# Analysis Task

You have been asked to investigate whether any locations appear to be associated with poor train service performance. To help answer this question you have been given access to two data sources (see below). Using only this data prepare a short presentation (10-15 minutes) of your findings suitable for a non-technical audience. Also be prepared for some technical questions and to show any code you have written. Complete it in your own time, feel free to use the internet, but please write your own code where appropriate. You may work in any language and present your work however you think best. Credit will be given for impactful graphs, well-presented stats and good coding.

## Data Source 1: TRUST Move

"TRUST move" is a collection of messages from the TRUST (Train Running Under System TOPS) system. This feed delivers live data about trains to many other systems across the country. Seven files are provided in the sample, taken from seven consecutive dates (1st-7th June 2018 inclusive). Each message corresponds to a train arriving, departing or terminating at a location on the railway. Only trains which traverse a particular area of the Western Region are included in the sample.

The date in the filename refers to the date the messages relate to. The meaning of each column in the raw data in order is as follows:

* **Trust ID**: A unique ID for the train
* **TIPLOC**: TIming Point LOCation. A string up to 7 characters long that refers to a location on the railway, such as a station, junction or siding (e.g. PADTON for Paddington)
* **Event Time**: The time the event occurred, in YYYYMMDDHHMMSS format
* **Event Type**: The type of event that occurred: A = Arrival, D=Departure, T=Terminate (last stop of the train's journey)
* **Lateness**: Difference between the train's actual and planned arrival in minutes. Negative lateness means the train was early
* **Direction**: "U" (up) or "D" (down). "Up" direction means the train is travelling in the direction of a major location, usually London. "Down" direction means the opposite
* **Platform**: Platform used at the location, if available. Should be treated as a string
* **Planned Time**: The time the movement was scheduled to occur. May not always be populated
* **Message Time**: The time the message was recorded in TRUST, in HHMMSS format.

## Data Source 2: Locations

The file `locations.csv` contains information on each location on the network. The meaning of the columns is as follows:

* **Name**: The human-readable name of the location
* **TIPLOC**: Location TIPLOC - see above
* **Station**: True if the location is a station, otherwise false.
* **Longitude**: GPS Longitude of the location
* **Latitude**: GPS Latitude of the station