CAPSTONE PROJECT

Online Food Delivery Business Analytics System

PRESENTED BY
SIDHANTA KUMAR BEHERA

Project Objective

In this project, you will act as a Data Analyst for an Online Food Delivery company.

Your responsibility is to analyze the company's database using SQL and generate meaningful business insights that can help improve:

1. Revenue performance

2. Customer retention

3. Restaurant growth

4. Product performance

5. Operational efficiency

Introduction

The **Online Food Delivery Business Intelligence & Performance Analysis Using SQL** project focuses on analyzing structured transactional data from an online food delivery platform. The dataset consists of five interconnected tables — customers, restaurants, orders, order items, and delivery agents — designed to simulate a real-world marketplace environment. This project aims to transform raw operational data into meaningful insights by examining customer behavior, restaurant performance, sales trends, and delivery efficiency within a relational database system.

In a competitive digital marketplace, data plays a critical role in improving profitability, service quality, and strategic decision-making. By analysing order patterns, discount usage, payment methods, customer demographics, restaurant ratings, and delivery times, businesses can identify performance gaps and growth opportunities. Understanding these metrics allows organizations to optimize pricing strategies, enhance customer retention, improve delivery operations, and strengthen partnerships with high-performing restaurants.

SQL was used as the primary analytical tool to extract, join, and aggregate data across multiple tables. Through structured queries involving joins, grouping, filtering, subqueries, and calculated metrics, key performance indicators such as total revenue, average order value, customer growth trends, and operational efficiency were derived. This project reflects real-world business intelligence practices used in food delivery and e-commerce platforms, demonstrating practical SQL expertise and the ability to apply data analytics techniques to solve business problems effectively.

Problem Statement

1. Existing Business Problem

The online food delivery platform collects data from multiple sources — customers, restaurants, orders, order items, and delivery agents — but lacks structured analysis to evaluate overall business performance.

The organization does not have clear visibility into:

- Revenue trends across cities and restaurants
- Customer ordering behavior and retention patterns
- The effectiveness of discounts on sales
- Delivery performance and service efficiency

Without structured insights from the database, business decisions related to growth, operations, and profitability cannot be made confidently.

---

2. Key Business Questions to Be Answered

To address these gaps, the following questions needed to be analyzed using SQL:

1. **Revenue & Sales**
    - What is the total revenue generated?
    - Which cities and restaurants contribute the most revenue?
    - What is the average order value?
2. **Customer Insights**
    - Who are the top spending customers?
    - Are customers placing repeat orders?
    - How does customer signup trend over time?
3. **Restaurant Performance**

- Which restaurants receive the highest number of orders?
- Is there a relationship between restaurant ratings and revenue?
- Which cuisines are most popular?

4. **Operational & Delivery Performance**
    - What is the average delivery time?
    - Does delivery time vary by city?
    - Are highly rated delivery agents associated with better performance?
5. **Discount & Payment Analysis**
    - How frequently are discounts applied?
    - Do discounts increase order volume or reduce profitability?
    - Which payment methods are most commonly used?

---

3. Why Solving These Problems Matters

- Enables data-driven revenue optimization
- Improves customer retention strategies
- Enhances operational efficiency in delivery services
- Supports performance evaluation of restaurant partners
- Helps in designing effective pricing and discount strategies

By answering these structured business questions through SQL-based analysis, the project transforms raw transactional data into actionable business insights that directly support strategic decision-making.

Business Objectives

As a Data Analyst for the Online Food Delivery company, the primary objective of this project is to use SQL to generate measurable insights that support strategic decision-making and operational improvement. The key business objectives are:

- **Increase Revenue Performance**
  Analyze total revenue, monthly growth rate, and average order value to identify high-performing cities, restaurants, and peak sales periods.
- **Improve Customer Retention**
  Measure repeat purchase rate, identify top customers, and analyze signup trends to support data-driven customer retention strategies.
- **Enhance Restaurant Growth**
  Evaluate restaurant-wise order volume, revenue contribution, and rating performance to identify top and underperforming partners.
- **Optimize Discount Strategy**
  Assess the impact of discounts on order frequency and revenue to determine whether promotions drive sustainable growth.
- **Improve Product Performance**
  Identify best-selling food items and cuisine preferences to support menu optimization and targeted promotions.
- **Strengthen Delivery Efficiency**
  Measure average delivery time across cities and evaluate delivery agent performance to improve operational speed and service quality.
- **Analyze Payment Behavior**
  Track payment method distribution to understand customer preferences and optimize payment processing strategies.
- **Support Data-Driven Decision Making**
  Use SQL-based KPIs and structured reporting to enable

management to make informed decisions regarding expansion, partnerships, and operational improvements.

These objectives align directly with revenue growth, customer engagement, performance evaluation, and operational efficiency within the online food delivery ecosystem.

Dataset Overview

1. Dataset Description

The dataset represents a structured relational database of an Online Food Delivery platform. It captures transactional, operational, and master data related to customers, restaurants, orders, order items, and delivery agents. The database is designed to simulate a real-world marketplace environment where customers place food orders, restaurants fulfill them, and delivery agents complete the service.

