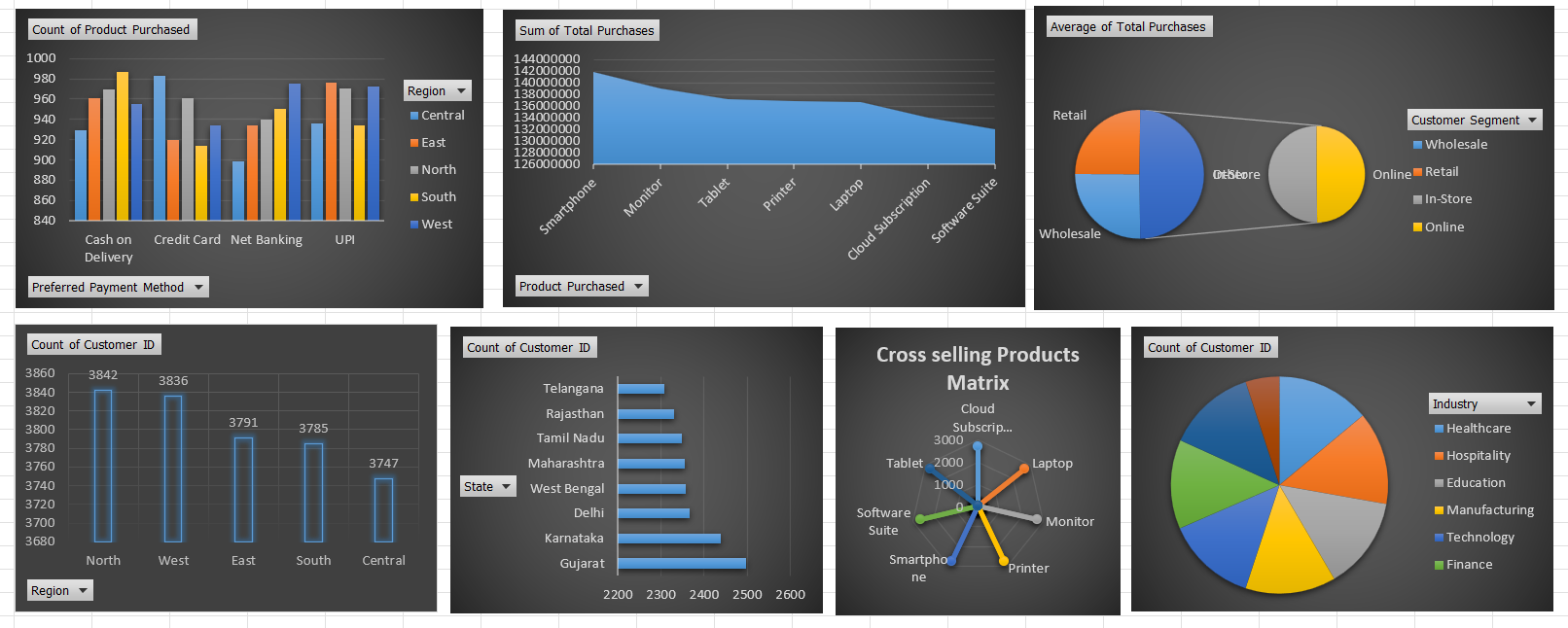
**Sales Analysis Report**

**Project Title:** Sales Performance and Customer Insights Analysis  
**Analyst:** Gokul S  
**Tools Used:** Excel, Power Query

**1. Project Overview**

This project aims to analyse customer purchasing behaviour, regional and industrial distribution, and sales performance across different product categories. The dataset includes customer demographics, transaction data, product details, and regional attributes. A cross-selling product matrix was also created to examine co-occurrence in product purchases.



**2. Data Description**

The dataset consists of the following key columns:

* **Customer Attributes:** Customer ID, Name, Email, Phone, Address, City, State, Zip Code, Country, Signup Date, Last Purchase Date, Is Active
* **Transaction Attributes:** Product Purchased, Quantity Purchased, Product Price, Total Purchases, Number of Orders, Total Product Revenue
* **Other Attributes:** Preferred Payment Method, Customer Segment, Industry, Region, Sales In charge

**3. Data Extraction and Transformation (ETL Process)**

The raw dataset required significant cleaning and transformation before meaningful analysis could begin. The following steps were taken using **Excel Power Query Editor** to prepare the data:

**a. Email Format Correction**

* Used the **Split Column** feature on the "Email" field to detect invalid concatenations or missing delimiters.
* Applied **Power Query M Code** to reconstruct improperly formatted email addresses.
* Ensured all entries included valid domain formats (e.g., @gmail.com, @yahoo.com, etc.).
* Removed any whitespace or hidden special characters that were breaking the structure of emails.

