

Amazon Logistics Optimization SQL Project

BY : ASWANI KASIREDDY

Project Introduction:

Amazon, a global e-commerce leader, handles millions of daily shipments across various regions. The logistics network is vast, comprising warehouses, fulfillment centers, and last-mile delivery partners. With increasing order volumes, delays and inefficiencies in delivery routes can significantly affect customer satisfaction and operational costs.

PROJECT OBJECTIVE:

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

Task 1: Data Cleaning & Preparation

```
USE Amazon_Logistics;
# 1. Identify and delete duplicate Order_ID records.
SELECT Order_ID, COUNT(*) AS count FROM orders GROUP BY Order_ID HAVING count > 1;

SELECT * FROM (
    SELECT*, 
        ROW_NUMBER() OVER(PARTITION BY Order_ID ORDER BY Order_Date) AS rn
    FROM orders
) t
WHERE rn = 1;

DELETE FROM orders WHERE Order_ID IN (
    SELECT Order_ID FROM (
        SELECT Order_ID, ROW_NUMBER() OVER(PARTITION BY Order_ID ORDER BY Order_ID) AS row_num
        FROM orders
    ) t
    WHERE row_num > 1
);

# 2. Replace null Traffic_Delay_Min with the average delay for that route.
SELECT * FROM routes WHERE Traffic_Delay_Min IS NULL;

UPDATE routes r
JOIN (
    SELECT Route_ID, AVG(Traffic_Delay_Min) AS avg_delay
    FROM routes
    WHERE Traffic_Delay_Min IS NOT NULL
    GROUP BY Route_ID
) t ON r.Route_ID = t.Route_ID
SET r.Traffic_Delay_Min = t.avg_delay
WHERE r.Traffic_Delay_Min IS NULL;

# 3. Convert all date columns into YYYY-MM-DD format using SQL functions.
ALTER TABLE orders
MODIFY COLUMN Order_Date DATE,
MODIFY COLUMN Expected_Delivery_Date DATE,
MODIFY COLUMN Actual_Delivery_Date DATE;

ALTER TABLE shipmenttracking
MODIFY COLUMN Checkpoint_Time DATE;

# 4. Ensure that no Actual_Delivery_Date is before Order_Date (flag such records).
SELECT COUNT(*) AS invalid_delivery_count FROM orders WHERE Actual_Delivery_Date < Order_Date;

SELECT Order_ID, Order_Date, Actual_Delivery_Date,
CASE WHEN Actual_Delivery_Date < Order_Date THEN 'Invalid' ELSE 'Valid' END AS status
FROM orders WHERE Actual_Delivery_Date < Order_Date;
```

The screenshot shows the MySQL Workbench interface with the following details:

- MySQL Workbench Version:** Local instance 3306 - Warning - not supported
- Session:** # Task 1: Data Cleaning & Preparation
- Object Info:** Shows the current object being edited is the script for Task 1.
- Connection Details:** Name: Local instance 3306, Host: localhost, Port: 3306, User: root, Current User: root@localhost, SSL cipher: SSL not used.
- Server:** Product: MySQL Community Server - GPL, Version: 9.4.0, Connector: C++ 9.4.0.
- Action Output:** Displays the execution log with the following entries:

Time	Action	Response	Duration / Fetch Time
10:20:01	USE Amazon_Logistics	0 row(s) affected	0.00025 sec
10:20:06	USE Amazon_Logistics	0 row(s) affected	0.00030 sec
10:20:11	SET SQL_SAFE_UPDATES = 0	0 row(s) affected	0.00049 sec
13:45:13	UPDATE routes r JOIN (SELECT Route_ID, AVG(Traffic_Delay_Min) AS avg_delay FROM routes WHERE Traffic_Delay_Min IS NOT NULL GROUP BY Route_ID) t ON r.Route_ID = t.Route_ID SET r.Traffic_Delay_Min = t.avg_delay WHERE r.Traffic_Delay_Min IS NULL;	0 row(s) affected Rows matched: 0 Changed: 0 Warnings: 0	0.033 sec
13:45:19	ALTER TABLE shipmenttracking MODIFY COLUMN Checkpoint_Time DATE	Error Code: 1146. Table 'amazon_logistics.shipmenttracking' doesn't exist	0.0072 sec
13:45:29	ALTER TABLE orders MODIFY COLUMN Order_Date DATE, MODIFY COLUMN Expected_Delivery_Date DATE, MODIFY COLUMN Actual_Delivery_Date DATE	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.029 sec
13:46:37	ALTER TABLE shipmenttracking MODIFY COLUMN Checkpoint_Time DATE	Error Code: 1146. Table 'amazon_logistics.shipmenttracking' doesn't exist	0.0010 sec
13:46:39	ALTER TABLE shipment_tracking MODIFY COLUMN Checkpoint_Time DATE	1040 row(s) affected Records: 1040 Duplicates: 0 Warnings: 0	0.037 sec
13:46:44	ALTER TABLE shipment_tracking MODIFY COLUMN Checkpoint_Time DATE	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.0064 sec
- Status:** Query Completed

