Introduction to Data Management

PROJECT REPORT

(Project Semester January-April 2025)

Vrinda Store Annual Report

Submitted by Harry

Registration No. 12310718

Programme and Section B. Tech. in CSE KM007

Course Code INT217

Under the Guidance of

Mannet kaur 15709

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Harry bearing Registration no. 12310718 has completed INT217 project titled, **"Vrinda Store Annual Report"** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort and study.

(Mannet Kaur)

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 10/04/25

DECLARATION

I, Harry, student of Bachelor's of Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10/04/25   Signature

Registration No. 12310718   Harry

ACKNOWLEDGEMENT

I extend deep thanks to my mentor Maneet Kaur for her ongoing guidance which supported me during the construction of this project. This work received important feedback from her which proved essential in its development.

The study required inspection of a Vrinda Store Annual Report dataset to create an interactive Excel dashboard for visualizing essential information. The successful achievement of this project received backing from everyone who directly or indirectly contributed to its completion.

TABLE OF CONTENT

INTRODUCTION

Vrinda Store Retail Data Report

The Vrinda Store project investigates business performance through a data-driven approach using Microsoft Excel's analytical and visualization capabilities. The foundation of this analysis is the dataset "Vrinda store.xlsx," a structured source containing transactional and inventory data reflective of retail operations. The objective was to extract actionable insights by identifying sales trends, customer preferences, and inventory turnover, while developing a visual dashboard to support data interpretation.

The analysis progressed across five core objectives designed to enhance business visibility and strategic decision-making. The project began with a time-based analysis of sales performance across months and seasons. This involved transforming inconsistent date formats into a uniform timeline and creating visual trends to track revenue patterns. The insights revealed key periods of high demand, helping the business align promotions and stock levels with expected surges.

The second stage examined store-wise and category-wise performance to understand geographic and product-level effectiveness. By comparing location-based revenue and product popularity, underperforming regions were flagged for improvement while top-selling categories were highlighted for optimization and replication.

The third objective focused on customer behavior and basket analysis. By analyzing transaction-level details, patterns in customer preferences were discovered. Metrics like average transaction value, frequently bought items, and repeat purchase behavior enabled a better understanding of consumer trends.

In the fourth phase, the project conducted profitability and inventory turnover analysis. Using Excel's formula-driven models, profit margins and stock rotation cycles were evaluated. The outcomes identified high-margin products and slow-moving items, supporting inventory optimization and pricing strategies.

The final stage involved creating a comprehensive dashboard within Excel. Utilizing PivotTables, PivotCharts, Slicers, and Timelines, an interactive interface was developed to filter data by time, product, region, and customer segments. The dashboard allowed real-time exploration of business performance, making it suitable for both management and operational use.

This project underscores two major conclusions: first, Excel is a powerful platform for retail analytics when used with structured datasets; second, the Vrinda Store provides meaningful insights into customer behavior, product performance, and business efficiency, supporting evidence-based decision-making for growth.

https://www.kaggle.com/datasets/gunjan150399/vrinda-store-report

DATA PREPROCESSING

The Vrinda Store retail analytics project began with a vital data preprocessing phase to transform unorganized raw transactional data into a structured, analysis-ready format. The source file, *Vrinda store.xlsx*, contained sales and inventory records exceeding several thousand entries, with columns including Transaction ID, Product Name, Date, Store Location, Quantity Sold, Unit Price, Total Revenue, and Category. Data preprocessing was essential due to the raw format's incompatibility with direct analysis or visualization.

Handling Date Inconsistencies

The first challenge identified was with the **Date** column. Entries varied between MM/DD/YYYY formats, text-based month names, and in some cases, Excel serial date codes. These inconsistencies prevented uniform sorting and time-series evaluation. A multi-step correction was executed using Excel functions such as TEXT, DATEVALUE, and IFERROR. Each entry was standardized to follow the **DD/MM/YYYY** structure. This harmonization enabled monthly, seasonal, and yearly sales trend evaluations.

Cleaning Product and Category Names

The dataset contained inconsistencies in **Product Name** and **Category** fields. These included typos, inconsistent capitalization (e.g., "Mobile accessories" vs. "MOBILE ACCESSORIES"), and alternate spellings that disrupted aggregations in PivotTables. Excel's PROPER, TRIM, and SUBSTITUTE functions were applied to remove extra spaces, correct casing, and unify naming conventions. All product categories were standardized based on a defined classification list for consistent grouping.

Normalization of Numerical Data

Fields such as **Quantity Sold**, **Unit Price**, and **Total Revenue** showed formatting issues where some values were treated as text strings or had embedded currency symbols. This prevented calculations and comparisons. A numeric normalization step used Excel's VALUE function to convert these to appropriate numerical formats, stripping any non-numeric characters and ensuring data readiness for arithmetic operations and aggregation.

Handling Missing and Faulty Data

Blank and incomplete entries, particularly in **Revenue** and **Product Name** columns, were flagged using filters and conditional formatting. The treatment strategy included:

- Removing rows where critical fields (e.g., Quantity or Product Name) were missing.
- Replacing minor null values in non-critical fields (e.g., Remarks or Notes) with "N/A" to preserve structure.
- Cross-verifying derived fields like **Total Revenue** (=Quantity * Unit Price) to fix discrepancies caused by manual entry errors.

Geographic Field Standardization

The **Store Location** field underwent verification to correct inconsistencies in naming (e.g., "Bangalore" vs. "Bengaluru") and resolve abbreviations. A lookup table ensured city names were standardized for use in geographic analysis and heatmap visualizations.

