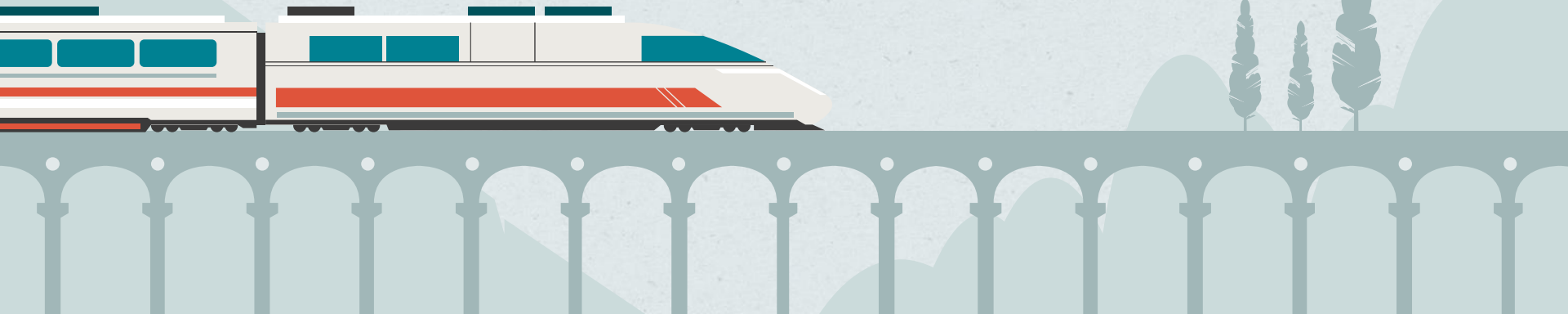
UK TRAIN RIDE REPORT

Presented by Rana Basak
Data Analyst



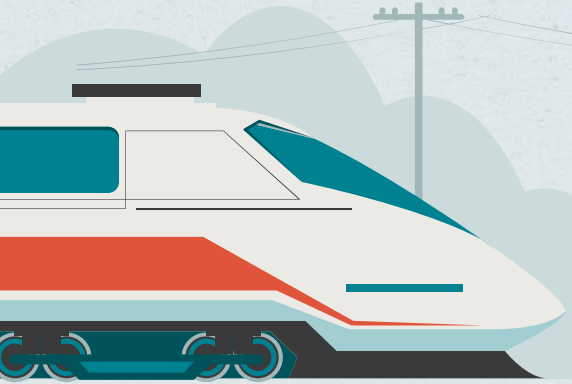
INTRODUCTION

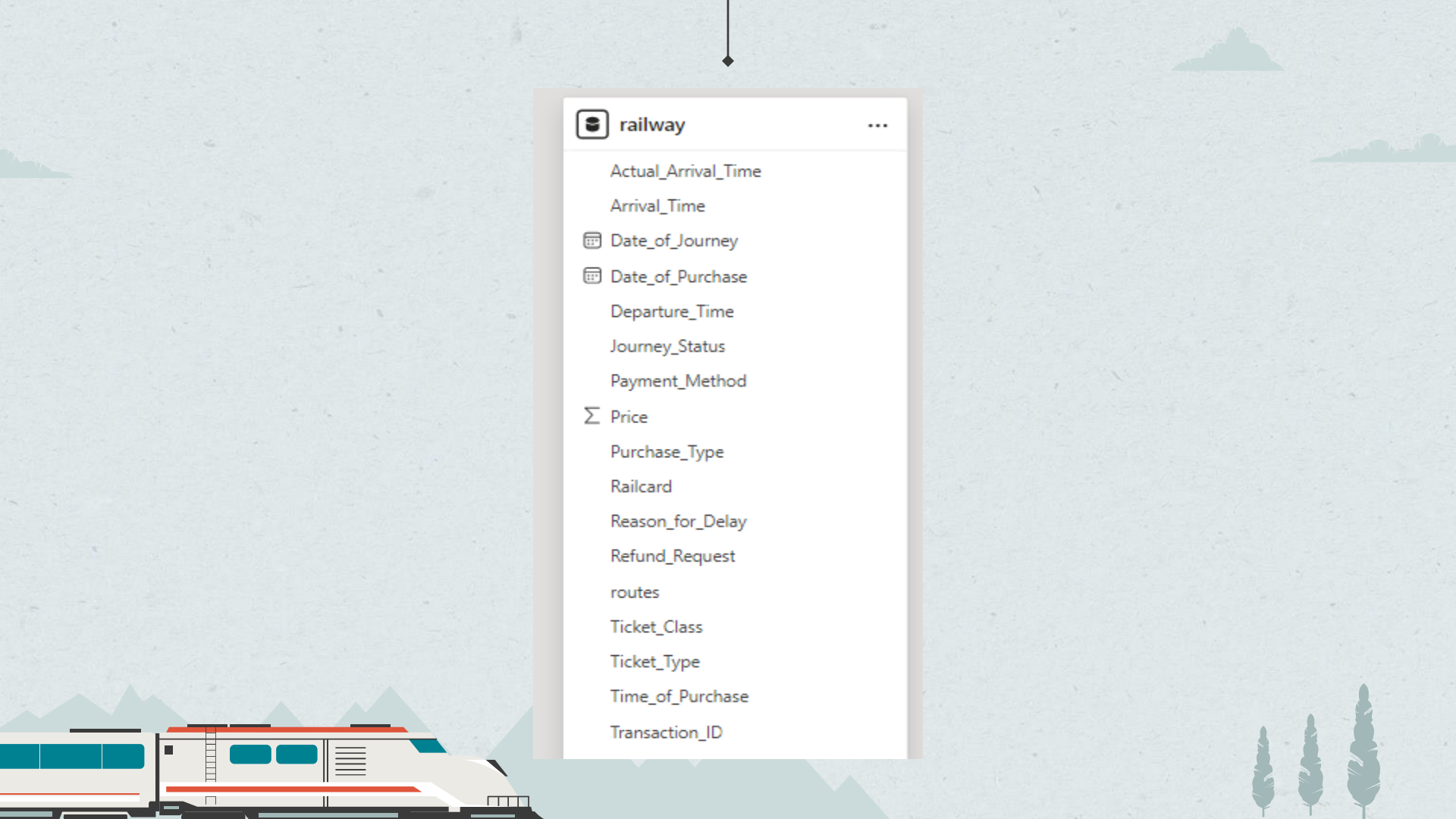
- Hello everyone, my name is Rana Basak.
- I’m a Data Analyst with a Passion for uncovering insights from data.
- In this project, I’ve undertaken a detailed analysis of train rides, utilizing SQL from basic to advanced levels.
- This presentation will walks you through the methods and insights gained from this analysis, showcasing how SQL can be powerfull tool for driving data- drivin decisions.

Table of contents

01

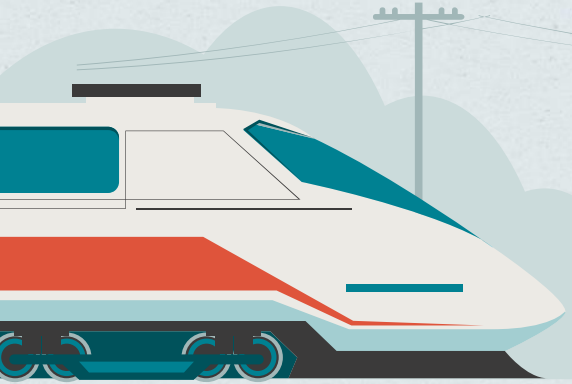
Data Model




railway
Actual_Arrival_Time
Arrival_Time
Date_of_Journey
Date_of_Purchase
Departure_Time
Journey_Status
Payment_Method
Price
Purchase_Type
Railcard
Reason_for_Delay
Refund_Request
routes
Ticket_Class
Ticket_Type
Time_of_Purchase
Transaction_ID