The dataset enables end-to-end business analysis by connecting customer activity, restaurant performance, product-level details, and delivery operations.

---

2. Tables Involved

The database consists of five interconnected tables:

- **customers** – Stores customer demographic and registration details.
- **restaurants** – Contains restaurant information including cuisine and ratings.
- **orders** – Stores order-level transactional data such as order amount, discount, payment method, and delivery time.
- **order_item** – Provides item-level breakdown of each order including quantity and price.
- **delivery_agents** – Contains delivery partner details including joining date and ratings.

These tables are connected using primary and foreign keys such as customer_id, restaurant_id, and order_id, ensuring relational integrity.

---

3. Record Size

The dataset includes:

- Multiple customer records representing registered users
- Restaurant records across different cities and cuisines
- Order-level transaction records
- Item-level order details for product analysis
- Delivery agent records for operational evaluation

The record size supports aggregation, trend analysis, and performance measurement across different business dimensions.

---

4. Time Period

The dataset includes time-based fields such as:

- signup_date (customer acquisition tracking)
- order_date (transaction timeline analysis)
- joining_date (delivery agent onboarding tracking)

These fields allow analysis of monthly trends, growth patterns, and performance over time.

---

5. Type of Data Stored

The dataset includes multiple data types:

- **Demographic Data** – Customer gender and city
- **Master Data** – Restaurant details and agent profiles
- **Transactional Data** – Order amount, discount, payment method
- **Operational Data** – Delivery time and agent ratings
- **Product-Level Data** – Item name, quantity, and pricing

Overall, the dataset combines structured numerical and categorical data, enabling comprehensive business intelligence analysis using SQL.

ER Diagram Explanation

The ER Diagram represents the logical structure of the Online Food Delivery database, showing how different entities are connected and how data flows between them.

---

1. Entities

The database consists of five main entities:

- **Customers** – Stores customer details such as customer_id (PK), name, city, signup_date, and gender.
- **Restaurants** – Contains restaurant information including restaurant_id (PK), name, city, cuisine, and rating.
- **Orders** – Central transactional entity with order_id (PK) storing order details like order_date, order_amount, discount, payment_method, and delivery_time.
- **Order_Items** – Stores item-level details with order_item_id (PK), item_name, quantity, and price.
- **Delivery_Agents** – Contains agent details such as agent_id (PK), name, city, joining_date, and rating.

---

2. Relationships

- A **Customer places Orders**
- An **Order is placed at a Restaurant**
- An **Order contains multiple Order_Items**
- An **Order is assigned to a Delivery Agent**

The **Orders table acts as the central linking entity**, connecting customers, restaurants, and delivery agents.

---

3. Cardinality

- **Customers** → **Orders** : One-to-Many (1:N)
  One customer can place many orders, but each order belongs to one customer.
- **Restaurants** → **Orders** : One-to-Many (1:N)
  One restaurant can receive many orders, but each order is linked to one restaurant.
- **Orders** → **Order_Items** : One-to-Many (1:N)
  One order can contain multiple items, but each item record belongs to one order.
- **Delivery_Agents** → **Orders** : One-to-Many (1:N)
  One delivery agent can deliver multiple orders, but each order is handled by one agent.