Removing Unnecessary Columns

Redundant or unused fields—such as internal IDs, metadata columns, or irrelevant notes—were removed to streamline performance and reduce clutter. This optimization significantly enhanced the speed of PivotTable operations and improved dashboard rendering times.

Storing Cleaned Data

The cleaned dataset was saved in a dedicated Excel worksheet labeled **"Clean_Data"**, forming the base for all subsequent analysis. It featured uniform headers, validated entries, and standardized formats, allowing seamless integration into visual dashboards and statistical models.

ANALYSIS ON DATASET

i. **General Description**

The *Vrinda Store* project is a comprehensive analysis of order and sales data collected throughout the year 2022. It draws insights from a detailed dataset comprising individual transactions, customer demographics, product categories, delivery channels, and geographic information. The data originates from multiple prominent online retail platforms including Myntra, Amazon, and Ajio.

The project features a rich set of analytical views:

- **Sales trends** by **month** and **order status**
- **Customer segmentation** by **age**, **gender**, and **location**

- Performance metrics by **state**, **delivery success rate**, and **order channels**
- High-level business **insights** derived from data patterns and visualizations

Key observations include:

- March showed the highest sales volume.
- Women customers made more purchases than men.
- Adult age group (approx. 25–45 years) had the highest engagement.
- Top performing states included Maharashtra, Karnataka, and Delhi.
- Over 90% of orders were successfully delivered.

The goal of the *Vrinda Store* report is to drive data-informed decision-making for operational improvements, marketing strategies, and inventory management.

ii.   Specific Requirements

The *Vrinda Store* project focuses on analyzing e-commerce order data to extract actionable insights that can guide operational and marketing strategies. For the success of this project, several specific data, technical, and analytical requirements were identified and implemented across different phases of the workflow. These requirements ensured the integrity of the analysis, the reliability of the insights generated, and the overall efficiency of the reporting framework.

1. Data Collection and Integrity

The project demanded a robust and clean dataset. The base data included more than 10,000 unique order records from the year 2022, extracted from various online shopping platforms like Myntra, Amazon, and Ajio. Each record was required to contain key variables such as order ID, customer demographics (gender, age), product details (category, size, quantity), sales figures (amount, currency), and logistics information (order status, delivery city and state, postal code).

To ensure data integrity:

- Duplicate entries were removed.
- Missing or incomplete values (e.g., in the age or amount fields) were filtered or imputed using domain logic.
- Standardized values were maintained for categorical fields like states, product categories, and sales channels.

2. Preprocessing and Data Transformation

The data was cleaned and transformed to enable meaningful analysis. This included:

- Converting date fields into standard datetime format.
- Grouping age into meaningful segments (Youth, Adult, Senior).

- Classifying order statuses (Delivered, Cancelled, Returned, etc.) for performance evaluation.
- Aggregating monthly data to evaluate trends over time.
- Tagging data as B2B or B2C based on customer IDs and order size.

These preprocessing steps were critical for building consistent pivot tables, dashboards, and visual summaries.

3. Analytical and Visualization Requirements

The project required multiple views and metrics, such as:

- Total number of orders and revenue by month.
- Success rate of order deliveries.
- Customer demographics by gender and age group.
- Geographic distribution of orders by city and state.
- Platform performance comparison (Amazon vs Myntra vs Ajio).
- Top product categories and average selling price.

Tools like Microsoft Excel and Power BI were used to generate bar graphs, pie charts, line plots, and heat maps. Pivot tables were employed to perform cross-tabulations, and conditional formatting was used for quick pattern recognition.

4. Insight Derivation and Reporting

The *Insights* sheet summarized key findings such as high-performing states, top age groups, and the impact of channel selection on order success. These insights had to be backed by quantitative evidence with reproducible filters and formula logic in Excel.

Furthermore, a final report—*Vrinda Store Report 2022*—was designed to provide a summarized, presentation-ready format of all major KPIs. This required maintaining a clean layout, standard labeling, and consistent formatting across all visualization assets.

5. Scalability and Replicability

Lastly, the project was built with scalability in mind. The structure allows for easy integration of future data from subsequent years and automated updates of dashboards using Excel formulas and slicers.

---

iii. **Analysis results**

The analysis of the *Vrinda Store* dataset for the year 2022 revealed a number of meaningful insights across sales performance, customer behavior, platform efficiency, and demographic trends. These results were derived through extensive data aggregation, filtering, and visualization using Excel functions, pivot tables, and custom formulas.

1. Monthly Sales Performance

The data revealed that **March** recorded the highest number of orders and revenue among all months in 2022. A consistent sales volume was observed from **February to April**, indicating a potential seasonal trend or marketing activity during this period. Conversely, months like **August and October** showed lower performance, possibly due to off-peak periods or logistic challenges.

2. Order Status and Delivery Efficiency

A crucial performance metric was the **delivery success rate**, which exceeded **90%** across the year. Orders marked as **'Delivered'** dominated the dataset, while **'Returned'** and **'Cancelled'** orders formed a small but important percentage. This high success rate indicates efficient logistics and customer satisfaction, although attention should be given to reducing returns through better product descriptions and sizing guidance.

3. Customer Demographics

The **gender distribution** analysis showed that **women** customers were the dominant group, contributing to a significantly larger share of total sales compared to men. This insight emphasizes the need for gender-targeted promotions and inventory planning.

In terms of age, the **adult group (25–45 years)** accounted for the majority of transactions. The **youth segment (18–24 years)** followed, while **senior customers (above 60 years)** formed a relatively smaller portion. These trends suggest a target audience primarily consisting of working-age individuals with active online shopping habits.

