**SQL Project Report – Indian Railways Fare & Route Analysis**

**Author:** *SAHIL SHAIKH*  
**Date:** *JULY 2025*

**1. Project Overview**

The objective of this project is to analyze Indian Railways fare and route data to identify trends in pricing, popular routes, and class-based revenue distribution.  
SQL was used to query the dataset, and insights were drawn from the results.

**Dataset Details:**

* **Source:** *[“Provided CSV file in Dataset Folder downloaded from Kaggle”]*
* **Rows:** *Number of rows*
* **Columns:** *Number of columns*
* **Key Fields:** fromStnCode, toStnCode, classCode, totalFare, distance, trainNumber, timeStamp

**2. Tools & Technologies Used**

* Oracle SQL Developer
* SQL (SELECT, GROUP BY, ORDER BY, UNION, Aggregate Functions)
* CSV dataset

**3. Queries & Insights**

**Query 1 – Top 5 Most Expensive Routes**

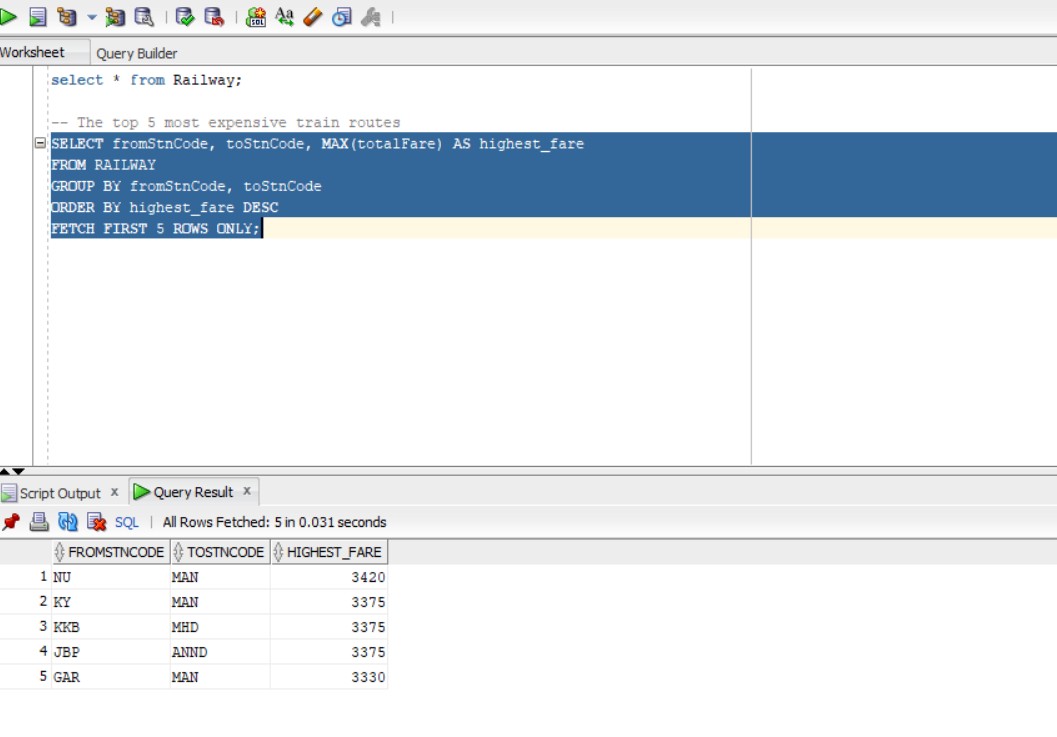
SELECT fromStnCode, toStnCode, MAX(totalFare) AS highest\_fare

FROM RAILWAY

GROUP BY fromStnCode, toStnCode

ORDER BY highest\_fare DESC

FETCH FIRST 5 ROWS ONLY;

**Result Screenshot:**  
**

**Insight:**  
These routes represent premium travel sectors, possibly due to long distances, higher-class coaches, or seasonal pricing.