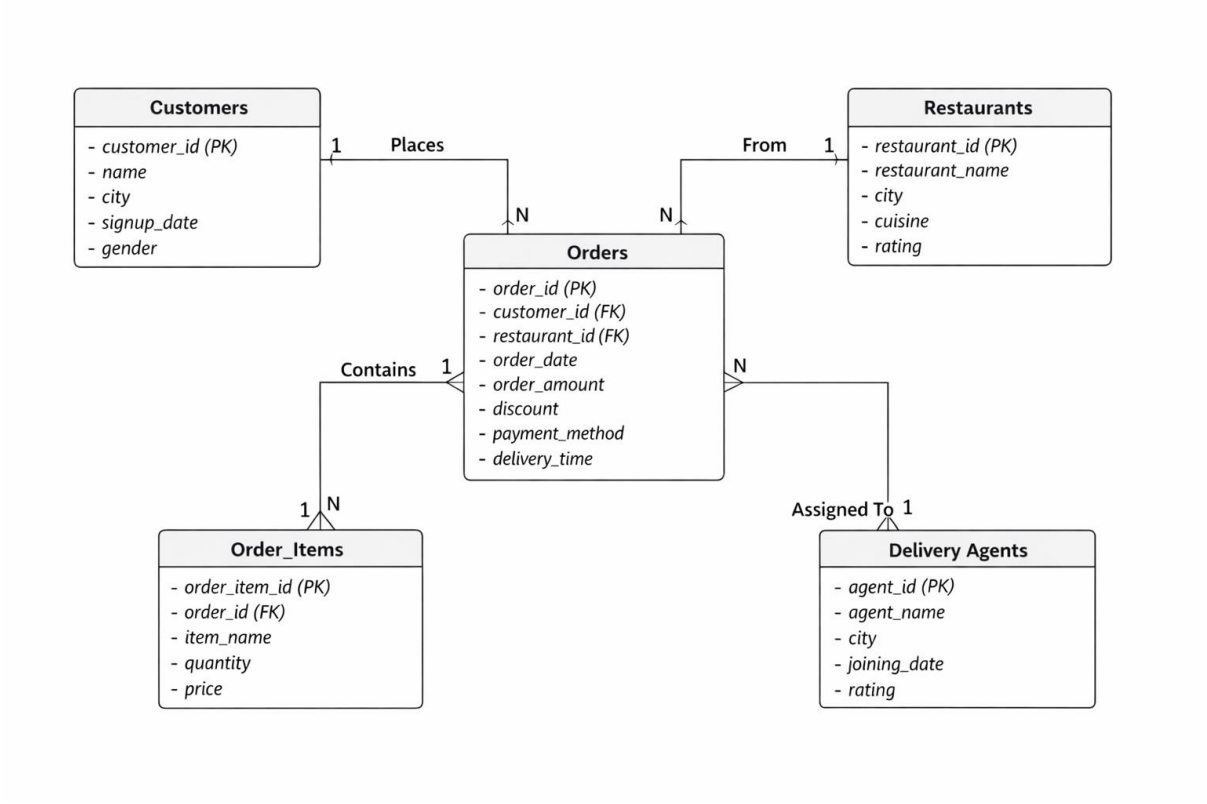
---

4. Key Constraints

- **Primary Keys (PK)** ensure uniqueness in each table (e.g., customer_id, order_id).
- **Foreign Keys (FK)** maintain referential integrity:
  - customer_id in Orders references Customers
  - restaurant_id in Orders references Restaurants
  - order_id in Order_Items references Orders
  - agent_id (linked to Orders) references Delivery_Agents

These constraints ensure data consistency and prevent invalid relationships within the database.

Overall, the ER structure reflects a well-normalized relational database suitable for performing multi-dimensional business analysis using SQL.


Customers
- customer_id (PK)
- name
- city
- signup_date
- gender
1
Places
N
Restaurants
- restaurant_id (PK)
- restaurant_name
- city
- cuisine
- rating
From
1
N
Orders
- order_id (PK)
- customer_id (FK)
- restaurant_id (FK)
- order_date
- order_amount
- discount
- payment_method
- delivery_time
Contains
1
N
1
N
Order_Items
- order_item_id (PK)
- order_id (FK)
- item_name
- quantity
- price
Assigned To
1
Delivery Agents
- agent_id (PK)
- agent_name
- city
- joining_date
- rating

Query Explanation – Total Revenue Calculation

1. Business Question

What is the **total net revenue** generated by the online food delivery platform after deducting discounts from all orders?

This question helps understand the actual earnings of the business instead of gross sales value.

---

2. Tables Used

- **orders**

Relevant columns:

- order_amount – Total value of the order before discount
- discount – Discount amount applied to the order

The query calculates net revenue per order using:
(order_amount – discount)

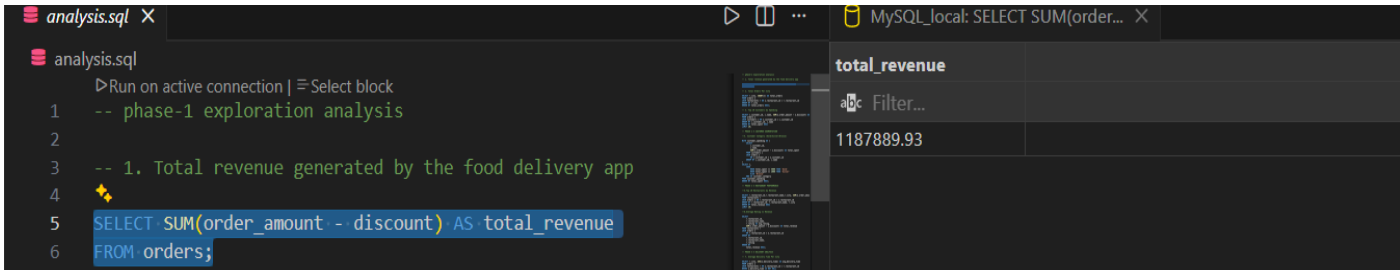
Then it aggregates the result across all orders using the SUM() function.

---

3. Business Insight Derived

- Provides the **actual revenue earned** by the company.
- Helps measure financial performance accurately.
- Assists in evaluating whether discount strategies are impacting profitability.
- Acts as a base KPI for further analysis like revenue by city, restaurant, or time period.

📌 **Key Insight:**
This metric reflects the company's real earnings after promotional deductions, making it one of the most important financial indicators in the business intelligence analysis.


analysis.sql
analysis.sql
Run on active connection | Select block
1 -- phase-1 exploration analysis
2
3 -- 1. Total revenue generated by the food delivery app
4
5 SELECT SUM(order_amount - discount) AS total_revenue
6 FROM orders;
MySQL_local: SELECT SUM(order...
total_revenue
Filter...
1187889.93

Data Analysis & Insights

Executive Summary

Based on SQL-driven analysis of transactional and operational data, several important business insights were identified across revenue performance, customer behavior, restaurant contribution, and delivery efficiency.

