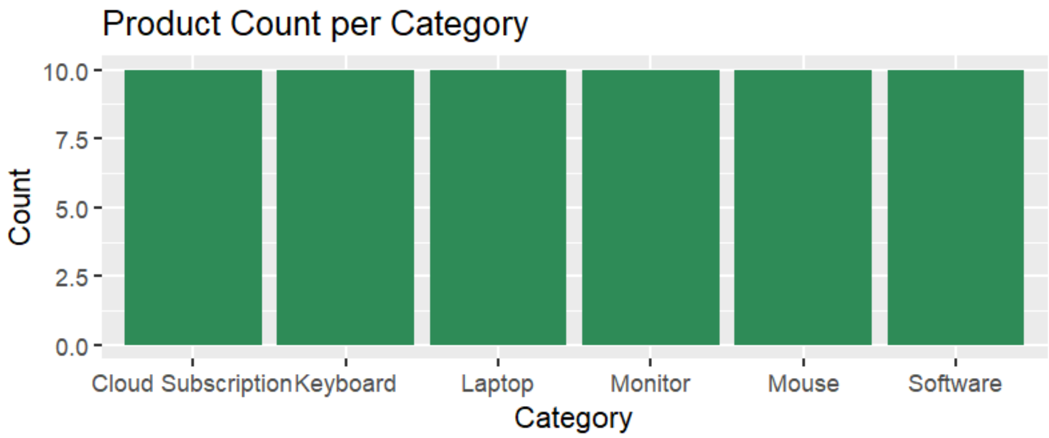
3. Pricing Strategy Alignment:

The observed skew in selling price distribution suggests a need to re-evaluate the product portfolio. Management may consider introducing additional mid-range products or adopting differentiated pricing models that better reflect product value and market demand.

4. Operational Efficiency:

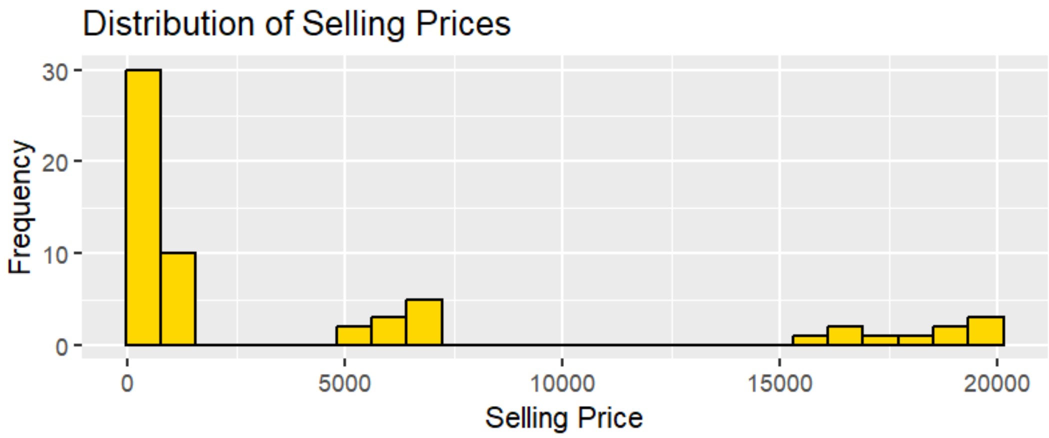
Although consistent markup levels may reflect disciplined pricing practices, it is advisable to further investigate whether similar consistency is present within cost management. Conducting comprehensive cost–benefit analyses across all product lines will help validate operational efficiency and highlight areas for improvement.

In summary, data-driven decision-making in these areas will enable management to more effectively allocate resources, enhance competitiveness, and support long-term organizational objectives.



(Figure 1.4)

The “Product Count per Category” bar chart displays an identical product count—ten—for each category: Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software. This even distribution creates an appearance of balance across categories, but it falls short in delivering substantive analytical value from an industrial engineering standpoint. The chart omits critical information regarding production output, consumer demand, and inventory efficiency—metrics essential for process optimization and strategic resource allocation. The uniformity could point to oversimplified data or aggregation, which restricts the chart’s capacity to offer meaningful operational insights. For upper management, the visualization provides little strategic direction, as it fails to identify high-performing areas or indicate where improvements are necessary. To increase its relevance and utility, incorporating variables such as production volume, sales performance, or cost efficiency would enable more informed decisions for both engineering and managerial purposes.

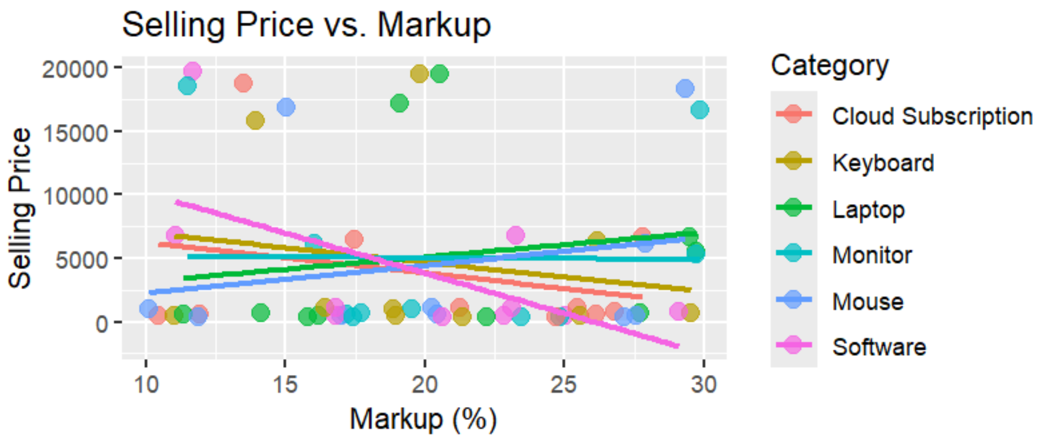


(Figure 1.5)

The “Distribution of Selling Prices” histogram presents a distinctly right-skewed pattern, with most products clustered at the lower end of the price spectrum, while only a handful extend into the higher price brackets. This indicates that the company’s offerings are predominantly low-cost, with premium-priced products representing a minority.

From an industrial engineering standpoint, such a distribution suggests a strategic emphasis on affordability and mass-market reach, rather than targeting the high-value or luxury segments. The notable spread in prices also points to considerable variability in product types or quality, which introduces complexity in production planning and cost control.

For upper management, the visualization draws attention to potential pricing imbalances and highlights areas for strategic refinement—such as expanding the presence of higher-margin items or diversifying price points. To deepen its managerial utility, the analysis could be broadened to incorporate data on demand patterns, profit margins, or price comparisons across different product categories. This would yield a more nuanced and actionable understanding of the company’s pricing architecture and competitive position.



(Figure 1.4)

The scatter plot titled “Selling Price vs. Markup” presents data on the relationship between selling price and markup percentage across six product categories: Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software. Each category is assigned a unique color, making patterns more accessible, and the inclusion of trend lines provides a visual cue regarding the nature and strength of these relationships.

Analysis reveals that for categories such as Laptops, Monitors, and Mice, there is a discernible positive correlation between markup and selling price; in other words, higher-priced products within these categories are generally associated with higher markups. Conversely, the Software and Cloud Subscription categories deviate from this trend, demonstrating either a weak or negative correlation between selling price and markup. This suggests that pricing within these categories may be governed by different factors, or that the relationship between price and markup is less straightforward.

From an industrial engineering perspective, this visualization offers valuable insight into pricing efficiency and the alignment of cost and performance across various product lines. By highlighting the differing responses to changes in markup among product categories, the chart supports more informed decisions regarding cost optimization, pricing strategies, and production planning. For upper management, the visualization serves as a strategic tool for identifying patterns of profitability and inconsistencies in pricing, potentially revealing areas where adjustments could enhance overall revenue. Nevertheless, the notable dispersion of data points underscores considerable variability, indicating the necessity for further analysis involving cost, demand, and sales data to ensure that pricing strategies are well-aligned with market dynamics and operational realities.

Sales Data Analysis

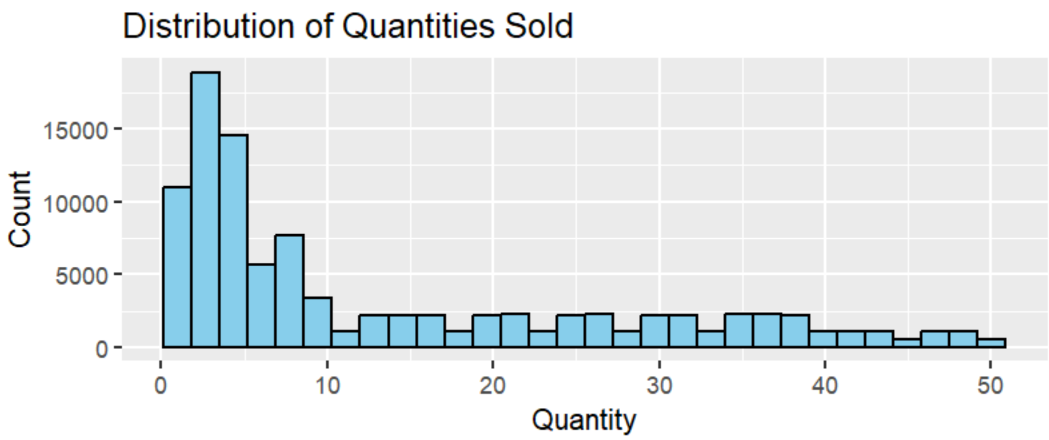
The sales dataset, comprising 100,000 records from both 2022 and 2023, offers substantial insights into organizational performance. Representing 5,000 unique customers and 60 distinct products, the data reflects significant diversity in both clientele and product range. The mean order quantity sits at 13.5 units; however, the standard deviation is notably high (SD = 13.76), indicating a wide range of purchasing behaviours among customers.

Temporal analysis reveals that most transactions occur during standard business hours and are distributed relatively evenly throughout the year. The average picking time is 14.7 hours, and the mean delivery time is 17.5 hours, each accompanied by considerable standard deviations. This observed variability in operational efficiency highlights potential areas for further investigation and process optimization within the organization.