Insights:

- No duplicate Order_ID records were found, ensuring the dataset maintains data integrity.
- All missing values in Traffic_Delay_Min were successfully replaced with the average delay calculated per route, improving completeness.
- All date columns were standardized into YYYY-MM-DD format, ensuring uniformity across records for time-based analysis.
- Validation confirmed that no Actual_Delivery_Date occurs before Order_Date, indicating no timeline inconsistencies.
- The dataset is now fully cleaned and ready for accurate analysis in subsequent tasks.

MySQL Workbench

Local instance 3306 - Warning - not supported

Administration Schemas

Result Grid Filter Rows: Search Export:

SCHEMAS

Filter objects

Amazon_Logistics

- Tables
- Views
- Stored Procedures
- Functions

Object Info Session

Connection Details

Name: Local instance 3306
Host: localhost
Port: 3306
Login User: root
Current User: root@localhost
SSL cipher: SSL not used

Server

Product: MySQL Community Server - GPL
Version: 9.4.0
Connector

Version: C++ 9.4.0

Result Grid

Order_ID	Customer_ID	Warehouse_ID	Route_ID	Order_Date	Expected_Delivery_Da...	Actual_Delivery_Da...	Delivery_Status	rn
O0001	C0080	W004	R005	2025-06-08	2025-06-11	2025-06-11	On Time	1
O0002	C0566	W009	R004	2025-06-22	2025-06-27	2025-06-29	Delayed	1
O0003	C0899	W008	R013	2025-06-04	2025-06-09	2025-06-11	Delayed	1
O0004	C0311	W001	R005	2025-06-28	2025-07-05	2025-07-07	Delayed	1
O0005	C0304	W006	R012	2025-06-04	2025-06-09	2025-06-10	Delayed	1
O0006	C0719	W004	R002	2025-07-17	2025-07-19	2025-07-19	On Time	1
O0007	C0652	W001	R020	2025-06-14	2025-06-17	2025-06-18	Delayed	1
O0008	C0111	W006	R008	2025-06-23	2025-06-30	2025-07-02	Delayed	1
O0009	C0138	W010	R011	2025-07-11	2025-07-16	2025-07-18	Delayed	1
O0010	C0096	W009	R017	2025-06-14	2025-06-21	2025-06-21	On Time	1
O0011	C0422	W008	R002	2025-05-26	2025-06-02	2025-06-02	On Time	1
O0012	C0838	W009	R014	2025-07-16	2025-07-20	2025-07-21	Delayed	1
O0013	C0543	W007	R002	2025-07-16	2025-07-18	2025-07-20	Delayed	1
O0014	C0385	W001	R007	2025-07-07	2025-07-11	2025-07-11	On Time	1
O0015	C0108	W001	R006	2025-06-18	2025-06-24	2025-06-24	On Time	1
O0016	C0077	W003	R005	2025-06-26	2025-06-29	2025-07-01	Delayed	1
O0017	C0703	W006	R014	2025-07-11	2025-07-17	2025-07-17	On Time	1
O0018	C0089	W006	R004	2025-05-23	2025-05-25	2025-05-25	On Time	1
O0019	C0391	W005	R003	2025-06-29	2025-07-03	2025-07-03	On Time	1
O0020	C0878	W009	R007	2025-06-17	2025-06-19	2025-06-20	Delayed	1
O0021	C0541	W004	R002	2025-07-05	2025-07-10	2025-07-12	Delayed	1
O0022	C0534	W002	R015	2025-06-22	2025-06-25	2025-06-27	Delayed	1
O0023	C0371	W008	R009	2025-06-10	2025-06-14	2025-06-15	Delayed	1
O0024	C0491	W004	R002	2025-06-19	2025-06-21	2025-06-21	On Time	1
O0025	C0867	W002	R010	2025-06-29	2025-07-02	2025-07-02	On Time	1
O0026	C0199	W007	R005	2025-05-22	2025-05-29	2025-05-29	On Time	1
O0027	C0709	W003	R007	2025-07-06	2025-07-11	2025-07-11	On Time	1
O0028	C0826	W008	R019	2025-07-08	2025-07-13	2025-07-13	On Time	1
O0029	C0900	W005	R017	2025-05-25	2025-05-30	2025-06-01	Delayed	1
O0030	C0426	W010	R010	2025-06-26	2025-07-02	2025-07-02	On Time	1
O0031	C0333	W004	R008	2025-05-26	2025-06-01	2025-06-01	On Time	1
O0032	C0186	W007	R016	2025-05-22	2025-05-24	2025-05-24	On Time	1
O0033	C0307	W008	R019	2025-06-15	2025-06-20	2025-06-22	Delayed	1
O0034	C0449	W002	R019	2025-06-24	2025-06-28	2025-06-28	On Time	1

