

**Sales analysis report**

Sales Performance & Customer Insights



dATA ANALYTICS

GODABA VANDANA

**OVERVIEW**

This project focuses on deriving actionable insights from customer and sales data as part of a Business Analyst Internship assignment for Inkarp Instruments. The primary objective is to clean and validate the provided dataset—ensuring data quality by correcting invalid email formats, handling missing values, removing duplicates, and standardizing phone numbers and zip codes according to Indian formatting standards. Once cleaned, the data is analyzed to uncover patterns in customer distribution across regions, states, and industries, calculate average purchases by customer segments, and identify top-performing products by revenue. Optional analysis explores cross-selling opportunities to boost sales further. The findings are presented using impactful visualizations such as bar charts and heatmaps, supporting a concise summary report and presentation for submission.

The Process

**1.Data Acquisition from Inkarp:**

The dataset was initially acquired from a provided source file, which served as the base for the analysis. The data was downloaded and imported into Excel for further cleaning, transformation, and insights generation..

**2. Data Transformation and Enhancement:**  
Data transformation involved using Excel formulas (like IF, TEXT, TRIM, LEN, etc.) and Power Query to clean and restructure the dataset. This included correcting email formats, standardizing phone numbers, removing duplicates, handling missing values, and ensuring consistency in zip code and address formatting, thus preparing the data for accurate analysis.

**3. Connecting with Analytical Tools:**  
Excel acted as the primary analytical tool. Using Power Query for advanced data shaping and transformation, and Excel’s built-in functions and pivot tables, the data was connected and analyzed to derive insights. No external tools were used, but Excel's capabilities were maximized for professional-level analysis.

**4. Problem Statement Solution in Excel:**All tasks mentioned in the assignment—like customer segmentation by region and industry, calculating average purchases, and identifying top-performing products—were addressed directly within Excel using formulas, filters, pivot tables, and calculated fields. Each solution was placed in dedicated Excel sheets with clear headings and summaries.

**5.Exploratory Data Analysis (EDA):**  
Exploratory Data Analysis was conducted by summarizing data through pivot tables, charts, and aggregations. Patterns in customer behavior, product sales, and geographic distribution were explored. EDA helped identify trends and outliers, supporting decisions like which customer segments or regions contribute most to sales.

**6.Creation of Visual and Insightful PowerPoint:**  
Charts and visuals created in Excel (bar charts, pie charts, heatmaps, etc.) were exported or embedded into a PowerPoint presentation. These visuals supported key findings and provided an easy-to-understand summary for stakeholders, making the data story clear and actionable.

**7.Detailed Documentation:**  
A detailed write-up accompanied the final deliverables, explaining the steps taken—data cleaning, analysis logic, tools used, and insights drawn. The documentation ensured that reviewers could understand your workflow, interpret the analysis, and appreciate the results even without interacting directly with the Excel file.

**Objective:**

To clean, analyze, and visualize customer and sales data to uncover meaningful business insights for Inkarp Instruments. The primary goal is to ensure data quality and use it to support data-driven decisions related to customer segmentation, sales performance, and product strategy.

**Project Scope & Key Tasks:**

**1. Data Cleaning & Validation**

* Remove duplicates, fix invalid email formats, and handle missing values.
* Standardize phone numbers and zip codes using Power Query and Excel formulas.

**2. Data Analysis**

* Segment customers by region, state, and industry.
* Calculate average total purchases per customer segment.
* Identify top-performing products based on revenue.

**3. Exploratory Data Analysis (EDA)**

* Use pivot tables, Excel formulas, and charts to explore patterns in customer behavior and sales trends.

**4. Visualization & Reporting**

* Create charts (bar, pie, etc.) directly in Excel to represent insights.
* Summarize findings in a PowerPoint presentation using visualizations.

**5. Optional (Advanced Analysis)**

* Explore cross-selling opportunities based on purchase behavior and product associations.

**6. Documentation**

* Provide a detailed report summarizing the approach, methodologies used, key findings, and visual insights.

**Success Metrics:**

* **Data Quality Improvements:** Percentage reduction in duplicates, invalid formats, and missing values**.**
* **Insight Accuracy:** Valid and meaningful segmentation and product analysis that aligns with business logic.
* **Visualization Clarity:** Use of clear, readable charts and tables to communicate findings effectively.
* **Task Completion:** All assignment deliverables (cleaned dataset, visualizations, report) are submitted accurately and on time.

**Significance of Sales and Customer Data:**

The **Sales and Customer Data** provided in this project is highly significant for understanding the core business operations and customer dynamics of Inkarp Instruments. It serves as a foundational asset for multiple business functions, including marketing, sales, and strategic planning. Here are the key reasons why this dataset is important:

**Customer Insight & Segmentation:**  
The dataset allows the business to analyze customer demographics such as location (state, region) and industry. This helps in identifying high-value customer segments, tailoring marketing strategies, and improving customer targeting efforts.

