

TCS's Data Visualization Virtual Internship Report

Internship Program: TCS Virtual Internship Program - Data Visualization:
Empowering Business with Business Insights via Forage

Role: Data Consultant Intern

Completion Date: June 30, 2025

This report outlines the scope, methodology, findings, and conclusions of a Data Visualization Virtual Internship focused on providing actionable insights to an online retail store's CEO and CMO. The primary objective was to analyze transactional sales data to identify key revenue drivers, customer behavior patterns, and market opportunities to inform strategic planning and business expansion. The internship involved rigorous data cleaning, strategic problem-solving for executive-level questions, and the development of impactful data visualizations using Power BI. Key accomplishments include enhanced data quality, the provision of actionable business insights, and the application of advanced data analytics techniques.

1. Introduction and Business Problem:

The online retail store engaged a data consultant to review its extensive dataset and provide strategic insights for the CEO and CMO. The core business problem revolved around understanding major contributing factors to revenue to facilitate strategic planning for the upcoming year and guide business expansion. The leadership sought metrics from both operational and marketing perspectives, with a particular interest in identifying well-performing areas and demographic-based insights for expansion.

Key Objectives for the Report:

- Evaluate current business performance.
- Suggest metrics for expansion decisions.
- Address specific questions from the CEO (revenue trends, regional performance, customer value) and CMO (customer acquisition, repeat orders, product demand).

2. Data Cleaning and Preparation:

A critical phase of the internship involved meticulous data cleaning and preparation, acknowledging that robust insights hinge on clean data. The initial analysis of the online retail dataset (541,910 rows, 8 columns) revealed significant discrepancies.

Initial Data State:

- Total Rows: 541,910
- Total Columns: 8
- Missing Values:

- Description: 1,454 blank rows
- CustomerID: 135,080 blank rows

Data Transformation Steps (Executed using Power Query in Power BI):

1. InvoiceNo and StockCode:
 - Transformation: Data type changed to TEXT to accommodate alphanumeric values.
 - Cleaning: Values trimmed to remove leading/trailing whitespace.
2. Description:
 - Transformation: Data format changed to lowercase for standardization.
3. Quantity:
 - Validation: Observed negative values (indicating returns) and zero values.
 - Cleaning: Rows where Quantity was less than 1 were removed to ensure all sales data represented actual positive unit sales. This was crucial for accurate revenue and quantity calculations.
4. InvoiceDate and InvoiceTime:
 - Derivation: Split from the original InvoiceNo column (which contained both date and time information).
 - Formatting: InvoiceDate formatted as "YYYY-MM-DD", InvoiceTime as "HH:MM:SS" for better temporal analysis.
5. UnitPrice:
 - Validation: Identified rows where UnitPrice was less than \$0 (erroneous entries or credits).
 - Cleaning: Rows where UnitPrice was less than \$0 were removed to prevent skewed revenue calculations.
6. CustomerID:
 - Handling Missing Values: This column had the most significant number of missing values. To preserve neighboring valuable data, blank CustomerID entries were first trimmed and then filled with "No CustomerID".
 - Transformation: Data type changed to TEXT anticipating potential alphanumeric characters.
7. Country:
 - Validation: No missing data or observed discrepancies.
 - Format: Maintained its PROPER case format (e.g., "United Kingdom").

This rigorous data preparation phase ensured that all subsequent calculations and visualizations were based on a clean, reliable, and accurate dataset, adhering to the principle of "clean up before attempting any question."

3. Key Business Questions and Analytical Approach:

The analysis was driven by specific questions tailored for the CEO and CMO, aiming to provide actionable insights for strategic decision-making.

3.1 Questions of Interest to the CEO

1. Regional Revenue Performance:

- Question: "Which region is generating the highest revenue, and which region is generating the lowest?"
- Importance: Crucial for understanding fundamental income sources and identifying areas for strategic focus (maximizing high-performing regions, improving low-performing ones).

2. Monthly Revenue Trend & Seasonality:

- Question: "What is the monthly trend of revenue, which months have faced the biggest increase/decrease? Which months generated the most revenue? Is there a seasonality in sales?"
- Importance: Provides insights into revenue fluctuations, impact of internal changes (e.g., new product launches), and identification of seasonal peaks for strategic planning and forecasting.