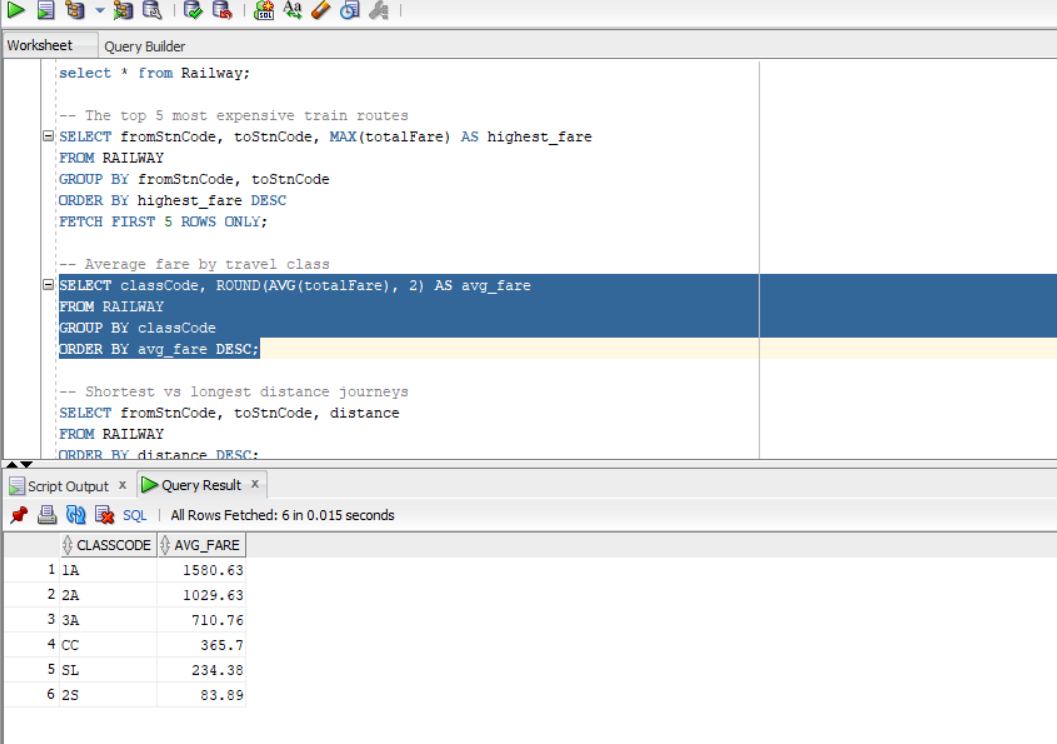
**Query 2 – Average Fare by Travel Class**

SELECT classCode, ROUND(AVG(totalFare), 2) AS avg\_fare

FROM RAILWAY

GROUP BY classCode

ORDER BY avg\_fare DESC;

