**Ecommerce Data Summary**

The ecommerce analysis presents a comprehensive overview of customer behavior, sales distribution, and operational trends based on transactional data stored in a MySQL database. The analysis combines SQL queries and Python (pandas, matplotlib, seaborn) for data visualization, trend identification, and performance metrics. And effectively handled issues of data duplication using DISTINCT and validated the integrity of results by cross-checking metrics.

Key performance indicators (KPIs) such as total sales, average order values, customer retention, and product-wise revenue contribution were derived, making the analysis valuable for e-commerce decision-makers. Visuals were used to complement insights for business clarity.

**Analysis Highlights**

**1. Customer Demographics :**

* **Unique cities** identified where customers are located using:
  + **SELECT DISTINCT customer\_city FROM customers**
* **Customer distribution by state** reveals the geographical reach and market segmentation.

**2. Order Trends :**

* **Orders in 2017:** Queried orders based on the purchase timestamp.
* **Monthly orders in 2018** helped in understanding seasonality.

**3. Payment Insights :**

**Percentage of Installment Payments:**

* Calculated what percentage of payments were made in installments.
* Example result: 99.99**%** of customers used installment payments.

**4.** **Sales Analysis :**

**Total sales per product category:**

* Used clean joins with DISTINCT on order\_items and payments to avoid duplication.
* Categories like "Furniture", "Clothing", etc., ranked based on revenue.
* **% of revenue per category:** Showed category contribution using:
  + **ROUND(SUM(payment\_value) \* 100 / total\_revenue, 2)**

**5. Customer Behavior :**

* **Average products per order by city :** Grouped customers by city and averaged product counts.
* **Retention rate** **:**
  + Calculated based on repeat purchases within 6 months.
  + **Result :** 0% retention — indicating customers did not return in that period (needs business attention).
* **Top 3 spenders per year :** identified using window functions and ranking logic.

**6. Seller & Revenue Metrics :**

* **Revenue by seller** and **ranking** based on cumulative payment values.
* **Moving average of order values** across customer order histories revealed spending consistency.
* **Cumulative monthly sales** showcased trends and helped estimate growth.

**7. Correlation & Trends :**

* **Price vs. frequency correlation**: Attempted to understand if expensive products are bought less often.
* **Year-over-year growth**: Computed percentage increase in revenue between years.

**Visuals & Enhancements :**

* Used **bar plots**, **line plots**, and **legends** with improved readability.
* Included **percentage annotations** and value labels on bars for clarity.
* Filtered legends and reduced font sizes to make graphs more presentable.

**Best Practices Followed :**

* Handled **data duplication** using DISTINCT in SQL joins.
* Aggregated values carefully to prevent **inflated metrics**.
* Used **CTEs** and **window functions** for clean logic and readability.