UK TRAIN RIDES

Analysis – SQL CASE STUDY

CREATED BY –
YASH GUPTA

INTRODUCTION

Mock train ticket data for National Rail in the UK, from Jan to Apr 2024, including details on the type of ticket, the date & time for each journey,the departure & arrival stations, the ticket price, and more.

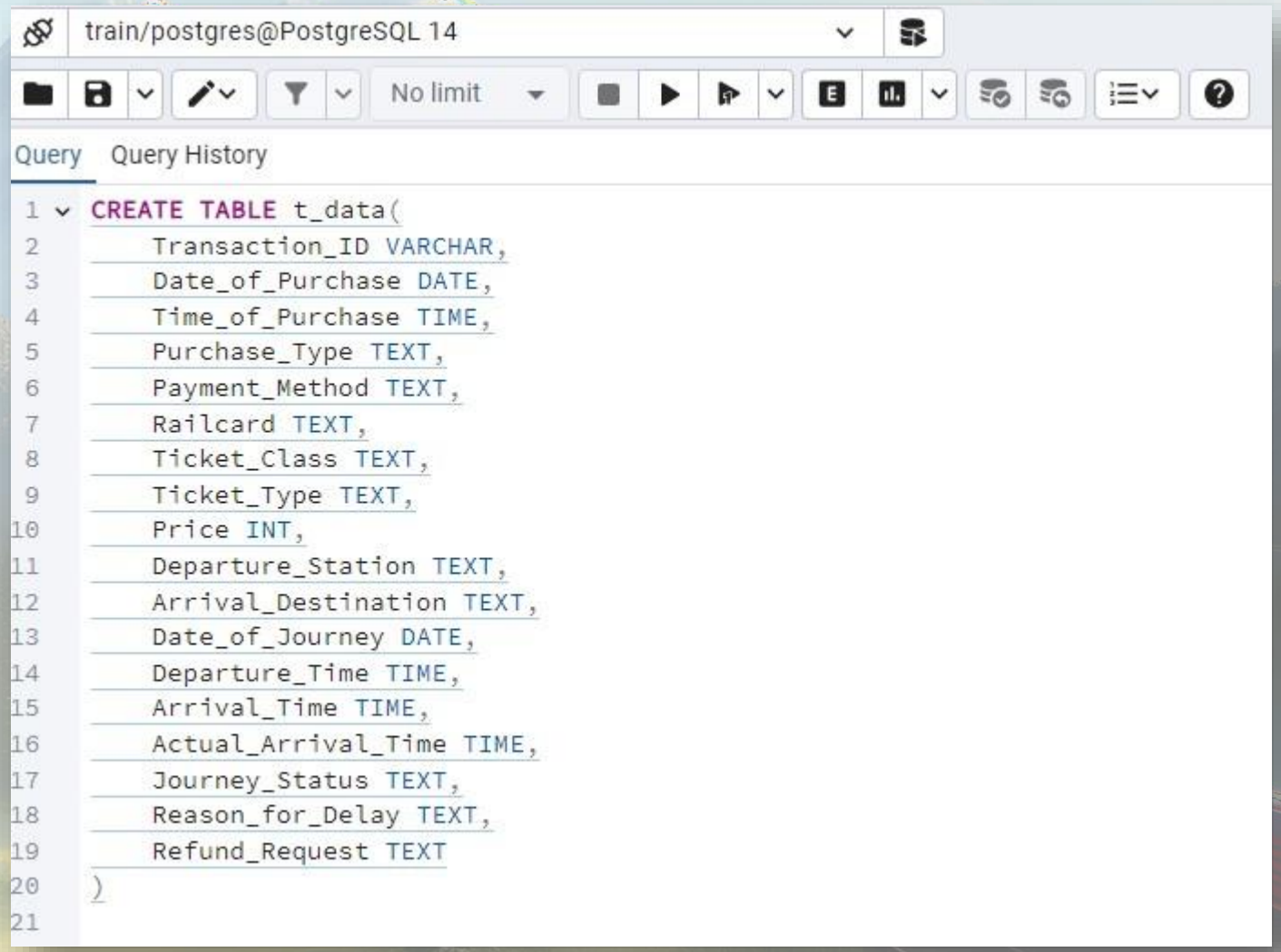
OBJECTIVES

1. What are the most popular routes?
2. What are the peak travel times?
3. How does revenue vary by ticket types and classes?
4. What is the on-time performance?
5. What are the main contributing factors?

DATASET

- 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

CREATE TABLE IN SQL DATABASE



Q1 What are the most popular routes?