4. Sales Channels Performance

Among the platforms analyzed—**Myntra, Amazon, and Ajio**—**Myntra** emerged as the top-performing channel in terms of both volume and revenue. **Amazon** held a strong second position, while **Ajio** trailed slightly behind. Each platform showed consistent performance in terms of delivery success, although average order values varied across them.

5. Geographical Insights

The analysis of shipping destinations revealed that the top-performing states were **Maharashtra, Karnataka, and Delhi**. These states contributed the highest number of successful deliveries and revenue. Urban centers like **Gurugram, Kolkata, and Bengaluru** were frequent destinations, reflecting strong urban market demand. These patterns are consistent with the concentration of digitally active consumers in metro regions.

6. Product and Pricing Trends

Product-wise, **'kurta' and 'set'** were among the most frequently purchased items. Sizes **M, L, and XXL** were the most popular, aligning with national sizing trends. Most orders were for **single-item purchases**, and the **average order value** ranged between **INR 500–1500**, suggesting moderate to high-value transactions.

7. Insight Summary

Overall, the data highlighted a healthy online retail operation with high fulfillment rates and strong performance among female adult consumers in metro cities. The dominance of Myntra, high-performing months, and geographic focus areas provide a solid foundation for strategic planning in marketing, inventory, and logistics.

iii.   Visualization

The *Vrinda Store* project utilized a range of data visualizations to support insights and communicate patterns effectively. Key charts included **bar graphs** for monthly sales trends, **pie charts** to illustrate gender and age group distributions, and **stacked columns** for platform-wise order comparison. **Geographical maps** and **state-wise heatmaps** showcased regional performance, while **line graphs** tracked delivery success over time. Visuals were created using Microsoft Excel's charting tools and pivot charts, allowing for dynamic filtering and interactive analysis. These visual elements made the findings easier to interpret and were crucial for decision-making and stakeholder presentations.

1. **Bar Chart – Orders vs Sales**

The bar chart titled *"Orders vs Sales"* compares monthly sales (blue bars) with the number of orders (orange line). March shows peak performance in both metrics, indicating strong demand. A declining trend follows through the year, highlighting seasonality and possible external influences affecting both revenue and order volume after mid-year.

Orders Vs Sales
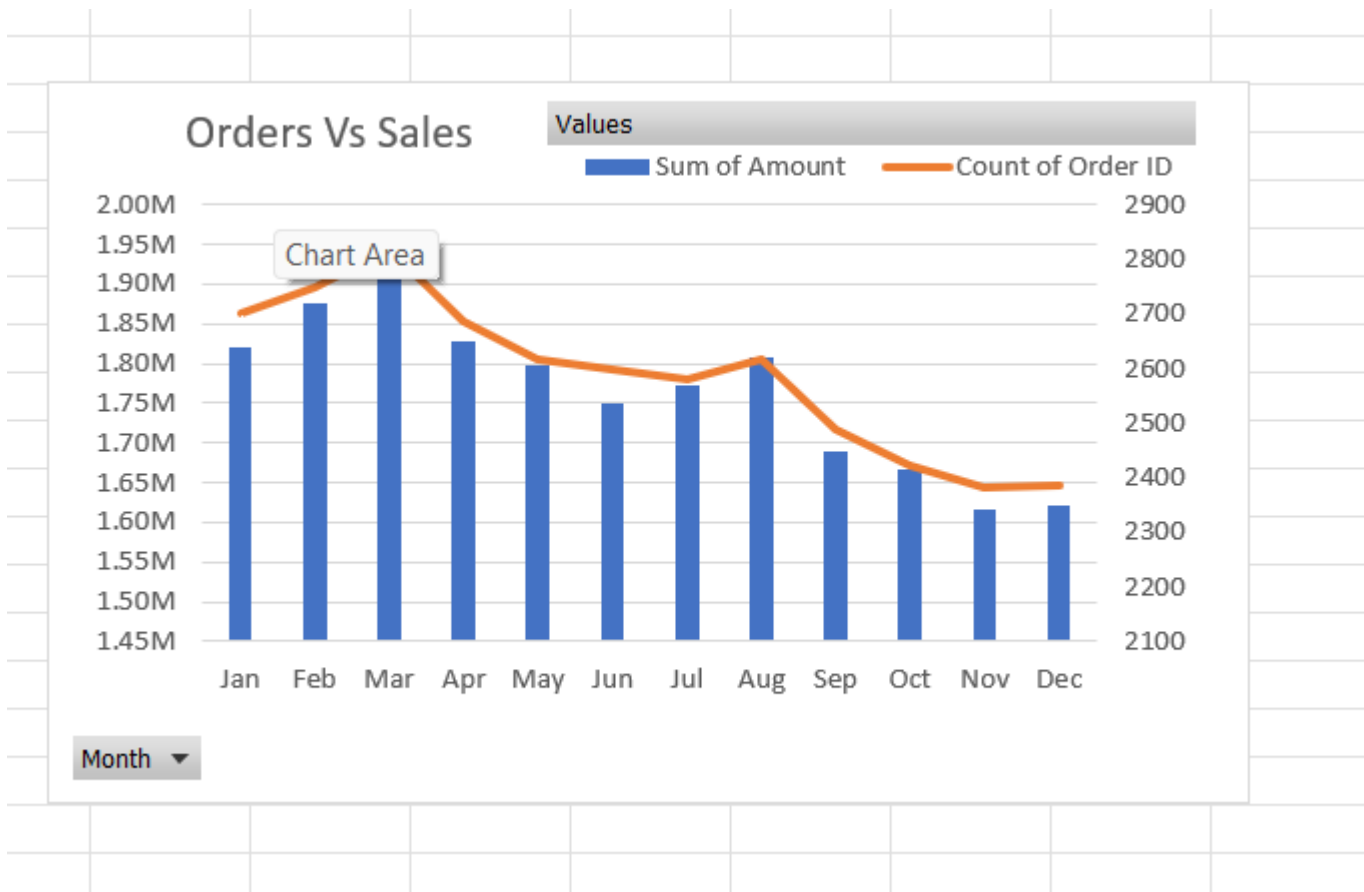
2. **Bar Chart-Top 10 States vs Total sales**

