Sigma Labs

Railway Tracker

Problem Statement

What problem are you aiming to solve?

Most people hold the view that the railways in the country are the worst in the UK. We've all been stuck on a crowded, **delayed** train, or been stood on a platform in the cold looking at a display listing **cancelled** trains, but how severe is the actual problem?

Elevator Pitch

A short description of your project

This project would take advantage of public APIs from national rail operators to calculate performance metrics that would be compared, giving an insight into delay and cancellations in the UK.

Data Sources

What data sources will you use? How will you access them? How often do they update?

- The Realtime Trains API
- The National Rail APIs

Proposed Solution & Functionality

What features will your project have? How will you solve the problem? Write as much as you can here - the more detail the better.

For a selected set of stations in the south and north of England , fetch data for every train operating through that station in the last 24 hour period and store it for analysis. For each station, identify

- % of trains cancelled.
- % of trains delayed by 5 minutes.
- Average delay across all trains and average delay across all trains delayed > 1 minute.
- Incidents XML (Service Disruption)

Additionally, track rail disruption in realtime and provide SMS/email alerts for specific lines/stations.

Planned outputs

List the user-facing outputs of your project.

- Regular summary reports in PDF format (stored in S3, automatically emailed to subscribers).
- A dashboard with the latest statistics (metrics outlined above as well as highest overall delay, comparisons to the previous day, seven day rolling averages)
- Automatic alerts for delays/cancellations on particular train lines

Tools & Technology Stack

A list of the tools, languages and libraries you expect to use.

Code to be written in Python, likely requiring the requests, boto3 and a SQL library. RDS, ECS, emailing, archiving to be hosted by AWS.

Deadlines

Is this project possible to complete within two weeks for a team of 3-5 people?

- ☑ ~~Yes~~
- ☐ No

Stretch Goals

If you have more time, what other features or problems would you like to solve related to this project?

- Use the international APIs for international comparison (data output will be different, frequency of data refresh may be different and time for which data remains available).
- Look at other elements of journey quality - coach occupancy, working toilets, etc.
- Integrate another data source of population data. Determine whether the seats/hour/capita is higher or lower in different parts of the country.
- Estimated compensation tracker; given the number of carriages and the delay of a particular service, how much money might be refunded?