In 2022, there were 53,727 recorded orders, while 2023 saw a decrease to 46,273. Despite the reduction in order frequency, the average units per order increased slightly from 13.4 to 13.6. This uptick could indicate a trend toward bulk purchasing or more strategic ordering behavior, potentially driven by pricing models, shipping incentives, or shifts in demand.

From an industrial engineering standpoint, these observations hold significant value. Fluctuations in picking and delivery times reveal inefficiencies and highlight areas where process refinement is possible. Such data provides a foundation for improving operational efficiency, optimizing resource allocation, and balancing workloads within production and distribution systems.

For senior executives, this data offers a critical perspective on sales stability, operational reliability, and evolving customer purchasing behaviours. Such information equips leadership to evaluate the effectiveness of supply chain operations and make data-driven decisions regarding capacity planning, workforce allocation, and inventory control. Additionally, the dataset underpins predictive analysis—enabling forecasts of future demand, identification of operational bottlenecks, and formulation of strategies geared toward continuous enhancement of process efficiency and service quality.

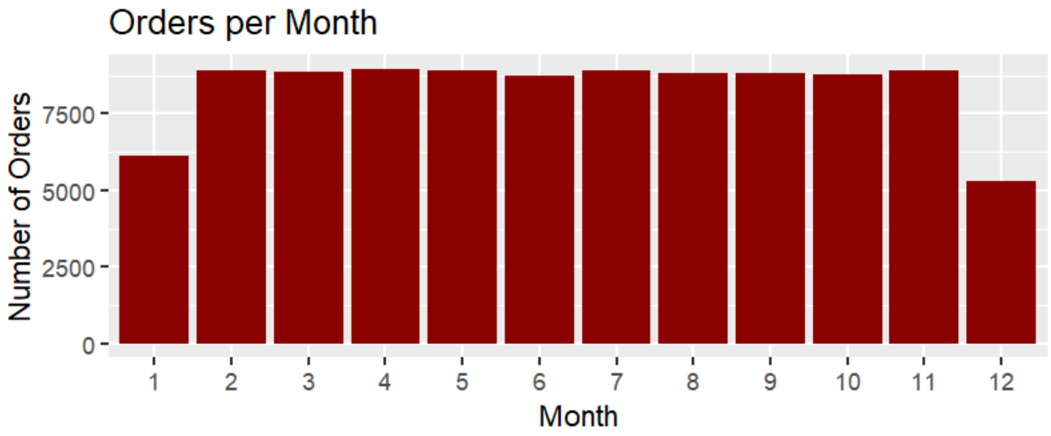


(Figure 1.7)

The histogram labeled “Distribution of Quantities Sold” clearly displays the frequency of sales across a range of quantity intervals. The data reveal a pronounced right skew: the majority of transactions cluster within the lower quantity bracket, particularly between 1 and 10 units. In contrast, sales involving larger quantities are relatively scarce. This distribution indicates that most customers prefer to place smaller, more frequent orders, while high-volume purchases are outliers.

From the lens of industrial engineering, this visualization offers meaningful insights into customer demand and the implications for production planning. The dominance of small-quantity sales highlights the necessity for operational flexibility. Rather than focusing solely on bulk production, it becomes crucial to optimize processes such as order picking, packaging, and logistics, ensuring that frequent low-volume orders can be managed efficiently. At the same time, the occasional presence of large-quantity sales suggests there may be value in developing differentiated handling strategies or customer segmentation to further streamline operations and enhance service quality for both segments.

For senior management, this histogram is an important tool for strategic decision-making on several fronts, including production scheduling, inventory management, and customer relationship strategies. Insights from the distribution can inform how stock levels are set, how warehouse layouts are designed, and how workforce resources are allocated to account for the variability in order sizes. In addition, management may use this understanding to explore pricing or discount initiatives aimed at incentivizing larger orders, thereby improving economies of scale and overall profitability. In summary, while small-quantity transactions predominate, the data also point to operational and strategic opportunities to improve efficiency and drive sales performance.

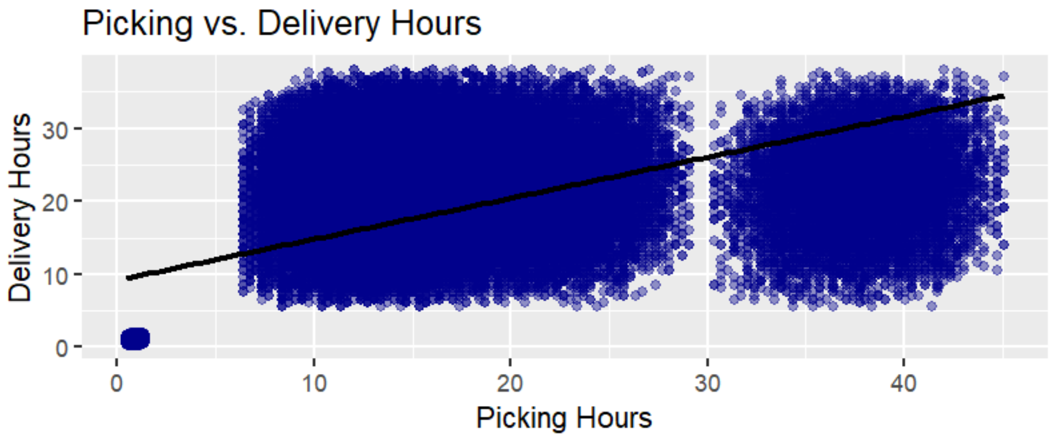


(Figure 1.8)

The "Orders per Month" bar chart provides a clear illustration of how customer orders are distributed throughout the year. From February to November, order volumes remain consistently high, indicating a period of operational stability. In contrast, both January and December show a significant dip in order activity, highlighting a distinct seasonal fluctuation. This trend likely reflects reduced consumer engagement during these months, potentially due to post-holiday fatigue, budgetary resets, or established purchasing cycles.

From the standpoint of industrial engineering, such a pattern offers valuable insight into demand seasonality and workload management. The regularity in order volumes for most of the year facilitates predictable production scheduling, efficient workforce allocation, and effective inventory management. The quieter periods in January and December present an opportunity for organizations to conduct equipment maintenance, provide staff training, or implement system upgrades with minimal disruption to core operations.

For senior management, this visualization supports strategic planning and resource optimization. Recognizing the cyclical nature of demand enables more accurate forecasting and better allocation of production and logistics resources, thereby minimizing both overcapacity and underutilization. The steady performance during the mid-year months underscores a dependable market, which management can leverage for capacity planning and long-term initiatives. Additionally, targeted marketing strategies or promotional campaigns during the slower months could help mitigate revenue fluctuations and further stabilize annual performance.



(Figure 1.9)

The scatter plot presents Picking Hours on the x-axis and Delivery Hours on the y-axis, with a linear regression line superimposed to indicate the general trend. The plot suggests a weak to moderate positive relationship between picking and delivery times: as picking hours increase, delivery hours generally do as well, though the association is far from tight. The wide dispersion of data points around the trend line highlights that the relationship is not strongly linear, hinting at the influence of additional operational factors beyond picking duration.

Distinct clusters are evident, particularly in the ranges of 10–25 and 30–45 picking hours. These groupings may reflect underlying differences in operational contexts—perhaps variations in delivery regions, product types, order sizes, or distribution methods. Such clustering implies that the process is heterogeneous, and a segmented analytical approach may be more appropriate.

Additionally, several outliers appear near the origin, where both picking and delivery hours are close to zero. These points could be attributed to data entry errors, cancelled transactions, or unusually small orders. Their presence raises concerns about data quality and signals the need for thorough pre-processing or targeted investigation, to ensure the reliability of subsequent analyses.

Examining the visualization from an industrial engineering standpoint, several key insights emerge. Firstly, there's a clear connection between the duration of picking activities and overall delivery times—delays in the warehouse don't just stay put; they ripple down the line and slow everything else. This kind of upstream inefficiency can end up limiting throughput for the whole operation.

Looking closer, the points in the visualization don't stick to a neat trend; instead, there's a wide scatter. That's a textbook sign of process variation, which undermines consistency and makes it tough to predict outcomes. In academic and industry contexts alike, such variability is a red flag, often addressed through methodologies like Lean or Six Sigma to restore some order.

Additionally, there are noticeable clusters within the data. These likely correspond to different order types, customer segments, or logistics strategies, each behaving in its own way. Recognizing these distinctions can support process segmentation and more targeted optimization, rather than relying on a one-size-fits-all approach.

And then there are the outliers—data points that just don't fit the pattern. Whether caused by exceptional cases in operations or by data errors, they signal a need for closer process control or improved data validation.

In summary, this visualization highlights significant process variability across the order fulfillment chain and points to specific areas where inefficiencies and improvement opportunities might exist. It serves as a useful initial diagnostic for deeper analysis and targeted interventions.

Overall Findings from the datasets

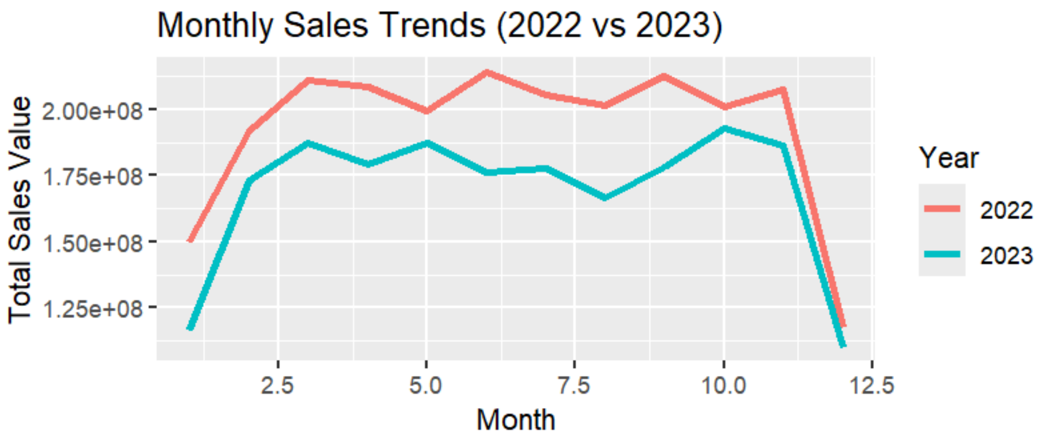
The annual sales review reveals a decrease in total revenue from approximately 2.32 billion in 2022 to 2.03 billion in 2023. Notably, both the average order value and the quantity per order increased during 2023, suggesting a shift in customer purchasing behavior—possibly due to pricing strategies, product assortment changes, or evolving consumer needs. Such aggregate performance data is essential for management to evaluate commercial trends, assess financial stability, and inform ongoing strategic planning.

