

Creating the table

```
CREATE TABLE TrainRides (  
    Transaction_ID INT PRIMARY KEY,           -- Unique identifier for ticket purchase  
    Date_of_Purchase DATE,                   -- Date the ticket was purchased  
    Time_of_Purchase TIME,                   -- Time the ticket was purchased  
    Purchase_Type VARCHAR(20),               -- Ticket purchased online or station  
    Payment_Method VARCHAR(20),              -- Payment method used  
    Railcard VARCHAR(20),                    -- Railcard type  
    Ticket_Class VARCHAR(10),                -- Ticket class (Standard or First)  
    Ticket_Type VARCHAR(20),                 -- Ticket type (Advance, Off-Peak,  
Anytime)  
    Price DECIMAL(10, 2),                    -- Final ticket cost  
    Departure_Station VARCHAR(100),          -- Departure station name  
    Arrival_Destination VARCHAR(100),        -- Arrival station name  
    Date_of_Journey DATE,                   -- Date of the train journey  
    Departure_Time TIME,                    -- Train departure time  
    Arrival_Time TIME,                      -- Scheduled train arrival time  
    Actual_Arrival_Time TIME,               -- Actual train arrival time  
    Journey_Status VARCHAR(20),              -- Journey status  
    Reason_for_Delay VARCHAR(100),           -- Reason for delay or cancellation  
    Refund_Request VARCHAR(3),               -- Whether a refund was requested  
);
```

Alter the table

```
ALTER TABLE TrainRides  
ALTER COLUMN Transaction_ID TYPE TEXT;
```

```
ALTER TABLE TrainRides  
ALTER COLUMN Ticket_Class TYPE TEXT;
```

```
ALTER TABLE TrainRides  
ALTER COLUMN Ticket_Type TYPE TEXT;
```

Uploading the Dataset

```
COPY TrainRides(Transaction_ID, Date_of_Purchase, Time_of_Purchase, Purchase_Type,  
Payment_Method, Railcard,  
Ticket_Class, Ticket_Type, Price, Departure_Station, Arrival_Destination, Date_of_Journey,  
Departure_Time, Arrival_Time, Actual_Arrival_Time, Journey_Status, Reason_for_Delay,  
Refund_Request)  
FROM 'C:\Users\Administrator\Downloads\UK+Train+Rides\railway.csv'  
DELIMITER ','  
CSV HEADER;
```

```
SELECT * FROM TrainRides;
```

Total number of transactions, total revenue, average ticket price, and breakdown by purchase type.

```
SELECT
    Purchase_Type,
    COUNT(*) AS Total_Transactions,
    SUM(Price) AS Total_Revenue,
    AVG(Price) AS Avg_Ticket_Price
FROM TrainRides
GROUP BY Purchase_Type
ORDER BY Total_Revenue DESC;
```

Result:

purchase_type	total_transactions	total_revenue	avg_ticket_price
Online	18521	382754	20.66595
Station	13132	359167	27.35052

Analyze ticket sales and revenue by ticket type, railcard type, and class.

```
SELECT
    Ticket_Type,
    Railcard,
    Ticket_Class,
    COUNT(*) AS Total_Tickets_Sold,
    SUM(Price) AS Revenue,
    AVG(Price) AS Average_Ticket_Price
FROM TrainRides
GROUP BY Ticket_Type, Railcard, Ticket_Class
ORDER BY Revenue DESC;
```

Result:

ticket_type	railcard	ticket_class	total_tickets_sold	revenue	average_ticket_price
Advance	None	Standard	10494	189496	18.05756
Off-Peak	None	Standard	5364	139473	26.00168
Anytime	None	Standard	3075	132430	43.06667
Advance	None	First Class	1131	50337	44.50663
Off-Peak	None	First Class	558	35627	63.84767
Advance	Adult	Standard	2420	27855	11.51033
Anytime	None	First Class	296	26334	88.96622
Off-Peak	Adult	Standard	1208	22645	18.74586
Anytime	Adult	Standard	728	17551	24.10852
Advance	Disabled	Standard	1489	15217	10.21961
Anytime	Disabled	Standard	493	14090	28.58012
Advance	Senior	Standard	1394	9820	7.044476

Off-Peak	Disabled	Standard	793	9746	12.29004
Advance	Adult	First Class	291	8525	29.29553
Anytime	Senior	Standard	554	7397	13.35199
Off-Peak	Senior	Standard	583	6802	11.66724
Advance	Disabled	First Class	172	5316	30.90698
Anytime	Adult	First Class	83	5236	63.08434
Anytime	Disabled	First Class	56	4673	83.44643
Off-Peak	Adult	First Class	116	4518	38.94828
Off-Peak	Disabled	First Class	86	3236	37.62791
Advance	Senior	First Class	170	2708	15.92941
Anytime	Senior	First Class	55	1598	29.05455
Off-Peak	Senior	First Class	44	1291	29.34091