Result 2 Read Only

Export resultset canceled

Task 2: Delivery Delay Analysis

-- 1. Calculate delivery delay for each order

```
SELECT
    Order_ID,
    Warehouse_ID,
    Route_ID,
    Expected_Delivery_Date,
    Actual_Delivery_Date,
    DATEDIFF(Actual_Delivery_Date, Expected_Delivery_Date) AS Delay_Days
FROM orders
ORDER BY Delay_Days DESC;
```

-- 2. Top 10 delayed routes by avg delay

```
SELECT
    Route_ID,
    ROUND(AVG(DATEDIFF(Actual_Delivery_Date, Expected_Delivery_Date)), 2) AS Avg_Delay_Days,
    COUNT(*) AS Total_Orders
FROM orders
GROUP BY Route_ID
ORDER BY Avg_Delay_Days DESC
LIMIT 10;
```

- 3. Rank orders within each warehouse by delay

```
SELECT
    Order_ID,
    Warehouse_ID,
    Route_ID,
    DATEDIFF(Actual_Delivery_Date, Expected_Delivery_Date) AS Delay_Days,
    RANK() OVER (PARTITION BY Warehouse_ID ORDER BY DATEDIFF(Actual_Delivery_Date, Expected_Delivery_Date) DESC) AS Delay_Rank
FROM orders
ORDER BY Warehouse_ID, Delay_Rank;
```

Insights:

- Delivery delays mostly ranged between **1–2 days** across orders.
- **Top delayed routes** were identified as **R012, R020, and R017** with the highest average delays.
- Several other routes (**R015, R011, R010, R007**) also showed recurring delays.
- **Warehouse W001** recorded the **highest number of 2-day delays**, marking it as a critical bottleneck.
- **Warehouse W010** showed consistent delays, affecting overall delivery timelines.
- Delay ranking placed **W001, W002, and W003** among the top warehouses with severe delays.
- Insights that indicated the need to optimize **R012, R020, R017 routes** and improve **W001 & W010 warehouse operations**.

MySQL Workbench

Local instance 3306 - Warning - not supported

Administration Schemas SQL File 7* SQL File 5* SQL File 6* SQL File 8* SQL File 7* SQL File 11* SQL File 12* SQL File 13* >

Result Grid Filter Rows: Search Export:

Amazon_Logistics

Order_ID	Warehouse_ID	Route_ID	Expected_Delivery_Date	Actual_Delivery_Date	Delay_Days
O000	W008	R013	2025-06-08	2025-06-11	2
O0005	W001	R005	2025-07-05	2025-07-07	2
O0008	W006	R008	2025-06-30	2025-07-02	2
O0009	W010	R011	2025-07-16	2025-07-18	2
O0013	W007	R002	2025-07-18	2025-07-20	2
O0016	W003	R005	2025-06-29	2025-07-01	2
O0021	W004	R002	2025-07-10	2025-07-12	2
O0022	W000	R015	2025-06-25	2025-06-27	2
O0023	W008	R019	2025-06-20	2025-06-22	2
O0037	W007	R002	2025-06-12	2025-06-14	2
O0043	W006	R018	2025-06-18	2025-06-28	2
O0044	W005	R007	2025-07-23	2025-07-25	2
O0046	W007	R014	2025-05-31	2025-06-02	2
O0059	W000	R011	2025-06-21	2025-06-23	2
O0065	W001	R013	2025-07-17	2025-07-19	2
O0069	W008	R017	2025-07-13	2025-07-15	2
O0074	W009	R007	2025-06-07	2025-06-19	2
O0075	W007	R012	2025-07-05	2025-07-07	2
O0081	W010	R009	2025-06-14	2025-06-16	2
O0086	W007	R015	2025-06-08	2025-06-08	2
O0090	W010	R016	2025-07-12	2025-07-14	2
O0101	W007	R011	2025-07-25	2025-07-27	2
O0102	W000	R010	2025-07-14	2025-07-05	2
O0104	W010	R016	2025-06-18	2025-06-20	2
O0107	W002	R006	2025-06-02	2025-06-02	2
O0108	W006	R009	2025-07-17	2025-07-19	2
O0115	W005	R020	2025-06-28	2025-06-30	2
O0116	W009	R004	2025-07-20	2025-07-22	2
O0117	W007	R019	2025-07-15	2025-07-17	2
O0120	W005	R001	2025-07-24	2025-07-26	2
O0122	W009	R010	2025-06-09	2025-06-11	2
O0138	W010	R012	2025-06-18	2025-06-20	2
O0140	W004	R009	2025-06-21	2025-06-23	2
O0141	W008	R011	2025-06-02	2025-06-14	2