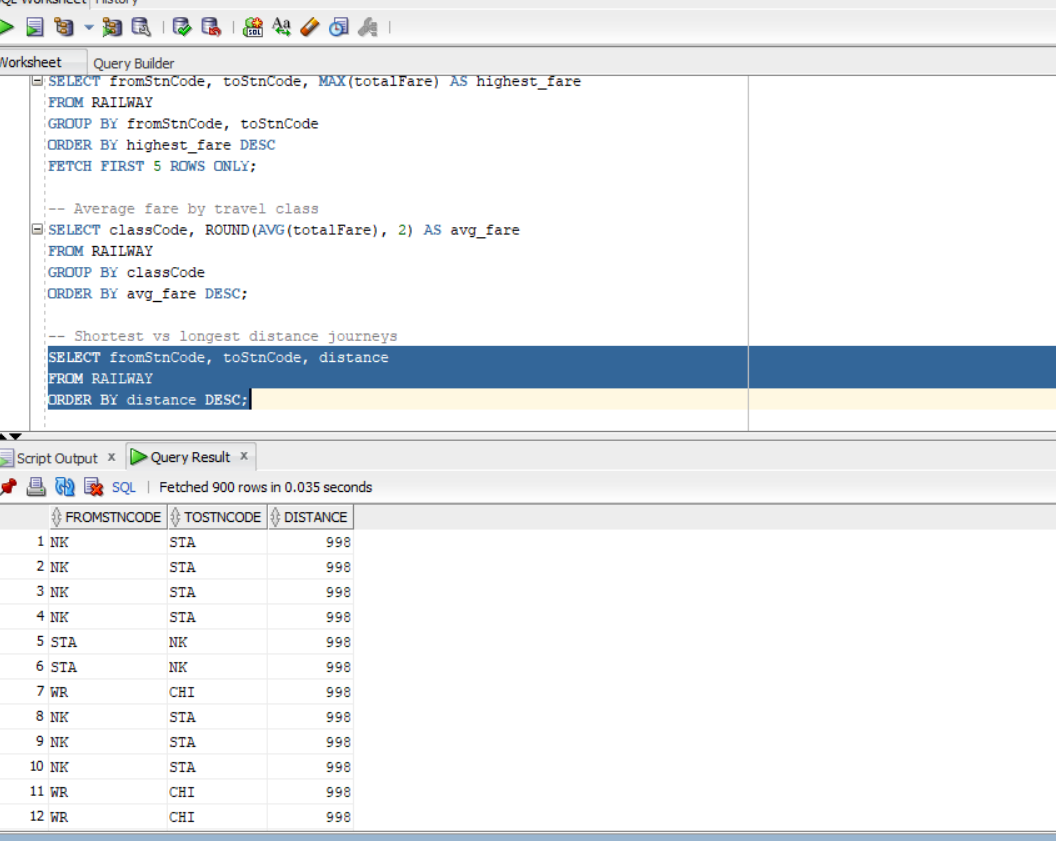
**Result Screenshot:**  
**

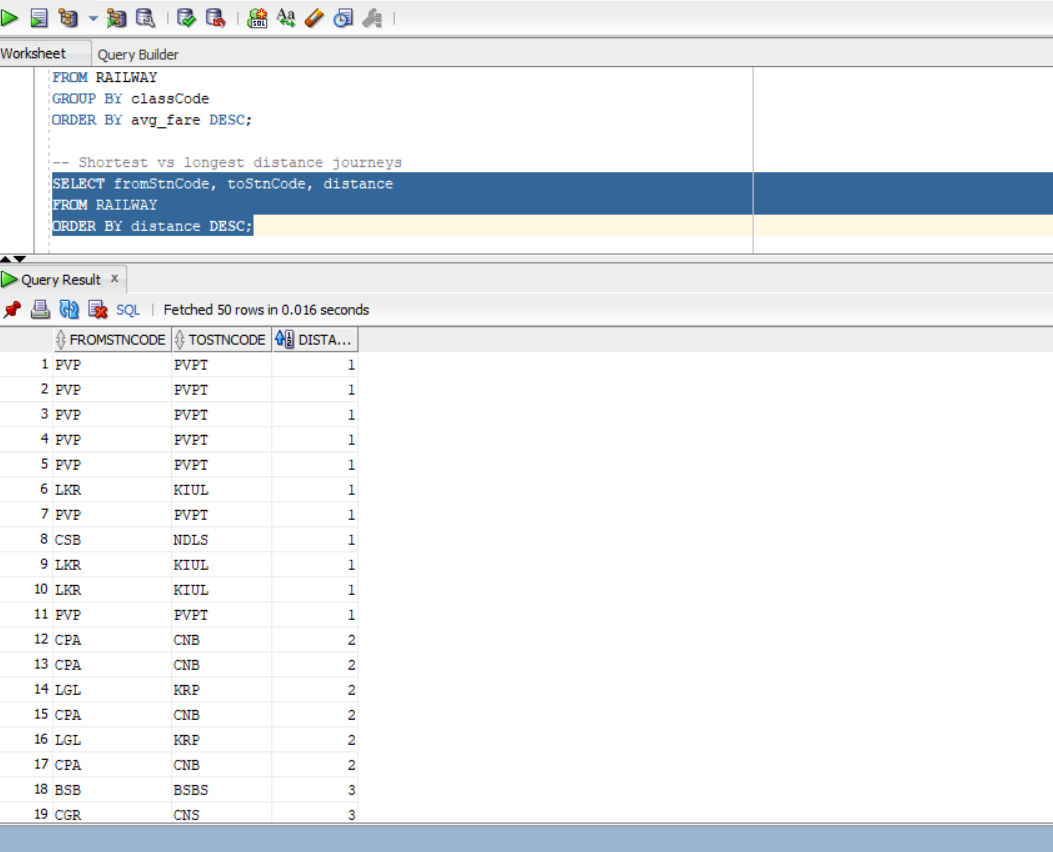
**Insight:**  
Higher-class compartments like 1A and 2A have significantly higher fares compared to sleeper classes.