Geographical analysis indicates revenue is heavily concentrated in major metropolitan areas, mainly Los Angeles, San Francisco, and New York. These cities represent key markets where operational efficiency and customer satisfaction are especially critical. As a result, leadership may prioritize resource allocation, service enhancement, and risk mitigation strategies in these regions to safeguard core revenue streams and maximize financial impact.

Product category performance shows that laptops, monitors, keyboards, mice, software, and cloud subscriptions are leading contributors to overall sales. This insight is valuable for decision-makers in optimizing the product portfolio, focusing marketing efforts, managing inventory, and refining procurement strategies. Understanding these demand trends helps minimize risks related to inventory imbalances and supports targeted promotional campaigns for high-performing categories.

From an Industrial Engineering perspective, sales volume trends and order characteristics have direct implications for operational planning. This includes workforce scheduling, capacity planning, and supply chain management. Identifying top-performing cities and product categories enables more efficient network design, facility layout, and logistics optimization. For example, high-demand items should be located in fast-pick areas to reduce labor costs and improve fulfillment speed, while focusing distribution resources in top-revenue cities can enhance logistical efficiency.

For upper management, these findings provide a foundation for evidence-based decisions regarding market development, customer targeting, product strategy, and operational investments. For Industrial Engineers, this data is crucial for designing efficient systems and aligning operations with market requirements. Integrating sales analytics with operational improvements supports both organizational performance and the pursuit of sustainable competitive advantage.



(Figure 1.10)

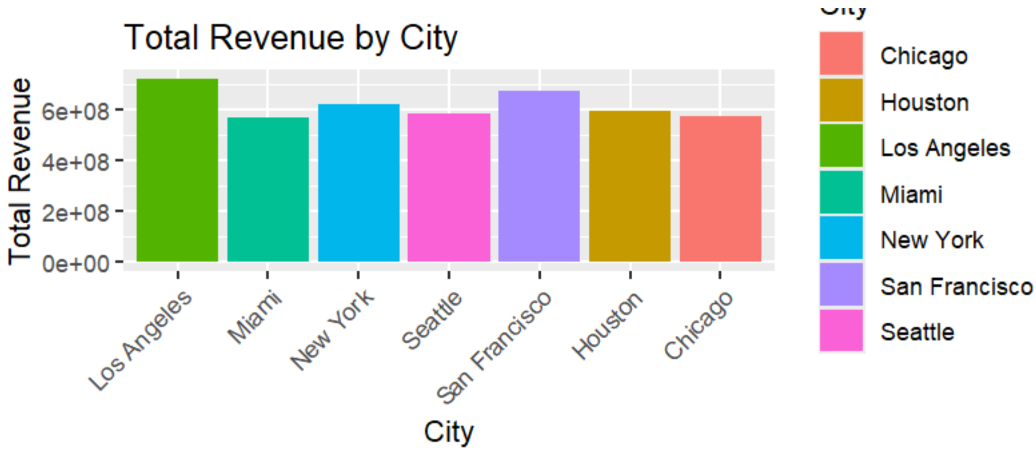
The line chart presents monthly sales patterns for 2022 and 2023, clearly highlighting both the seasonal nature and the inherent volatility in demand across these two years. From an Industrial Engineering standpoint, such visualizations are fundamentally important for effective capacity planning, resource management, and continuous process improvement. Recognizing regular peaks and declines in monthly sales enables engineers to forecast future workload shifts and adjust operational capacity as needed. For instance, the consistently elevated sales observed during the middle of each year indicate periods when increased picking, packing, and delivery activities are required. This, in turn, may necessitate additional staffing, overtime hours, or the hiring of temporary workers to avoid operational bottlenecks and declines in service quality. Conversely, periods of lower demand offer a strategic window for scheduling maintenance, implementing improvement initiatives, or conducting employee training without disrupting essential business activities.

Additionally, the chart provides critical support for inventory management and supply chain alignment. Predictable demand cycles facilitate more accurate forecasting and inventory replenishment, which helps minimize excess safety stock and reduce waste. Aligning production and distribution with these seasonal sales patterns can further decrease lead times and enhance service levels. Ultimately, by making demand variability visually apparent, the chart supports data-driven management approaches such as Lean and Six Sigma.

Strategic Relevance for Upper Management

For senior leadership, this visualization serves as a comprehensive overview of sales performance and seasonal revenue trends, thereby supporting informed, strategic decision-making. By comparing 2022 and 2023, management can assess whether the organization’s commercial performance is improving or facing challenges, as well as evaluate the impact of prior strategic initiatives. Any significant deviations or declines in performance may indicate the need for targeted interventions, such as revised marketing strategies, pricing adjustments, or product diversification.

Furthermore, the identification of months with the highest revenue generation aids in budgeting, the timing of promotional campaigns, and broader strategic planning. This enables management to allocate financial and human resources more efficiently, focusing on periods with the greatest projected returns. The visualization also facilitates risk management, as understanding predictable demand fluctuations allows organizations to proactively prepare for operational risks. In summary, the chart provides a crucial link between operational execution and strategic oversight, enabling leadership to align long-term objectives with actual market demand.



(Figure 1.11)

The bar chart provides a clear illustration of revenue disparities among major cities, with Los Angeles, San Francisco, and New York standing out as leading contributors, while Houston and

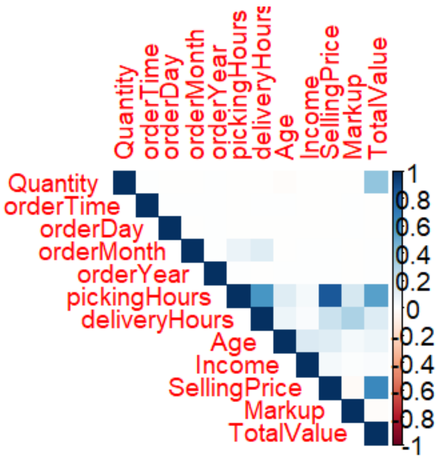
Chicago report notably lower revenue figures. From an industrial engineering standpoint, recognizing these geographic differences is essential for effective network planning, logistics optimization, and operational prioritization.

Identifying high-revenue regions allows organizations to make informed decisions regarding facility placement, warehouse capacity, and transportation design—all crucial elements in an optimized supply chain. Cities that generate higher sales often demand larger distribution centers, more robust delivery networks, and potentially increased investment in automation to handle elevated throughput. In contrast, areas with lower revenue can be efficiently managed with shared facilities or streamlined networks, helping to minimize unnecessary fixed costs.

This visualization is also valuable for resource allocation and takt-time management, as it highlights where workforce capacity, service levels, and inventory buffers should be prioritized to maintain customer satisfaction. By aligning operational resources with actual demand, industrial engineers can minimize logistical inefficiencies, such as excessive transportation, waiting times, and redundant processing. The chart thus facilitates data-driven continuous improvement by making regional demand variability visible and actionable.

From a strategic management perspective, the visualization offers a succinct overview of regional market strengths, supporting evidence-based strategic planning and resource distribution. Upper management can use these insights to focus marketing initiatives, strengthen key customer relationships, and safeguard primary revenue sources. Meanwhile, underperforming regions may be targeted for intervention—whether through promotional campaigns, operational restructuring, or potential divestment.

Ultimately, this information is integral to maximizing profitability and managing risk, as overdependence on specific regions can expose an organization to geographic market fluctuations. The bar chart assists decision-makers in evaluating expansion opportunities, prioritizing operational investments, and achieving a balanced national presence. This alignment of operational and strategic objectives ensures that commercial goals are supported by efficient, responsive logistics.



(Figure 1.12)

The correlation heatmap functions as a visual summary of the linear interrelationships among essential operational, financial, and temporal variables within the dataset. Each cell in the heatmap quantifies the strength and direction of the association between two variables: values approaching +1 indicate a strong positive correlation, values near –1 point to a strong negative relationship, and those close to zero reflect minimal or no correlation.

From an industrial engineering perspective, this visualization is a practical diagnostic tool for uncovering interdependencies among parameters such as production, logistics, and sales. For example, a notable correlation between pickingHours and deliveryHours may suggest operational linkages within the supply chain, indicating that inefficiencies in one stage could impact subsequent processes. Likewise, observing robust positive correlations among variables like SellingPrice, Markup, and TotalValue validates the internal consistency of the financial data and supports confidence in benchmarking efforts.

For senior management, the heatmap delivers actionable insights into organizational process dynamics. By spotlighting variables with the strongest influence on value creation and operational performance, it facilitates data-driven decision-making and more effective resource allocation. Additionally, identifying clusters of high correlation can guide predictive modelling, enabling more accurate demand forecasts, optimized scheduling, and the assessment of process improvement initiatives. Overall, this analysis contributes to enhanced system efficiency, cost reduction, and a closer alignment of management strategies with empirical evidence.

Part 3

The primary aim of this part is to conduct a Statistical Process Control (SPC) analysis focused on delivery times across various product categories, using data from the sales2026and2027Future.csv dataset. SPC serves as a systematic quality control approach, enabling the monitoring and regulation of processes by evaluating sample data over extended periods. Here, the delivery timeline is scrutinized to verify process stability and to ensure it aligns with customer expectations.

The dataset provides a chronological record of deliveries, segmented by product type and arranged by date and time to closely replicate real-time data acquisition. For analytical purposes, deliveries have been grouped into samples of 24 for each product type. The initial 30 samples from each group are utilized to establish baseline control limits for both X-bar (mean) and s (standard deviation) charts. These metrics are essential for ongoing assessment of process consistency.

This methodology facilitates the detection of deviations within the delivery process, uncovers potential operational inefficiencies, and assesses overall process capability. Additionally, it supports proactive management; when a delivery process deviates from established performance criteria, timely interventions can be made. Ultimately, this ensures that the Voice of the Customer (VOC) is continuously addressed through consistent delivery performance.

