UK TRAIN RIDES

Analysis – SQL CASE STUDY

CREATED BY – RANA BASAK

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?

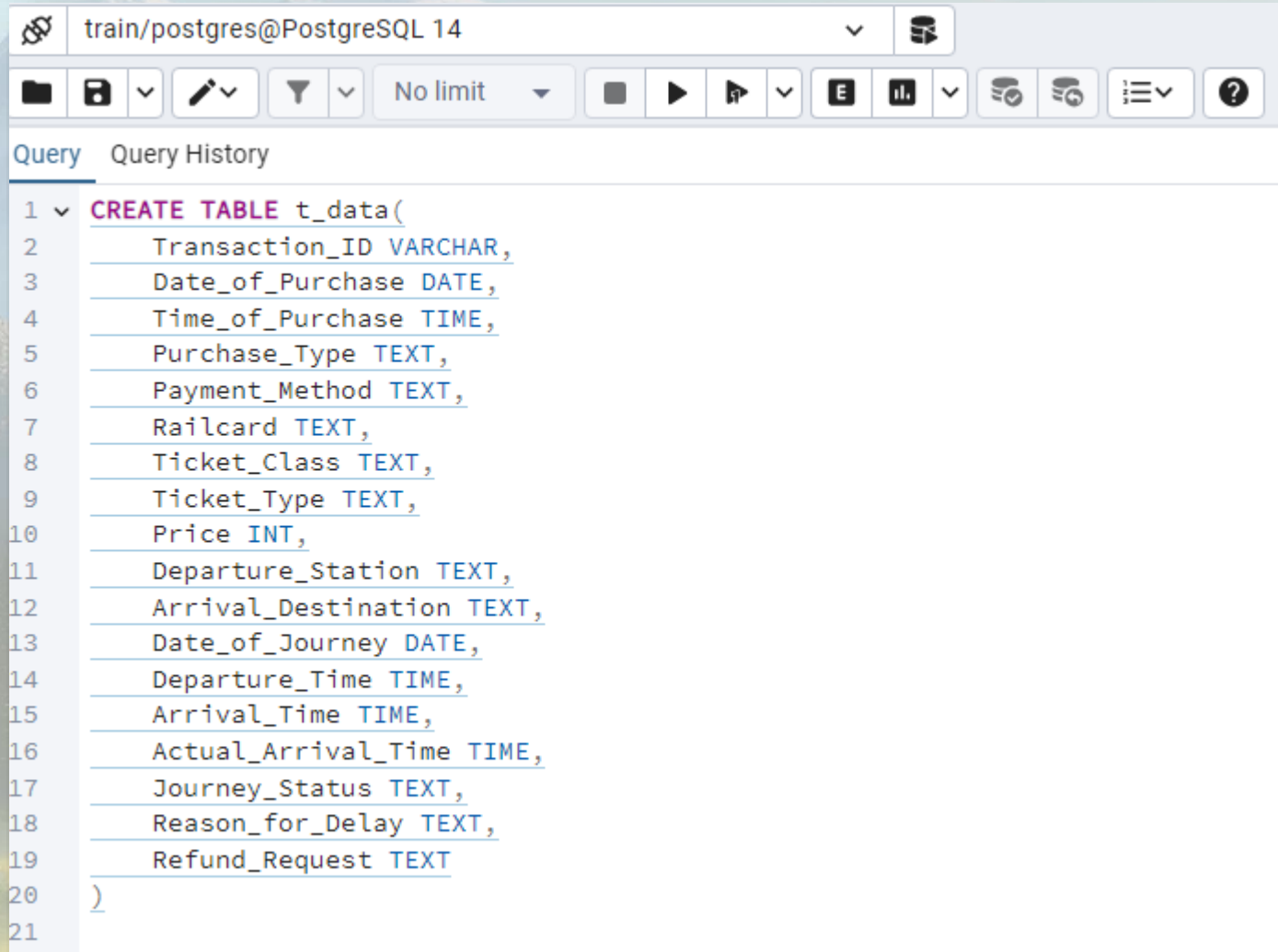5. What are the main contributing factors?
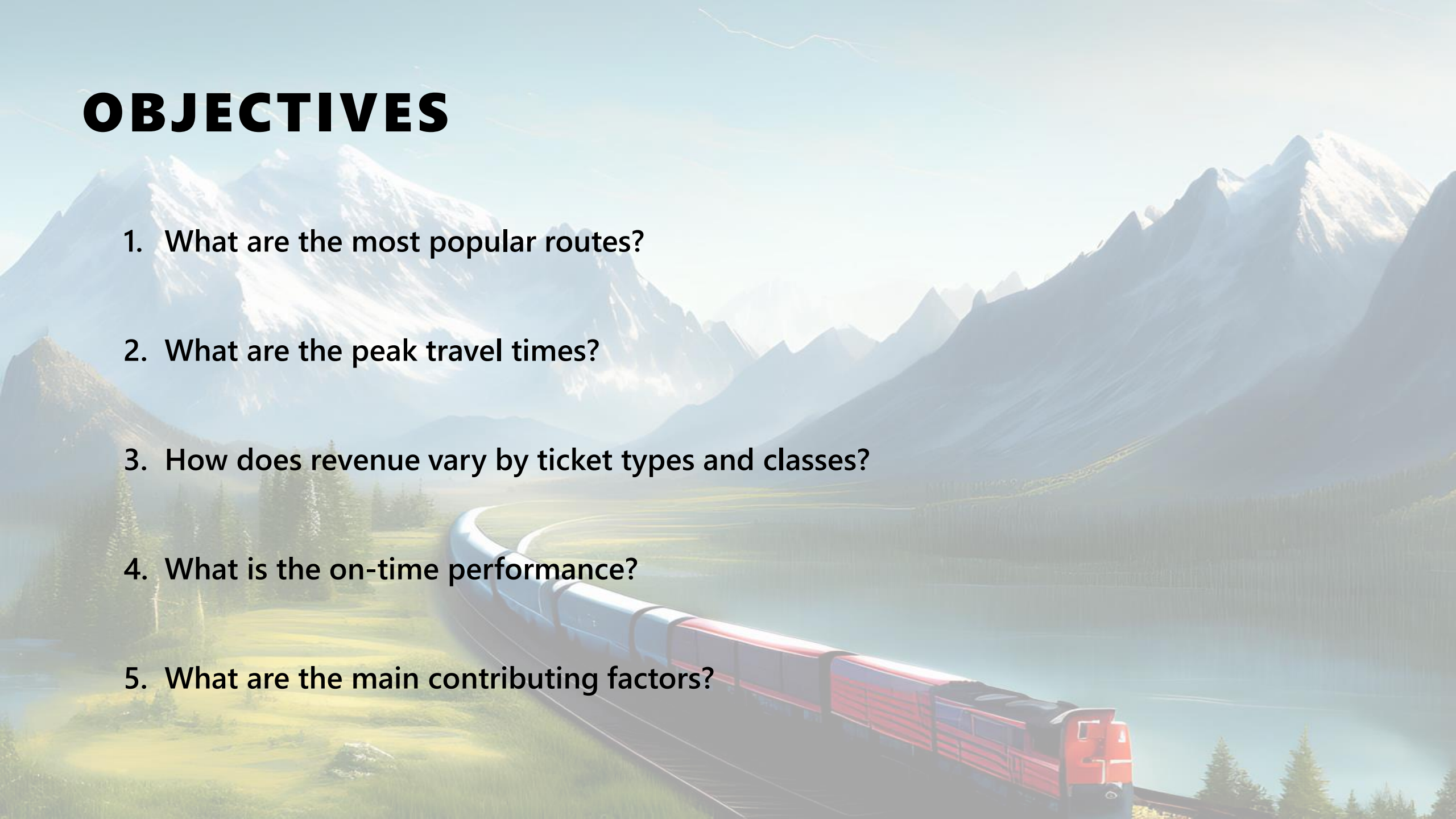
DATASET