No limit

Query | Query History

```sql
WITH NEW_D AS (
    SELECT transaction_id,CONCAT(departure_station,' to ',arrival_destination) as routes
    FROM t_data)

select routes ,
    rank () over (order by count(routes)) as ranking ,
    count(transaction_id)
    from NEW_D
    GROUP BY routes
    order by ranking desc
    limit 5
```

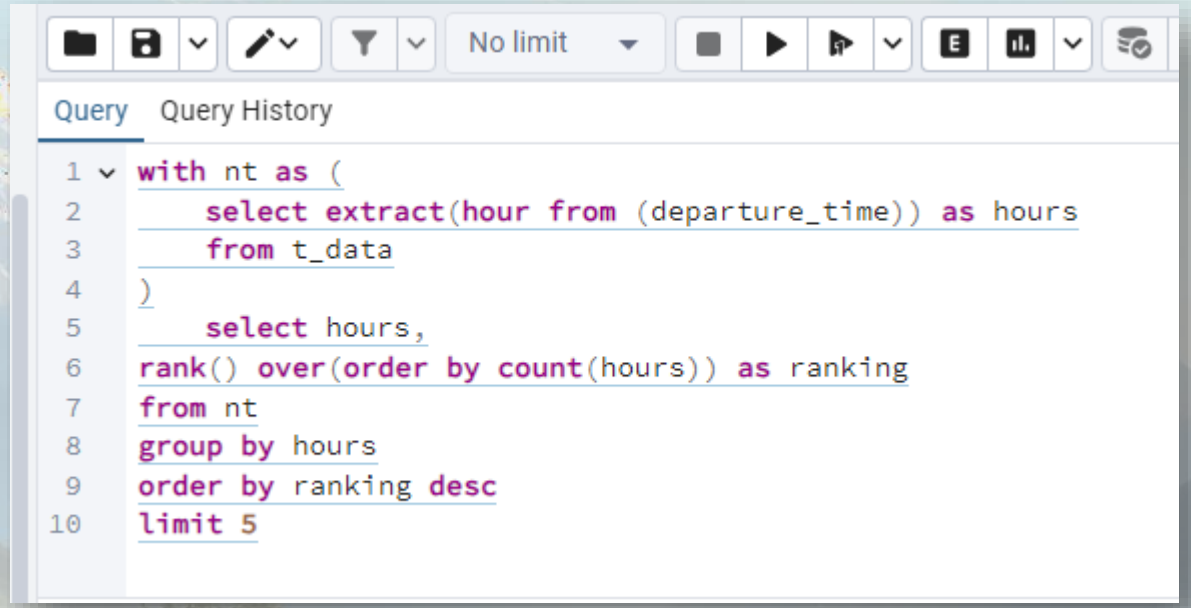
Output

Data Output | Messages | Notifications

SQL

	routes text	ranking bigint	count bigint
1	Manchester Piccadilly to Liverpool Lime Street	65	4628
2	London Euston to Birmingham New Street	64	4209
3	London Kings Cross to York	63	3922
4	London Paddington to Reading	62	3873
5	London St Pancras to Birmingham New Street	61	3471

Q2 What are the peak travel times?



Output

Data Output | Messages | Notifications

	hours numeric	ranking bigint
1	18	24
2	6	23
3	17	22
4	7	21
5	16	20

Q3 How does revenue vary by ticket type and class?

No limit

Query | Query History

```
1  select ticket_class,ticket_type , sum(price) as revenue
2  from t_data
3  group by ticket_class,ticket_type
4  order by revenue desc
```

Output

Data Output | Messages | Notifications

SQL

	ticket_class text	ticket_type text	revenue bigint
1	Standard	Advance	242388
2	Standard	Off-Peak	178666
3	Standard	Anytime	171468
4	First Class	Advance	66886
5	First Class	Off-Peak	44672
6	First Class	Anytime	37841

Q4 What is the on time performance?

No limit

Query | Query History

```
1  SELECT journey_status,count(journey_status) as total
2  from t_data
3  group by journey_status
4  order by total desc
```

Output

Data Output | Messages | Notifications

SQL

	journey_status text	total bigint
1	On Time	27481
2	Delayed	2292
3	Cancelled	1880

Q5 What are the main contributing factors?

No limit

Query | Query History

```
1  select reason_for_delay, count(reason_for_delay) as number_of_delay
2  from t_data
3  group by reason_for_delay
4  order by number_of_delay desc
5  limit 5
```