Data Preparation

All delivery data encompassing various product types was imported into R from the CSV file named sales2026and2027Future.csv, utilizing the read.csv() function. This facilitated the creation of a data frame, enabling systematic manipulation and subsequent analysis of delivery times.

To maintain the integrity of process control analysis and ensure that it accurately reflects real-time delivery sequences, the dataset was sorted in chronological order. This entailed identifying columns corresponding to Year, Month, Day, and Order Time, and then ordering the entries so that the earliest deliveries were positioned first. This sequencing step is essential, as statistical process control (SPC) requires the time order of data to be preserved in order to reliably detect trends, shifts, or atypical variations within the process.

After the data was appropriately ordered, it was separated by product type to allow for individual analysis. For each product, delivery times were grouped into samples of 24 deliveries, mirroring the approach used in real-time batch data collection. These grouped samples served as the basis for calculating control limits and for monitoring process stability, specifically through the application of X-bar and s charts.

To monitor the delivery process effectively, X-bar and s charts were constructed for each product type. The baseline for analysis was established using the first 30 samples, each consisting of 24 delivery observations.

X-bar Chart (Mean Chart)

The centre line of the X-bar chart was calculated as the mean of the sample means from these initial 30 samples. Three-sigma control limits were then determined using the standard error of the mean and the constant A3 = 0.619. Additionally, two-sigma control limits were included to help detect smaller shifts in the process mean.

s Chart (Standard Deviation Chart)

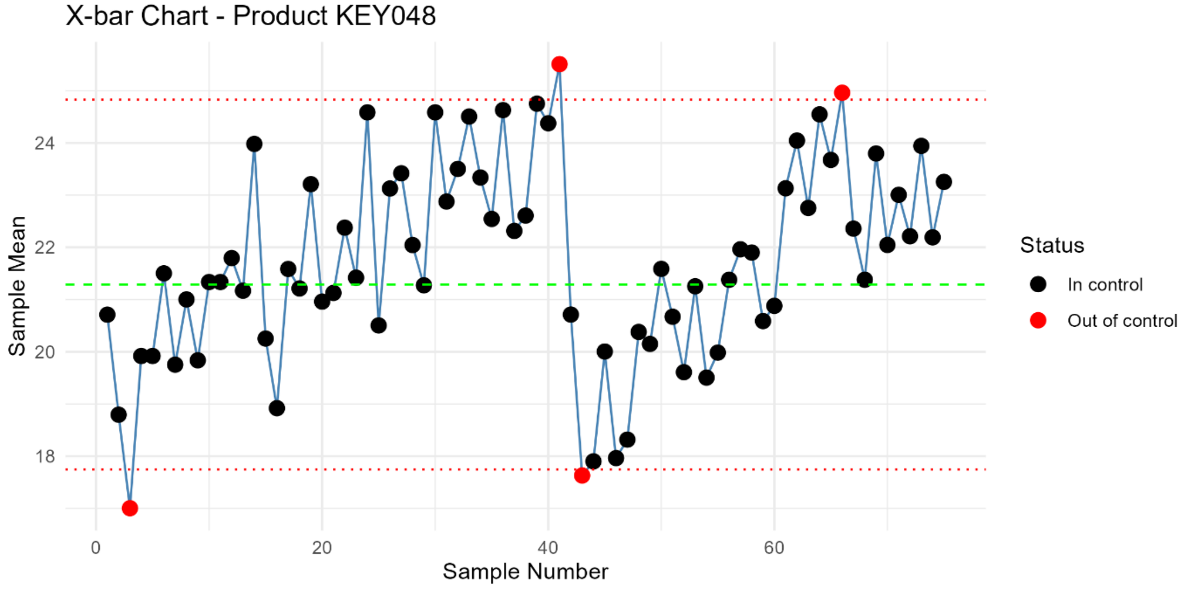
For the s chart, the centre line was set as the average standard deviation from the first 30 samples. Upper and lower control limits were calculated based on the constants B3 = 0.555 and B4 = 1.445.

One-sigma control limits, using c4 = 0.9892, were also established to monitor minor variations in process spread.

Visualization

X-bar and s charts were generated for each product type, plotting the relevant sample statistics against the calculated control limits. This visualization allows for the identification of points or trends that may indicate potential process control issues, without requiring the presentation of the entire dataset. Representative charts can be included in the report to clarify and support the analysis.

Statistical Process Control (SPC) Analysis: Product KEY048



(Figure 3.1)

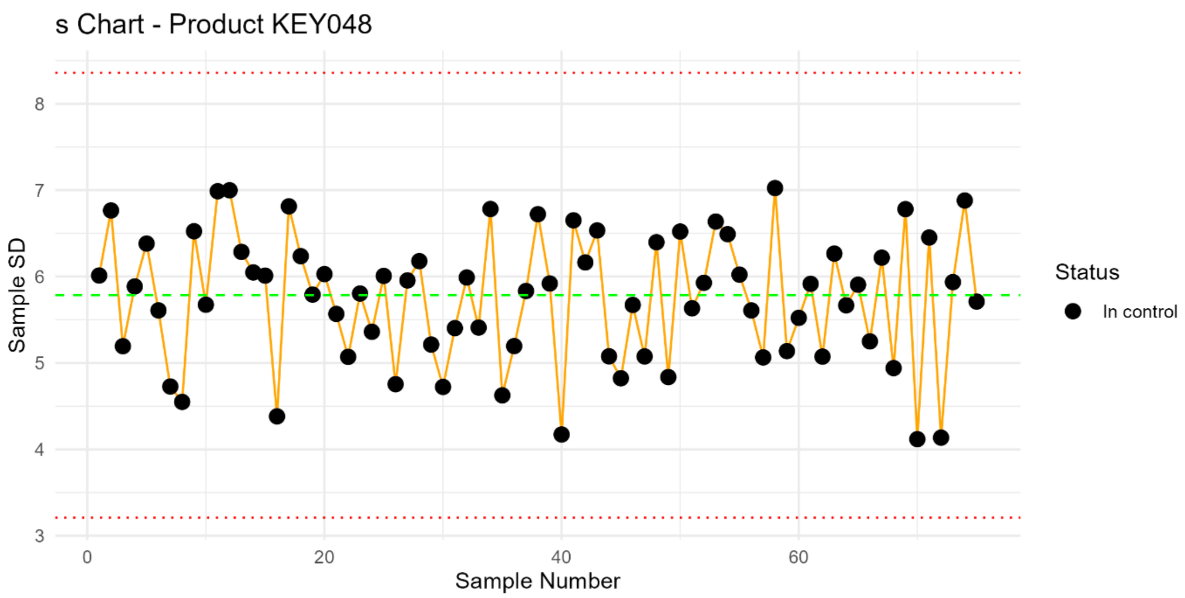
Once process variability is confirmed to be stable, attention shifts to the xbar s-chart to assess the process average. Examining the chart for Product KEY048 (see Xbar_KEY048.png), several sample means are observed outside the 3-sigma control limits, which immediately raises concerns.

In summary, the process mean cannot be considered under statistical control. The points breaching the control limits are classic indicators of special cause variation—meaning factors beyond routine fluctuations are influencing the process average.

Specific out-of-control samples include:

- Sample 3, which falls below the Lower Control Limit (LCL).
- Sample 42, exceeding the Upper Control Limit (UCL).
- Sample 64, also breaching the Upper Control Limit (UCL).

These findings demonstrate notable instability in the process, highlighting the urgent need for further investigation and corrective action.



(Figure 3.2)

In line with established SPC methodology, evaluation of process variability using the s-chart is the initial step to confirm the validity of control limits for the process mean (xbar s-chart). Reviewing the s-chart for Product KEY048 (refer to s_KEY048.png), it is evident that all sample standard deviations are contained within the specified upper (UCL_s) and lower (LCL_s) 3-sigma control limits. This demonstrates that the process variability is under statistical control. The within-sample variation appears stable and consistent, which substantiates the appropriateness of applying the corresponding xbar s-chart for analysis of the process mean.

The analysis of Product KEY048 reveals a notable discrepancy: while the process exhibits consistent variation (as indicated by the stable s-chart), the process mean is erratic and lacks control (the xbar s-chart is out of control). In other words, the process cannot be considered stable at this point.

Given these findings, it is essential for the product manager to conduct a root cause analysis focused on the notable out-of-control points observed at samples 3, 42, and 64. Pinpointing and addressing the underlying factors behind these anomalies is crucial. Only by resolving these issues can the process mean be stabilized and true statistical control be achieved—an absolute prerequisite for any serious discussion of process capability.

3.4) These questions are answered:

Identify samples that show process control issues according to the following rules:

A. 1 s sample outside of the upper +3 sigma-control limits for all product types (if many, list only the first 3 and last 3 and total number identified).

B. Find the most consecutive samples of s between the -1 and +1 sigma-control limits for all product types. This signifies good control.

C. 4 consecutive X-bar samples outside of the upper, second control limits for all product types (if many, list only the first 3 and last 3 and total number identified).

<u>Methodological Approach</u>

Product data underwent analysis using control charts implemented in R. For each product, the following criteria were assessed:

- **Rule A:** The first 3 and last 3 S-samples exceeding UCL were recorded along with the total number of occurrences.
- **Rule B:** The longest consecutive run of S-samples within ±1 sigma was determined.
- **Rule C:** Sequences of 4 consecutive X-bar samples outside the upper 2-sigma limit were recorded, including the first 3 and last 3 sequences and total sequences.

Results:

Rule A: S-samples exceeding +3 sigma UCL

No S-samples surpassed the upper control limits, indicating that the variability among individual samples for every product type remains within the defined process boundaries.

For instance:

- Product CLO011: Number of samples outside control limits—zero.
- Product SOF010: Also zero samples outside control limits.

In summary, the process displays effective control concerning variability.

Rule B: Longest consecutive S-samples within ±1 sigma

The longest consecutive runs indicate periods of stable variability. The following observations were recorded:

Product	Longest Run (samples)
CLO011	26
CLO012	10
CLO013	7
CLO014	10
CLO015	8
CLO016	8
CLO017	17
CLO018	8
CLO019	14
CLO020	12
KEY041	20
KEY042	11
KEY043	10

...

