Executive Summary — Retail Sales Analysis (Novus Retail Pvt. Ltd.)

1) Project purpose

Turn Novus Retail's raw 2022 transaction log into validated, SQL-driven business intelligence that reveals revenue drivers, customer segments, product performance, and operational patterns to enable inventory, marketing and finance decisions.

---

2) Methodology — Succinct

- **Data preparation (SQL):** created a project schema and table(s), corrected data types, removed duplicate rows, handled/null-treated incomplete records, normalised date/time fields, and derived key fields (for example Total_Sale where needed, Year/Month, Shift buckets, and age groups).

- **Exploratory analysis (SQL):** implemented 10 targeted analytic queries (documented as comments in Retail Sales Data Exploration & Analysis.sql) to profile revenue, time patterns, category performance, customer value, demographics, and data quality checks (duplicates, nulls, and Quantity × Price_Per_Unit vs Total_Sale mismatches).

- **Outputs:** a set of reusable SQL views/queries that produce the KPIs and tables required to power a dashboard and management summaries.

Note: the SQL files contain the full query logic used to produce the results (aggregation, ranking, window functions, CASE buckets). Use those scripts to re-run or refresh numbers.

---

3) Headline KPIs (what to expect in the dashboard)

(These are the KPI definitions produced by the SQL scripts; the actual numeric values are available when the scripts are executed against the cleaned CSV.)

- **Total Revenue (2022)** — SUM(Total_Sale) across all validated transactions.

- **Total Transactions / Unique Orders** — COUNT(DISTINCT Transaction_Id).

- **Unique Customers** — COUNT(DISTINCT Customer_Id).

- **Average Order Value (AOV)** — AVG(Total_Sale).

- **Top 5 Categories by Revenue** — Category ranked by SUM(Total_Sale).

- **Top Customers (by revenue)**— Customer_Id, SUM(Total_Sale).

- **High-value transactions** — transactions where Total_Sale exceeds the chosen threshold (e.g., 1000).

- **Data Quality Indicators** — # duplicates removed, # rows with nulls handled, % transactions with Total_Sale mismatch.

---

4) Key findings

- **Revenue concentration & top categories:** A small number of categories contribute the majority of revenue (SQL: category-level SUM(Total_Sale) and cumulative contribution). These are priority areas for inventory and promotional focus.

- **Seasonality & monthly peaks:** Sales show clear month-to-month variation; the SQL time-series queries identify peak months and months with weaker demand — useful for seasonal stocking and promotions.

- **Time-of-day & shift patterns:** Transaction counts and revenue cluster in distinct shift buckets (morning / afternoon / evening) — operational staffing and fulfilment capacity should align with these peaks.

- **Customer value distribution:** A Pareto-like distribution appears: the top X% of customers generate disproportionate revenue (SQL: customer LTV and ranking). These customers are candidates for loyalty and retention programs.

- **Demographic preferences:** Age groups and gender segments display different category preferences and average spends (SQL: GROUP BY Gender, Age_Group, Category). Use these insights to sharpen marketing creatives and audience targeting.

- **High-value orders & exceptions:** The dataset contains a measurable count of high-value transactions; these may represent bulk/wholesale buyers or require fraud/fulfilment attention.

- **Data quality impact:** The raw file contained duplicates and nulls that materially affected counts and aggregates. Cleaning reduced noise but also changed sample size — always present cleaned vs raw counts in reports to show the effect of cleaning.

- **Pricing / calculation mismatches:** A non-trivial subset of rows exhibited mismatches between Quantity × Price_Per_Unit and recorded Total_Sale — these were flagged for review and either corrected or excluded depending on the case.

---

5) Business implications & recommended actions (quick wins → strategic)

Quick wins (immediate, high ROI)

1. **Prioritise inventory for the top 3 revenue-driving categories** (as identified by the SQL category ranking). Ensure safety stock before identified peak months.

2. **Target top customers (top 5–10%) with retention offers** — small increases in retention among top customers will disproportionately increase revenue.

3. **Operational capacity alignment** — schedule fulfilment and staffing to match the shift/time-of-day revenue peaks found in the analysis.

Medium-term (process + policy changes)
4. **Fix pricing / data-entry errors** where Total_Sale mismatches appear — implement a validation rule at data intake (or mark and exclude such rows from reporting until resolved).
5. **Run occasion-based promotions** ahead of months identified as seasonal peaks; use A/B tests to measure lift.

6. **Introduce high-value-customer playbook** (personalised offers, priority fulfilment, account manager for B2B-like buyers).

Strategic (longer-term)

7. **SKU-level margin tracking** (augment dataset with SKU and cost fields) to move from revenue-only to profitability optimization (similar approach suggested in the Ecommerce reference).

8. **Automate data pipeline & dashboard refresh** (materialized views for heavy aggregates, scheduled SQL jobs) so leadership has timely insights.

9. **Forecasting and replenishment models** using the monthly trends detected in the SQL time-series outputs.

---

6) Deliverables produced (and where to find them)

- **Data cleaning SQL script** — Retail Sales Analysis Data Cleaning.sql (duplicates removal, null handling, type corrections, derived fields).

- **Exploratory & analytic SQL script** — Retail Sales Data Exploration & Analysis.sql (10 questions implemented as queries / views; produce the KPIs listed above).

- **Suggested SQL views** for direct dashboard consumption (monthly revenue, category performance, top customers, shift sales, high-value txns, data quality summary).

---

7) Data quality notes (important)

- The cleaning script documents the **number of duplicate rows removed** and **number of rows with nulls** in critical fields (date/time, customer id, price fields). Always present "cleaned N / raw N" when showing KPIs.

- Flagged transactions with Total_Sale calculation mismatches require domain decisions (fix price, correct quantity, or exclude).

---

8) Next steps & handover checklist

1. **Execute the analytic SQL scripts against the cleaned dataset** to populate final KPI numbers and export them to the dashboard tool of choice (Power BI / Tableau / Excel).

2. **Populate dashboard mock-up** with the headline KPIs and the following visuals: KPI cards, monthly trend line, category bar chart (with cumulative %), top customers table, demographic heatmap, shift-level line/bar, high-value transactions table, and data quality widget. (The Ecommerce summary provides a strong visual template for layout and recommendations.)

3. **Review flagged mismatches** with domain stakeholders (finance/sales) and decide correction vs exclusion rules.

4. **Schedule a handover** (walkthrough of SQL scripts, views, and the dashboard wireframe) and provide short documentation for scheduled refresh.

---

9) Conclusion (one paragraph)

The SQL work you completed transforms noisy transactional data into a repeatable analytical pipeline that answers the project's core business questions: who buys, what they buy, when they buy, and which categories drive revenue. Cleaning exposed key data-quality gaps (duplicates and mismatches) that were resolved or flagged, and the 10 analytic queries produce the KPIs required for inventory, marketing, and finance decisions. Implementing the recommended quick wins and automating the pipeline will convert these insights into measurable revenue and margin improvements.