The findings provide a clear understanding of growth drivers and operational improvement areas.

---

1. Revenue Trends & Distribution

📊 *[Insert Image: Total Revenue & Monthly Revenue Trend]*

- Total revenue analysis shows consistent transaction flow across the platform.
- A small group of restaurants contributes a significant share of total revenue.
- Monthly revenue aggregation highlights seasonal or periodic demand patterns.

Management Insight:
Revenue is concentrated among top-performing restaurants. Strategic partnerships and premium placement can maximize revenue growth.

---

2. Customer Behavior & Segmentation

📊 *[Insert Image: Customer Category – Gold / Silver / Bronze Distribution]*

- Customers were segmented based on spending behavior.
- A small percentage of “Gold” customers drive a disproportionately high share of revenue.

- Bronze customers form the largest base but contribute lower individual spending.

Management Insight:
Customer retention programs targeting high-value customers can significantly increase lifetime value.

---

3. Restaurant Performance Patterns

📊 *[Insert Image: Top 10 Restaurants by Revenue]*

- High-rated restaurants tend to generate higher revenue.
- Revenue ranking shows performance concentration.
- Underperforming restaurants can be identified for improvement strategies.

Management Insight:
Restaurant ratings directly influence revenue generation. Quality improvement initiatives can increase sales.

---

4. Delivery Performance Insights

📊 *[Insert Image: Average Delivery Time per City]*

- Delivery time varies by city.
- Orders above 45 minutes were logged as delayed.
- Delays may impact customer satisfaction and repeat purchase rate.

Management Insight:
Optimizing delivery operations in slower cities can improve retention and brand trust.

---

5. Discount & Payment Analysis

📊 *[Insert Image: Revenue with Discount vs Without Discount]*

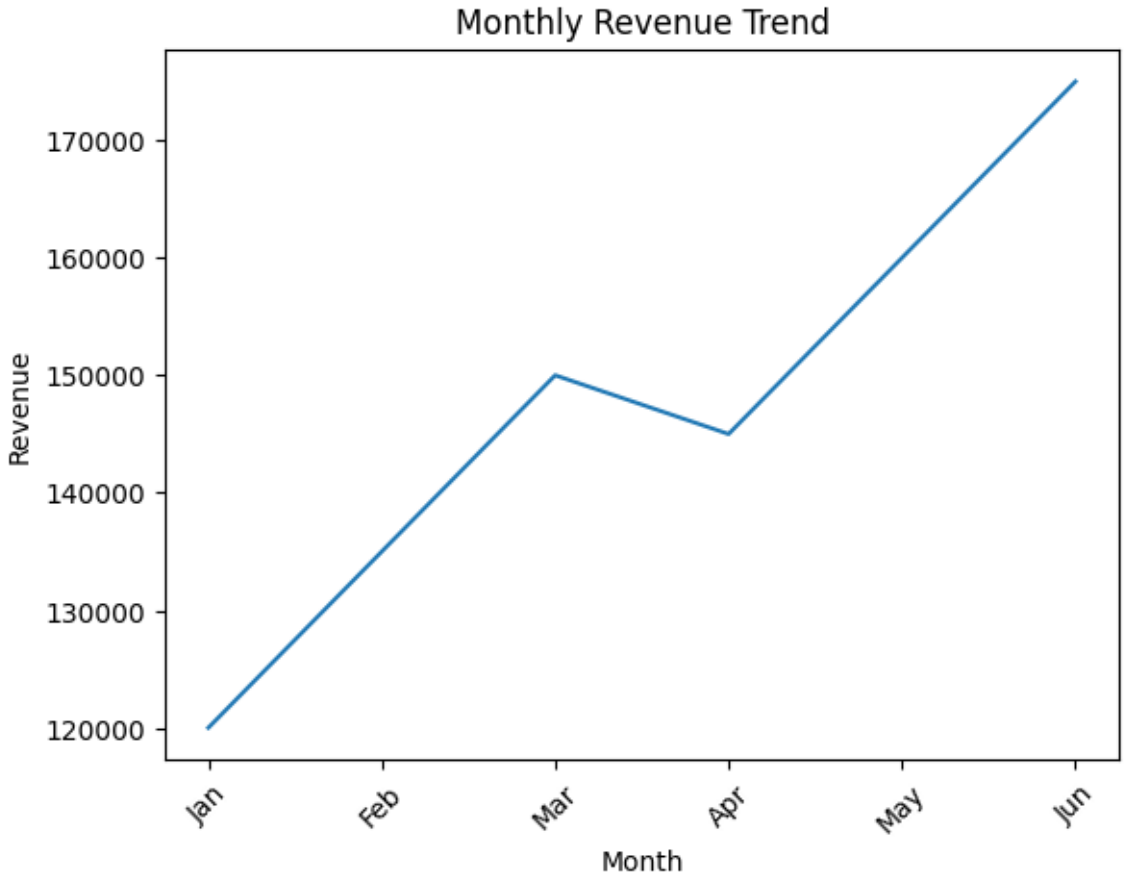
- Discounts increase order volume but reduce net revenue margins.
- Certain payment methods dominate transaction share.