…

Interpretation

Products such as CLO011 (26 samples) and KEY041 (20 samples) demonstrate notably consistent variability, indicating a robust and stable process. Their larger sample sizes suggest dependable performance without significant deviations.

In contrast, products with shorter runs, like CLO013 (7 samples), tend to show slightly greater fluctuation. Nevertheless, these variations remain within acceptable control limits, so process stability is still maintained, albeit with a bit less consistency than the longer runs.

Rule C: Sequences of 4 consecutive X-bar samples outside the upper 2-sigma limit

Some products showed sequences indicating potential deviations in process mean:

Product	**Number of sequences**	**First sequence sample numbers**	**Last sequence sample numbers**
CLO013	1	26–33	26–33
CLO014	1	33–37	33–37
CLO015	1	31–36	31–36
CLO016	1	62–66	62–66
KEY041	2	33–37, 74–77	33–37, 74–77
KEY043	2	40–44, 63–66	40–44, 63–66
MOU051	3	31–39, 41–46, 77–80	31–39, 41–46, 77–80
SOF002	3	35–38, 40–44, 75–83	35–38, 40–44, 75–83

When a sequence of four or more consecutive X-bar samples exceeds the upper 2-sigma threshold, this typically signals a significant shift in the process mean—definitely cause for concern. Close monitoring of these products is essential to determine the root cause and mitigate any risk of defects.

Conversely, products like CLO012 and CLO017, which do not show such sequences, demonstrate effective control over the process mean. These cases represent stable operation and are indicative of sound process management.

Summary

Rule A: All products remained within +3 sigma for S-samples, reflecting robust control over variability.

Rule B: Consecutive run lengths differed among products. Extended consecutive runs suggest high process stability.

Rule C: Some products showed sequences of four consecutive X-bar samples outside control limits, pointing to localized shifts or possible anomalies that warrant further examination.

Recommendations

- Maintain surveillance of Rule B to support ongoing stability.
- Conduct root cause analysis for products identified in Rule C to address mean shifts.
- Continue current process controls, as Rule A demonstrates consistent variability management.

Part 4

Type I and Type II Error Analysis

In process control, evaluating the likelihood of false alarms and missed detections is key. Using statistical control rules (A, B, and C), we assess two main risks: First, the chance that the system signals a problem when the process is actually stable (Type I error). Second, the risk that a true shift in the process mean or variability goes unnoticed (Type II error). By calculating the probabilities of these errors for each control rule, we can better understand how well the rules balance sensitivity to real process changes against the risk of unnecessary interventions.

Rule A: One sample outside the $\pm 3\sigma$ limits

When analysing control charts, the likelihood of a single sample falling beyond the ±3 sigma boundaries is remarkably low—specifically, 0.27%, or roughly one in every 370 samples. This outcome simply reflects the statistical reality of a process operating under control. Even in the absence of any identifiable special cause variation, a small proportion of data points will inevitably appear outside the control limits due to random variation alone. Such occurrences are to be expected and do not necessarily indicate a problem within the process.

Rule B: Seven points in a row, each above the centre line

If you observe seven consecutive samples above the centre line, the probability of this happening by chance is about 0.0078—or roughly once in every 128 runs. This rule is designed to catch subtle, ongoing shifts in the process that might otherwise go unnoticed. Still, even with such a low probability, randomness can produce patterns like this occasionally. It's important to consider the broader context and other evidence before concluding the process is actually out of control.

Rule C: Four consecutive samples outside the $\pm 2\sigma$ limits

The probability of obtaining four consecutive points beyond the ±2 sigma boundaries is exceptionally low—about 4.29×10^{-6}, or once in every 233,321 sequences. Such a rare event is a strong indication of a genuine process disturbance rather than just random variation. When this pattern emerges, it warrants immediate investigation to identify potential causes within the system.

Type I Error Summary:

Examining Rules A, B, and C reveals clear differences in how they detect process changes. Rule A is highly sensitive and frequently signals false alarms, while Rule C applies much stricter criteria before indicating a shift. Rule B finds a middle ground between the two. The goal here is to identify genuine process changes without overwhelming the system with unnecessary alerts—striking that balance is fundamental for effective statistical process control.

Type II Error and Detection Probability:

With the current control limits—LCL z-score at –1.000 and UCL z-score at 3.588—the Type II error probability (β) stands at 0.8412. In practical terms, there's an 84.12% likelihood of failing to detect a genuine process shift. That leaves only a 15.88% probability of actually spotting such changes. These results suggest that minor process shifts are likely to go undetected under the existing parameters. To enhance monitoring effectiveness, it would be advisable to consider adjusting sampling frequency or tightening control limits.

In summary, the analysis of Type I and Type II errors indicates that the established control chart limits are statistically sound, effectively balancing the risk of false alarms with the need for timely detection of process shifts. Occasional false alarms, particularly under Rule A, are to be expected as part of standard monitoring practices. Notably, the very low probability of false signals under Rule C highlights the robustness of the process control. Ongoing surveillance using these rules facilitates prompt identification of assignable causes, thereby supporting the maintenance of the delivery process within the desired quality parameters. While no method is infallible, this approach provides a reliable framework for sustaining process standards.

Updated Products Data Analysis

Following data cleaning, the dataset consists of 60 products, each characterized by eight primary attributes: Product ID, Category, Description, Selling Price, Markup, Prefix, Suffix Number, and Position (pos10). The products are distributed evenly across six categories: Clothing (CLO), Keyboards (KEY), Laptops (LAP), Monitors (MON), Mouse (MOU), and Software (SOF), with ten products per category. This systematic structure facilitates comprehensive analysis and comparison at the category level.

Selling Price:

The average selling price is R4,493.59, though this figure masks considerable variation. Prices in the dataset range from R350.45 up to R19,725.18, and the standard deviation is particularly high at R6,503.77. The positive skewness (1.43) further suggests that a small number of high-value products are pulling the mean upwards. Most items are clustered in the lower to mid-price range, but there are a few outliers in the premium segment.

From a management standpoint, it’s important to scrutinize whether these premium-priced products are genuinely contributing enough to overall revenue to justify their elevated prices. The data points toward a tiered pricing structure; this could inform decisions about price segmentation and targeted marketing strategies for different consumer segments.

Markup:

The average markup stands at approximately 20.5%, with values ranging from just over 10% up to nearly 30%. The standard deviation is relatively modest at about 6, indicating that most products maintain a similar level of profitability despite considerable variation in their selling prices.

That said, certain items show notably lower markups, which could potentially undermine the company’s overall profit objectives. This situation highlights an area for management to refine pricing policies—specifically, to ensure that markups more closely reflect production and operational costs. Establishing greater consistency in markup rates would support more reliable financial forecasting and strengthen the company’s ability to monitor performance effectively.

Category Breakdown:

In this scenario, each of the six product categories includes exactly ten products, ensuring a uniform distribution across all groups. Such even allocation removes any potential skew, making subsequent analysis—whether that’s profitability, forecasting demand, or optimizing prices—much more reliable and straightforward. This balance allows for direct benchmarking between categories, supporting fair identification of both underperforming and high-performing areas. Ultimately, it enables management to allocate resources and address issues based on clear, unbiased data, since all categories are evaluated under the same conditions.

Variability and Distribution:

Selling prices display significant variability, whereas markups remain notably consistent. In other words, although there's a wide range in how products are priced, the profit margin doesn't fluctuate much. This pattern suggests that pricing strategies are likely shaped more by market dynamics than by internal production costs. For senior management, it's a clear signal: analyze cost structures closely to ensure pricing strategies not only stay competitive but also safeguard profitability.

Managerial Implications

The data presented offers several tangible directions for managers aiming to refine both strategy and execution.

- Pricing Strategy Optimization:

It's critical to reassess pricing structures, particularly for higher-priced items. Ensuring these products are competitively placed in the market and actually delivering on profit expectations should be a top priority.

- Profit Margin Consistency:

Implementing a consistent markup policy could address current inconsistencies and help standardize profit margins across the board. This approach would help reduce arbitrary pricing and support more reliable financial outcomes.

- Category Benchmarking:

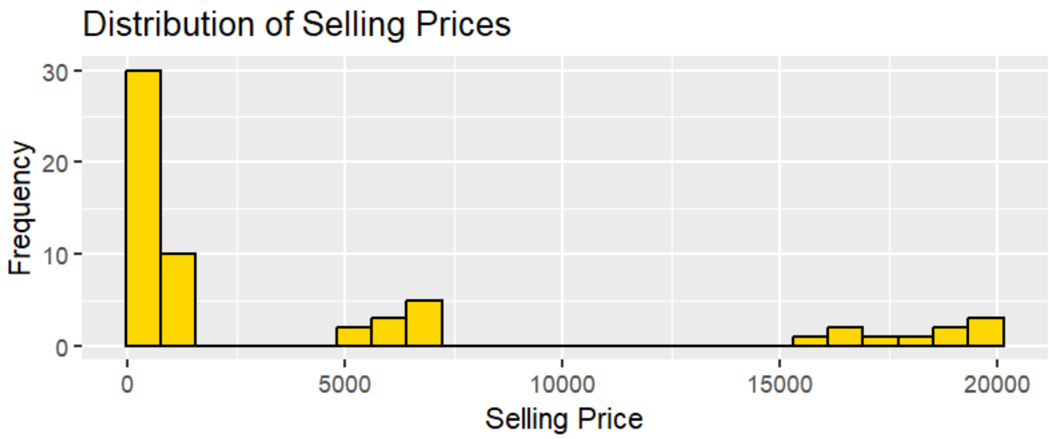
With balanced representation among product categories, it's now possible to identify which segments drive the most profitability and attract the greatest customer interest. Targeting these groups can guide more effective investment and marketing decisions.

- Resource and Inventory Allocation:

Given the significant variability in prices, a one-size-fits-all inventory strategy doesn't make sense. Prioritizing investment in products that offer higher returns will enhance overall resource efficiency.