- 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
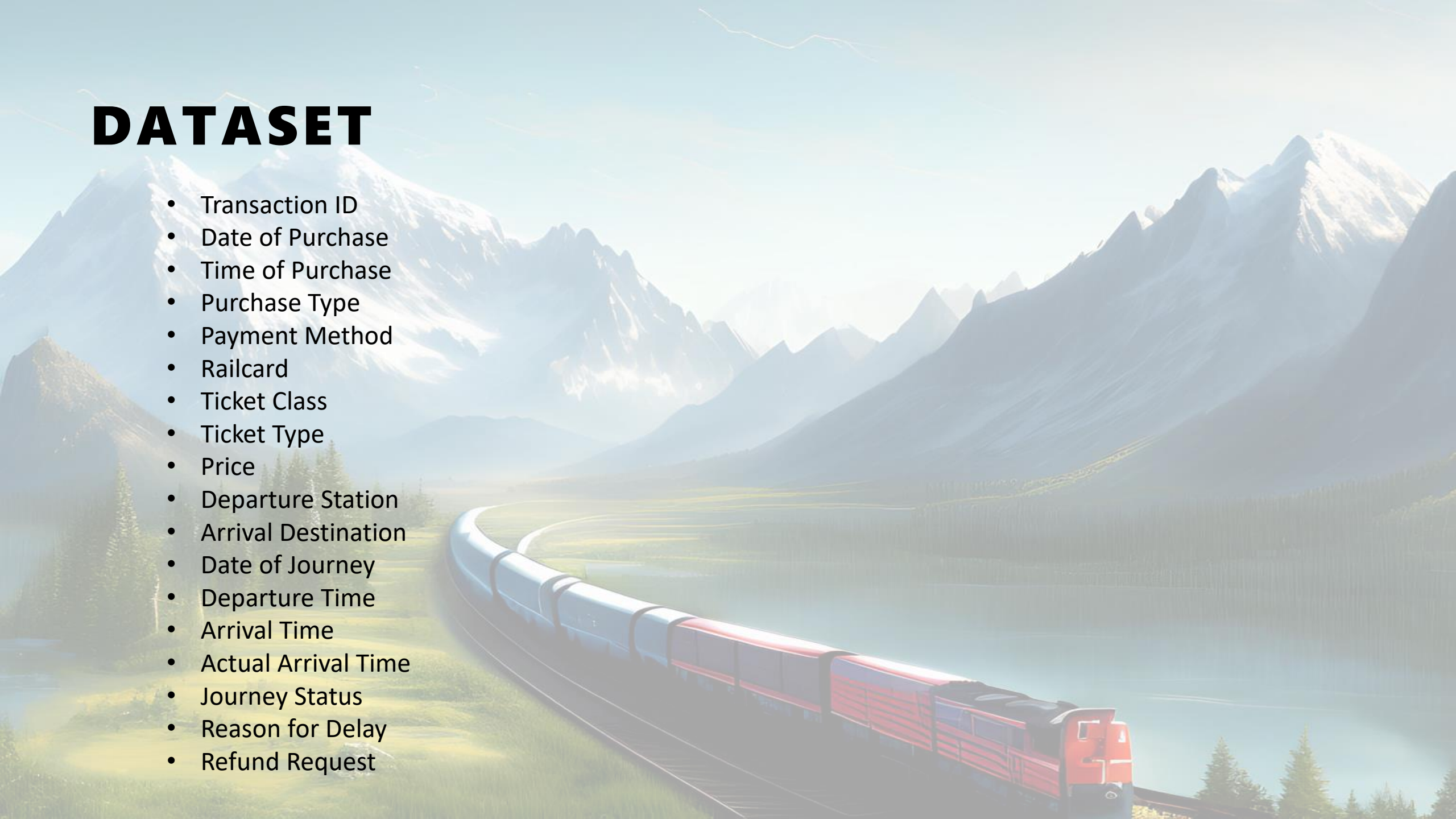
CREATE TABLE IN SQL DATABASE