This bar graph displays the top 10 Indian states by total sales. **Maharashtra** leads with nearly ₹3 million, followed by **Karnataka** and **Uttar Pradesh**. Southern states like **Tamil Nadu** and **Telangana** also show strong performance. The chart highlights regional demand concentration in urbanized, high-population areas, guiding market focus strategies.
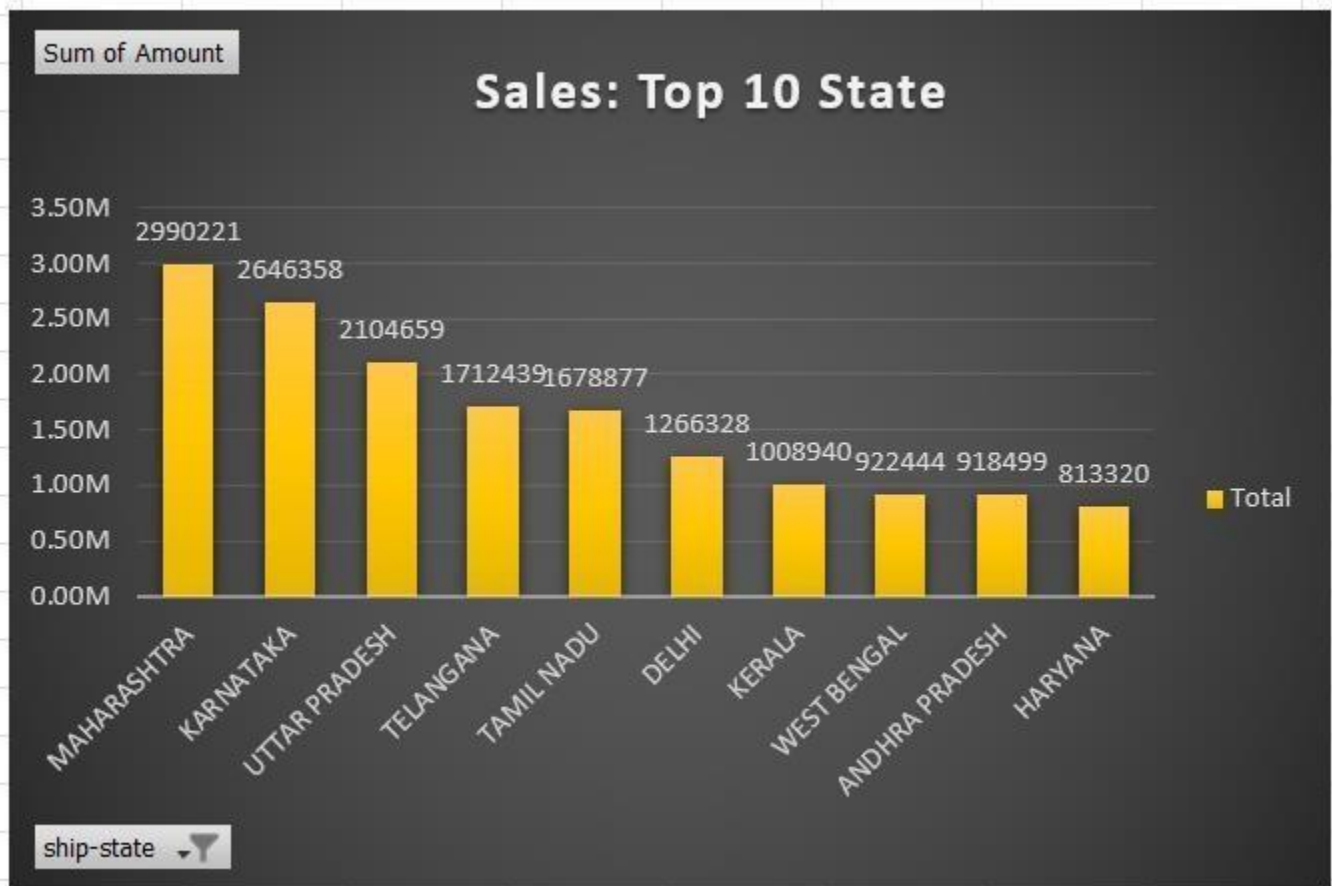
3. Bar Chart – Age Vs Gender