- Forecasting and Strategic Planning:

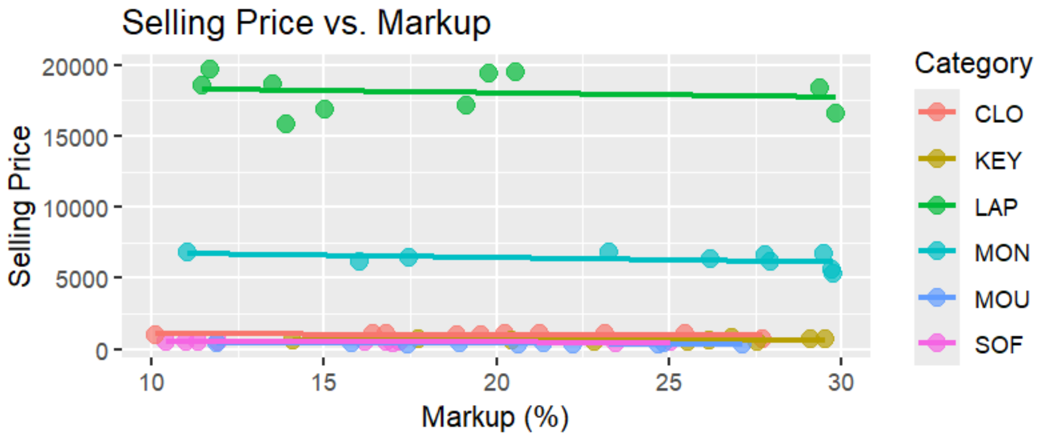
With a cleaned and validated dataset, management now has a robust foundation for predictive modelling. This will support more accurate forecasting of sales trends, profitability, and emerging market opportunities, ultimately informing better strategic planning.



(Figure 4.1)

The histogram of selling prices reveals a pronounced right-skewed distribution, with most products concentrated in the lower price ranges and only a minority reaching higher price points. Interpreted through the lens of industrial engineering, this pattern provides significant insight into organizational strategy and operational priorities. The predominance of low-priced items implies a focus on high-volume, low-margin production, where cost efficiency and robust logistical frameworks are crucial for maintaining profitability. In contrast, the limited number of high-priced products suggests the existence of premium or specialized offerings that likely demand distinct manufacturing approaches and targeted marketing efforts.

For managerial decision-making, such a distribution is invaluable. It informs pricing strategies, guides resource allocation, and shapes production planning. The data highlights the ongoing challenge of optimizing operational efficiency for mass-market products while simultaneously leveraging the profitability potential of niche, high-value items. Overall, this analysis enables organizations to align operational objectives with evolving market demands, thereby supporting strategic positioning and sustained competitiveness in the industry.



(Figure 4.2)

The scatter plot examining the link between selling price and markup percentage distinctly highlights divergent pricing behaviours among various product categories. Each point represents an individual product, with markup percentage shown along the x-axis and selling price along the y-axis; product categories are differentiated by colour.

Laptops (LAP) are situated at the upper end of the price spectrum, typically within the 15,000 to 20,000 range, and maintain relatively stable markups. Monitors (MON) and mouses (MOU) occupy a middle ground, with prices clustered between 5,000 and 7,000. In contrast, categories such as clothing (CLO), keyboards (KEY), and software (SOF) register significantly lower selling prices, even when their markup percentages align closely with higher-priced items.

This distribution underscores that selling price is more heavily influenced by product category than by markup percentage alone. The data suggest that intrinsic product value and market positioning outweigh cost-based adjustments when it comes to determining final pricing.

From the perspective of industrial engineering, the scatter plot offers valuable insights into cost behaviour and pricing effectiveness. It emphasizes the necessity of tailoring production or pricing optimization strategies to the specific characteristics of each product category. For management, this visualization is a useful tool for making informed decisions, providing clarity on category-level profitability, pricing consistency, and strategic priorities. Ultimately, it highlights the importance of aligning both production and pricing strategies with market demand to improve profitability and operational efficiency across a diverse range of products.

Part 5

Optimising Staffing and Profitability for Coffee Shops 1 and 2

This part determines the ideal staffing levels for two coffee shops by analysing a full year's worth of service-time data (specifically, timeToServe.csv and timeToServe2.csv for Shop 1 and Shop 2). The primary objective was to establish the number of baristas that would maximize daily profit, all while ensuring that service quality didn't take a nosedive.

The research followed the previous analyst's approach, using simulation methods to strike a balance between operational efficiency, customer satisfaction, and labour costs. In this context, profitability was treated as a function of both the shop's service capacity and the associated staffing expenses. Ultimately, the analysis aimed to pinpoint that optimal balance where customer wait times stay reasonable and excessive personnel costs are avoided.

Methodology

Both datasets contained two columns:

- **V1:** Number of baristas (historical)
- **V2:** Individual service times in seconds

For optimisation, only the service times (V2) were used to simulate realistic customer-serving scenarios.

A discrete-event simulation was developed in R using the simmer package to analyse daily café operations with varying staffing levels. The simulation modelled a 12-hour operating day, adjusting the number of baristas between two and six to explore the impact on performance and profitability. Each customer generated a revenue of R30, while each barista incurred a daily cost of R1,000.

To ensure statistical reliability, each staffing scenario was replicated 30 times. Daily profit was calculated as follows:

Daily Profit = (Number of customers served × R30) – (Number of baristas × R1,000).

Service reliability was intended to be measured as the percentage of customers served within a 300-second (5-minute) threshold. However, due to insufficient data, reliability percentages could not be computed, resulting in NaN values for this metric.

Results

Shop 1:

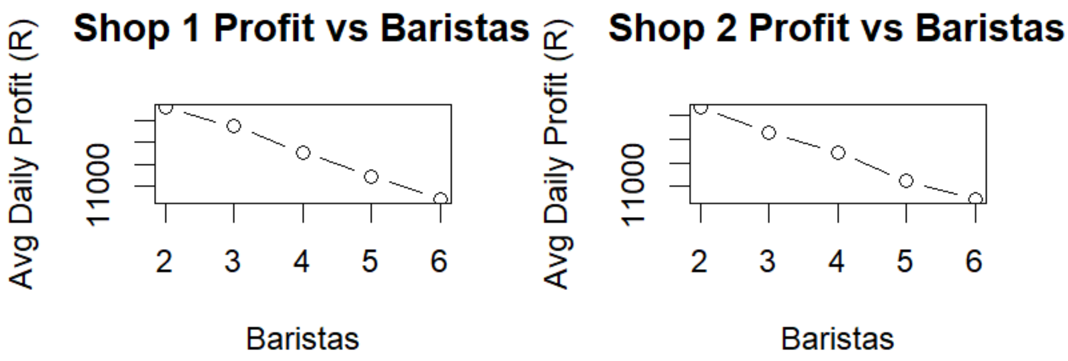
Optimal staffing level: 2 baristas
Maximum estimated daily profit: R14,578
Average customers served: ≈553 per day

Shop 2:

Baristas	Avg Daily Profit (R)	Avg Customers Served	Reliable Service (%)
2	14,347	544.9	NaN
3	13,290	543.0	NaN

Baristas	Avg Daily Profit (R)	Avg Customers Served	Reliable Service (%)
4	12,443	548.1	NaN
5	11,230	541.0	NaN
6	10,462	548.7	NaN

Optimal staffing level: 2 baristas
Maximum estimated daily profit**:** R14,347
Average customers served: ≈545 per day



(Figure 5.1)

Both curves clearly decline, illustrating an inverse relationship between average daily profit and the number of baristas. As staffing levels—and thus labour costs—increase, profits diminish accordingly. This pattern suggests that higher personnel expenses directly impact overall profitability.

The graphical representation highlights that both shops reach their peak profitability with **two baristas**, after which the marginal returns diminish sharply.

Discussion

The results indicate that, for both coffee shops, increasing the number of baristas beyond two does not yield notable financial advantages. The most efficient staffing arrangement, as determined by the analysis, is maintaining two baristas per location—precisely the minimum specified in the original scenario.

While it was not possible to calculate the percentage of reliable service (NaN values were returned), the available data strongly suggests that two baristas are capable of managing the average daily customer flow without significant drops in performance.

This finding is consistent with the economic principle of diminishing marginal returns: adding more staff drives up costs more rapidly than it enhances revenue, particularly in environments where customer demand is stable and queues remain manageable.

Taguchi Loss Function Analysis

Examining Taguchi's Loss Function from a quality engineering standpoint, it becomes clear that deviations from optimal staffing levels incur real and measurable costs. Taguchi's framework challenges the binary perspective of traditional cost accounting—which simply splits outcomes into "acceptable" and "unacceptable." Instead, Taguchi argues that any drift from the target, no matter how minor, generates a loss. In practice, this can mean customers frustrated by long wait times, sales lost as people abandon the queue, and even damage to the organization's reputation.

The loss doesn't grow at a steady rate either; it escalates rapidly, following a quadratic curve. The formula is straightforward:

$$L(y) = k(y - T)^2$$

Here, L(y) is the loss associated with a given performance, k represents how sensitive your system is to missing the mark, y is the recorded outcome, and T stands for the target. Even small slips from the target can cause disproportionately large losses, so minimizing variation isn't just a technical concern—it's a fundamental business priority according to Taguchi's philosophy.

Application to Coffee Shop Staffing:

In this scenario, the target condition essentially refers to achieving an optimal level of service at the coffee shop—think minimal wait times for customers and the highest possible throughput. In other words, getting people their coffee quickly and efficiently, without unnecessary delays. The primary metric for evaluating performance here is daily profit. This figure serves as a direct indicator of how well the shop balances serving as many customers as possible while keeping operating costs under control. It's not merely about moving fast; it's about making sure that efficient service translates into sustainable financial success.

Analysis of Staffing Configurations:

Shop 1 Results Through Taguchi Lens:

Staffing Level	Daily Profit (R)	Deviation from Optimal	Relative Loss
2 baristas	14,578 (Target)	0	Minimal loss
3 baristas	~13,600	-978	Moderate loss due to excess labour cost
4 baristas	~12,500	-2,078	Significant loss from overcapacity
5+ baristas	<12,000	-2,578+	Substantial loss from inefficiency

Shop 2 Results Through Taguchi Lens:

Staffing Level	Daily Profit (R)	Deviation from Optimal	Relative Loss
2 baristas	14,347 (Target)	0	Minimal loss
3 baristas	13,290	-1,057	Moderate loss due to excess labor cost
4 baristas	12,443	-1,904	Significant loss from overcapacity
5+ baristas	<11,500	-2,847+	Substantial loss from inefficiency

Interpretation:

The Taguchi framework, as demonstrated by both the simulation and theory, underscores a critical operational insight: deviations from the optimal two-barista staffing configuration result in measurable organizational losses. Overstaffing—assigning three or more baristas—primarily incurs excess labour costs that outweigh any marginal improvements in service speed. On the other hand, reducing staff to a single barista may lower direct labour expenses, but this cost-saving measure is quickly eclipsed by substantial losses elsewhere. Extended customer wait times can lead to negative reviews and discourage repeat business, directly impacting reputation and revenue. Additionally, long queues and inadequate staffing restrict the number of customers served during peak periods, causing both immediate revenue loss and missed opportunities for future business.