02

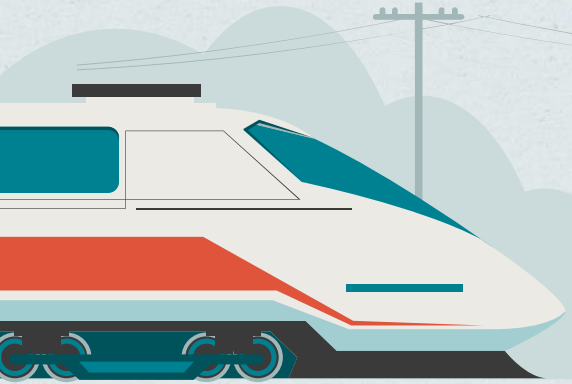
Skilled Applied



- Imported data from CSV file into PostgreSQL database.
- Use Window Functions for calculate required context.
- Applied Group By to aggregate data for insightful analysis.
- Used Order By to short data for better clairty.
- Implemented CTE to solve complex questions.

03

SQL Queries



Retrive the total Number of Transaction and total revenue

Query | Query History

```
1  SELECT COUNT(*) AS TOTAL_TRANSACTION
2  FROM
3      T_DATA
4
```

Data Output | Messages

	total_transaction bigint
1	31653

Query | Query History

```
1  SELECT SUM(Price) AS TOTAL_REVENUE
2  FROM
3      T_DATA
4
```

Data Output | Messages

	total_revenue bigint
1	741921

Retrive the total number of journey and total railcards

Query | Query History

```
1  SELECT COUNT(journey_STATUS) AS TOTAL_JOURNEY
2  FROM
3      T_DATA
4  WHERE
5      JOURNEY_STATUS != 'Cancelled';
```

Data Output | Messages

	total_journey bigint
1	29773

Query | Query History

```
1  SELECT
2      COUNT(RAILCARD) AS TOTAL_RAILCARDS
3  FROM
4      T_DATA
5  WHERE
6      RAILCARD != 'None';
```

Data Output | Messages

	total_railcards bigint
1	10735

Identify the most popular routes

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;
```

Data Output | Messages | Notifications

	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

Identify the peak travel times

Query | Query History

```
1  WITH
2      NT AS (
3          SELECT EXTRACT( HOUR FROM (DEPARTURE_TIME)) AS time
4          FROM T_DATA
5      )
6  SELECT
7      time,
8      RANK() OVER ( ORDER BY COUNT(time)) AS RANKING
9  FROM
10     NT GROUP BY time
11 ORDER BY
12     RANKING DESC;
```

Data Output | Messages | Notifi

	time numeric	ranking bigint
1	18	24
2	6	23
3	17	22
4	7	21
5	16	20
6	8	19
7	13	18
8	9	17
9	15	16
10	11	15

Revenue vary by ticket type and class

Query | Query History

```sql
SELECT
    TICKET_CLASS,
    TICKET_TYPE,
    SUM(PRICE) AS TOTAL_REVENUE
FROM
    T_DATA
GROUP BY
    TICKET_CLASS,
    TICKET_TYPE
ORDER BY
    TOTAL_REVENUE DESC
```

Data Output | Messages | Notifications

	ticket_class text	ticket_type text	total_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

On time performance of trains

Query | Query History

