# Analysis of M1 Speed Data for Delivery Lorry Route Optimization

**1. Introduction**

In the modern era of e-commerce and rapid urbanization, businesses need to be competitive by reducing operational costs while providing good service quality (Chou, Hsia and Lan, 2017). One of ways to reduce costs is through efficient logistics and delivery schedules. One important aspect of optimizing delivery operations is the management of truck routes. The allocation of resources, reduction in consumption of fuel, minimization of delivery time, and reduction of environmental impacts are all key goals.

To achieve these goals, companies can use data driven approaches and technologies, which would provide efficient planning and punctual delivery (Zhou et al., 2021). An example of low-cost tracking technology would be GPS devices installed on delivery trucks. This would allow the company to obtain regarding route travelled, fuel consumed, time taken in real time (Stopher et al., n.d) from which secondary data such as traffic patterns, congestion and route efficiency can be extracted.

This report presents the statistical analysis of speed data on the M1 motorway, collected from Traffic England. The goal is to obtain an efficient sampling strategy to determine average speeds along the M1, which is important for optimizing delivery routes. This analysis would also allow companies to find alternative routes in real time based on historical data thereby ensuring that minimum delays in deliveries.

**2. Data Collection**

**2.1 Sampling Strategy**

Sampling strategy plays a crucial role in analysis of speed data for the truck route optimization. With the vast amount of data generated by delivery trucks travelling the M1 highway, strategic sampling techniques are essential to avoid wrong methodology which can affect the output (Lopez and Whitehead, 2013) and extract insights while managing computational resources effectively.

For the sampling strategy, the report has considered factors such as time, location, and changes in speed along the M1. Based on these factors, spatial, random and temporal sampling approaches have been taken. Each are explained below,

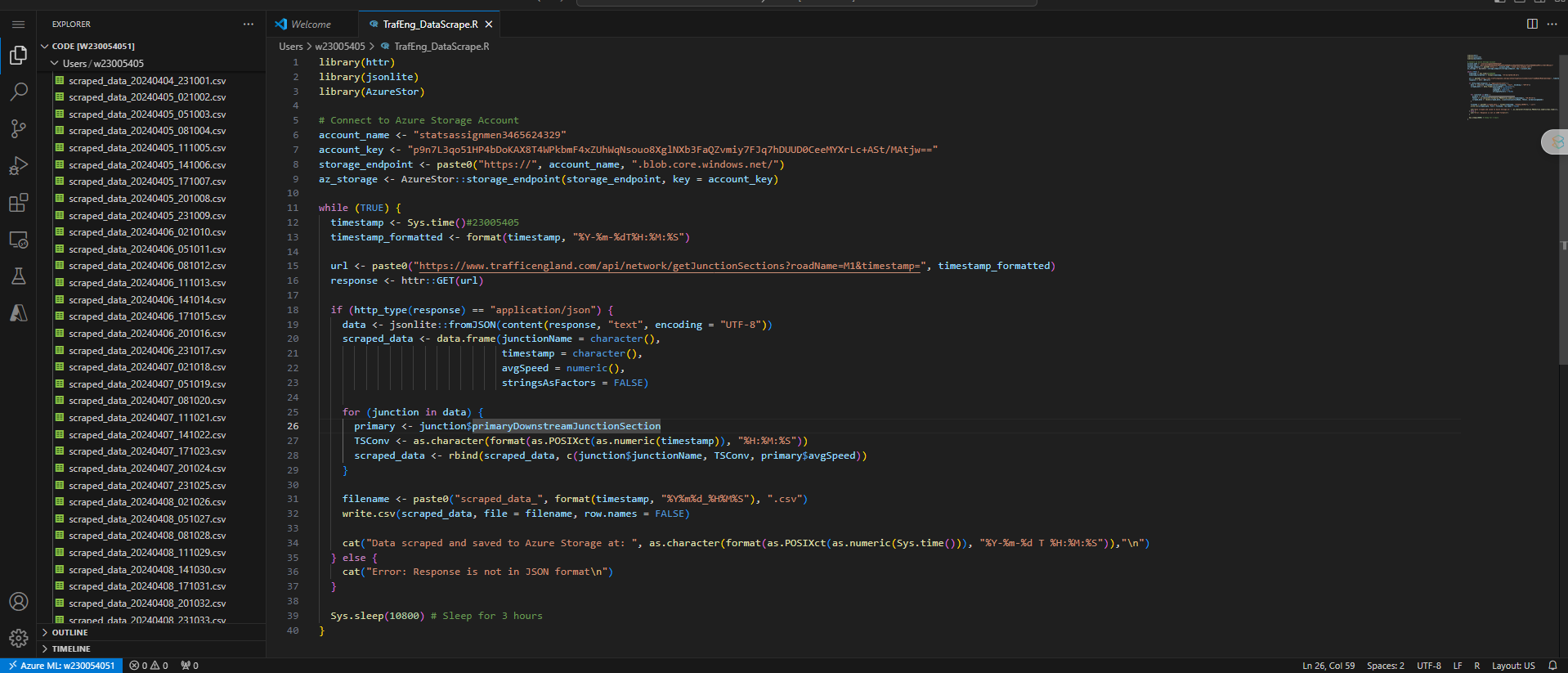
* **Spatial Sampling**: The study will divide the M1 highway into various geographical sections and the average speed data is captured across all sections to ensure comprehensive coverage.
* **Random Sampling**: Random data points are chosen from the dataset to get an unbiased representation of the traffic conditions throughout the day. This is help reduce sampling bias in turn improving accuracy and predictions.
* **Temporal Sampling**: The data points are taken over different time points across multiple weeks at three-hour intervals.