Result 1 Read Only

Query Completed

MySQL Workbench

Local instance 3306 - Warning - not supported

Administration Schemas SQL File 7* SQL File 5* SQL File 6* SQL File 8* SQL File 7* SQL File 11* SQL File 10* SQL File 12* SQL File 13* >

Result Grid Filter Rows: Search Export:

Amazon_Logistics

```

16
17 -- 2. Top 10 delayed routes by avg delay
18 • SELECT
19   Route_ID,
20   ROUND(AVG(DATEDIFF(Actual_Delivery_Date, Expected_Delivery_Date)),2) AS Avg_Delay_Days,
21   COUNT(*) AS Total_Orders
22 FROM orders
23 GROUP BY Route_ID
24 ORDER BY Avg_Delay_Days DESC
25 LIMIT 10;
26

```

Route_ID	Avg_Delay_Days	Total_Orders
R012	1.08	12
R050	1.06	17
R017	1.00	14
R015	1.00	6
R005	0.80	15
R004	0.80	20
R013	0.76	17
R011	0.70	20
R010	0.70	20
R007	0.69	16

Result 3 Read Only

Query Completed

MySQL Workbench

Local instance 3306 - Warning - not supported

Administration Schemas SQL File 7* SQL File 5* SQL File 6* SQL File 8* SQL File 7* SQL File 11* SQL File 10* SQL File 12* SQL File 13* >

Result Grid Filter Rows: Search Export:

Amazon_Logistics

Order_ID	Warehouse_ID	Route_ID	Delay_Days	Delay_Rank
O0263	W001	R004	2	1
O0004	W001	R005	2	1
O0274	W001	R004	2	1
O0001	W001	R004	2	1
O0276	W001	R006	2	1
O0065	W001	R013	2	1
O0238	W001	R005	2	1
O0220	W001	R017	2	1
O0222	W001	R020	2	1
O0185	W001	R019	1	10
O0047	W001	R020	1	10
O0193	W001	R003	1	10
O0007	W001	R020	1	10
O0057	W001	R002	1	10
O0160	W001	R018	1	10
O0003	W001	R008	1	10
O0042	W001	R020	1	10
O0223	W001	R020	1	10
O0091	W001	R015	0	19
O0015	W001	R006	0	19
O0187	W001	R017	0	19
O0131	W001	R010	0	19
O0070	W000	R008	0	19
O0195	W001	R002	0	19
O0014	W001	R007	0	19
O0289	W001	R011	0	19
O0290	W001	R004	0	19
O0002	W001	R019	0	19
O0215	W001	R016	0	19
O0214	W001	R010	0	19
O0054	W001	R013	0	19
O0210	W001	R013	0	19
O0018	W001	R001	0	19
O0260	W001	R008	0	19
O0199	W002	R001	0	1
O0022	W002	R015	0	1

Result 5 Read Only

Query Completed

Task 3: Route Optimization Insights

-- 1. For each route, calculate:

```
# Average delivery time (in days)
SELECT Route_ID, AVG(DATEDIFF(Actual_Delivery_Date, Order_Date)) AS avg_delivery_days
FROM orders
```

```
GROUP BY Route_ID;
```

```
# Average traffic delay
```

```
SELECT Route_ID, ROUND(AVG(Traffic_Delay_Min),2) AS avg_traffic_delay
FROM routes
GROUP BY Route_ID;
```

```
# Distance-to-time efficiency ratio: Distance_KM / Average_Travel_Time_Min.
```

```
SELECT Route_ID, Distance_KM / Average_Travel_Time_Min AS efficiency_ratio
FROM routes;
```

-- 2. Identify 3 routes with the worst efficiency ratio.

```
# (Assuming worst = lowest efficiency ratio)
```

```
SELECT Route_ID, Distance_KM / Average_Travel_Time_Min AS efficiency_ratio
FROM routes
ORDER BY efficiency_ratio ASC LIMIT 3;
```

-- 3. Find routes with >20% delayed shipments.

```
SELECT
    Route_ID,
    ROUND(SUM(CASE WHEN Delivery_Status = 'Delayed' THEN 1 ELSE 0 END) * 100.0 / COUNT(*),2) AS delayed_percent
FROM orders
GROUP BY Route_ID
HAVING delayed_percent > 20;
```

