Logistics Optimization for Delivery Routes – Flipkart

Project Overview

Flipkart, one of India's largest e-commerce giants, delivers millions of orders every day across metros, Tier-2, and Tier-3 cities through its logistics arm, Ekart Logistics.

The logistics network includes regional fulfillment centers, sorting hubs, and last-mile delivery partners across India.

As order volumes rise - especially during sales and festive seasons - delays, route inefficiencies, and traffic disruptions can significantly affect both customer satisfaction and operational costs.

Current Challenges

Flipkart's logistics team currently faces challenges in:

- Identifying the root causes of delivery delays (e.g., congestion, hub processing issues).
- Optimizing delivery routes for faster, more cost-efficient fulfillment.
- Improving shipment efficiency and agent performance using data-driven insights.

The logistics data, stored in relational databases, can be analyzed using SQL to extract meaningful patterns and performance metrics. These insights can help Flipkart improve route planning, reduce delivery delays, and enhance warehouse and agent efficiency.

Project Objective

Built a SQL-driven Logistics analytics system to analyze delays, optimize routes, and enhance shipment efficiency by leveraging queries and aggregations. The project aims to answer key business questions, uncover inefficiencies, and recommend actionable improvements based on data analysis.

Dataset Description

The dataset includes the following key tables:

1. Orders Table

The Orders dataset contains order-level delivery details, including warehouse, route, agent, delivery dates, status, and order value. It is used to analyze delivery performance, agent efficiency, and order value trends across regions.

2. Routes Table

The Routes dataset contains route-level transportation details, including start and end locations, distance, travel time, and delay information. It helps analyze route efficiency, traffic impact, and overall travel performance across different delivery routes.

3. Warehouses Table

The Warehouse dataset provides warehouse-level information, including warehouse ID, name, city, processing capacity, and average processing time. It is used to evaluate warehouse performance, operational capacity, and efficiency across different locations.

4. Delivery Agents Table

The Delivery dataset contains delivery agent performance details, including their ID, name, assigned route, average speed (in km/h), delivery efficiency (in %), and years of experience. It can be used to analyze agent productivity and route performance trends.

5. Shipment Tracking Table

The Shipping dataset includes key details such as shipment ID, route, delivery speed, efficiency, and agent experience. It is useful for analyzing shipping performance, optimizing routes, and improving overall delivery operations.

Tasks to be Performed

Task 1: Data Cleaning & Preparation

- Identified and deleted duplicate Order_ID records.
- Replaced null Traffic_Delay_Min with the average delay for that route.
- Converted all date columns into YYYY-MM-DD format using SQL functions.
- Ensured that no Actual_Delivery_Date is before Order_Date (flagged such records).

Task 2: Delivery Delay Analysis

- Calculated delivery delay (in days) for each order.
- Found Top 10 delayed routes based on average delay days.
- Used window functions to rank all orders by delay within each warehouse.

Task 3: Route Optimization Insights

For each route, calculated:

- Average delivery time (in days).

- Average traffic delay.
- Distance-to-time efficiency ratio: Distance_KM / Average_Travel_Time_Min.

Additional analyses:

- Identified 3 routes with the worst efficiency ratio.
- Found routes with >20% delayed shipments.
- Recommended potential routes for optimization.

Task 4: Warehouse Performance

- Found the top 3 warehouses with the highest average processing time.
- Calculated total vs. delayed shipments for each warehouse.
- Used CTEs to find bottleneck warehouses where processing time > global average.
- Ranked warehouses based on on-time delivery percentage.

Task 5: Delivery Agent Performance

- Ranked agents (per route) by on-time delivery percentage.
- Found agents with on-time % < 80%.
- Compared average speed of top 5 vs bottom 5 agents using subqueries.
- Suggested training or workload balancing strategies for low performers.

Task 6: Shipment Tracking Analytics

- For each order, listed the last checkpoint and time.
- Found the most common delay reasons (excluding None).
- Identified orders with >2 delayed checkpoints.

Task 7: Advanced KPI Reporting

Calculated KPIs using SQL queries:

- Average Delivery Delay per Region (Start_Location).
- On-Time Delivery % = (Total On-Time Deliveries / Total Deliveries) × 100.
- Average Traffic Delay per Route.

Task 8: PPT Presentation

- Presented analysis and findings by copying all the queries and result tables from the previous steps into a PowerPoint presentation.
- Copied and pasted SQL queries and their corresponding results for Tasks 1 to 7.
- Ensured that the tables are formatted clearly, and the queries are concise.

Expected Outcomes

This project will deliver:

1. A comprehensive SQL-based analytics framework for logistics optimization.
2. Identification of performance bottlenecks in warehouses, routes, and delivery agents.
3. Data-driven recommendations for route optimization and delay reduction.
4. KPI dashboards and PowerPoint presentations for stakeholder communication.
5. Actionable insights to improve customer satisfaction and reduce operational costs.