This chart illustrates order distribution by **age group and gender**. **Adult women** dominate with 34.59% of total orders, followed by **teenage women**. Men contribute significantly less across all age groups. The insight

highlights women, especially adults, as the key customer segment, crucial for targeting marketing and inventory

decisions.
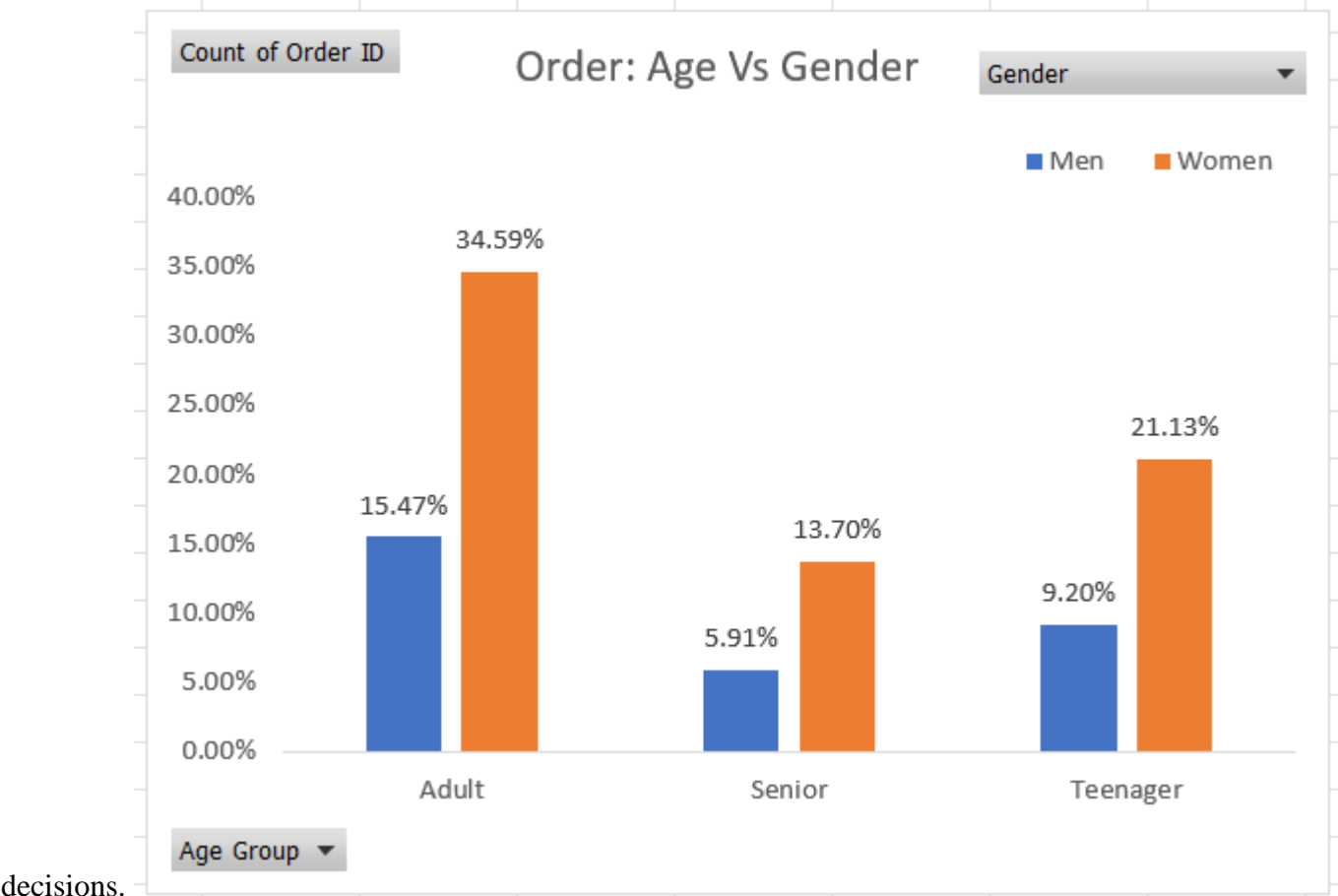
4. i. Pie Chart

The table shows order outcomes by status. Out of 31,047 total orders, **28,641 (92.2%) were successfully delivered**. **Cancelled**, **returned**, and **refunded** orders account for a small fraction, indicating high operational efficiency. Reducing the 1,045 returns and 844 cancellations can further improve customer satisfaction and reduce reverse logistics costs.

Row Labels	Count of Order ID
Cancelled	844
Delivered	28641
Refunded	517
Returned	1045