-- 4. Recommend potential routes for optimization.

```
# Optimize R020 (Juliestad to South Jonathanshire) with dynamic rerouting and top-performing agents (>90% on-time) due to 70.59% delayed shipments and 5.88-day average delivery.
```

```
# Shorten R017 (Port Tammybury to Abbottside) route or automate warehouse sorting for 57.14% delays and 6.29-day average delivery.
```

```
# Schedule R007 (Aaronhaven to Jeanneberg) deliveries during off-peak hours to reduce 51-minute traffic delay and 48.75% delays.
```

```
# Reroute R009 (Bellfurt to Lake Marvin) via less congested paths to address 46-minute traffic delay and 50.00% delays.
```

```
# Assign faster vehicles to R006 (Lake Robinhaven to East Christopherville) for its 0.1607 efficiency ratio and 40.00% delays.
```

Administration **Schemas** **SQL File 7*** **SQL File 5*** **SQL File 6***

SCHEMAS

Filter objects

- Amazon_Logistics
 - Tables
 - Views
 - Stored Procedures
 - Functions
- sys

Result Grid Filter Rows: Search

Route_ID	efficiency_ratio
R001	0.6136792452830189
R002	2.156179775280899
R003	0.5815789473684211
R004	4.278125
R005	0.37382198952879586
R006	0.16065573770491803
R007	0.28250950570342204
R008	0.24372093023255814
R009	0.6047846889952153
R010	0.6686046511627907
R011	0.9556179775280899
R012	0.9036363636363637
R013	0.9324999999999999
R014	2.27818181818183
R015	0.30086956521739133
R016	0.15120481927710844
R017	0.328
R018	0.5366037735849056
R019	3.7693877551020405
R020	0.5502923976608187

Object Info Session

Connection Details

Name: Local instance 3306
 Host: localhost
 Port: 3306
 Login: root
 User: root@localhost
 Current User: root@localhost
 SSL cipher: SSL not used

Server

Product: MySQL Community Server - GPL
 Version: 9.4.0

Connector

Version: C++ 9.4.0

Administration **Schemas** **SQL File 7*** **SQL File 5*** **SQL File 6***

SCHEMAS

Filter objects

- Amazon_Logistics
 - Tables
 - Views
 - Stored Procedures
 - Functions
- sys

Result Grid Filter Rows: Search

Route_ID	efficiency_ratio
R001	0.6136792452830189
R002	2.156179775280899
R003	0.5815789473684211
R004	4.278125
R005	0.37382198952879586
R006	0.16065573770491803
R007	0.28250950570342204
R008	0.24372093023255814
R009	0.6047846889952153
R010	0.6686046511627907
R011	0.9556179775280899
R012	0.9036363636363637
R013	0.9324999999999999
R014	2.27818181818183
R015	0.30086956521739133
R016	0.15120481927710844
R017	0.328
R018	0.5366037735849056
R019	3.7693877551020405
R020	0.5502923976608187

Object Info Session

Connection Details

Name: Local instance 3306
 Host: localhost
 Port: 3306
 Login: root
 User: root@localhost
 Current User: root@localhost
 SSL cipher: SSL not used

Server

Product: MySQL Community Server - GPL
 Version: 9.4.0

Connector

Version: C++ 9.4.0

Administration **Schemas** **SQL File 7*** **SQL File 5*** **SQL File 6***

SCHEMAS

Filter objects

- Amazon_Logistics
 - Tables
 - Views
 - Stored Procedures
 - Functions
- sys

Result Grid Filter Rows: Search

Route_ID	delayed_percent
R005	46.67
R004	50.00
R013	52.94
R012	66.67
R002	31.82
R020	70.59
R008	31.25
R011	45.00
R017	57.14
R014	38.46
R007	43.75
R006	36.36
R003	40.00
R015	50.00
R009	37.50
R010	45.00
R019	30.77
R016	25.00
R001	40.00
R018	40.00

Object Info Session

Connection Details

Name: Local instance 3306
 Host: localhost
 Port: 3306
 Login: root
 User: root@localhost
 Current User: root@localhost
 SSL cipher: SSL not used

Server

Product: MySQL Community Server - GPL
 Version: 9.4.0

Connector

Version: C++ 9.4.0

Task 4: Warehouse Performance

-- 1. Find the top 3 warehouses with the highest average processing time.

```
SELECT Warehouse_ID, ROUND(AVG(Processing_Time_Min),2) AS avg_processing_time  
FROM warehouses  
GROUP BY Warehouse_ID  
ORDER BY avg_processing_time DESC  
LIMIT 3;
```

-- 2. Calculate total vs. delayed shipments for each warehouse.

```
SELECT Warehouse_ID, COUNT(*) AS total_shipments, SUM(CASE WHEN Delivery_Status = 'Delayed' THEN 1 ELSE 0 END) AS delayed_shipments  
FROM orders  
GROUP BY Warehouse_ID;
```