**Sales Performance Monitoring:**  
By tracking total purchases and product performance, the dataset helps measure how different products and segments contribute to overall revenue. This is essential for inventory management, product development, and strategic sales planning.

**Data Quality Assessment:**  
The presence of invalid formats, missing values, and inconsistencies highlights the need for robust data cleaning processes. Improving data quality not only ensures better analysis but also prepares the organization for more advanced tools like CRM systems or predictive analytics.

**Cross-Selling Opportunities:**  
With enough depth, the dataset can reveal patterns of customer purchasing behavior that point toward potential cross-sell or upsell opportunities, supporting revenue growth.

**Strategic Decision Support:**  
The cleaned and analyzed data becomes a reliable basis for informed decision-making, whether it’s launching a new product, expanding into new regions, or adjusting pricing strategies.

**Data Dictionary:**

**Customer\_ID**: A unique identifier assigned to each customer to differentiate them in the dataset.

**Customer\_Name**: The name of the customer or the organization making the purchase.

**Email**: The customer’s email address, which is used for communication and marketing; validated for proper format.

**Phone\_Number**: The contact number of the customer, standardized to follow thecorrect Indian format (10 digits starting with 6–9).

**Address**: The detailed street address of the customer, which may include building name, street, and locality.

**City**: The city where the customer is located.

**State**: The Indian state in which the customer resides or operates.

**Region**: The geographical region of the customer, such as North, South, East, or West India.

**Zip\_Code**: The postal PIN code of the customer's address, standardized for consistency.

**Industry**: The type of industry or sector the customer belongs to, such as Healthcare, Education, or Research.

**Total\_Purchase**: The total monetary value of all purchases made by the customer over a period of time.

**Product\_Name**: The specific product purchased by the customer.

**Product\_Category**: The broader category or type that the product belongs to, e.g., Lab Equipment or Medical Devices.

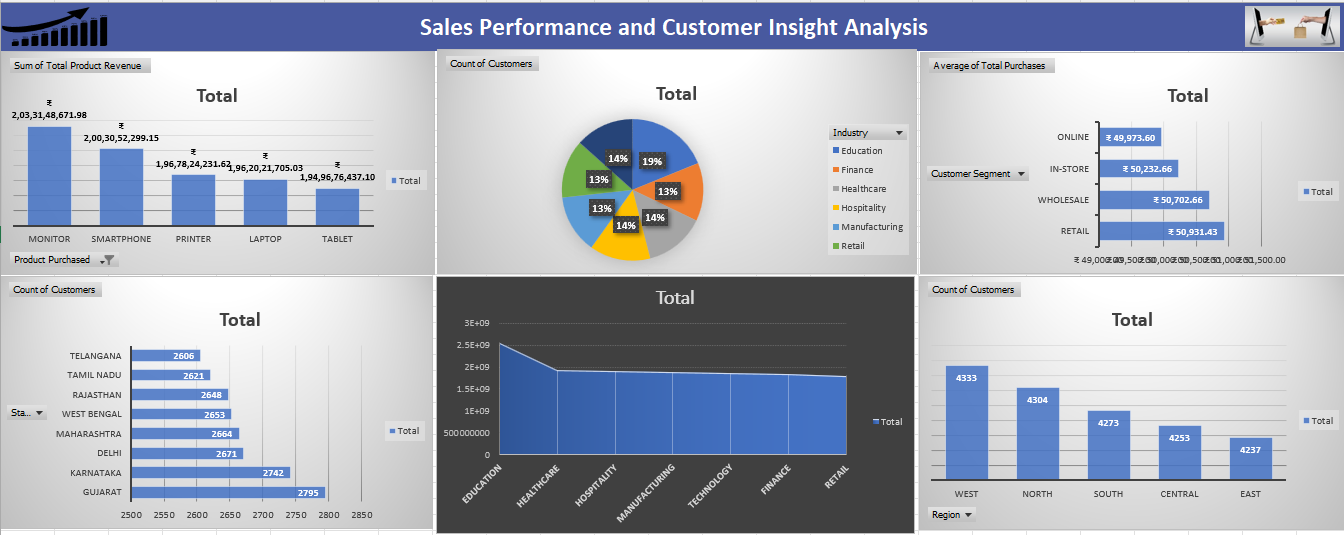
**Purchase\_Date**: The date when the purchase transaction occurred.