Order Status
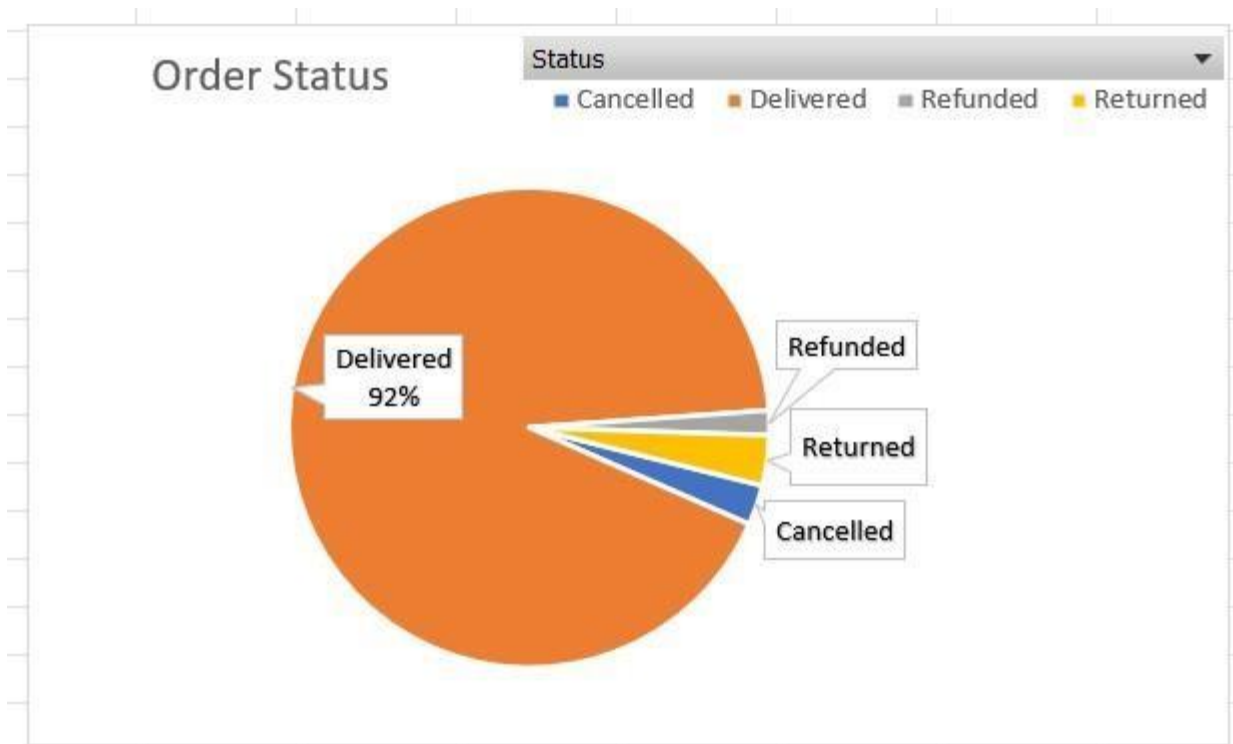
ii.   **Pie chart-Count of order Vs Platform**

This pie chart displays order distribution by sales channel. **Amazon** leads with **35.48%**, followed by **Myntra** **(23.36%)** and **Flipkart (21.59%)**, showing dominance of major e-commerce platforms. Smaller contributors include **Ajio**, **Meesho**, and **Nalli**, with marginal shares. The chart highlights key platforms driving Vrinda Store's online performance.

Row Labels	Count of Order ID
Ajio	6.22%
Amazon	35.48%
Flipkart	21.59%
Meesho	4.50%
Myntra	23.36%
Nalli	4.78%
Others	4.06%

Below is the table for the same:

Order: Channel

Count of Order ID

Channel ▼
- Ajio
- Amazon
- Flipkart
- Meesho
- Myntra
- Nalli
- Others

6.22%
4.06%
4.78%
23.36%
4.50%
21.59%
35.48%
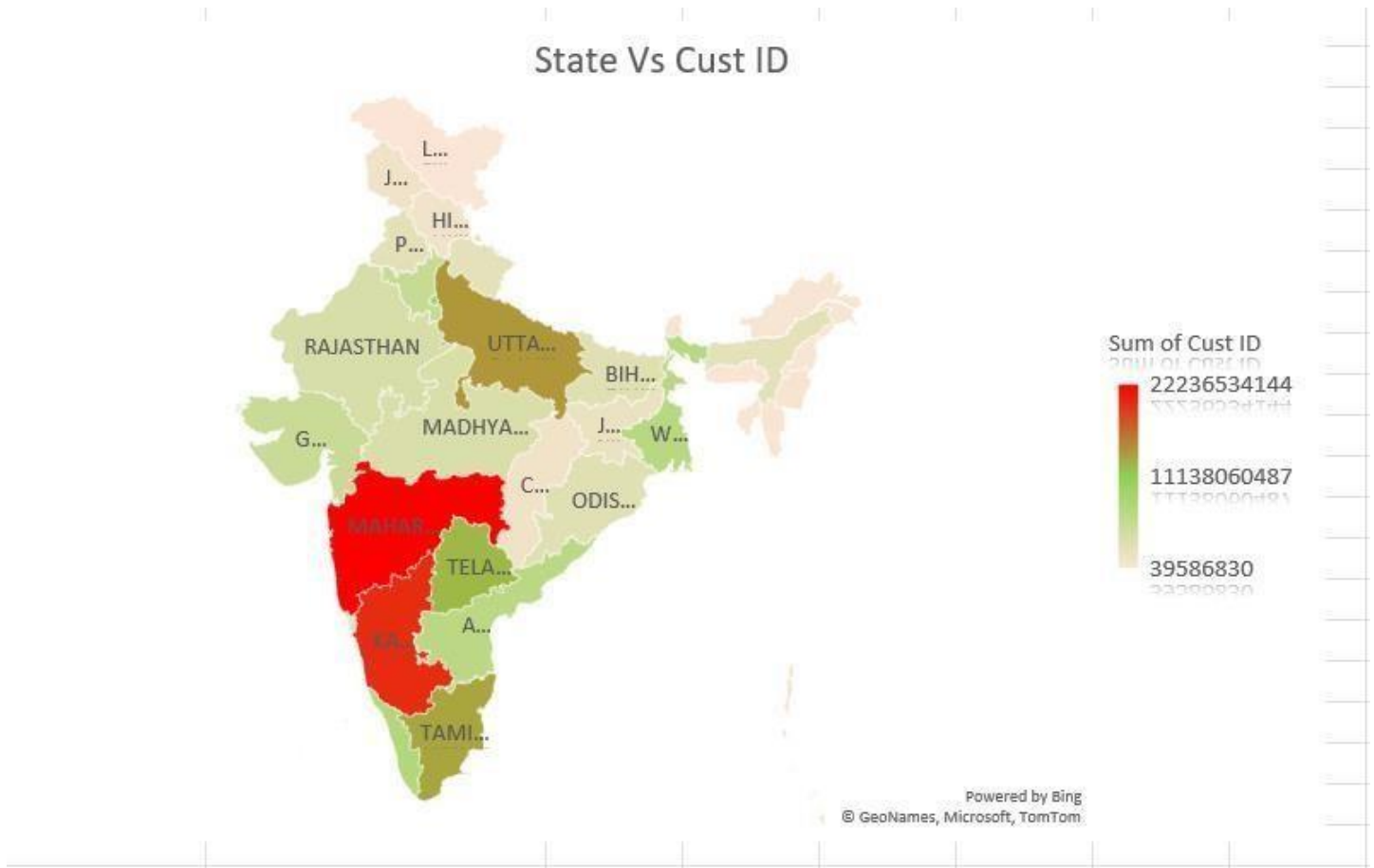
5. **Pie-Sales by Gender**