-- 3. Use CTEs to find bottleneck warehouses where processing time > global average.

```
WITH global_avg AS (SELECT AVG(Processing_Time_Min) AS avg_proc FROM warehouses)  
SELECT w.Warehouse_ID, w.Processing_Time_Min  
FROM warehouses w, global_avg g
```

```
WHERE w.Processing_Time_Min > g.avg_proc;
```

-- 4. Rank warehouses based on on-time delivery percentage.

```
SELECT Warehouse_ID,  
    ROUND(((COUNT(*) - SUM(CASE WHEN Delivery_Status = 'Delayed' THEN 1 ELSE 0 END)) * 100.0 / COUNT(*)),2) AS on_time_percent,  
    RANK() OVER (ORDER BY ((COUNT(*) - SUM(CASE WHEN Delivery_Status = 'Delayed' THEN 1 ELSE 0 END)) * 100.0 / COUNT(*)) DESC) AS on_time_percent_rank  
FROM orders  
GROUP BY Warehouse_ID;
```

Insights:

- **W010, W009, and W007** recorded the **highest processing times** (171, 170, 166 mins).
- **W010 (67% delayed)** and **W001 (53% delayed)** are the **worst-performing warehouses**.
- **W003 (57% delayed)** and **W005 (56% delayed)** also showed severe inefficiencies.
- **Bottlenecks:** W001, W007, W008, W009, W010 all exceed global average processing times.
- **Best performers:** **W008 (76% on-time)** and **W006 (68% on-time)** lead in reliability.
- **Worst performer:** **W010 (33% on-time)** ranks last in overall delivery performance.

SQL File 7* SQL File 5* SQL File 6* SQL File 8*

```

1
2
3 -- 1. Find the top 3 warehouses with the high
4 • SELECT Warehouse_ID, ROUND(AVG(Processing_Tim
5   FROM warehouses
6   GROUP BY Warehouse_ID

```

100% 31:2

Result Grid Filter Rows: Search Export:

Warehouse_ID	avg_processing_ti...
W010	171.00
W009	170.00
W007	166.00

SQL File 7* SQL File 5* SQL File 6* SQL File 8*

```

13
14 -- 2. Calculate total vs. delayed shipments f
15 • SELECT Warehouse_ID, COUNT(*) AS total_shipme
16   FROM orders
17   GROUP BY Warehouse_ID;

```

100% 1:13

Result Grid Filter Rows: Search Export:

Warehouse_ID	total_shipments	delayed_shipments
W004	31	12
W009	33	13
W008	21	5
W001	34	18
W006	37	12
W010	30	20
W007	38	15
W003	23	13
W005	25	14
W002	28	10

SQL File 7* SQL File 5* SQL File 6* SQL File 8*

Limit to 1000 rows

```

19
20 -- 3. Use CTEs to find bottleneck warehouses
21 • WITH global_avg AS (SELECT AVG(Processing_Tim
22   SELECT w.Warehouse_ID, w.Processing_Time_Min
23   FROM warehouses w, global_avg g
24   WHERE w.Processing_Time_Min > g.avg_proc;
25
26

```

100% 23:17

Result Grid Filter Rows: Search Export:

Warehouse_ID	Processing_Time_Min
W001	142
W007	166
W008	161
W009	170
W010	171

SQL File 7* SQL File 5* SQL File 6* SQL File 8*

Limit to 1000 rows

```

-- 4. Rank warehouses based on on-time delivery
29 • SELECT Warehouse_ID,
30   ROUND(((COUNT(*) - SUM(CASE WHEN Deliver
31   RANK() OVER (ORDER BY ((COUNT(*) - SUM(C
32   FROM orders
33   GROUP BY Warehouse_ID;
34

```

100% 22:29

Result Grid Filter Rows: Search Export:

Warehouse_ID	on_time_percent	on_time_percent_rank
W008	76.19	1
W006	67.57	2
W002	64.29	3
W004	61.29	4
W009	60.61	5
W007	60.53	6
W001	47.06	7
W005	44.00	8
W003	43.48	9
W010	33.33	10

Task 5: Delivery Agent Performance

-- 1. Rank agents (per route) by on-time delivery percentage.

```
SELECT Agent_ID, Route_ID, On_Time_Percentage,  
RANK() OVER (PARTITION BY Route_ID  
ORDER BY On_Time_Percentage DESC) AS On_Time_Percentage_rank  
FROM deliveryagents;
```

-- 2. Find agents with on-time % < 80%.

```
SELECT Agent_ID, Route_ID, On_Time_Percentage  
FROM deliveryagents  
WHERE On_Time_Percentage < 80;
```

