National Rail

Train Sales Dashboard

by Rolanda Azeem

Table of Contents

Project Description

Project Objectives

- Assess traveler behavior
- Explore operations performance
- Assess income streams

Data Source

The data for this project was sourced from Maven Analytics for educational purposes.

Data Structure

- Transaction ID - Unique identifier for an individual train ticket purchase
- Date of Purchase - Date the ticket was purchased
- Time of Purchase - Time the ticket was purchased
- Purchase Type - Whether the ticket was purchased online or directly at a train station
- Payment Method - Payment method used to purchase the ticket (Contactless, Credit Card, or Debit Card)
- Railcard - Whether the passenger is a National Railcard holder (Adult, Senior, or Disabled) or not (None). Railcard holders get 1/3 off their ticket purchases.
- Ticket Class - Seat class for the ticket (Standard or First)
- Ticket Type - When you bought or can use the ticket. Advance tickets are 1/2 off and must be purchased at least a day prior to departure. Off-Peak tickets are 1/4 off and must be used outside of peak hours (weekdays between 6-8am and 4-6pm). Anytime tickets are full price and can be bought and used at any time during the day.
- Price - Final cost of the ticket
- Departure Station - Station to board the train
- Arrival Destination - Station to exit the train
- Date of Journey - Date the train departed
- Departure Time - Time the train departed
- Arrival Time - Time the train was scheduled to arrive at its destination (can be on the day after departure)

- Actual Arrival Time - Time the train arrived at its destination (can be on the day after departure)
- Journey Status - Whether the train was on time, delayed, or canceled
- Reason for Delay - Reason for the delay or cancellation
- Refund Request - Whether the passenger requested a refund after a delay or cancellation

Design

Data Cleaning
- Column data types changed
- Changed some column names to better reflect content
- Reason for delay column contains both "Weather" and "Weather Conditions" which I made "Weather"

Data Transformation
- Additional columns were created for specific data to facilitate analysis
- Created columns
    - DepartureTimeOfDay : {6am - 12pm : morning; 12pm-5pm : afternoon; 5pm-8pm : evening; 8pm-12am : night; 12am-6am : Dawn}
    - JourneyDayOfWeek : the day of week the journey occurred
    - DayOfMonth: the day in the month journey occurred
    - TimeDifference - the difference between actual arrival time and scheduled arrival time
    - DaysToJourney: days between purchase and travel
    - DepartureHour: hour of day train departed

- Created a day of week sort table to sort the bars in graph.

Visualization Tools

Cards
Used to display key metrics and summary statistics.

Bar and Column Chart
Used to compare frequency of categories at a glance.

Clustered Bar Chart
Used where the differences in categories within a measure needed to be shown.

Area Chart
Used to visualize trends over time.

Pie and Doughnut Chart
Used to show proportions of a whole.

Combo Chart
Used to show patterns between 2 different matrices using different scales.

Tables
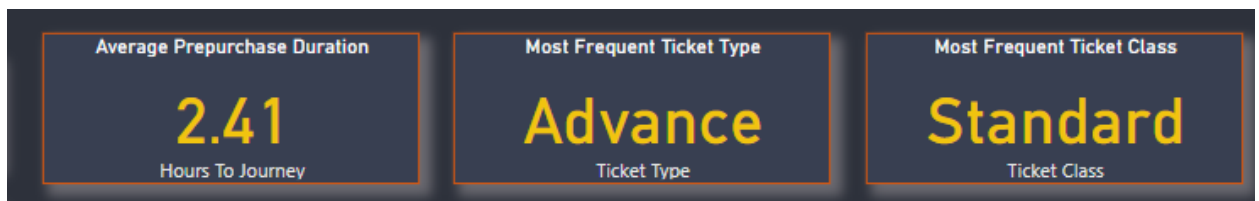Used to display relevant data in a structured manner.