**Quantity\_Purchased**: The number of units of the product bought by the customer in a given transaction.

**Revenue**: The total amount of revenue generated from the sale, typically calculated as Quantity × Unit Price.

This data dictionary provides a clear understanding of the key fields within the Sales and Customer dataset used for analysis. It outlines essential information such as customer identifiers, contact details, geographic location, and industry classification. Additionally, it includes sales-related fields like product details, purchase quantities, total revenue, and transaction dates. These well-defined columns form the foundation for data cleaning, transformation, and in-depth analysis, enabling meaningful insights into customer behavior, sales performance, and business opportunities.

Sales ANALYSIS dashboard



**DATA CLEANING AND VALIDATION PROCESS:**

**1. Identify and correct invalid email formats:**

**Inserting @ before the domain names:**

The first and most important step in your email cleaning process because emails without @ are invalid and cannot be used for communication or analysis. So I have replaced domain name with @domainname.

**Correcting Domain Name Errors (Missing Dots in ".com"):**

I have used nested formula to detect and correct common domain errors like "@gmailcom" (missing dot). It replaces them with the correct format such as "@gmail.com", "@yahoo.com", or "@outlook.com".

**Inserting Missing Username Before Email Domain:**

This step ensured that emails missing both @ and a username were properly reconstructed, making them valid and useful for communication or marketing. It also showed smart use of name data to fill in gaps in the email column.

Example: Original Email: @gmail.com

Customer Name (B2): Vandana Sharma

Result: [vandana@gmail.com](mailto:vandana@gmail.com)

**2. Detect and remove duplicate records:**

I used **Power Query** in Excel, which allowed me to efficiently identify and eliminate rows with identical values across key columns. After loading the dataset into Power Query, I selected the relevant columns and used the **"Remove Duplicates"** option to ensure only unique records were retained. To validate the results, I applied the **UNIQUE() function** in Excel to cross-check and confirm that all duplicate entries had been successfully removed. This process helped improve data quality and ensured accurate analysis.

**3. Validate phone numbers to ensure they follow the correct Indian format:**

* Checked if the Phone Number is 10 Digits Long:  
  I have used the LEN() function to ensure the number has exactly 10 digits.
* Ensure the Number Starts with 6, 7, 8, or 9:  
  I have used the LEFT() function to extract the first digit and check if it's between 6 and 9
* Remove Extra Characters (like spaces, dashes, brackets):  
  Use the SUBSTITUTE() function to clean unwanted symbols.
* Combine Cleaning and Validation in One Formula:  
  Clean the number and check if it’s valid in a single step by using if function.
* Highlight Invalid Numbers with Conditional Formatting (Optional):  
  Use Excel’s **Conditional Formatting** to automatically highlight numbers that do not meet the criteria for easy review.
* Removing Extra Symbols or Characters from Phone Numbers in Excel:

To standardize phone numbers and prepare them for validation, I removed unnecessary characters such as **parentheses (), dashes -, and spaces** using Excel formulas like “substitue function”. These extra symbols often appear due to inconsistent data entry and can prevent proper analysis or validation.

**DATA ANALYSIS**

**1.How does customer distribution vary across different regions?**

Based on the bar chart showing customer distribution across regions, we can conclude the following:

The West region has the highest number of customers, totaling 4,333, followed closely by the North region with 4,304. The South and Central regions are relatively balanced, with 4,273 and 4,253 customers respectively. The East region has the lowest customer count at 4,237, though the difference among regions is not very large. Overall, customer distribution is fairly even across all regions, with a slight dominance in the western and northern parts of the country.

4o

**2.How does customer distribution vary across different States?**

Based on the chart showing customer distribution by state, **Gujarat** has the highest customer count at **2,795**, followed by **Karnataka** with **2,742**. **Delhi**, **Maharashtra**, and **West Bengal** also show strong customer presence, each with over **2,650** customers. **Telangana** has the lowest customer count at **2,606**, but the variation between states is relatively small. Overall, the distribution is quite balanced, with only a marginal difference between the highest and lowest values. This suggests a widespread customer base across multiple key Indian states.

**3.How does customer distribution vary across different Industries?**

Based on the customer distribution across industries, the **Education sector** has the highest number of customers with **4,032**, significantly ahead of all other sectors. **Hospitality**, **Healthcare**, and **Technology** follow closely, each with just under **3,000** customers. **Finance**, **Retail**, and **Manufacturing** have relatively similar customer counts, ranging from **2,847 to 2,863**. This indicates that while customer presence is fairly balanced across most industries, Education stands out as the most dominant sector. The variation reflects differing levels of demand or market reach among industries.

