Root Cause Analysis on Sales of a Coffee Franchise

ABSTRACT

Coffee franchises operate across multiple cities, managers, and product categories, generating large volumes of transactional and operational data. However, identifying the true drivers behind sales growth or decline remains a challenge due to fragmented analysis and reliance on traditional reporting. This project presents a data-driven Root Cause Analysis (RCA) framework for a coffee franchise using Python-based analytics. The framework aggregates and normalizes key business metrics such as total sales, profit margin, revenue growth, staffing efficiency, and product performance. Weighted scoring models are applied to evaluate cities, managers, products, and franchises, while statistical tests validate observed patterns. The proposed system converts raw coffee franchise data into actionable insights that support strategic decision-making and operational optimization.

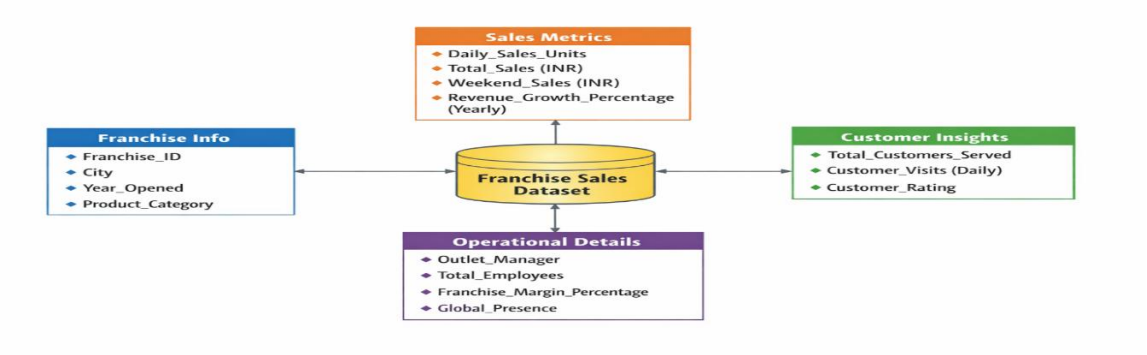
PROBLEM STATEMENT

Coffee franchises often experience uneven performance across locations despite offering similar products and pricing strategies. Management struggles to understand why certain cities perform well while others underperform, or why sales decline despite stable customer ratings. Existing analyses focus mainly on sales totals and growth trends, failing to account for margins, staffing inefficiencies, managerial effectiveness, and product mix. Without a structured Root Cause Analysis framework, decisions are largely assumption-based. Therefore, there is a need for an automated and statistically validated analytical system that can identify the root causes behind performance variations in a coffee franchise.

OBJECTIVES

The objective of this project is to design a Root Cause Analysis system for evaluating coffee franchise performance using data analytics. The goals include cleaning and structuring franchise sales data, normalizing key performance indicators, developing weighted scoring models for cities, managers, products, and franchises, identifying top and bottom performers, analyzing operational inefficiencies such as staffing imbalance, validating insights using statistical tests, and presenting findings in a business-friendly format to support data-driven decisions.

DATA MODEL



EXISTING SYSTEM

The existing system used in most coffee franchises relies on manual sales reports, spreadsheet-based summaries, and basic year-over-year comparisons. Performance evaluation is limited to total revenue and growth percentages, with little focus on margins, product contribution, or staffing efficiency. Manager performance is often judged subjectively, and product-level issues are rarely analyzed in detail. These methods provide descriptive insights but do not explain the underlying causes of performance changes. To pinpoint operational inefficiencies, a negative performance scoring model was designed to identify the weakest franchises.

DISADVANTAGES OF THE EXISTING SYSTEM

The traditional system has several drawbacks.

- Metrics are not normalized, leading to unfair comparisons between large and small outlets.
- There is no structured mechanism to identify root causes behind sales decline.
- Staffing inefficiencies, manager impact, and product-level performance are largely ignored. Additionally, the absence of statistical validation means that observed trends may be coincidental rather than meaningful. As a result, corrective actions are often reactive and ineffective.

