**✅ 📄 Expanded Summary (Professional Paragraph)**

In this end-to-end data analysis project, I used advanced SQL to explore and analyze customer and loan data for a financial institution. The objective was to extract insights that could help the business improve loan recovery rates, identify high-risk customers, and optimize repayment strategies. I performed descriptive, diagnostic, and predictive-style analysis using queries that included aggregations, conditional logic (CASE WHEN), subqueries, and window functions. Key tasks included profiling defaulters, analyzing EMI repayment behavior, calculating risk scores, and implementing the Pareto 80/20 principle to identify customers contributing to the majority of unpaid dues. Additionally, I segmented customers by gender, marital status, and employment type to understand trends across demographics. The output provides a comprehensive view of credit performance and can support strategic decision-making around credit policies, customer targeting, and default mitigation. This project showcases my ability to translate raw data into actionable business intelligence using SQL, and is designed to demonstrate both technical proficiency and domain understanding to potential employers.

**Key Points**

**1. Descriptive & Aggregation Analysis**

* Counted total number of **loans issued** to understand lending volume.
* Aggregated total **loan amount** and **EMI amount** by gender and employment type to spot lending trends.
* Grouped customers by **repayment status** to analyze credit behavior (Defaulted, Late, On Time).
* Identified the most common **loan purposes** and most frequent **recovery actions** used.

**2. Customer Segmentation & Profiling**

* Segmented customers by **marital status**, **gender**, and **employment type** for behavioral comparison.
* Performed **gender-wise defaulter count** to evaluate if risk varies across male/female borrowers.
* Found **top 10 customers** with the highest unpaid dues for direct recovery targeting.
* Identified customers who **paid 0 EMIs** or had **very low EMI compliance (<30%)** — critical for early intervention.

**3. Risk & Recovery Insights**

* Calculated **recovery rate** per customer: (EMIs Paid / Total EMIs) to measure loan performance.
* Designed a **simple risk score** formula: (Due Amount / Loan Amount \* 100) to rank customers by risk.
* Profiled combinations of **employment type + loan purpose** most prone to default — helpful for risk-based pricing or approval decisions.

**4. Temporal & Behavioral Trends**

* Extracted **month-wise trend** of loan defaults using DATE\_PART() — to identify seasonal or cyclical risk patterns.
* Compared **default, late, and on-time payment rates** across **marital statuses** using multiple CASE WHEN conditions — useful for socio-demographic insights.

**5. Advanced Business Insights (Pareto Analysis)**

* Applied the **Pareto 80/20 Rule**: Identified the smallest segment of customers (e.g., ~20%) responsible for the largest share (~80%) of unpaid dues — critical for resource allocation in recovery.

**Technical Highlights (What Tools/Skills You Used)**

* SQL concepts: JOIN, GROUP BY, HAVING, CASE WHEN, ROUND(), DATE\_PART(), WINDOW FUNCTIONS
* Cleaned and merged datasets using **customer\_id** for relational analysis
* Used **window functions** to calculate cumulative sums for Pareto insights
* Performed **cross-table joins** between loan and customer data to enable multi-dimensional analysis

**📊 Business Value Created**

* Identified high-risk segments and behaviors contributing to most unpaid loans
* Provided actionable insights to **optimize recovery strategy** and **reduce future defaults**
* Enabled data-driven **credit decisioning and segmentation**
* Helped visualize **operational bottlenecks** in EMI collection and customer follow-up