Matrix
Used to summarize, aggregate and place data in a hierarchy.
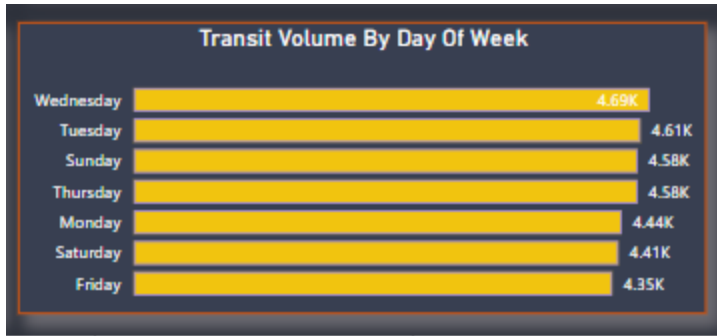
Filter
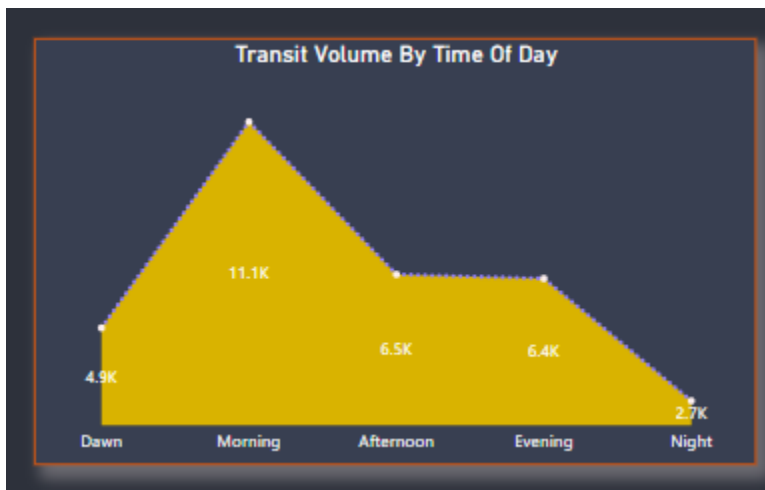Used to focus on specific subsets of data during analysis

Findings

- Though most people purchase Standard tickets in advance, the tickets are usually purchased just a few hours ahead of the journey.
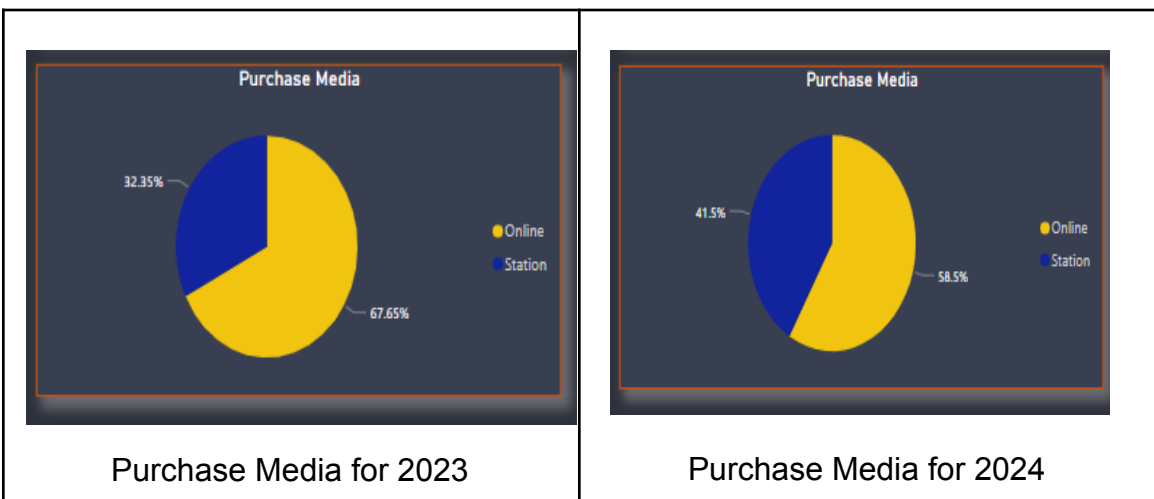

- Throughout the week, we get about the same patronage volume.

Transit Volume By Day Of Week

| Day	Volume
Wednesday	4.69K
Tuesday	4.61K
Sunday	4.58K
Thursday	4.58K
Monday	4.44K
Saturday	4.41K
Friday	4.35K

- As expected, mornings are peak time for train rides as most people have tightly scheduled mornings to get to work and appointments.

Transit Volume By Time Of Day

Dawn: 4.9K, Morning: 11.1K, Afternoon: 6.5K, Evening: 6.4K, Night: 2.7K

- We saw a surprising decrease in the use of online channels from 2023 to 2024.