**b. Phone Number Standardization**

* Filtered out invalid or incomplete phone numbers.
* Applied **M code filtering** logic to keep only phone numbers starting with digits **6, 7, 8, or 9**, aligning with standard Indian mobile number conventions.
* Removed any non-numeric characters using transformation steps (Text. Select) to retain only digit sequences.

**c. Column Splitting and Cleanup**

* Used **Split by Delimiter** and **Split by Number of Characters** on address fields to extract city, state, and zip code if needed.
* Removed duplicates and nulls from customer identifier columns to maintain data integrity.
* Converted data types (e.g., dates, currency, quantities) into consistent formats suitable for visualizations.

**d. Loading**

* Once transformations were complete, the cleaned data was loaded into the Excel Data Model.
* From there, it was connected to Power BI for dashboard creation and visualization.

This ETL pipeline ensured that the analysis was based on clean, structured, and reliable data, avoiding inconsistencies that could mislead insights.

**4. Key Insights**

**A. Customer Segment Analysis**

* Retail customers exhibit the highest average total purchase value.
* Online customers show the lowest average total spend.
* Retail and Wholesale have higher average values than In-store and online, suggesting a higher ticket size per transaction.

**B. Product Performance**

* **Smartphones** lead in total product revenue.
* Other high-performing products include Monitors, Tablets, and Laptops.
* Products are being sold independently without any co-purchase pattern (as per cross-selling matrix).

**C. Industry Distribution**

* Top customer bases come from **Education**, **Manufacturing**, and **Healthcare** sectors.
* Industries such as **Retail** and **Finance** have the lowest customer counts.

**D. Geographic Insights**

* **State-Wise Distribution:** Maharashtra and Gujarat hold the largest customer shares.
* **Region-Wise Distribution:** The **West** and **North** regions have the most customers, with Central and East regions lagging.

A pie chart with different colored sections

AI-generated content may be incorrect.

**E. Payment Preferences**

* **UPI** and **Credit Cards** are the most used methods across all regions.
* **Cash on Delivery** and **Net Banking** are also commonly used but less preferred.

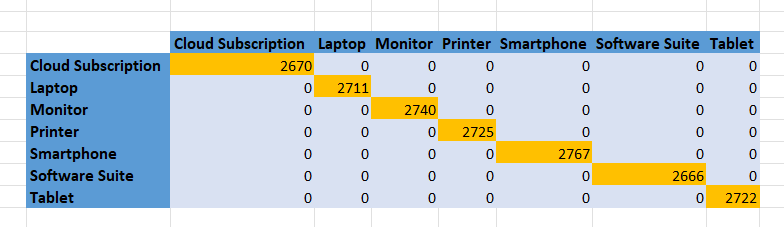
**5. Cross-Selling Product Matrix**

| **Product** | **Observations** |
| --- | --- |
| Laptop, Smartphone, Tablet, etc. | No co-occurrence: each purchase event involved a single product only. |
| Highest Sold Product | **Smartphones** with 2916 unique purchases. |
| Cross-Selling | No evidence of customers buying multiple products together. |

The matrix clearly shows a lack of bundling or cross-promotion. This may indicate an opportunity for introducing combo offers or bundled sales strategies.

**6. Recommendations**

* **Introduce Bundling:** Since no cross-selling is evident, consider introducing bundled deals (e.g., Laptop + Monitor).
* **Focus on High Revenue Segments:** Continue targeting **Retail** and **Wholesale** for higher ticket purchases.
* **Target High Potential Regions:** Boost marketing efforts in **West** and **North** regions.
* **Payment Flexibility:** Keep offering UPI and digital methods prominently.
* **Industry Targeting:** Focus campaigns on **Education** and **Manufacturing** sectors.



**7. Conclusion**

This analysis provides a clear understanding of the customer landscape, purchasing behaviour, and product performance. The insights derived from segmentation and cross-selling can help tailor marketing strategies, optimize sales channels, and improve overall business decisions.