TARGET DATASET EXPLORATION: AN SQL AND PYTHON APPROCH

"NAVIGATING THE FUTURE OF ONLINE SHOPPING"

PROJECT OBJECTIVE

The primary goal of this project is to conduct a comprehensive data analysis of a large-scale e-commerce dataset to uncover actionable business intelligence.

- **Extract Fundamental Metrics:** Identify core performance indicators like total sales, order volume, and geographic distribution.
- **Analyze Behavioral Trends:** Understand customer purchasing patterns, payment preferences (installments), and seasonal trends.
- **Identify Strategic Correlations:** Determine the relationship between pricing and sales volume to optimize inventory and pricing strategies.
- **Evaluate Business Health:** Use advanced analytics to measure customer retention and yearly revenue growth.

TECHNICAL APPROCH

The project utilized a dual-tool methodology to leverage the high-speed processing of SQL and the analytical flexibility of Python.

1.**Database Management (MySQL):**
- **Data Ingestion:** Handled over 1 million rows of data using LOAD DATA LOCAL INFILE for high-speed importing.
- **Complex Querying:** Utilized Joins, CTEs (Common Table Expressions), and Window Functions (RANK, AVG OVER) to process multi-layered data relationships**.**
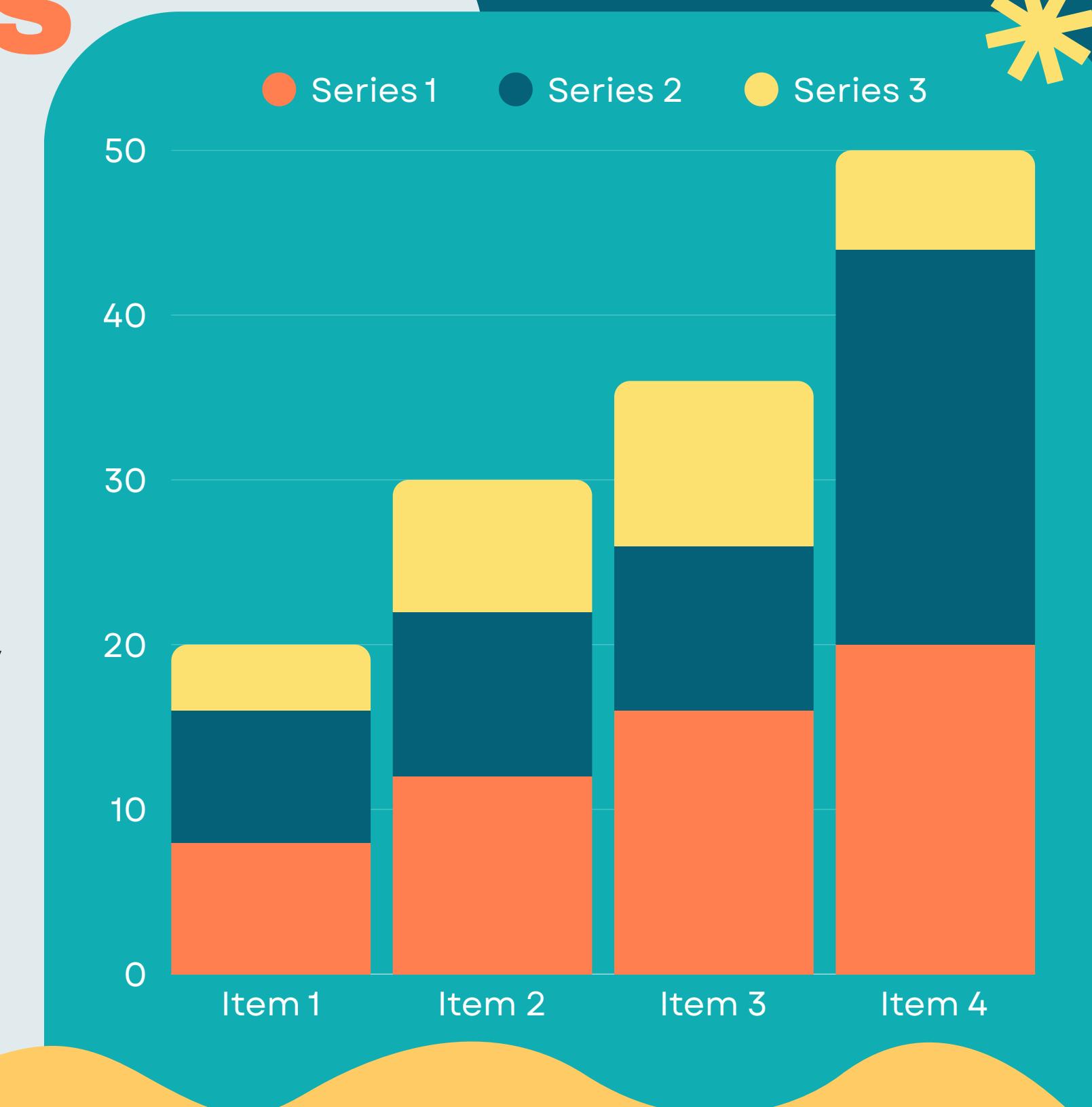
2.**Python Analysis (Google colab):**
- **Manipulation:** Used Pandas for merging SQL outputs with supplementary CSV data and performing mathematical correlations.