Purchase Media for 2023 — Online 67.65%, Station 32.35%

Purchase Media for 2024 — Online 58.5%, Station 41.5%

- Naturally, our cheapest tickets get the most patronage, and subsequently, generates the most revenue.
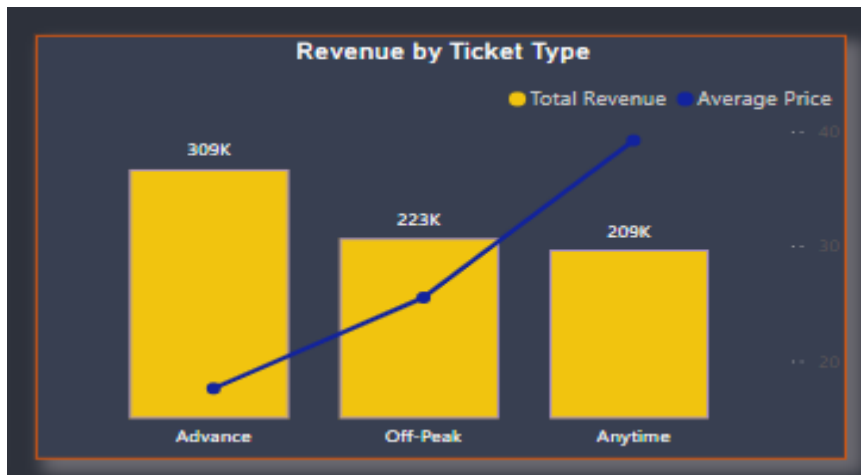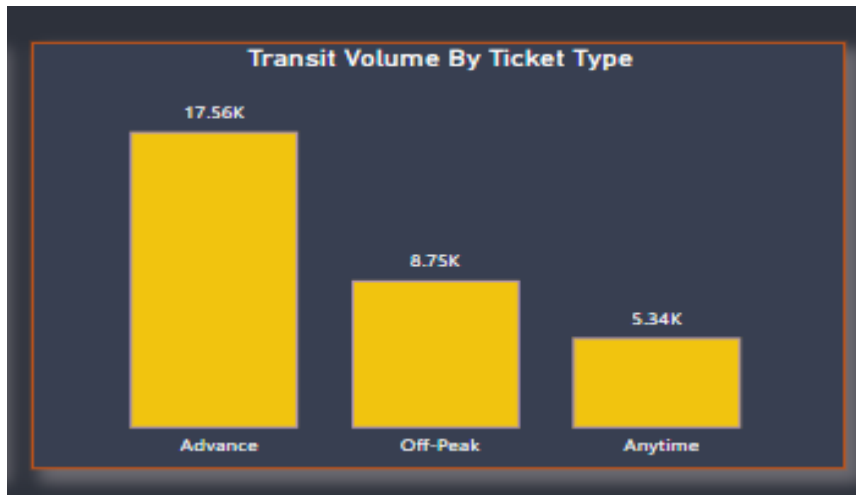
- Generally, a small percentage of the journeys were either delayed or canceled.
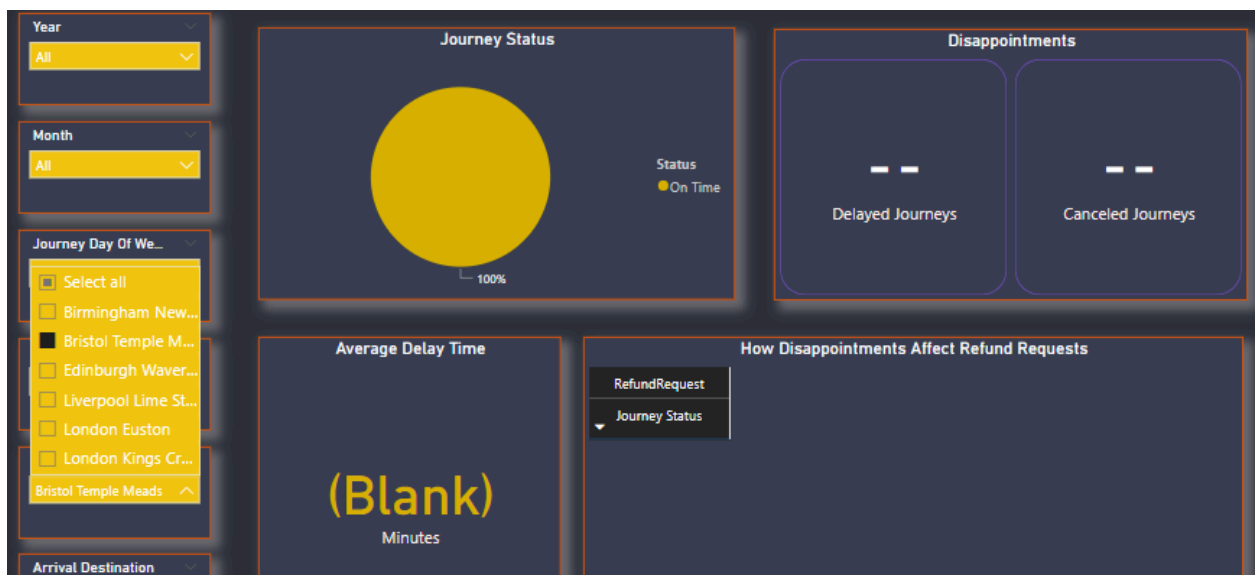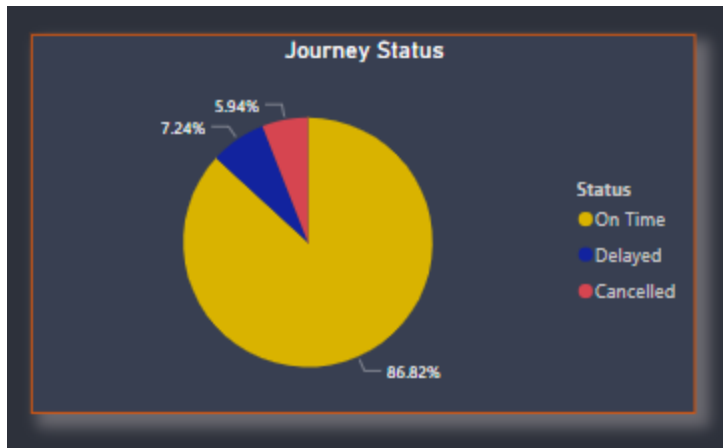