```
1  SELECT
2      JOURNEY_STATUS,
3      COUNT(JOURNEY_STATUS) AS TOTAL
4  FROM
5      T_DATA
6  GROUP BY
7      JOURNEY_STATUS
8  ORDER BY
9      TOTAL DESC
```

Data Output | Messages | Notifications

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

Delayed Reasons

Query | Query History

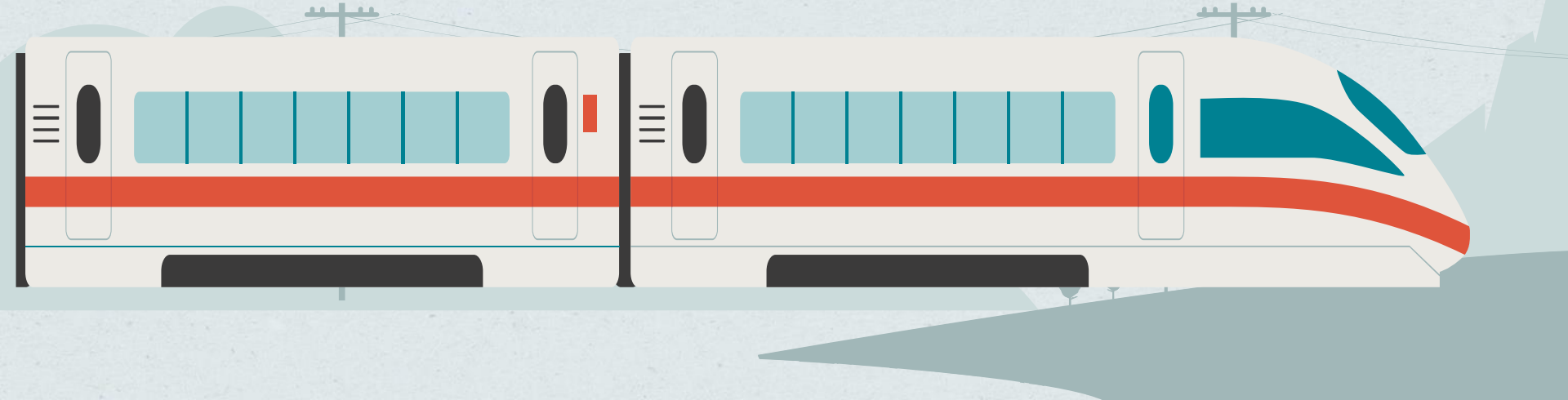
```
1  SELECT
2      REASON_FOR_DELAY,
3      COUNT(REASON_FOR_DELAY) AS NUMBER_OF_DELAY
4  FROM
5      T_DATA
6  GROUP BY
7      REASON_FOR_DELAY
8  ORDER BY
9      NUMBER_OF_DELAY DESC
```

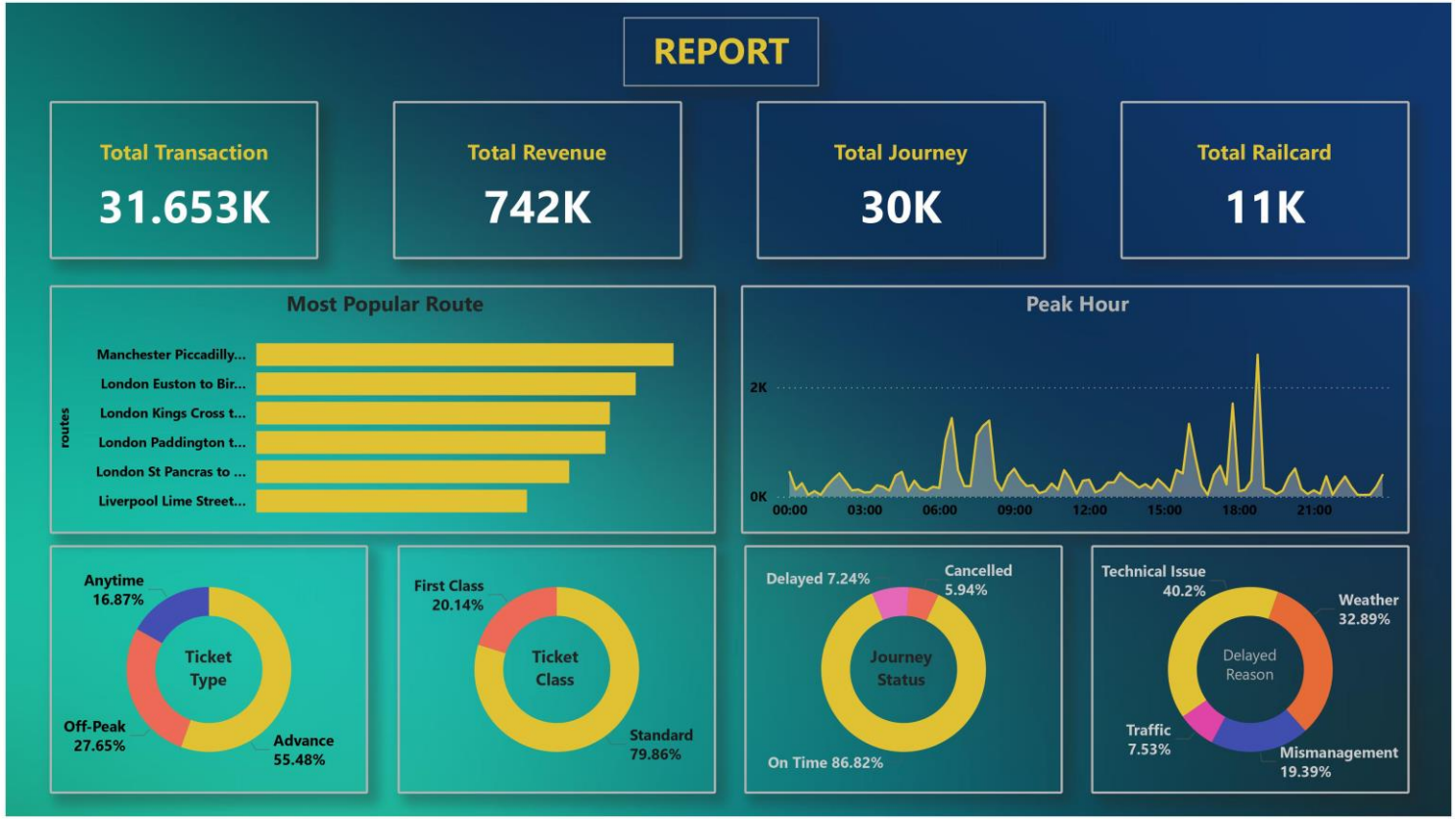
Data Output | Messages | Notifications

	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
6	Staff Shortage	399
7	Weather Conditions	377
8	Traffic	314

04

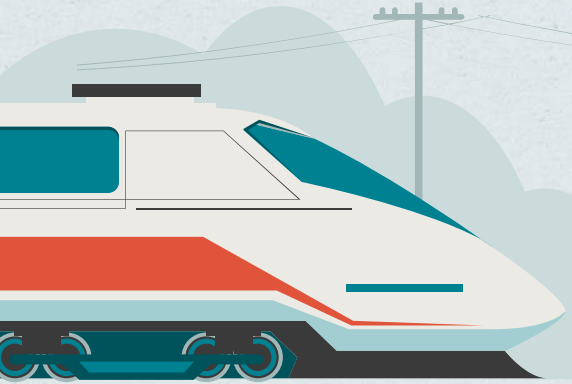
Dashboard



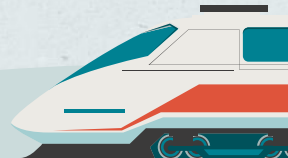

REPORT
Total Transaction
31.653K
Total Revenue
742K
Total Journey
30K
Total Railcard
11K
Most Popular Route
routes
Manchester Piccadilly...
London Euston to Bir...
London Kings Cross t...
London Paddington t...
London St Pancras to ...
Liverpool Lime Street...
Peak Hour
2K
0K
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00
Anytime
16.87%
Ticket
Type
Off-Peak
27.65%
Advance
55.48%
First Class
20.14%
Ticket
Class
Standard
79.86%
Delayed 7.24%
Cancelled
5.94%
Journey
Status
On Time 86.82%
Technical Issue
40.2%
Weather
32.89%
Delayed
Reason
Traffic
7.53%
Mismanagement
19.39%

05

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



THANKS!

Rana Basak

Ranabasak293@gmail.com