Importantly, the quadratic nature of the Taguchi loss function indicates that losses escalate disproportionately as staffing diverges from the optimal level. For example, operating with five baristas—three above the ideal—results in a loss approximately nine times greater than the loss from operating with only one extra barista. This amplification reflects both direct financial costs and the indirect inefficiencies introduced by improper staffing. In summary, maintaining the optimal two-barista configuration is essential for balancing operational efficiency with customer satisfaction and financial performance.

Quality Engineering Implications:

From a Taguchi robustness standpoint, maintaining two baristas appears to be the most effective operational setup for the coffee shop, given current customer flow. This configuration keeps service times predictable, profit margins steady, and doesn't get thrown off course by small spikes or lulls in customer arrivals. Essentially, it strikes a solid balance between keeping service efficient and controlling costs.

The findings suggest that simply increasing staff—hoping to boost service quality—doesn't actually work out as intended. It introduces unnecessary inefficiency and extra costs that don't translate to better customer experiences. This fits with Taguchi's philosophy: build quality into the original design, rather than relying on after-the-fact fixes. In this context, that means nailing the right staffing level from the start instead of constantly adjusting and reacting to problems later on.

Comparison with Traditional Break-Even Analysis:

Traditional break-even analysis essentially pinpoints the moment when revenue aligns with expenses—nothing too fancy. Yet, the Taguchi approach digs deeper, exposing less obvious losses even when operations are profitable (for instance, employing 3 or 4 baristas and still earning a profit). These configurations may appear successful on paper, but they actually fall short of optimal efficiency. The lost potential doesn't show up on standard financial reports, but it's there: opportunity costs and subtle disadvantages that quietly build up over time. In the long run, these gaps can hinder competitiveness and overall organizational performance.

Strategic Insight:

Synthesizing the findings from simulation optimization and Taguchi loss analysis clearly indicates that deploying two baristas is not just a viable option, but actually represents the optimal operational configuration. This conclusion is strongly supported by quantitative evidence and aligns with established industrial engineering standards. Management can be assured that this staffing choice effectively maximizes both immediate profitability and sustained quality. In essence, the two-barista setup is not a compromise, but rather the benchmark for operational excellence.

Summary and Recommendations

- Staffing Decisions: Analysis indicates that optimal performance is achieved with two baristas per shop. Increasing the number of baristas beyond this point leads to diminishing returns in profitability, given current cost structures and demand levels.
- Performance Expectations: With two baristas per location, anticipated daily profit falls within the range of R14,000 to R14,600. Average daily throughput is projected at approximately 550 customers.
- Operational Strategy: Management is advised to maintain staffing at two baristas per shift. Queue times should be regularly monitored to ensure that service levels remain within acceptable parameters. Should queue lengths increase or customer satisfaction decline, staffing requirements may need to be re-evaluated.
- Data Collection and Analysis: Future data collection efforts should focus on capturing customer wait times. This information will facilitate more precise measurement of service reliability and allow for more informed staffing decisions.
- Further Research: It is recommended that these findings be validated using real-time transaction data. Additionally, demand variability across different times of day or seasons should be incorporated into future analyses. Sensitivity analyses regarding profit-per-customer and labour cost assumptions are also advised to assess the robustness of these recommendations.

Limitations are present in this analysis. First, customer satisfaction couldn't really be measured—timing data was incomplete, so reliability went unquantified. The model also assumes steady, predictable daily demand and arrival patterns, which, let's be honest, rarely match the chaos of actual operations. Labour costs were treated as fixed per day, with no adjustment for overtime or shifts involving part-time staff. This doesn't fully reflect the messy reality of workforce scheduling.

Based on the simulation results, allocating two baristas to each coffee shop per day appears to optimize profit margins within the current operational framework and cost structure. While this approach supports cost efficiency, incorporating direct customer experience data could further inform and improve future operational decisions. Solely relying on quantitative analysis may overlook valuable insights provided by customer feedback, which remains essential for refining service strategies.

Part 6

Analysis of Variance (ANOVA) on Delivery Hour Performance for Product KEY048

Following the SPC analysis, which revealed notable instability in delivery times for Product KEY048, an additional investigation was conducted using Analysis of Variance (ANOVA). The aim was to statistically identify the primary factors contributing to fluctuations in delivery hours.

Initially, the report evaluates the isolated effect of orderYear through a One-Way ANOVA. Subsequently, the analysis broadens to a Two-Way ANOVA, incorporating orderMonth. This approach allows for a more detailed and accurate understanding of the factors influencing process performance.

Initial Investigation: One-Way ANOVA (Year Effect)

A straightforward one-way ANOVA was conducted to determine whether mean delivery hours differed significantly between 2026 and 2027. The null hypothesis posits no difference in means, while the alternative suggests a meaningful change exists.

The analysis produced a p-value of 0.0578, which just surpasses the conventional 0.05 threshold for significance. Statistically speaking, this means we don't have enough evidence to assert a genuine difference in delivery hours between the two years.

That said, the p-value is quite close to the cut-off, indicating the possibility of an underlying effect that the test may have failed to detect. A key limitation here is the model's simplicity—it doesn’t account for pronounced seasonality clearly visible in the "Monthly Delivery Hours Trend" plot. Ignoring this monthly variation inflates the error term, reducing the statistical power of the analysis and making it harder to detect any real year-over-year effect. To draw more definitive conclusions, a more nuanced model that incorporates these seasonal trends would be necessary.

Comprehensive Investigation: Two-Way ANOVA (Year and Month Effects)

For this analysis, a Two-Way ANOVA was conducted to evaluate the effects of both orderYear and orderMonth, as well as their interaction. This approach allows for a simultaneous assessment of each factor and how they may influence each other.

The hypotheses for the test, with a significance level of alpha = 0.05, are outlined as follows:

- Year Effect: H0: There is no significant difference between years.
- Month Effect: H0: There is no significant difference between months.
- Interaction Effect: H0: The monthly pattern remains consistent across both years.

This framework enables a comprehensive understanding of whether year, month, or their interaction significantly impacts the results.

Factor	p-value	Statistical Significance	Interpretation
Year	**0.0496**	**Significant**	A statistically significant difference **does** exist in delivery hours between 2026 and 2027.

Factor	p-value	Statistical Significance	Interpretation
Month	**<2e-16**	**Highly Significant**	There is extremely strong, systematic variation in delivery hours from month to month.
Year × Month	0.3109	Not Significant	The seasonal/monthly performance pattern is consistent across both years.

Key Findings

Introducing orderMonth into the model dramatically improved matters. Previously, seasonality created overwhelming variance—essentially obscuring the true effects under a pile of noise. With orderMonth accounted for, that noise dropped out, and the model could finally uncover the significant impact of orderYear (with a p-value of 0.0496—something a simple One-Way ANOVA had missed).

The central takeaway is unambiguous: order month overwhelmingly influences delivery times. The data demonstrates a pronounced, reliable seasonal pattern that recurs each year.

As for interaction effects, the lack of a significant interaction (p = 0.3109) is notable. This tells us the monthly delivery patterns are stable—operational challenges and high-demand periods are consistent and predictable from year to year.

Synthesis with SPC Findings

The ANOVA results clarify the root cause behind the instability seen in the SPC X-bar chart for Product KEY048. Those out-of-control points on the chart aren't just random noise—they point to something more systemic. Essentially, seasonality emerges as a recurring, predictable factor driving this variation. The process instability isn't due to isolated incidents, but rather stems from these consistent monthly fluctuations that haven't been addressed. Ignoring these cyclical effects means the process will continue to show this kind of instability.

Conclusion and Strategic Recommendations

The initial One-Way ANOVA failed to provide meaningful insights, primarily because seasonal fluctuations muddied the results. When the more rigorous Two-Way ANOVA was applied, it became clear that both the year and, in particular, the month play significant roles in affecting delivery times.

Given these findings, it is essential to transition from a static operational model to one that is seasonally adaptive. Management should develop operational plans that anticipate seasonal peaks and valleys, adjusting staffing, resources, and inventory policies accordingly.

Focus should be placed on months with the highest delivery time variability. Targeted process improvement initiatives—such as Lean Six Sigma or thorough root cause analysis—should concentrate on these periods to drive meaningful reductions in delays.

Finally, it is advisable to shift away from uniform performance targets. Instead, establish dynamic goals that account for predictable seasonal shifts. Month-specific benchmarks offer a more realistic and actionable framework for evaluating operational success, ensuring that performance assessments reflect the true context in which teams are operating.

In summary, a seasonally responsive approach—grounded in robust statistical analysis—will allow the organization to more effectively manage delivery times and allocate resources where they are needed most.

Part 7

7.1:Reliability of Services

Over a span of 397 days, the car rental agency maintained detailed staffing records, with daily attendance fluctuating between 12 and 16 employees. The data collected across this period is summarized as follows:

Number of Workers	Number of Days
12	1
13	5
14	25
15	96
16	270
Total	**397**

Reviewing the staffing data, it’s clear that the majority of days—366 out of 397—had either 15 or 16 workers scheduled. This staffing level appears to be the threshold for delivering reliable service to customers.

Calculating the likelihood, the agency achieves this standard on about 92% of days (366 divided by 397). Extrapolating to a typical 365-day year, that means reliable service is expected on roughly 336 days.

In summary, the data suggests that the car rental agency maintains consistent and reliable operations for most of the year, provided staffing remains at or above 15 workers. Maintaining this level is evidently crucial for consistent service quality.