Bristol Temple Meads Station shows an extraordinary efficiency with 100% of journeys from there being on time. However, this station has only 16 entries, too small a sample to base long term decision making on.
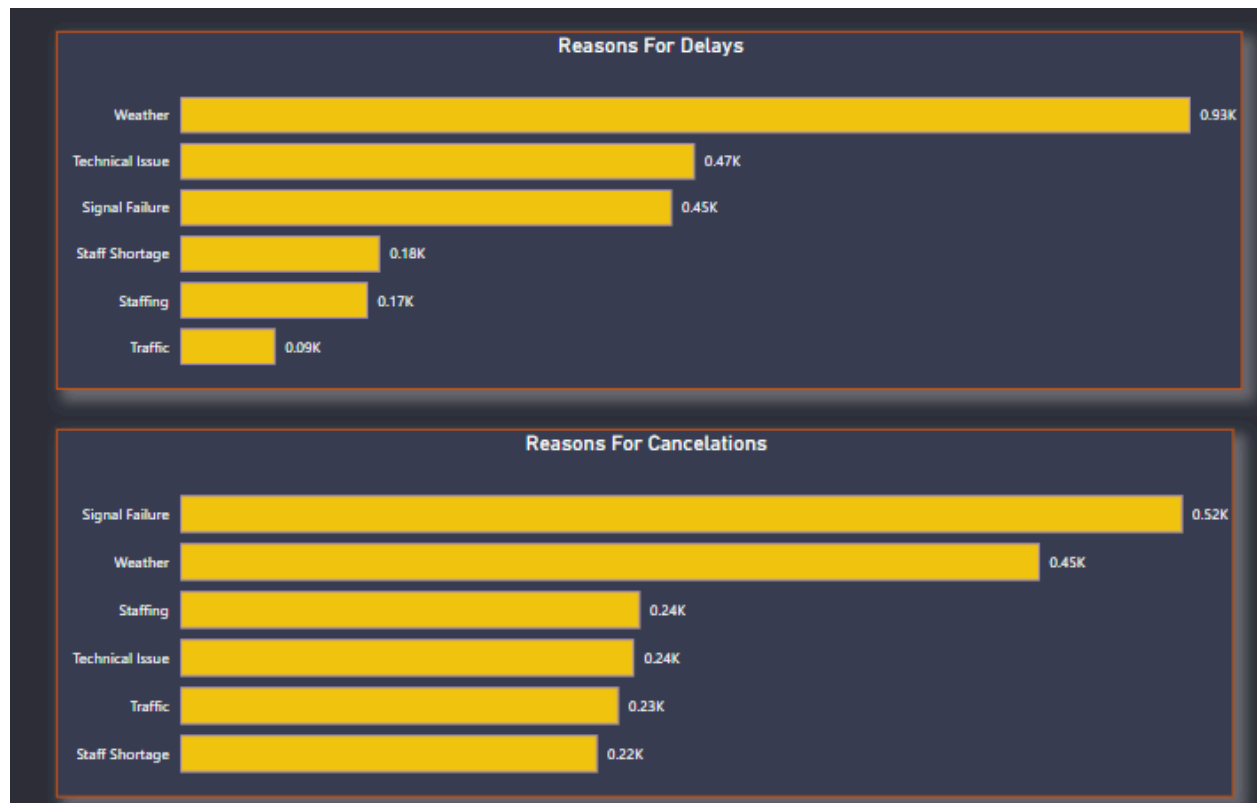
Whereas **Edinburgh Waverley** station always has delays averaging 15.27 minutes. This may be explained by the fact that this station very recently entered our data collection pool with only 51 entries so far.