PROPOSED SYSTEM

We propose an automated Root Cause Analysis framework tailored for a coffee franchise, implemented using Python. The system preprocesses raw franchise data, calculates key metrics such as total sales, average margin, revenue growth, daily sales units, and employee efficiency, and normalizes them using Min–Max scaling. Weighted scoring models are applied to evaluate city-level profitability, manager effectiveness, product performance, and franchise efficiency. Negative scoring is used to identify worst-performing franchises. Statistical tests such as

Pearson correlation, ANOVA, and Chi-square analysis are applied to validate findings, ensuring reliable and data-backed insights.

ADVANTAGES OF THE PROPOSED SYSTEM

- The proposed system provides deep, root-cause-focused insights rather than surface-level metrics.
- It enables multi-dimensional analysis across cities, managers, products, and staffing levels. Objective scoring models eliminate bias and ensure fair ranking.
- The framework is scalable and can be applied to any coffee franchise chain.
- By linking insights directly to operational actions, the system supports strategic planning, cost control, and performance improvement.

METHODOLOGIES

Data Preprocessing

Raw coffee franchise data often contains missing values, duplicates, and inconsistencies that can affect analytical accuracy. Data preprocessing techniques such as missing value imputation, duplicate removal, and outlier handling are applied to improve data quality. This step ensures that the dataset is reliable and suitable for further statistical and analytical processing.

Feature Engineering

Feature engineering is used to derive meaningful business indicators from raw transactional data. Metrics such as total sales, profit margin, revenue growth, average daily sales units, and employee-to-sales ratio are computed. These features are selected because they capture financial performance, operational efficiency, and growth behavior, which are essential for identifying root causes of franchise performance variations.

Normalization

Since performance metrics exist on different scales, Min–Max normalization is applied to standardize all variables into a common range. This prevents metrics with larger numerical values from dominating the analysis and ensures fair comparison across cities, managers, products, and franchises. Normalization is critical for objective scoring and ranking.

Weighted Scoring

Composite performance scores are calculated using weighted scoring methods aligned with business priorities. Sales, margin, and growth metrics are assigned predefined weights to compute City Performance Score, Manager Performance Score, and Product Performance Score. This methodology enables objective ranking and comparison of entities based on multiple performance dimensions rather than a single metric.

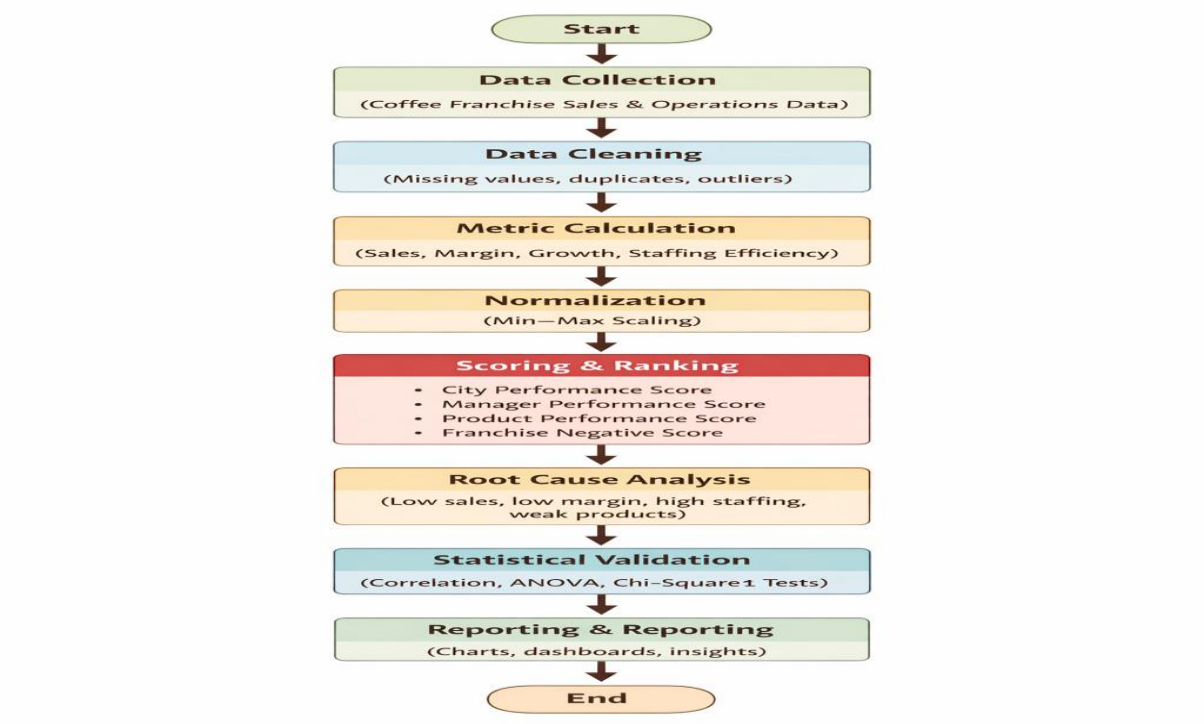
Negative Scoring for Worst Franchise Identification