- **Visualization:** Employed Matplotlib to translate raw data into 2D visual stories, including bar charts, line graphs, and geographic density maps.
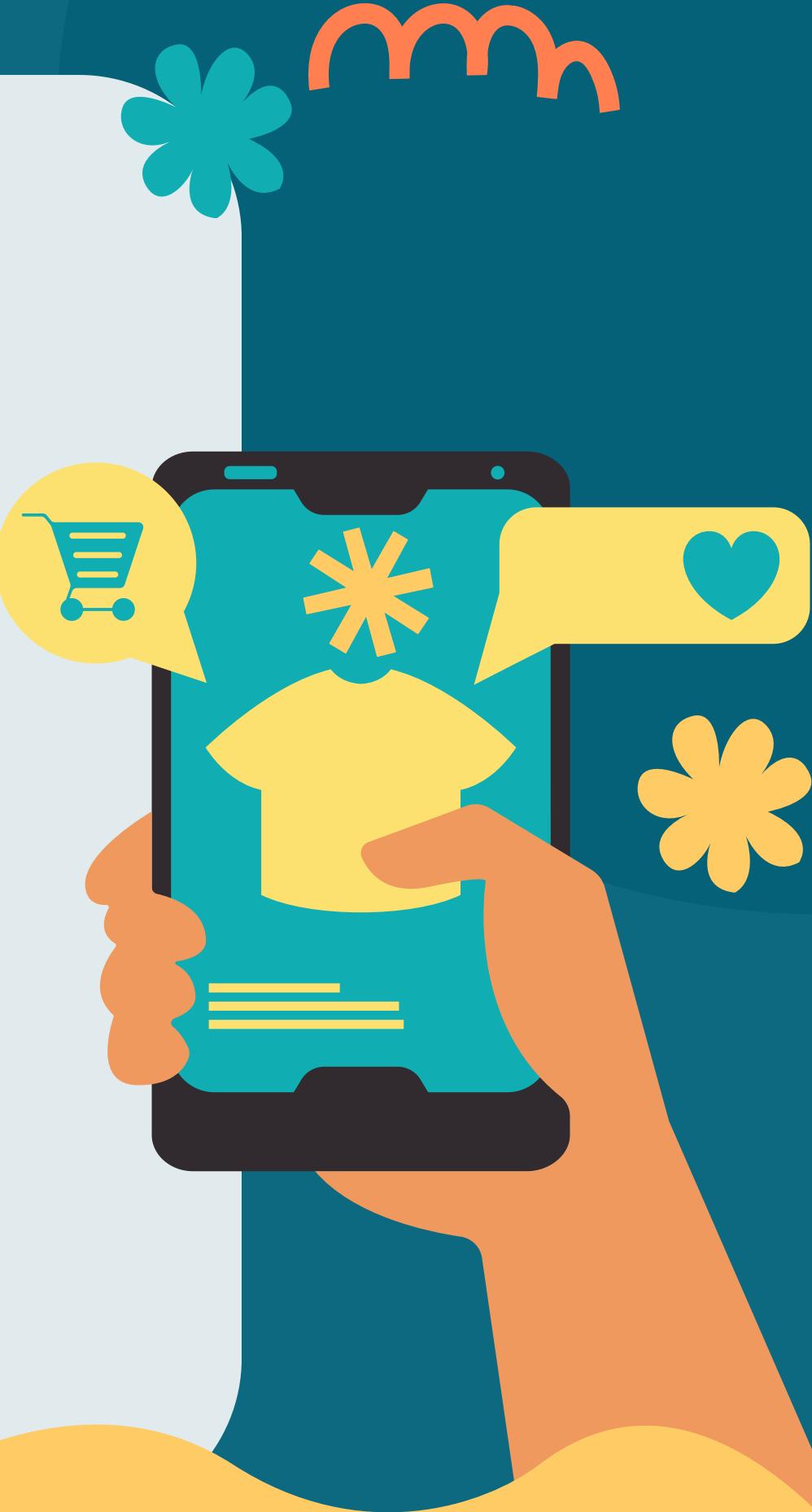
PROJECT INSIGHTS

- **Market Concentration:** Customer density is highest in specific geographic hubs. Strategic Recommendation: Focus marketing budget and logistics infrastructure on these top-performing states to reduce shipping costs and times.

- **Revenue Drivers:** The Top 10 Product Categories contribute to a disproportionately large percentage of total revenue. Focusing inventory expansion on these "star" categories is more efficient than broad-spectrum growth.

PROJECT INSIGHTS

- **Payment Behavior:** A significant portion of customers utilize installments (more than 1 payment). This indicates that offering flexible credit options is a critical factor in maintaining high conversion rates for the platform.
- **Growth Trends:** The 2018 Monthly Trends showed steady month-over-month growth, but also highlighted specific seasonal peaks. Identifying these peaks allows the business to prepare for high-volume shipping periods in advance.
- **Price Elasticity:** The Correlation between Price and Purchase Frequency showed that while lower-priced items sell more often, high-value items contribute significantly to the cumulative sales total, suggesting a balanced "Premium vs. Budget" inventory strategy is best.