**4. Calculate the average total purchases per customer segment.**

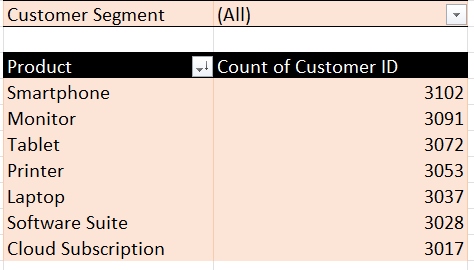
Based on the average total purchases per customer segment, the **Retail segment** has the highest average spending at **₹50,931.43**, followed closely by **Wholesale** with **₹50,702.66**. **In-Store** and **Online** customers have slightly lower average purchases, with **Online** being the lowest at **₹49,973.60**. The differences between segments are relatively small, indicating similar purchasing behavior across channels. However, **Retail and Wholesale customers tend to spend more on average**, possibly due to bulk or higher-value transactions. This insight can guide targeted marketing and sales strategies.

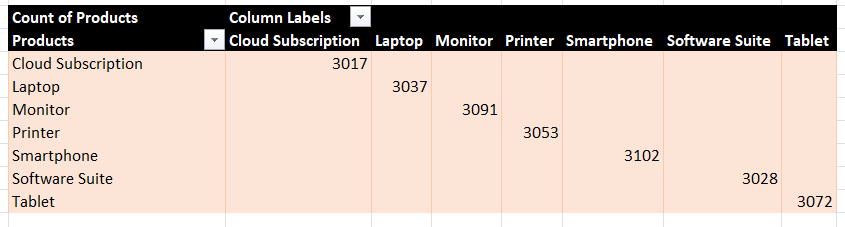
**5. Identify the top-performing products by revenue.**

The analysis of total product revenue reveals that **Monitor** is the top-performing product, generating the highest revenue of **₹2,03,31,48,671.98**. It is followed closely by **Smartphones** and **Printers**, both contributing over **₹196 crore** in revenue. **Laptops** and **Tablets** also show strong performance with comparable revenue figures. The narrow gap between the top five products suggests a well-diversified product portfolio. However, **Monitors lead in sales performance**, indicating higher demand or profit margins in that category.

**6. Analyze cross-selling possibilities to suggest potential opportunities for increasing sales.**

**Product Analysis:**

****

Product Popularity : We can easily identify the most popular products looking at the count of customers for each product. So here is the popular product is SmartPhone with 3102 sales. So that we can consider building or promoting the smartphone with the other products.

Although this table only reflects individual product sales, we can infer potential cross-sell opportunities based on **product relevance and popularity**.

The data shows that **Smartphones (3102)** and **Monitors (3091)** are the most purchased products, indicating high customer demand. Similarly, **Laptops (3037)**, **Printers (3053)**, and **Tablets (3072)** also have strong sales figures. Since many of these products are commonly used together (e.g., Laptops + Software Suite, or Monitors + Cloud Subscription), there is strong potential for cross-selling complementary items. For instance, customers purchasing Laptops could be targeted with bundled offers for Software Suites or Cloud Subscriptions. A deeper co-purchase analysis would better identify exact product pairings for effective cross-selling.

**Recommendations for Cross Selling:**

1. Identify Logical Product Pairs (Based on Use Cases):

Use product functionality and relevance to recommend logical pairs, even if you don’t have co-purchase data yet. For example:

* Laptops → Software Suite, Cloud Subscription
* Smartphones → Cloud Subscription, Tablets
* Monitors → Printers, Cloud Subscription
* Tablets → Software Suite

These are commonly used together, making them strong cross-sell candidates.

**2. Recommend Targeted Cross-Selling Campaigns**

Based on your sales data, you can recommend the following strategies:

**🔹 Bundle Offers**

* Create combo offers like:
  + “Buy a **Laptop**, get 30% off on **Software Suite**.”
  + “Purchase a **Monitor**, get **Cloud Subscription** free for 3 months.”

These drive higher cart value and encourage customers to purchase more.

**🔹 Email or SMS Promotions**

* Target existing customers who purchased a product with personalized messages:
  + “Hi [Customer Name], you recently purchased a Laptop. Boost your productivity with our premium Software Suite – now 25% off.”

**🔹 Upsell at Checkout**

* In online platforms or stores, recommend related products during the checkout process:
  + “People who bought this also bought...”
  + “Complete your setup with a Cloud Subscription.”

**Conclusion:**

This project successfully transformed raw customer and sales data into meaningful insights through systematic data cleaning, validation, analysis, and visualization. Key business questions were addressed, such as identifying regional and industry-wise customer distribution, average spending by segment, and top-performing products by revenue. Additionally, potential cross-selling opportunities were identified based on product popularity and logical pairings. Excel tools like formulas, pivot tables, Power Query, and chart visualizations played a vital role in efficiently cleaning and analyzing the data. The insights generated can directly support business decisions related to marketing strategies, customer targeting, and product bundling.