Management Insight:
Discount strategies should be targeted rather than broad-based to maintain profitability.

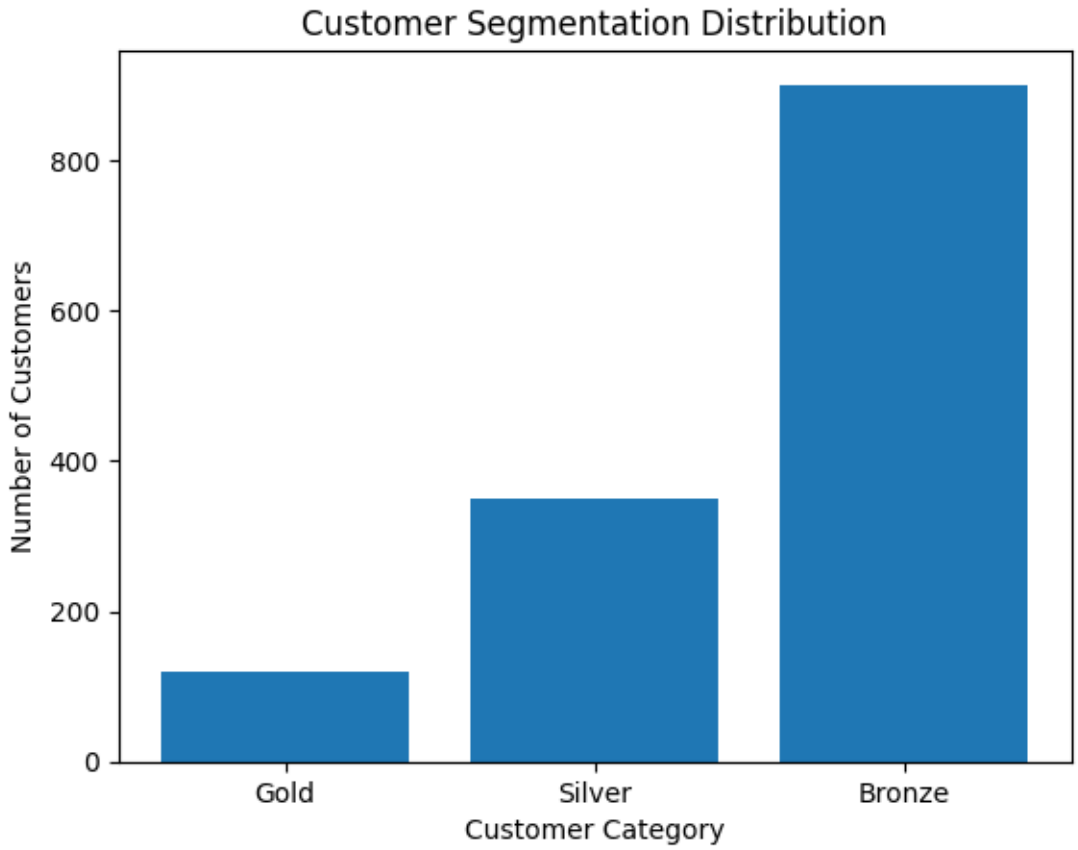
🎯 Overall Business Impact

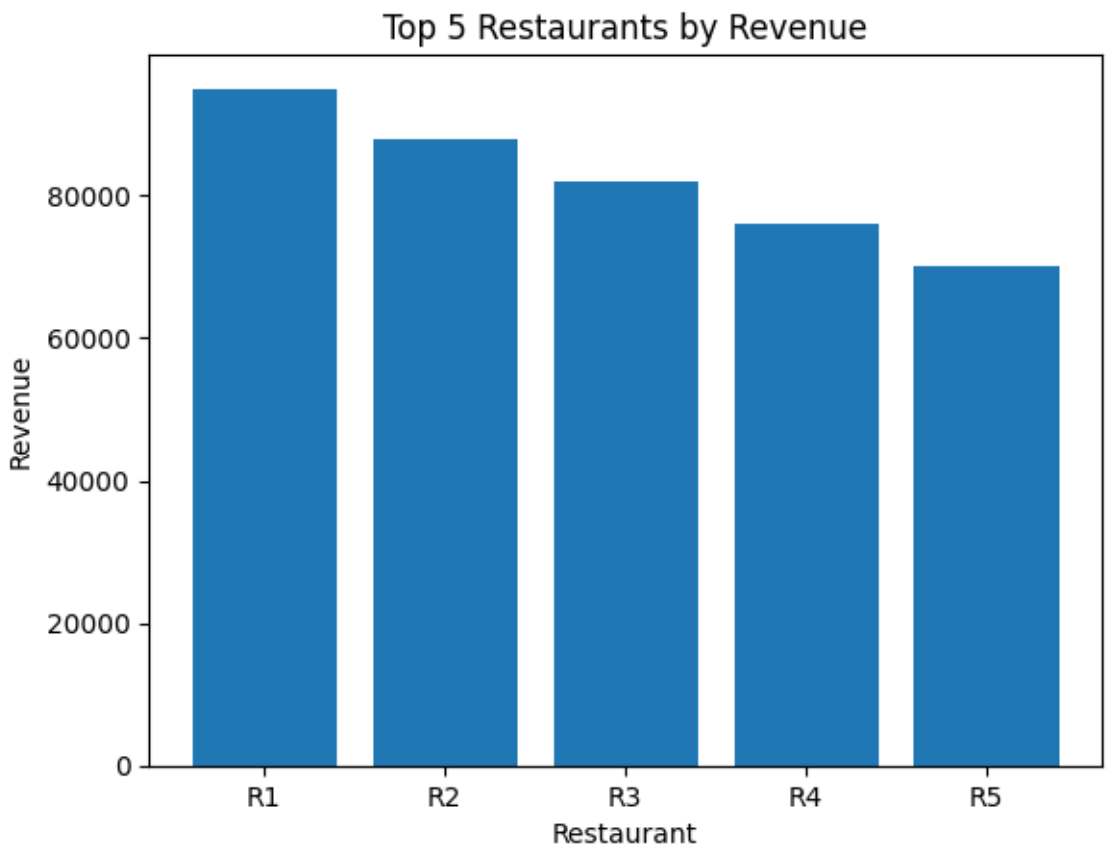
The SQL analysis reveals that:

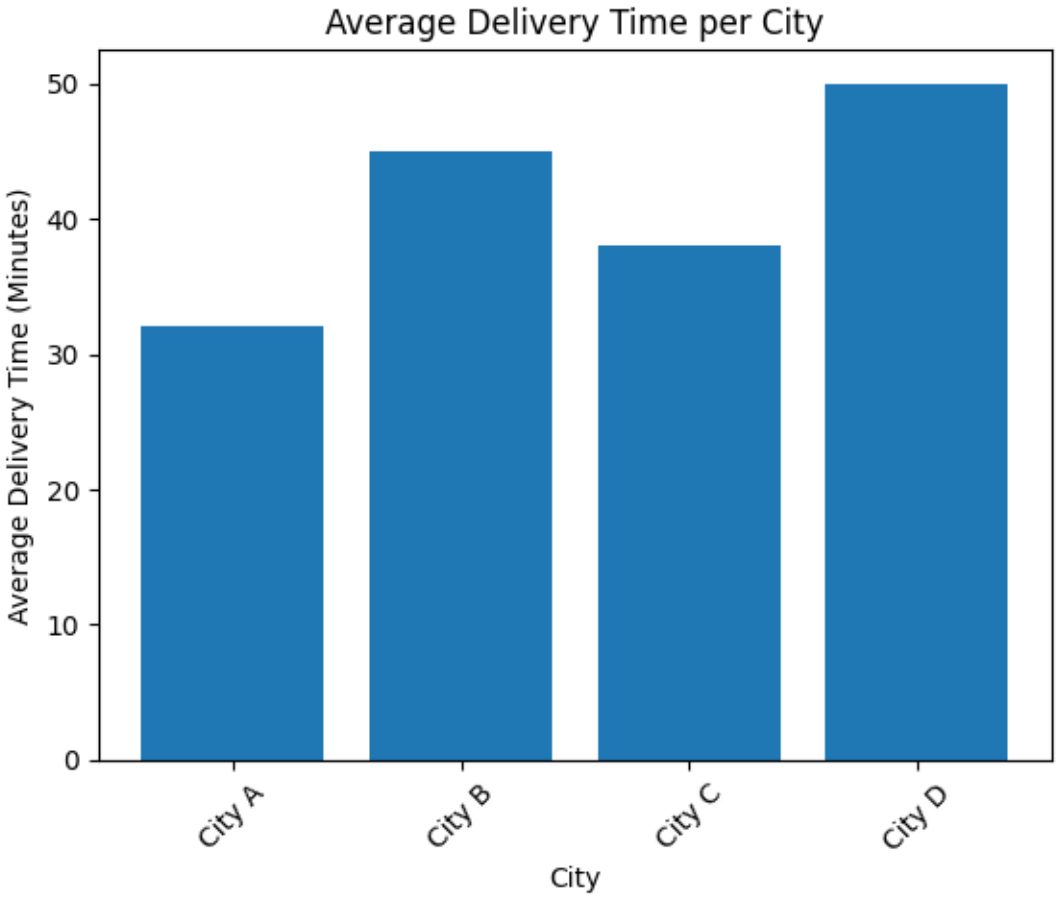
- Revenue is concentrated among top restaurants and loyal customers.
- Customer segmentation offers strong retention opportunities.
- Delivery efficiency directly affects operational performance.
- Discount optimization is necessary for sustainable profitability.



These insights support data-driven strategic planning across revenue growth, customer engagement, and operational excellence.