-- 3. Compare average speed of top 5 vs bottom 5 agents using subqueries.

```
SELECT  
AVG(top5.Avg_Speed_KM_HR) AS top5_avg_speed,  
AVG(bottom5.Avg_Speed_KM_HR) AS bottom5_avg_speed  
FROM (  
SELECT Avg_Speed_KM_HR  
FROM deliveryagents  
ORDER BY On_Time_Percentage DESC  
LIMIT 5  
) top5,  
(  
SELECT Avg_Speed_KM_HR  
FROM deliveryagents  
ORDER BY On_Time_Percentage ASC  
LIMIT 5  
) bottom5;
```

Insights:

- **A014 (99.9%), A007 (99.11%), and A004 (98.25%)** are the **best-performing agents** across routes.
- **R002, R012, R017, and R018** have multiple agents consistently delivering **>90% on-time**.
- **A010 (70.45%), A038 (71.8%), A011 (71.5%)** and others fall **below 80%**, marking them as underperformers.
- Nearly **40% of agents have <80% on-time rates**, showing widespread performance gaps.
- **Top 5 agents average 60 km/hr speed**, compared to **56.8 km/hr** for bottom 5, linking speed with punctuality.
- Training, route optimization, and performance monitoring are needed to improve **low-performing agents**.

Result Grid Filter Rows: Search Export:

Agent_ID	Route_ID	On_Time_Percentage	On_Time_Percentage_rank
A006	R001	81.56	1
A037	R001	78.94	2
A033	R002	98.01	1
A024	R002	97.16	2
A027	R002	73.8	3
A050	R002	72.75	4
A011	R002	71.5	5
A015	R003	83.31	1
A029	R005	81.71	1
A035	R005	74.5	2
A041	R006	91.47	1
A047	R006	84.62	2
A020	R006	77.65	3
A014	R007	99.9	1
A049	R007	99.82	2
A040	R007	77.83	3
A028	R007	74.47	4
A018	R007	73.21	5
A032	R008	99.32	1
A022	R008	87.72	2
A034	R008	81.67	3
A045	R009	95.08	1
A042	R010	93.37	1
A012	R010	80.85	2
A009	R012	92.27	1

Result Grid Filter Rows: Search

Agent_ID	Route_ID	On_Time_Percentage
A003	R014	74.15
A008	R018	72.39
A010	R019	70.45
A011	R002	71.5
A016	R012	78.39
A018	R007	73.21
A020	R006	77.65
A023	R017	74.21
A027	R002	73.8
A028	R007	74.47
A030	R014	76.6
A035	R005	74.5
A037	R001	78.94
A038	R018	71.8
A039	R016	72.6
A040	R007	77.83
A043	R016	79.51
A046	R020	75.46
A048	R012	77.92
A050	R002	72.75

100% 11:36

Result Grid Filter Rows: Search Export:

top5_avg_speed	bottom5_avg_speed
60.0000	56.8000

Task 6: Shipment Tracking Analytics

-- 1. For each order, list the last checkpoint and time.

```
SELECT s.Order_ID, s.Checkpoint, s.Checkpoint_Time  
FROM shipment_tracking s  
WHERE s.Checkpoint_Time = (SELECT MAX(Checkpoint_Time)  
FROM shipment_tracking  
WHERE Order_ID = s.Order_ID);
```

-- 2. Find the most common delay reasons (excluding None).

```
SELECT Delay_Reason, COUNT(*) AS count  
FROM shipment_tracking  
WHERE Delay_Reason != 'None'  
GROUP BY Delay_Reason  
ORDER BY count DESC;
```

-- 3. Identify orders with >2 delayed checkpoints.

```
SELECT Order_ID, COUNT(*) AS delayed_checkpoints  
FROM shipment_tracking  
WHERE Delay_Reason != 'None'  
GROUP BY Order_ID  
HAVING delayed_checkpoints > 2;
```

Insights:

- **Sorting Delay (272 cases)** is the **biggest cause of shipment delays**, followed by **Weather (263)** and **Traffic (247)**.
- Over **100+ orders had >2 delayed checkpoints**, showing repeated disruptions in delivery flow.
- Orders like **O0002, O0003, O0004, O0005** faced **3–5 delays each**, making them critical problem cases.
- **Final checkpoint visibility is intact** — most shipments are tracked up to their last checkpoint.
- **Recurring delays** highlight bottlenecks in **sorting centers and route congestion**, needing urgent optimization.