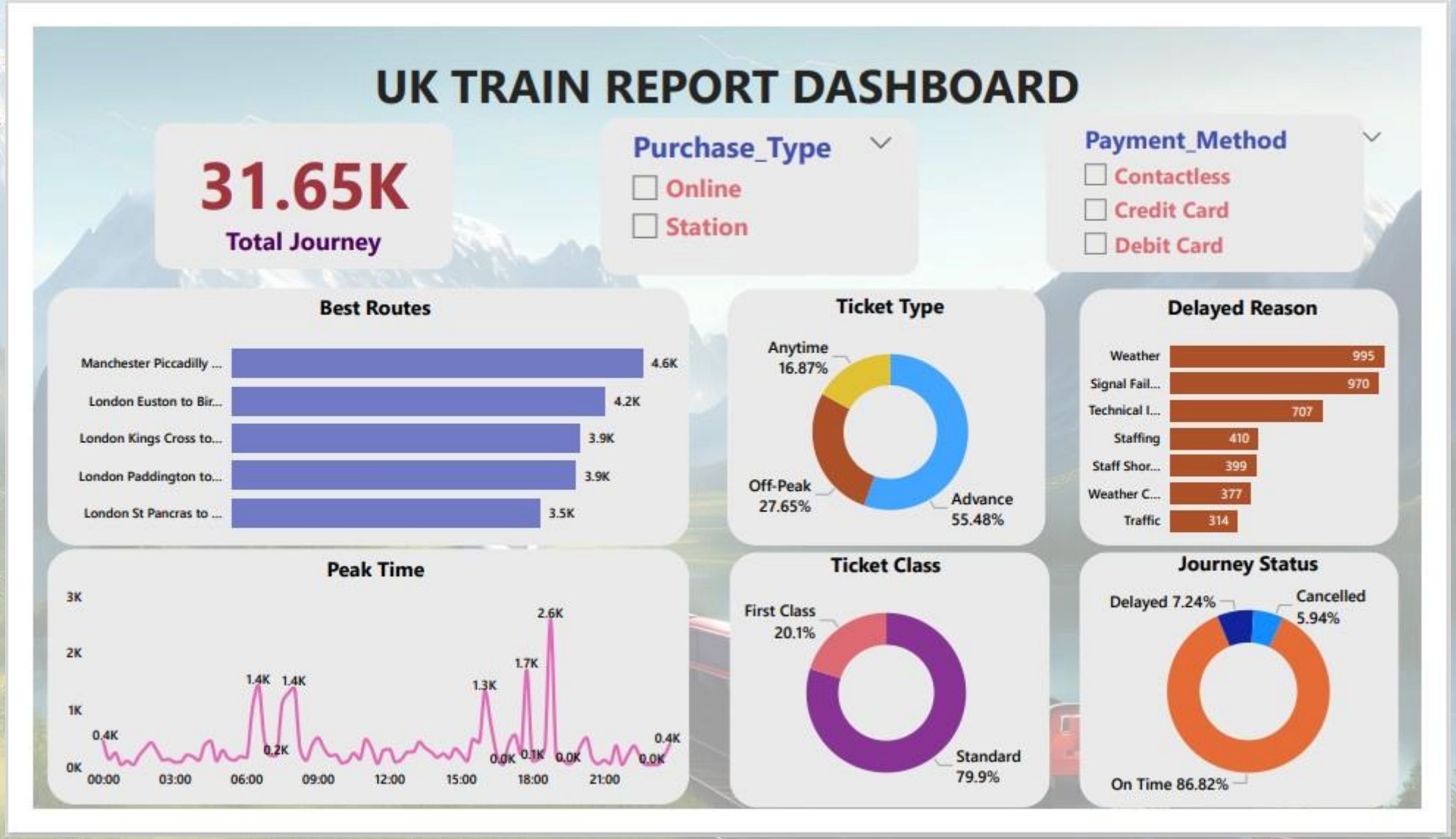
Output

Data Output | Messages | Notifications

SQL

	reason_for_delay text	number_of_delay bigint
1	Weather	995
2	Technical Issue	707
3	Signal Failure	523
4	Signal failure	447
5	Staffing	410

Dashboard


UK TRAIN REPORT DASHBOARD
31.65K
Total Journey
Purchase_Type
Online
Station
Payment_Method
Contactless
Credit Card
Debit Card
Best Routes
Manchester Piccadilly ...
4.6K
London Euston to Bir...
4.2K
London Kings Cross to...
3.9K
London Paddington to...
3.9K
London St Pancras to ...
3.5K
Ticket Type
Anytime
16.87%
Off-Peak
27.65%
Advance
55.48%
Delayed Reason
Weather
995
Signal Fail...
970
Technical I...
707
Staffing
410
Staff Shor...
399
Weather C...
377
Traffic
314
Peak Time
3K
2K
1K
0K
0.4K
1.4K
1.4K
0.2K
1.3K
0.0K
1.7K
0.1K
2.6K
0.0K
0.0K
0.4K
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00
Ticket Class
First Class
20.1%
Standard
79.9%
Journey Status
Delayed 7.24%
Cancelled
5.94%
On Time 86.82%

Insights

- Most popular route used by people is “Manchester Piccadilly to Liverpool Lime Street”
- Peak travel time in UK is between “6 am to 8 am” and “4 pm to 7 pm”
- Most revenue generated from “Standard class”
- On time performance of UK trains is 86%
- Maximum delayed reason is bad weather condition

SOFTWARE USES

- PostgreSQL
- Power BI

Thank you

Created by YASH GUPTA