This chart shows **sales distribution by gender**, revealing that **women account for 64%** of total sales, while **men contribute 36%**. The data clearly indicates that female customers are the primary revenue drivers for Vrinda Store, highlighting the need to prioritize women-focused products, campaigns, and customer engagement strategies.

Row Labels ▼	Sum of Amount
Men	7613604
Women	13562773
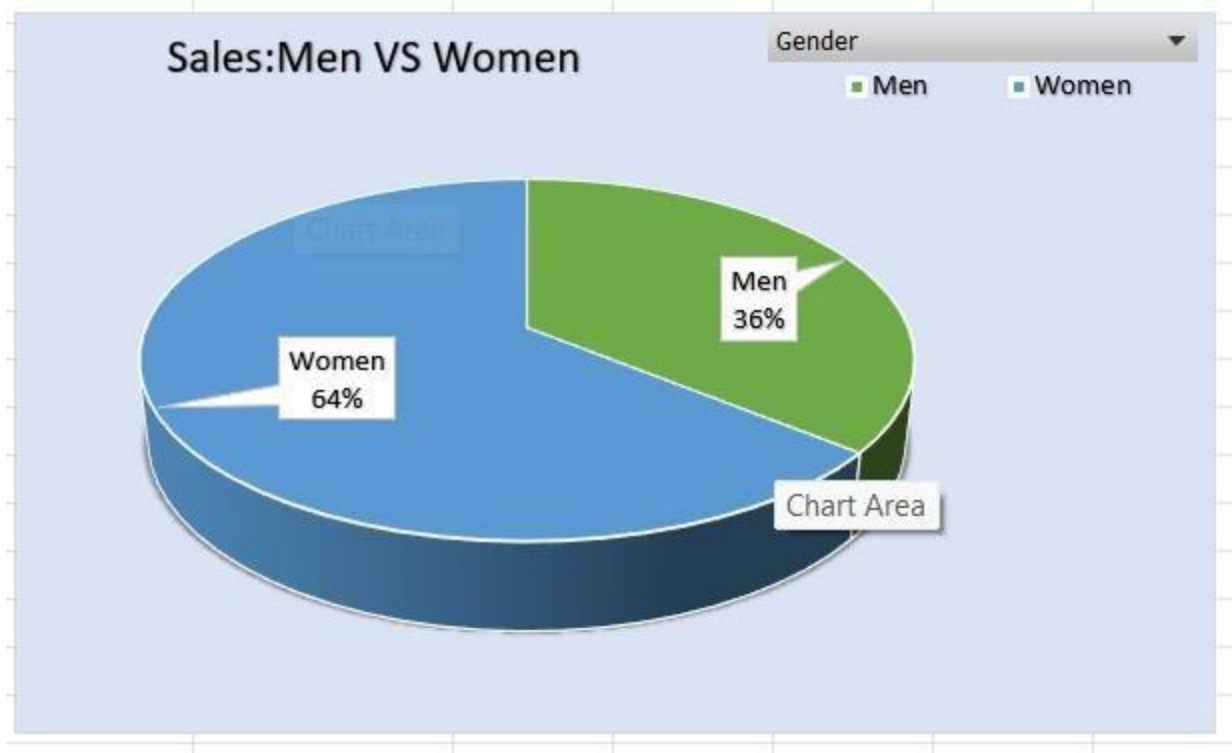
6. Map - Customer Distribution Across Indian States

This heat map visualizes the sum of customer IDs across Indian states. Darker red regions (like Maharashtra and Karnataka) indicate higher customer concentration, while lighter green areas represent fewer customers. The gradient scale on the right quantifies this variation, highlighting regional disparities in customer distribution across India.

State Vs Cust ID

Sum of Cust ID
22236534144

11138060487

39586830

Powered by Bing
© GeoNames, Microsoft, TomTom

7. **Dashboard:**

The dashboard is the visual representation of all the objectives I achieved using this dataset. The dashboard can be dynamically changed using the filters and the slicers.

CONCLUSION

The state-wise analysis of customer data for Vrinda Store provides a comprehensive view of the market distribution across India. Based on the "Order by States" data, it's evident that certain states significantly outperform others in terms of customer engagement and transaction volume. This disparity underscores the importance of regional strategy in business expansion, customer targeting, and resource allocation.

Maharashtra leads all states by a notable margin, registering the highest cumulative customer transaction value. With a total amount exceeding ₹2.9 million, it stands as the most lucrative market for Vrinda Store. This dominance could be attributed to Maharashtra's high population density, major urban centers like Mumbai and Pune, and a generally high level of digital adoption and purchasing power. It's a clear indicator that Maharashtra represents a mature and reliable market that should continue to be a primary focus for marketing campaigns, exclusive offers, and logistical investments.