train/postgres@PostgreSQL 14

Query   Query History

```sql
CREATE TABLE t_data(
    Transaction_ID VARCHAR,
    Date_of_Purchase DATE,
    Time_of_Purchase TIME,
    Purchase_Type TEXT,
    Payment_Method TEXT,
    Railcard TEXT,
    Ticket_Class TEXT,
    Ticket_Type TEXT,
    Price INT,
    Departure_Station TEXT,
    Arrival_Destination TEXT,
    Date_of_Journey DATE,
    Departure_Time TIME,
    Arrival_Time TIME,
    Actual_Arrival_Time TIME,
    Journey_Status TEXT,
    Reason_for_Delay TEXT,
    Refund_Request TEXT
)
```

Q1 What are the most popular routes?

```sql
1  WITH NEW_D AS (
2      SELECT transaction_id,CONCAT(departure_station,' to ',arrival_destination) as routes
3      FROM t_data)
4
5  select routes ,
6      rank () over (order by count(routes)) as ranking ,
7      count(transaction_id)
8      from NEW_D
9      GROUP BY routes
10     order by ranking desc
11     limit 5
```

Output

	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?

```sql
1  with nt as (
2      select extract(hour from (departure_time)) as hours
3      from t_data
4  )
5      select hours,
6  rank() over(order by count(hours)) as ranking
7  from nt
8  group by hours
9  order by ranking desc
10 limit 5
```

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?

```sql
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

	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?