- Delays at Edinburgh Waverley Station so far always result in requests for refunds which has cost us 2,093. Delays were sited to be caused by Staffing issues.
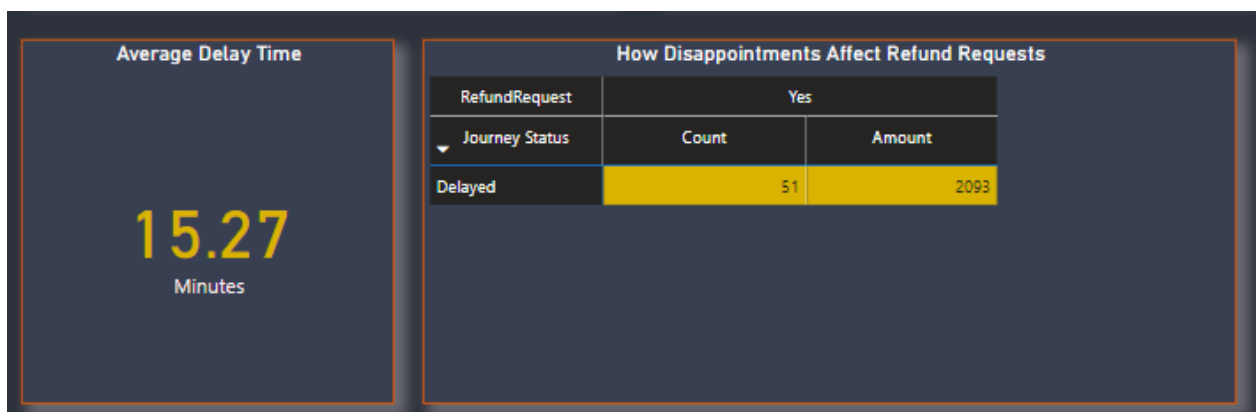
- Generally, Technical issues and Signal failure are 2 **operational** issues contributing vastly to delays and cancellations.

Reasons For Delays

Reason	Value
Weather	0.93K
Technical Issue	0.47K
Signal Failure	0.45K
Staff Shortage	0.18K
Staffing	0.17K
Traffic	0.09K

Reasons For Cancelations

Reason	Value
Signal Failure	0.52K
Weather	0.45K
Staffing	0.24K
Technical Issue	0.24K
Traffic	0.23K
Staff Shortage	0.22K

Conclusion

- We may find it profitable to put incentives in place for patrons to purchase non refundable tickets well ahead of time (not just a few hours).
- Though a small sample size, a closer look at how Bristol Temple Meads station is able to maintain its operational efficiency so far may prove beneficial to us.
- Edinburgh Waverley's staffing issues should be tackled promptly to correct the current state of consistent delays.
- A look at why there has been a decrease in online purchase in a more recent year may prove beneficial as this is contrary to what is expected as technology advances.