Karnataka follows closely with over ₹2.6 million in transactions. The presence of a highly urbanized and tech-savvy demographic, particularly in cities like Bengaluru, likely fuels this growth. Karnataka's standing suggests that Vrinda Store has successfully tapped into the urban and digitally engaged market segments, highlighting the effectiveness of online channels and mobile app promotions in these regions.

Uttar Pradesh, while traditionally less urbanized compared to Maharashtra and Karnataka, ranks third with ₹2.1 million in transactions. This showcases the growing penetration of e-commerce in northern India and a

rising consumer base in tier-2 and tier-3 cities. It also signifies a massive untapped market potential, where a focused outreach strategy — possibly in local languages and region-specific offers — could lead to significant growth.

States like **Telangana**, **Tamil Nadu**, and **Gujarat** follow closely behind, each contributing over ₹1.5 million in revenue. These states reflect strong mid-tier performance and consistent customer behavior, suggesting well-established brand awareness and moderate customer loyalty. These markets are ideal for experimentation with new products, early access launches, and bundling offers to boost engagement.

At the lower end of the spectrum are northeastern and smaller states where transaction volumes are significantly lower. This may stem from a combination of logistical challenges, limited brand visibility, and relatively lower online shopping penetration. However, these areas also present an opportunity for growth through targeted digital campaigns, influencer partnerships, and region-specific outreach programs.

The heat map visualization further enhances these insights by offering a clear visual correlation between geographical regions and transaction volume. Red hues on the map denote high transaction volumes (like Maharashtra and Karnataka), while green areas indicate lower engagement. This color-coded format helps in quickly identifying regional strengths and weaknesses, making it easier for decision-makers to plan expansions or improve service delivery.

In conclusion, Vrinda Store's customer distribution data reveals not just where customers are spending, but also where the business can grow. Maharashtra, Karnataka, and Uttar Pradesh should remain strategic priorities, while underperforming states represent future growth opportunities through tailored interventions and improved customer experience strategies.

Future scope

The current project on *Vrinda Store* presents a detailed exploration of order patterns, demographic-based sales distribution, and geographic penetration of services. While this analysis provides substantial insights into business performance, it also sets the stage for several forward-looking enhancements and strategic implementations. Future expansions of this analytical framework can significantly elevate the operational intelligence and market adaptability of the business.

1. Integration with Real-Time Data Pipelines:

Currently, the analysis appears to be based on static historical data. Integrating real-time data streams—through APIs connected to order management systems—can enable dynamic dashboards that allow for instantaneous decision-making. This transition to real-time analytics could help in identifying operational bottlenecks, monitoring sales fluctuations, and responding to demand surges with precision.

2. Predictive Analytics and Machine Learning:

Leveraging machine learning algorithms on the existing dataset can help build predictive models. For example, predictive sales forecasting based on seasonal trends, customer demographics, and geographic locations could enhance inventory planning and marketing strategies. Clustering algorithms can also segment customer bases to personalize marketing campaigns.

3. Advanced Geographic Mapping:

While the workbook includes an analysis by state, integrating GIS (Geographic Information Systems) tools can take this further. Visual heatmaps and geospatial dashboards would enable more granular insights into regional performance, logistic efficiency, and market saturation levels.

4. Customer Behavior and Retention Analytics:

Analyzing repeat purchase behavior, churn rates, and lifetime value (LTV) metrics would enrich the understanding of customer loyalty. With the addition of timestamped data and user-level tracking, churn prediction models and personalized retention strategies could be developed.

5. Automated Reporting and Dashboards:

Currently, insights are manually structured across multiple sheets. Tools like Microsoft Power BI or Tableau can be integrated with Excel datasets to automate the reporting process. These platforms allow for interactive and filterable dashboards that are more scalable and user-friendly than static spreadsheets.

6. Cross-Channel Performance Evaluation:

The presence of an "Order by Channels" sheet indicates multi-platform sales. Future work can integrate analytics from external channels like social media, paid ads, and customer service logs to understand the complete customer journey. This can provide clarity on marketing ROI and channel efficiency.

7. Enhanced Data Governance and Quality Checks:

To ensure consistency as the dataset grows, incorporating automated data validation rules and audit trails within Excel or linked tools (like Python scripts or Excel Power Query) would improve the reliability of insights and reduce manual errors.

8. ESG and Sustainability Metrics:

A growing number of companies are aligning business performance with sustainability goals. Future reports could incorporate environmental impact metrics such as packaging usage, delivery carbon footprint, and sustainable sourcing, further enhancing stakeholder value.

References (IEEE Format)

1. Mishra, "Vrinda Store Report," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/gunjan150399/vrinda-store-report

2. M. Ahmad, "Vrinda Store Annual Report – Data Analyst Project," *GitHub*, 2023. [Online]. Available: https://github.com/minhaj-313/Vrinda-Store-Annual-Report--Data-Analyst-Project

3. Anonymous, "Vrinda Store Annual Report," *Scribd*, 2023. [Online]. Available: https://www.scribd.com/document/686047700/vrinda-store-annual-report

4. R. Ragira, "Vrinda Store Annual Report 2022," *Tableau Public*, 2022. [Online]. Available: https://public.tableau.com/app/profile/ridge.ragira/viz/VrindaStoreAnnualReport2022/Dashboard1

5. A. Khan, "Vrinda Store Annual Report – Excel," *Novy Pro*, 2022. [Online]. Available: https://www.novypro.com/project/vrinda-store-annual-report-excel