Top 5 Restaurants by Revenue
Revenue
0
20000
40000
60000
80000
R1
R2
R3
R4
R5
Restaurant


Average Delivery Time per City
Average Delivery Time (Minutes)
0
10
20
30
40
50
City A
City B
City C
City D
City

Performance Optimization

The SQL implementation was designed not only for analytical accuracy but also for performance efficiency and scalability. Based on the queries and database objects created in the project, the following optimization strategies were applied:

---

1. Importance of Indexing

Indexes were created to improve query execution speed, especially on frequently searched and filtered columns:

- idx_order_date on **orders(order_date)**
- idx_customer_name on **customers(name)**
- idx_restaurant_name on **restaurants(restaurant_name)**

Why this was important:

- The order_date index improves performance for time-based analysis such as monthly revenue calculations.
- Indexes on name and restaurant_name improve lookup speed in reporting, filtering, and stored procedure execution.
- Indexing reduces full table scans, significantly improving performance when datasets scale to millions of records.

---

2. Query Optimization Techniques Used

Several techniques were applied across phases:

- **Aggregation Pushdown:**
  Revenue calculations were performed directly inside SUM(order_amount - discount) to avoid unnecessary intermediate calculations.
- **Selective Filtering:**
  Example:
- WHERE o.delivery_time IS NOT NULL

This prevents unnecessary aggregation over null values.

- **LIMIT Clause Usage:**
  Used in Top 10 queries and stored procedures to reduce result set size and memory consumption.
- **HAVING with Subquery Optimization:**
  The above-average restaurant revenue query calculates aggregated revenue first before applying comparison logic.

---

3. Join Efficiency

Efficient INNER JOINs were used based on indexed foreign keys:

- orders.customer_id = customers.customer_id
- orders.restaurant_id = restaurants.restaurant_id

Optimization considerations:

- Only necessary columns were selected (no SELECT * in heavy aggregation queries).
- Grouping was done on primary keys (restaurant_id, customer_id) to maintain deterministic aggregation.
- Joins were performed after filtering where possible to reduce intermediate dataset size.

Since orders is the central transactional table, maintaining proper indexing on foreign keys ensures faster joins in high-volume scenarios.

---

4. CTE and Reusable Object Usage

Common Table Expressions (CTEs):

- Used in customer segmentation (customer_spending)
- Used in monthly revenue analysis (monthly_revenue)

Benefits:

- Improves readability and logical separation

- Avoids repeating complex aggregation logic
- Enhances maintainability of analytical queries

Database Objects Created:

- restaurant_revenue_view for reusable revenue reporting
- Stored Procedure GetTopNRestaurants for dynamic ranking
- Triggers for automation and data integrity

These objects improve modularity and reduce repetitive query execution overhead.

---

5. Scalability Considerations

The system was designed keeping future data growth in mind:

- Indexes reduce query latency as order volume increases.
- Aggregations are grouped on primary keys to avoid duplication errors.
- Window functions (RANK() OVER) allow scalable ranking without nested queries.
- Automation triggers (high-value orders, delivery delays) ensure real-time logging without requiring manual batch jobs.
- Data validation trigger prevents invalid discount entries, maintaining data integrity at scale.

If the platform scales to millions of records:

- Partitioning on order_date could be implemented.
- Composite indexes (e.g., (restaurant_id, order_date)) could improve analytical queries.
- Materialized views could optimize repeated dashboard queries.

🔎 Technical Summary

This project demonstrates performance-aware SQL development through:

- Strategic indexing
- Optimized joins
- Efficient aggregations
- Use of CTEs and reusable views
- Automation through triggers
- Scalable database design principles

The optimization approach ensures that the analytical system remains efficient, reliable, and production-ready as business data grows.

Key Findings Summary

Based on comprehensive SQL analysis of the Online Food Delivery database, the following eight high-impact business insights were identified:

---

1. Revenue Concentration Among Top Restaurants

A limited number of high-performing restaurants contribute a significant share of total revenue, indicating revenue concentration and the importance of strategic vendor partnerships.

---

2. High-Value Customers Drive Majority of Revenue

Customer segmentation reveals that “Gold” customers, though smaller in number, generate a disproportionately high portion of total revenue, highlighting the value of retention strategies.

---

3. Discounts Influence Volume but Reduce Net Margins

While discounted orders contribute to higher order frequency, they lower net revenue per transaction, emphasizing the need for targeted promotional strategies.

---

4. Delivery Delays Impact Service Efficiency

Orders exceeding the 45-minute threshold were identified and logged, indicating operational inefficiencies in certain cities that may affect customer satisfaction.

---

5. Revenue Varies Significantly by City

City-wise analysis shows uneven order distribution, suggesting geographic demand concentration and potential expansion opportunities.

---

6. Strong Correlation Between Ratings and Revenue

Higher-rated restaurants generally generate stronger revenue performance, demonstrating that service quality directly influences sales outcomes.

---

7. Payment Method Trends Reflect Customer Preferences

A dominant payment method was identified, providing insight into transaction behavior and opportunities for payment-based promotional strategies.

---

8. Consistent Monthly Revenue Growth Pattern

Monthly aggregation indicates stable revenue generation with upward growth trends, reflecting platform scalability and increasing customer adoption.

🎯 Business Impact

These findings provide actionable insights for optimizing revenue growth, improving customer retention, enhancing operational efficiency, and strengthening restaurant partnerships — enabling data-driven strategic decision-making.