3. Top Customers and Revenue Contribution:

- Question: "Who are the top customers and how much do they contribute to the total revenue? Is the business dependent on these customers or is the customer base diversified?"
- Importance: Helps identify main revenue drivers, enables targeted strategies for high-value customers, and assesses customer base diversification to mitigate risks.

3.2 Questions of Interest to the CMO

1. Repeat Customer Percentage & Order Patterns:

- Question: "What is the percentage of customers who are repeating their orders? Are they ordering the same products or different?"
- Importance: Uncovers customer loyalty trends, informs strategies to target repeat customers with relevant offers, and identifies product demand trends.

2. Repeat Customer Order Frequency:

- Question: "For the repeat customers, how long does it take for them to place the next order after being delivered the previous one?"
- Importance: Determines customer re-ordering frequency, allowing for timely marketing interventions and incentivization strategies.

3. Revenue from Repeat Customers:

- Question: "What revenue is being generated from the customers who have ordered more than once?"
- Importance: Quantifies the financial impact of loyal customers, guiding marketing efforts to encourage continued high spending.

4. Most Frequent Repeat Customers & Revenue Contribution:

- Question: "Who are the customers that have repeated the most? How much are they contributing to revenue?"
- Importance: Identifies highly frequent buyers, differentiates between high-volume/low-value and low-volume/high-value repeat customers, and informs inventory and discount strategies.

4. Visualization Approach and Accomplishments

Power BI was chosen for its robust data modeling and interactive visualization capabilities. The approach emphasized clarity, stakeholder alignment, and actionable insights.

Core Measures Created:

- Total Revenue: `SUMX('online_retail', 'online_retail'[UnitPrice] * 'online_retail'[Quantity])`
- Total Quantity Sold: `SUM('online_retail'[Quantity])`

Visualization Principles:

- One Visual Per Tab: Each question's answer was presented on a dedicated page/tab.
- Tab Naming Convention: Explicitly named (e.g., "Question 1").
- Stakeholder-Centric Design: Visuals selected and configured to directly answer executive inquiries.

4.1 Key Visuals Developed and Insights Delivered

1. CEO - 2011 Monthly Revenue Trend:

- Visual: Line Chart.
- Insight: Clearly highlighted peak revenue months (e.g., November 2011), demonstrating seasonal patterns crucial for forecasting. (Resolution of month sorting issue ensures chronological accuracy).

2. CMO - Top 10 Countries by Revenue & Quantity (excluding UK):

- Visual: Clustered Column Chart.
- Insight: Showcased top international markets by both revenue and quantity simultaneously, providing a focused view for expansion strategies by explicitly excluding the domestic UK market.

3. CMO - Top 10 Customers by Revenue (descending, exclude blanks):

- Visual: Bar Chart (horizontal bars).
- Insight: Delivered a precise, ranked list of top customers by revenue, enabling the CMO to prioritize retention efforts on the most valuable client segments.

4. CEO - Demand by Region (Map Chart, exclude UK, no scroll/hover for core insight):

- Visual: Bubble Map.
- Insight: Provided an immediate and intuitive understanding of global demand hotspots (excluding the UK) based on Total Quantity Sold, allowing the CEO to quickly identify potential expansion territories without detailed numerical review.

5. Additional Analytical Findings and Strategic Implications:

Beyond the core questions, further analyses provided deeper insights into customer behavior and product performance.

5.1 Customer Segmentation (RFM Analysis)

- **Methodology:** Utilized Recency, Frequency, and Monetary (RFM) analysis to segment customers based on purchasing behavior.
- **Key Segments Identified:**
 - 'Best Customers' & 'Big Spenders': Consistently contribute most to sales revenue.
 - 'Look out buyers': A significant segment with potential for re-engagement or nurturing.
 - 'Potential to become Best Customers': Exhibit good frequency and monetary values, prime targets for engagement.
- **Strategic Implication:** Focusing on nurturing 'Potential to become Best Customers' and re-engaging 'Look out buyers' represents significant opportunities to increase customer lifetime value and overall sales.