**Query 3 – Shortest vs Longest Distance Journey**

SELECT fromStnCode, toStnCode, distance

FROM RAILWAY

**ORDER BY distance DESC;Result Screenshot:**  
**



**Insight:**  
Longest distance is 998km and shortest is 1km

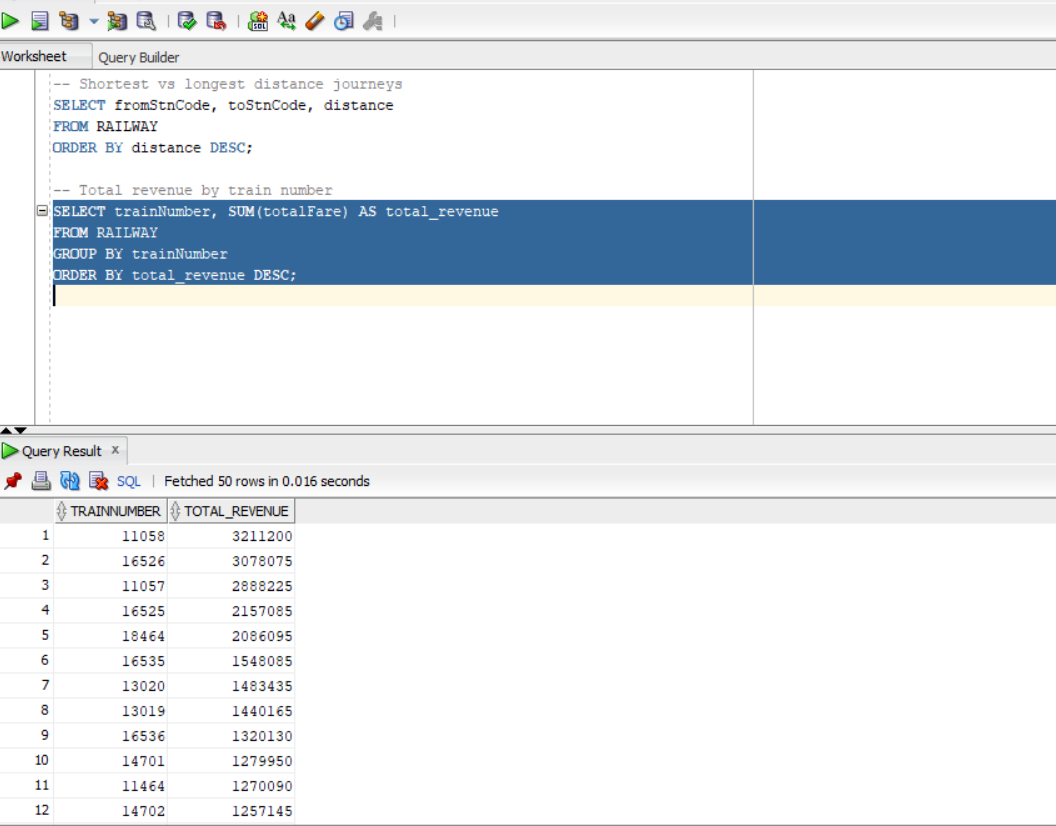
**Query 4 – Total Revenue by Train Number**

SELECT trainNumber, SUM(totalFare) AS total\_revenue

FROM RAILWAY

GROUP BY trainNumber

ORDER BY total\_revenue DESC;