```sql
SELECT journey_status,count(journey_status) as total
from t_data
group by journey_status
order by total desc
```

Output

journey_status	total
On Time	27481
Delayed	2292
Cancelled	1880

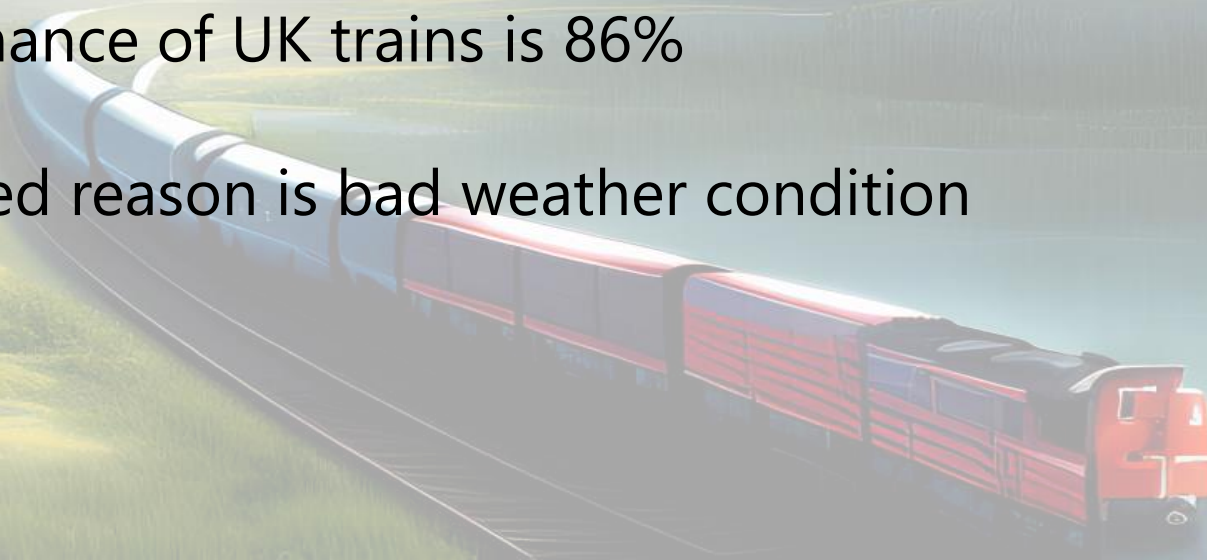# Q5 What are the main contributing factors?

Output

Dashboard
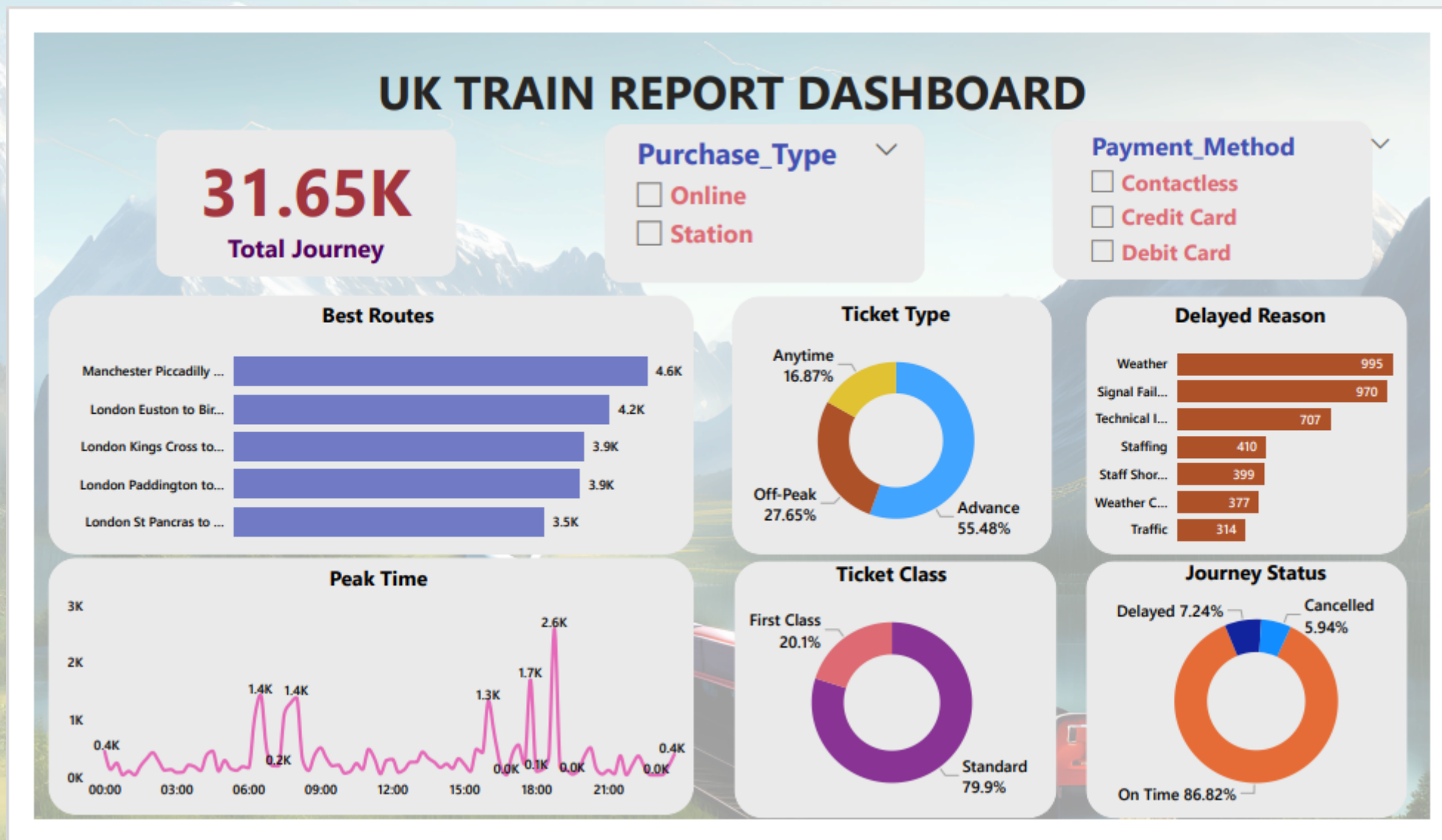
Insights

- 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 RANA BASAK