5.2 Purchase Frequency Analysis

- **Observation:** A steep decline in customer numbers after the first few purchases (1 to 5). A smaller, significant number make 5-10 purchases, with very few making more than 20.
- **Strategic Implication:** The vast majority of customers are infrequent purchasers. This highlights a substantial opportunity to implement strategies (e.g., loyalty programs, personalized recommendations, post-purchase follow-ups) to encourage repeat purchases and significantly boost revenue.

5.3 Sales Performance by Country

- **Top Performers (Sales Quantity & Revenue):** Netherlands, EIRE, Germany, France, Australia.
- **Bottom Performers (Sales Quantity):** Brazil, Czech Republic, Bahrain, USA, Saudi Arabia.
- **Revenue Distribution:** Box plot showed relatively low median sales revenue for most countries, but significant outliers indicating high-value transactions in certain regions.
- **Strategic Implication:**
 - **Strengthen Existing High-Performing Markets:** Double down on marketing, optimize logistics, and explore product range expansion in top countries.
 - **Investigate Underperforming Markets:** Conduct deeper dives to understand reasons for low performance (market size, competition, cultural differences, marketing gaps) to potentially reveal new growth avenues.
 - **Analyze Trends:** Understand monthly trends in top countries for inventory optimization and promotional planning.

5.4 Product Performance

- **High-Value Products:** Products like '22423' and 'DOT' generate significant revenue with high quantities, indicating strong demand. 'POST' emerged as a top-performing product by sales revenue.
- **Pareto Principle Application:** A small number of countries contribute a large percentage of total sales revenue, reinforcing the importance of top-performing markets.
- **Strategic Implication:**
 - **Leverage Top Products:** Prioritize these products in marketing, ensure consistent stock, and explore cross-selling/upselling opportunities (e.g., premium versions, bundles).
 - **Strategic Review of Underperforming Products:** Analyze reasons for low performance in bottom 10 products to inform decisions on discontinuation, re-evaluation, or targeted marketing.

6. Conclusion and Future Recommendations:

This virtual internship provided invaluable practical experience in data analysis, visualization, and strategic problem-solving. By transforming raw data into clear, actionable insights, the project demonstrated the power of data in driving informed business decisions and identifying lucrative opportunities for an online retail store.

Key Achievements:

- **Robust Data Quality:** Successfully implemented critical data cleaning steps, ensuring the reliability of all analyses.

- **Actionable Insights:** Delivered three distinct, clear, and actionable visuals directly addressing key strategic and marketing questions.
- **Improved Reporting Efficiency:** Structured the report for improved navigation and stakeholder comprehension.
- **Technical Proficiency:** Applied advanced Power BI features (DAX measures, Power Query transformations, diverse filtering, custom sorting).
- **Foundation for Expansion:** Provided crucial intelligence on sales trends, customer behavior, and geographical demand to guide planned business expansion strategies.

Recommendations for Future Growth: Based on the findings, the following areas are recommended for strategic focus:

1. **Customer Lifetime Value Enhancement:** Implement targeted campaigns to convert 'Potential to become Best Customers' and re-engage 'Look out buyers' through personalized offers and loyalty programs.
2. **Increased Purchase Frequency Initiatives:** Develop strategies (e.g., subscription models, personalized product recommendations based on past purchases, post-purchase follow-ups) to encourage repeat purchases from the large segment of infrequent buyers.
3. **Geographic Market Optimization:**
 - **Deepen Penetration in High-Performing Regions:** Invest further in marketing, localized promotions, and optimized logistics in countries like EIRE, Netherlands, and Germany.
 - **Strategic Evaluation of Underperforming Regions:** Conduct detailed market research into countries with low sales to understand barriers and identify specific, viable growth avenues.
4. **Product Portfolio Strategy:**
 - **Maximize Top Products:** Prioritize marketing and stock for top-performing products, exploring premium versions or bundles.
 - **Review Underperformers:** Analyze the profitability and demand for bottom-performing products to determine whether to discontinue, re-market, or innovate.

By focusing on these areas, the online retail store can strategically allocate resources, optimize marketing efforts, and ultimately drive significant growth and achieve its expansion goals.

Author: Himanshu Kumar

Gmail: ds.himanshu.kumar@gmail.com