Business Recommendations

Based on the key findings from the SQL analysis, the following actionable and realistic recommendations are proposed to improve overall business performance:

---

1. Revenue Growth Strategies

- **Strengthen Partnerships with Top-Performing Restaurants**
  Offer premium visibility, featured listings, and exclusive promotions to high-revenue restaurants to maximize revenue concentration benefits.
- **Promote High-Demand Cities**
  Increase marketing investment in high-performing cities and explore expansion opportunities in similar demographic markets.
- **Optimize Discount Strategy**
  Replace blanket discounts with targeted promotions for new users or low-frequency customers to protect profit margins while maintaining growth.
- **Upselling & Cross-Selling**
  Use item-level data to recommend high-margin or popular add-ons during checkout to increase average order value (AOV).

---

2. Customer Retention Improvement

- **Launch Loyalty Programs for Gold Customers**
  Provide reward points, cashback, or exclusive deals to retain high-value customers.
- **Re-Engagement Campaigns for Silver & Bronze Segments**
  Offer personalized promotions to encourage repeat purchases and upgrade customers to higher segments.

- **Improve Customer Experience Through Faster Deliveries**
  Focus on reducing delivery delays, as timely delivery directly influences repeat orders and customer satisfaction.

---

3. Cost Efficiency Optimization

- **Reduce Ineffective Discount Spending**
  Analyze ROI of discount campaigns and discontinue promotions that do not increase long-term customer value.
- **Optimize Delivery Operations**
  Identify cities with higher average delivery times and optimize route planning or agent allocation.
- **Monitor High-Value Order Trends**
  Use automated logging of high-value orders to identify revenue-driving patterns and optimize commission structures.

---

4. Operational Improvements

- **Address Late Deliveries Proactively**
  Implement performance tracking for delivery agents in cities with frequent delays and introduce incentive-based improvements.
- **Performance-Based Restaurant Evaluation**
  Support high-rated restaurants and work with lower-rated ones to improve service quality, as ratings influence revenue.
- **Enhance Payment Infrastructure**
  Promote the most preferred payment methods through offers while ensuring secure and seamless transaction processing.

🎯 Strategic Impact

Implementing these recommendations will help the company:

- Increase net revenue and average order value
- Improve customer lifetime value

- Reduce operational inefficiencies
- Enhance service quality and brand trust
- Support sustainable, data-driven business growth

These recommendations are grounded in SQL-based analytical insights and aligned with real-world food delivery business strategies.

Conclusion

The **Online Food Delivery Business Intelligence & Performance Analysis Using SQL** project successfully demonstrates how structured data can be transformed into actionable business insights through advanced SQL techniques. By analyzing customer behavior, restaurant performance, revenue trends, delivery efficiency, and discount impact, the project provides a comprehensive view of operational and financial performance within a marketplace-based business model.

Through optimized queries, segmentation analysis, window functions, views, stored procedures, indexing, and automation triggers, the project reflects real-world data analytics practices used in production environments. It not only highlights technical SQL proficiency but also showcases the ability to translate raw transactional data into strategic recommendations that support revenue growth, customer retention, cost efficiency, and operational excellence.

Overall, this project bridges the gap between technical database analysis and business decision-making, reinforcing the importance of data-driven strategy in modern digital platforms.

Future Scope

The current project establishes a strong SQL-based analytical foundation. However, several enhancements can further expand its impact and align it with enterprise-level data systems.

---

1. Automation Possibilities

- Schedule automated revenue and performance reports using database events or cron jobs.
- Implement automated alert systems for:
  - Revenue drops
  - High delivery delays
  - Unusual discount spikes
- Develop dashboard refresh pipelines for periodic reporting.

This would reduce manual monitoring and enable proactive business decisions.

---

2. Integration with BI Tools

The SQL database can be integrated with Business Intelligence tools such as:

- Power BI
- Tableau
- Looker

This would enable:

- Interactive dashboards
- Executive KPI tracking
- Real-time filtering by city, restaurant, or customer segment

- Visual trend analysis for decision-makers

---

3. Real-Time Analytics

Currently, analysis is batch-based. Future improvements could include:

- Real-time revenue tracking
- Live order monitoring dashboards
- Instant delivery delay alerts
- Dynamic restaurant ranking updates

This would enhance operational responsiveness and customer experience.

---

4. Advanced Data Modeling

The project can be extended with:

- Star schema or data warehouse design
- Fact and dimension modeling
- Partitioned tables for high-volume order data
- Materialized views for faster dashboard queries

This would make the system scalable for millions of transactions.

---

5. Machine Learning Integration

With structured historical data, predictive analytics can be introduced:

- Customer churn prediction
- Demand forecasting by city
- Dynamic pricing models
- Delivery time prediction models
- Recommendation systems for food items

This would shift the system from descriptive analytics to predictive and prescriptive analytics.

🎯 Final Note

Including the Future Scope section demonstrates that you understand how a SQL analytics project can evolve into a full-scale data platform. It shows forward-thinking capability — which is highly valued in Data Analyst roles.

For your portfolio → **Keep it. It strengthens your project significantly.**