**Result Screenshot:**  


**Insight:**  
Certain trains generate higher revenue due to route popularity and demand.

**Query 5 – Combine expenses with long-distance journeys**

SELECT fromStnCode, toStnCode, totalFare, distance

FROM RAILWAY

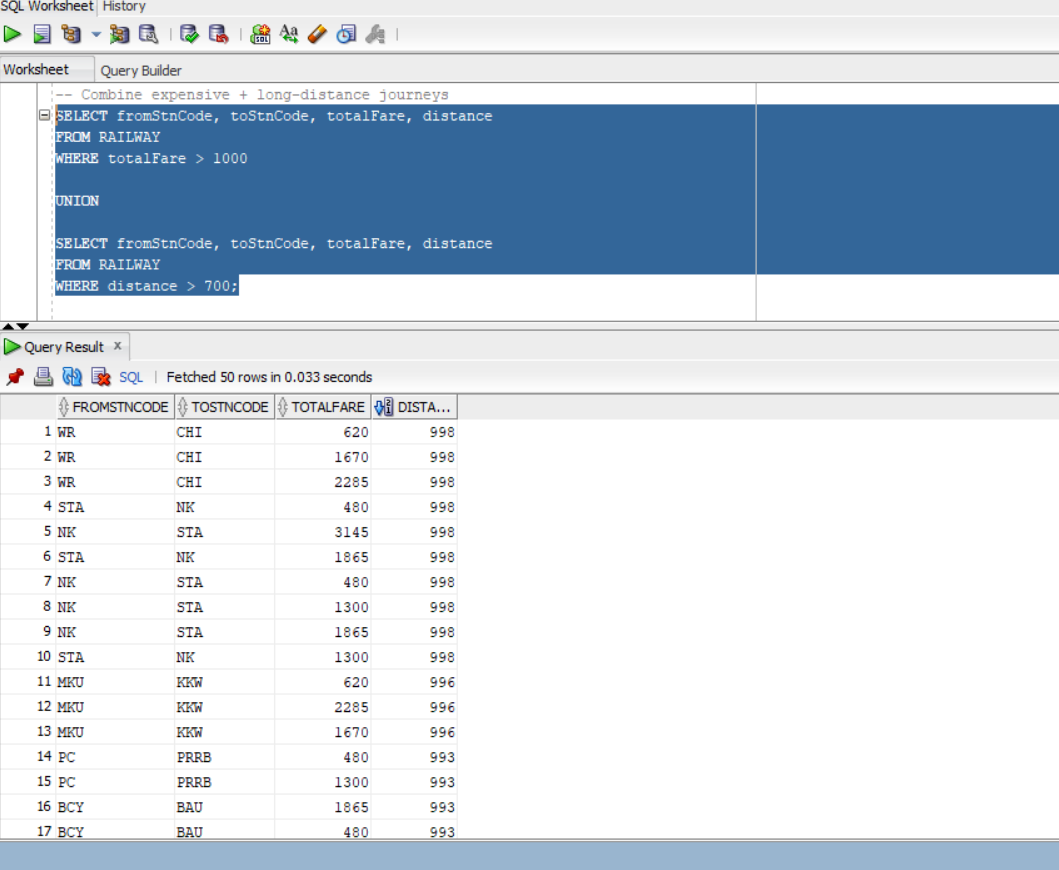
WHERE totalFare > 1000

UNION

SELECT fromStnCode, toStnCode, totalFare, distance

FROM RAILWAY

WHERE distance > 700;

**Result Screenshot:**  
**

**Insight:**  
By combining high-fare and long-distance routes using a UNION query, we identified premium train sectors that either generate high revenue due to fare pricing or cover major long-haul corridors. These routes highlight where Indian Railways can focus on improving onboard services, optimizing coach compositions, and potentially implementing dynamic pricing to balance affordability and profitability.

**4. Key Findings**

* Premium train routes and higher travel classes contribute the most to overall revenue.
* Certain long-distance routes have lower fare per km, indicating competitive pricing.
* Data can help optimize pricing and improve class allocation.

**5. Conclusion**

The SQL analysis successfully identified revenue-generating routes, high-cost travel sectors, and fare trends across classes. These insights can support better operational and pricing decisions for Indian Railways.