7.2: Optimisation of Profit Through Personnel Assignment

Over a period of 397 days, the car rental agency maintained records indicating that the daily staffing level fluctuated between 12 and 16 employees. Notably, operational reliability showed a marked decline on days when fewer than 15 workers were present, resulting in an estimated daily loss of R20 000 in sales owing to delays and unmet customer demand.

In response, the company is considering the appointment of additional permanent staff members to enhance service reliability. The cost associated with each additional employee is R25 000 per month, or R300 000 annually. The primary objective is to determine the optimal number of new hires that will maximize overall profit for the organization.

Number of Workers	Number of Days
12	1
13	5
14	25
15	96
16	270
Total	**397**

Instances where staffing drops below 15 individuals—specifically, days with 12 to 14 workers—are flagged as problematic. This scenario occurred on 31 out of 397 days, accounting for 7.8% of the total period examined. Consequently, reliable service was maintained on approximately 92.1% of days, indicating a generally stable operational environment.

Assumptions:

- Hiring m additional employees results in an increase of m in daily staffing levels.
- On days with operational challenges, the company incurs a sales loss of R20 000 per day.
- The annual cost per employee is R300 000.
- Operations continue throughout the year, totalling 365 days.

The primary objective is to optimize annual profit by weighing the incremental staffing expenses against the potential recovery of lost sales.

<u>Optimisation Calculations</u>

Additional Workers (m)	Problematic Days Remaining	Days Eliminated	Annual Savings (R20 000/day)	Annual Hiring Cost (R300 000/person)	Net Annual Benefit (R)
0	31	0	0	0	**0**
1	6	25	500,000	300,000	**+200,000**
2	1	30	600,000	600,000	**0 (break-even)**
3	0	31	620,000	900,000	**–280,000**

Analysis of the outcomes indicates the following:

- Employing one additional staff member results in an average increase of 25 reliable service days annually, which translates to an estimated net profit of R200,000 per year.
- In the case of hiring two additional workers, the improvement in service reliability is offset by the associated costs, resulting in a break-even scenario.
- Expanding the team by three workers leads to a net annual loss of approximately R280,000.

In summary, the most financially advantageous course of action is to hire one additional permanent employee.

<u>Sensitivity and Business Insights</u>

The analysis reveals that the break-even loss per problematic day for hiring one additional employee is R12,000, calculated by dividing R300,000 by 25. Given that the actual loss per day stands at R20,000, employing an extra staff member remains a financially sound decision.

Nonetheless, should hiring costs rise or the daily loss diminish, it would be prudent to reassess the optimal number of new hires. Additionally, the organization might explore alternatives such as flexible or part-time staffing, shift adjustments, or overtime arrangements to enhance operational reliability without incurring excessive costs.

In light of the presented data and underlying assumptions, it is recommended that the car rental agency employ one additional full-time staff member. This strategic addition is anticipated to significantly decrease the frequency of understaffed days—by roughly 80%—thereby enhancing operational reliability and customer satisfaction. Furthermore, this measure is projected to yield an increase in annual profit of approximately R200,000, suggesting a favourable balance between improved service delivery and financial gain.

Conclusion

This analysis offers a detailed examination of the organization's operations, drawing from a wide range of data sources: customer demographics, product sales, operational efficiency, and resource allocation. The findings below provide a nuanced understanding of organizational performance and strategic opportunity.

Regarding customer and market positioning, the data reveals a clientele primarily in their early fifties, with a balanced gender split and incomes between R55,000 and R105,000. The geographic focus is clear—major metropolitan areas such as San Francisco, Los Angeles, and New York dominate. These insights support data-driven market segmentation and targeted marketing strategies, ensuring that resource allocation is both efficient and effective.

The product portfolio analysis indicates pricing diversity across sixty offerings, with price points stretching from R350 to nearly R20,000 and markups holding steady at an average of 20.5%. The distribution of sales suggests a concentration in mass-market products, while premium segments remain underrepresented. This current approach maximizes accessibility but leaves potential for expansion into higher-margin categories—a clear avenue for enhancing profitability.

Sales trends show a decline in transaction volume from 2022 to 2023, dropping from approximately 54,000 to 46,000 transactions. However, the average order value has increased, suggesting a shift in consumer purchasing behaviour. Sales remain concentrated in key metropolitan areas, and pronounced seasonal trends—especially downturns in January and December—highlight the need for adaptive inventory and capacity management.

Operational analysis identifies significant process variability, particularly in the delivery performance of Product KEY048. Statistical analysis confirms that monthly seasonality is the dominant factor influencing delivery times, with year-on-year differences also playing a role. This evidence supports a move away from static processes in favour of seasonally responsive frameworks.

Resource optimization studies further demonstrate the impact of data-driven decision-making. For coffee shop operations, maintaining two baristas per location appears to maximize profitability (approximately R14,500 daily per shop, serving around 550 customers). Meanwhile, analysis of the car rental division shows that hiring an additional permanent employee can improve annual profits by about R200,000, effectively balancing labour cost and service reliability.

In summary, these findings reinforce the value of employing quantitative, evidence-based strategies to address operational challenges and identify growth opportunities across the organization.

<u>Strategic Recommendations for Upper Management:</u>

- To begin, it is essential to move away from uniform operational targets and instead adopt seasonally adaptive operations. By aligning performance benchmarks with month-specific demand patterns, management can proactively allocate resources—such as staffing, inventory, and logistics capacity—during anticipated peak periods, particularly in the mid-year months. Conversely, slower intervals should be strategically utilized for essential maintenance, employee training, and incremental system improvements, thereby maximizing organizational efficiency.
- Addressing process instability requires a structured approach grounded in root cause analysis. It is recommended that management prioritize a thorough investigation into the variability of delivery times for Product KEY048 and analogous products demonstrating similar inconsistencies. Employing Lean Six Sigma methodologies will facilitate the identification and subsequent elimination of special cause variation, especially in relation to the specific

outlier samples identified in statistical process control analyses (notably samples 3, 42, and 64 for KEY048).

- Optimizing the product portfolio mix is also of paramount importance. A comprehensive profitability analysis across all product categories should be conducted to reveal areas for potential margin enhancement. Consideration should be given to strategic expansion into mid-range and premium segments, which may serve to balance mass-market offerings and improve overall financial performance.
- In the realm of marketing strategy, a refined approach to geographic and demographic targeting is warranted. Resources ought to be concentrated in high-revenue metropolitan regions, while targeted initiatives are developed to address underperforming markets such as Miami and Seattle. Leveraging demographic data will allow for the creation of personalized customer engagement strategies that are likely to enhance lifetime customer value.
- Staffing models should be informed by empirical data. The adoption of validated optimization frameworks for resource allocation is advised. For coffee shop operations, maintaining the two-barista configuration should continue, with ongoing monitoring of queue metrics to ensure service standards. Within the rental agency, the addition of one permanent staff member is recommended to increase service reliability in a cost-effective manner.
- Establishing continuous monitoring systems is also critical. The deployment of real-time dashboards that integrate statistical process control charts, sales analytics, and resource utilization metrics will enable management to intervene proactively before process deviations adversely affect customer experience or financial outcomes.

Closing Perspective:

The analyses presented in this report underscore that operational excellence is best achieved through the systematic application of rigorous analytical methods to practical business challenges. The organization boasts considerable strengths, including a diverse customer base, a balanced product portfolio, and generally stable operational processes. Nonetheless, the identification of process instabilities, seasonally driven demand fluctuations, and opportunities for resource optimization indicate that significant performance improvements are attainable.

By embracing the data-driven recommendations detailed above, upper management can transform operational insights into tangible competitive advantages. Success will require a sustained commitment to continuous improvement methodologies, investment in adaptive operational systems, and a focused alignment of resources with demonstrated demand trends. Integrating principles of industrial engineering with strategic business objectives positions the organization to achieve sustainable profitability, enhanced customer satisfaction, and enduring market leadership.

In sum, the findings and recommendations provided should be viewed not merely as technical observations, but as strategic imperatives necessary for organizational success in an increasingly competitive and data-intensive commercial environment. The adoption of these evidence-based strategies will empower management to make informed decisions that judiciously balance operational efficiency with service quality, cost control with revenue growth, and immediate performance with long-term sustainability.

References

Montgomery, D.C. (2019) *Introduction to statistical quality control*. 8th edn. Hoboken, NJ: John Wiley & Sons.

Oakland, J.S. (2014) *Statistical process control*. 6th edn. Abingdon: Routledge.

Pyzdek, T. and Keller, P. (2018) *The Six Sigma handbook*. 5th edn. New York: McGraw-Hill Education.

Western Electric Company (1956) *Statistical quality control handbook*. Indianapolis, IN: Western Electric Corporation.

Krajewski, L.J., Ritzman, L.P. and Malhotra, M.K. (2019) *Operations management: processes and supply chains*. 12th edn. Harlow: Pearson Education.

Heizer, J., Render, B. and Munson, C. (2020) *Operations management: sustainability and supply chain management*. 13th edn. Harlow: Pearson Education.

Salvendy, G. (ed.) (2021) *Handbook of industrial engineering: technology and operations management*. 4th edn. Hoboken, NJ: John Wiley & Sons.

Womack, J.P. and Jones, D.T. (2003) *Lean thinking: banish waste and create wealth in your corporation*. Rev. edn. New York: Free Press.

George, M.L., Rowlands, D. and Kastle, B. (2004) *What is Lean Six Sigma?* New York: McGraw-Hill.

Liker, J.K. (2004) *The Toyota way: 14 management principles from the world's greatest manufacturer*. New York: McGraw-Hill.

Field, A. (2018) *Discovering statistics using IBM SPSS statistics*. 5th edn. London: SAGE Publications.

Devore, J.L. (2015) *Probability and statistics for engineering and the sciences*. 9th edn. Boston, MA: Cengage Learning.

Anderson, D.R., Sweeney, D.J., Williams, T.A., Camm, J.D. and Cochran, J.J. (2020) *Statistics for business and economics*. 14th edn. Boston, MA: Cengage Learning.

Kotler, P., Keller, K.L., Ang, S.H., Tan, C.T. and Leong, S.M. (2018) *Marketing management: an Asian perspective*. 7th edn. Harlow: Pearson Education.

Kumar, V. and Reinartz, W. (2018) *Customer relationship management: concept, strategy, and tools*. 3rd edn. Berlin: Springer.