100% 1:2

Result Grid Filter Rows: Search Export:

Order_ID	Checkpoint	Checkpoint_Time
O0001	Checkpoint 1	2025-06-09
O0001	Checkpoint 2	2025-06-09
O0002	Checkpoint 3	2025-06-23
O0002	Checkpoint 4	2025-06-23
O0003	Checkpoint 2	2025-06-05
O0003	Checkpoint 3	2025-06-05
O0003	Checkpoint 4	2025-06-05
O0004	Checkpoint 2	2025-06-29
O0004	Checkpoint 3	2025-06-29
O0004	Checkpoint 4	2025-06-29
O0004	Checkpoint 5	2025-06-29
O0005	Checkpoint 1	2025-06-04
O0005	Checkpoint 2	2025-06-04
O0005	Checkpoint 3	2025-06-04
O0005	Checkpoint 4	2025-06-04
O0005	Checkpoint 5	2025-06-04
O0006	Checkpoint 1	2025-07-18
O0006	Checkpoint 2	2025-07-18
O0007	Checkpoint 1	2025-06-15
O0007	Checkpoint 2	2025-06-15
O0007	Checkpoint 3	2025-06-15
O0008	Checkpoint 2	2025-06-24
O0009	Checkpoint 1	2025-07-11
O0009	Checkpoint 2	2025-07-11
O0009	Checkpoint 3	2025-07-11

100% 51:15

Result Grid Filter Rows: Search Export:

Order_ID	delayed_checkpoint
O0002	3
O0003	4
O0004	4
O0005	4
O0007	3
O0015	4
O0016	5
O0019	5
O0021	4
O0024	4
O0025	3
O0027	3
O0030	3
O0033	3
O0038	5
O0041	4
O0046	3
O0048	4
O0052	3
O0053	4
O0055	4
O0057	3
O0058	3
O0060	3
O0061	3
O0064	4

Result 11

100% 1:13

Result Grid Filter Rows: Search Export:

Delay_Reason	count
Sorting Delay	272
Weather	263
Traffic	247

Task 7: Advanced KPI Reporting

-- 1. Average Delivery Delay per Region (Start_Location).

```
SELECT r.Start_Location, AVG(DATEDIFF(o.Actual_Delivery_Date, o.Expected_Delivery_Date)) AS avg_delay
FROM orders o
LEFT JOIN routes r ON o.Route_ID = r.Route_ID
GROUP BY r.Start_Location;
```

-- 2. On-Time Delivery % = (Total On-Time Deliveries / Total Deliveries) * 100.

```
SELECT
    (SUM(CASE WHEN Delivery_Status = 'On Time' THEN 1 ELSE 0 END) / COUNT(*)) * 100
    AS on_time_percent
FROM orders;
```

-- 3. Average Traffic Delay per Route.

```
SELECT Route_ID,
    AVG(Traffic_Delay_Min) AS avg_traffic_delay
FROM routes
GROUP BY Route_ID;
```

Insights:

- Regions like **South Davidborough (0.44 avg delay)** and **Lake Kathrynchester (0.42 avg delay)** show the **least delivery delays**, while **West Calvinview (1.08)** and **Juliestad (1.05)** face the **highest delays**, making them critical focus areas.
- The overall **On-Time Delivery % is 56%**, highlighting that **nearly half of deliveries are delayed**, signaling operational inefficiency.
- Route R014 (58 min)** and **R007 (51 min)** have the **highest average traffic delays**, indicating frequent congestion issues.
- In contrast, **Routes R004 (3 min)** and **R018/R003 (11 min each)** show **minimal traffic delays**, serving as benchmarks for efficiency.
- Traffic bottlenecks in **R007, R014, and R009** should be prioritized for rerouting or scheduling changes to improve delivery efficiency.

100% 15:6

Result Grid Filter Rows: Search

Start_Location	avg_delay
New Jessica	0.8000
Ramosside	0.8000
South Johnville	0.7647
West Calvinview	1.0833
South Brianchester	0.5000
Juliestad	1.0588
South Davidborough	0.4375
Lowemouth	0.7000
Port Tammybury	1.0000
North Edward	0.6154
Aaronhaven	0.6875
Lake Robinhaven	0.5455
East Amy	0.5000
South Jenniferfurt	1.0000
Bellfurt	0.5000
North Aliceside	0.7000
Lake Oscarfurt	0.6154
Lake Kathrynchester	0.4167
New Aimeechester	0.6667
Lake Jessicastad	0.4667

100% 1:27

Result Grid Filter Rows: Search

Route_ID	avg_traffic_del...
R001	36.0000
R002	22.0000
R003	11.0000
R004	3.0000
R005	26.0000
R006	42.0000
R007	51.0000
R008	34.0000
R009	46.0000
R010	40.0000
R011	25.0000
R012	29.0000
R013	38.0000
R014	58.0000
R015	29.0000
R016	8.0000
R017	8.0000
R018	11.0000
R019	37.0000
R020	23.0000

100% 1:9

Result Grid Filter Rows: Search Export:

on_time_percent
56.0000

Task 8: PPT Presentation and Video Submission

Video Link :

https://drive.google.com/file/d/1dG63Qymg4oibHRyr844_87RILQJwdDUP/view?usp=sharing

Thank You