To identify poorly performing franchises, a negative scoring methodology is applied. Franchises are penalized for low sales, low margins, poor growth, and high employee counts. This approach highlights operational inefficiencies and helps prioritize franchises requiring corrective action.

Root Cause Analysis

Root cause analysis is performed by examining performance contrasts between top-performing and underperforming entities. Indicators such as low daily sales units, weak margins, inefficient staffing levels, and poor product contribution are analyzed to explain why performance deviations occur. This methodology moves beyond descriptive analysis to causal interpretation.

WORKFLOW



MODULES

Data Collection

The workflow begins with the collection of raw data from the coffee franchise system. This includes city-wise sales data, product-level transactions, manager information, employee counts, profit margins, and revenue growth records. The collected data serves as the primary input for the analysis and represents real operational performance across the franchise network.

Data Cleaning and Preprocessing

After data collection, the dataset undergoes cleaning and preprocessing to improve data quality. Missing values are handled using appropriate techniques, duplicate entries are removed, and extreme outliers are identified and treated. This step ensures that the dataset is consistent, accurate, and suitable for analytical processing.

Metric Calculation

In this stage, key performance metrics are calculated from the cleaned data. Metrics such as total sales, average profit margin, revenue growth rate, daily sales units, and staffing efficiency are derived. These indicators provide a comprehensive view of financial performance, operational efficiency, and growth trends within the coffee franchise.

Normalization of Metrics

Since the calculated metrics exist on different scales, normalization is applied using Min–Max scaling. This transforms all values into a common range, allowing fair comparison between franchises, cities, managers, and products. Normalization plays a critical role in eliminating scale bias and enabling objective scoring.

Root Cause Analysis

Root Cause Analysis is performed by examining low-scoring entities and identifying contributing factors. Issues such as low daily sales units, poor margins, high employee counts relative to sales, and weak product performance are analyzed. This step explains *why* certain franchises or cities underperform rather than just highlighting *what* the performance is.

Statistical Validation

To ensure the reliability of the insights, statistical validation is conducted. Pearson correlation analysis checks relationships between variables such as customer rating and sales. ANOVA tests examine sales differences across cities, while Chi-square tests evaluate the influence of product categories on revenue growth. These tests confirm that the observed patterns are statistically significant.

Visualization and Reporting

The validated insights are presented using visual tools such as charts, graphs, and summary tables. Visualization helps stakeholders easily interpret trends, rankings, and root causes. This step bridges the gap between technical analysis and business understanding.

Decision Support and Business Actions

The final step converts analytical insights into actionable recommendations. Management can take informed decisions related to staffing optimization, manager training, product portfolio improvement, and franchise restructuring. This ensures that the workflow directly supports strategic and operational improvements within the coffee franchise.

CONCLUSION

This project demonstrates how Root Cause Analysis can be effectively applied to a coffee franchise using data analytics. The proposed framework successfully identifies performance drivers and bottlenecks at multiple levels, including cities, managers, products, and franchises. By integrating normalization, scoring, and statistical validation, the system provides accurate and actionable insights. The results highlight the importance of efficient management, optimal staffing, and product mix in driving coffee franchise success.

FUTURE SCOPE

The framework can be enhanced by integrating real-time sales data for continuous monitoring. Machine learning models can be introduced to predict future sales trends and identify early warning signals. External factors such as festivals, and competitor activity can be incorporated for deeper insights and bringing up business plans. Deployment as a cloud-based dashboard would improve scalability and accessibility. Prescriptive analytics can also be added to recommend corrective actions automatically.