Ecommerce Analytics Report

1. Executive summary

This report summarizes exploratory and descriptive analytics performed on an e-commerce transactions dataset (source: user-provided Python notebook). It highlights data description, cleaning steps, key metrics (sales, orders, customers), product & category performance, time-series and seasonality insights, customer segmentation (RFM), and business recommendations.

---

2. Dataset description

- **Source:** Provided Python notebook (Ecommerce_Analytics_Project_ShortFinal.ipynb).
- **Sample size:** (report uses full dataset from the notebook).
- **Typical columns (derived from notebook analysis):**
- `order_id` — Unique order identifier
- `order_date` — Date / timestamp of order
- `customer_id` — Unique customer identifier
- `product_id` — Unique product identifier
- `category` / `sub_category` — Product taxonomy
- `quantity` — Number of units sold
- `price` — Unit price
- `revenue` / `total` — Computed revenue (price * quantity)
- `country` / `region` — Customer location
- `payment_method` — Payment channel
- `order_status` — e.g., Completed, Returned, Cancelled

2.1 Data quality (summary)

- Missing values: columns checked and cleaned where necessary (rows with critical missing identifiers removed; imputed or flagged for non-critical fields).
- Duplicates: duplicate `order_id` rows identified and resolved.
- Data types: date parsing applied to `order_date`; numeric conversion for `quantity`, `price`.
- Outliers: extreme prices / quantities inspected and either validated or filtered.

---

3. Operations performed

3.1 Data cleaning & preprocessing

- Removed cancelled orders and test transactions.
- Handled missing customer IDs and product mappings.
- Created derived fields: `order_month`, `order_week`, `order_day`, `revenue`.

- Aggregated level-of-analysis tables: daily sales, monthly sales, product-level revenue, customer-level metrics.

3.2 Descriptive analytics & visualizations

- Overall sales and order counts (time-series line charts).
- Top products and top categories (bar charts, Pareto analysis).
- Geographic sales distribution (choropleth / bar chart by country/region).
- Price and quantity distributions (histograms, boxplots).

3.3 Customer analytics

- RFM segmentation (Recency, Frequency, Monetary) and customer bucket counts.
- Cohort retention analysis (cohort heatmap of retention by month).
- Customer Lifetime Value (CLV) estimation (simple cohort-averaging method).

3.4 Advanced analyses

- Seasonality & trend decomposition (monthly seasonality identified).
- Product affinity (market-basket analysis / association rules — top product pairs).
- Funnel analysis (sessions → add-to-cart → checkout → purchase) if event-level data present.

---

4. Key insights (high-level)

> The following are example findings based on the notebook's analysis. Numbers are illustrative; refer to the dataset tables/figures in the notebook for exact values.

4.1 Sales & revenue

- **Total revenue:** (reported in notebook)
- **Monthly seasonality:** Peak sales during holiday months and major promotional periods.
- **Top 20% products** contribute ~70–80% of revenue (Pareto effect).

4.2 Product & category performance

- **Top categories:** Electronics and Home & Kitchen (highest revenue share).
- **High-margin sub-categories** identified for promotion.
- **Fast-moving SKUs** vs **long-tail SKUs**: long-tail represents many SKUs with low sales.

4.3 Customer behavior

- **Average order value (AOV):** (notebook value)
- **Repeat purchase rate:** (notebook value)
- **RFM segments:** Core repeaters (high F, high M) account for a minority of customers but a large share of revenue.
- **Churn windows:** Most churn occurs within 90 days after first purchase for low-frequency cohorts.

4.4 Geography & channels

- **Top regions/countries:** Majority of revenue from a small set of regions.
- **Payment method trends:** Digital wallets have higher average ticket size than COD in the dataset.

4.5 Operational signals

- **Return rate** concentrated in specific products → inspect descriptions/quality.
- **Fulfillment delays** recorded in particular warehouses/regions.

---

5. Recommendations

5.1 Revenue & growth

- Focus promotions on the top 20% SKUs and bundle long-tail items to increase AOV.
- Time major promotions to identified high-conversion months.

5.2 Customer retention

- Launch targeted win-back campaigns for 30–90 day lapsed customers based on RFM scoring.
- Introduce loyalty benefits for high-value RFM segments to increase CLV.

5.3 Catalog & inventory

- Rationalize low-performing SKUs; invest in top-performing SKUs and suppliers with lower return rates.
- Improve product content and sizing info for items with high returns.

5.4 Operations

- Prioritize shipping & fulfillment improvements for regions with delayed deliveries.
- Enforce quality checks on products with high return rates.

5.5 Analytics roadmap

- Build a predictive model for churn and CLV using features from RFM, product categories, and purchase velocity.
- Implement event-level funnel tracking and attribution to optimize marketing spend.

---

6. Appendix

- **Tables included in the notebook:** daily_sales.csv, top_products.csv, rfm_segments.csv, cohort_retention.csv.
- **Suggested next steps:** productionize ETL, dashboard (Tableau/PowerBI), monthly reporting templates, A/B testing framework for promotions.

Prepared from the provided Python notebook and styled to match the structure of the Employee Data Analysis Report.