Analyze train delays by reason and revenue loss due to refunds.

```

SELECT
    Journey_Status,
    Reason_For_Delay,
    COUNT(*) AS Total_Delayed_Or_Cancelled,
    AVG(EXTRACT(EPOCH FROM (Actual_Arrival_Time - Arrival_Time)) / 60) AS
Avg_Delay_Duration,
    SUM(CASE WHEN Refund_Request = 'Yes' THEN Price ELSE 0 END) AS
Revenue_Loss_Refunds
FROM TrainRides
WHERE Journey_Status IN ('Delayed', 'Cancelled')
GROUP BY Journey_Status, Reason_For_Delay
ORDER BY Total_Delayed_Or_Cancelled DESC;

```

Result:

journey_status	reason_for_delay	total_delayed_or_cancelled	avg_delay_duration	revenue_loss_refunds
Delayed	Weather	758	46.80738786	0
Delayed	Technical Issue	472	24.91737288	13227
Cancelled	Signal Failure	281	NULL	1548
Delayed	Signal Failure	242	71.68181818	308
Cancelled	Staffing	238	NULL	1637
Cancelled	Signal failure	238	NULL	1533
Cancelled	Weather	237	NULL	1556
Cancelled	Technical Issue	235	NULL	1819
Cancelled	Traffic	227	NULL	1676
Cancelled	Staff Shortage	216	NULL	1549
Delayed	Signal failure	209	28.80382775	1009
Cancelled	Weather	208	NULL	1219
	Conditions			
Delayed	Staff Shortage	183	74.93442623	871

Delayed	Staffing	172	25.90697674	7610
Delayed	Weather	169	30.5443787	0
	Conditions			
Delayed	Traffic	87	32.34482759	3140

Identify the most and least popular departure and arrival stations.

-- Top Departure Stations

```
SELECT
    Departure_Station AS Station,
    COUNT(*) AS Total_Departures
FROM TrainRides
GROUP BY Departure_Station
ORDER BY Total_Departures DESC
LIMIT 10;
```

Result:

station	total_departures
Manchester Piccadilly	5650
London Euston	4954
Liverpool Lime Street	4561
London Paddington	4500
London Kings Cross	4229
London St Pancras	3891
Birmingham New Street	2136
York	927
Reading	594
Oxford	144

-- Top Arrival Stations

```
SELECT
    Arrival_Destination AS Station,
    COUNT(*) AS Total_Arrivals
FROM TrainRides
GROUP BY Arrival_Destination
ORDER BY Total_Arrivals DESC
LIMIT 10;
```

Result:

station	total_arrivals
Birmingham New Street	7742
Liverpool Lime Street	5022
York	4019
Manchester Piccadilly	3968
Reading	3920

London Euston	1567
London St Pancras	749
Oxford	623
London Paddington	351
Leicester	337

Find peak times for ticket purchases and train departures.

-- [Peak Ticket Purchase Times](#)

```
SELECT
    EXTRACT(HOUR FROM Time_of_Purchase) AS Purchase_Hour,
    COUNT(*) AS Tickets_Purchased
FROM TrainRides
GROUP BY EXTRACT(HOUR FROM Time_of_Purchase)
ORDER BY Tickets_Purchased DESC;
```

Result:

purchase_hour	tickets_purchased
17	2740
20	2239
9	2070
7	2046
8	2008
6	1910
14	1869
5	1566
15	1468
18	1425
10	1187
19	1160
3	1107
16	1056
1	1032
12	1025
0	925
4	924
13	754
11	743
23	726
2	642
21	573
22	458

-- Peak Departure Times

```
SELECT
    EXTRACT(HOUR FROM Departure_Time) AS Departure_Hour,
    COUNT(*) AS Journeys_Departed
FROM TrainRides
GROUP BY EXTRACT(HOUR FROM Departure_Time)
ORDER BY Journeys_Departed DESC;
```

Result:

departure_hour	journeys_departed
18	3113
6	3112
17	2888
7	2795
16	2301
8	2179
13	1276
9	1230
15	1220
11	1143
20	1058
4	1041
2	942
14	855
0	853
22	788
12	773
5	725
1	644
23	641
21	570
3	543
10	525
19	438

Determine refund request rates based on journey status.

```
SELECT
    Journey_Status,
    COUNT(*) AS Total_Journeys,
    SUM(CASE WHEN Refund_Request = 'Yes' THEN 1 ELSE 0 END) AS Refund_Requests,
    ROUND(SUM(CASE WHEN Refund_Request = 'Yes' THEN 1 ELSE 0 END) * 100.0 /
```

```
COUNT(*), 2) AS Refund_Rate_Percentage
FROM TrainRides
GROUP BY Journey_Status;
```

Result:

journey_status	total_journeys	refund_requests	refund_rate_percentage
On Time	27481	0	0
Cancelled	1880	572	30.43
Delayed	2292	546	23.82