To improve the results and prediction of our algorithm, the above-mentioned sampling techniques can be merged with data updates in real time which allows the algorithm to get a better sense of the data. The dataset includes average speeds at various junction sections which will be used for analysis.

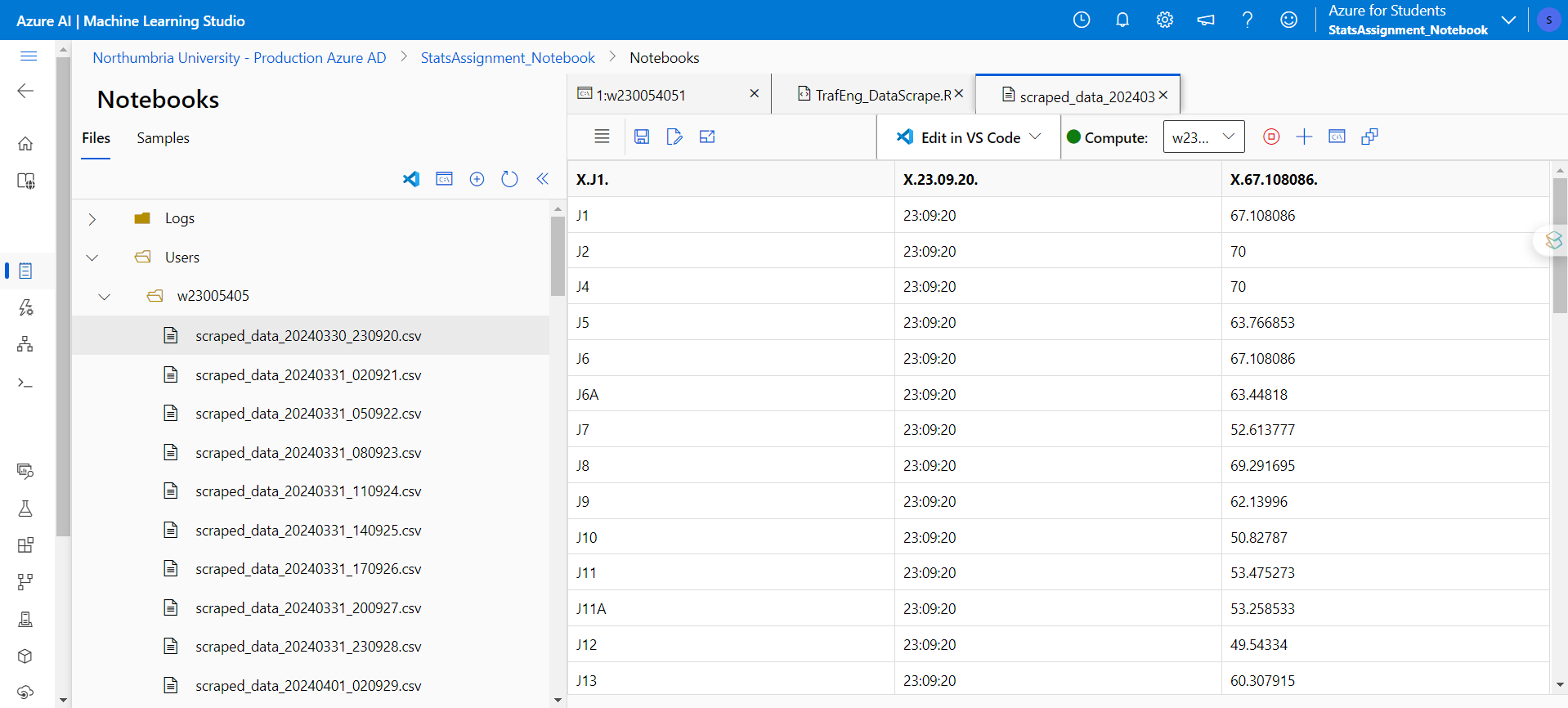
**2.2 Sending a request to collect data**

The dataset contains the average speed data for different junction sections of the M1 motorway. Each row corresponds to a specific junction and the average speed in miles per hour at a given time.

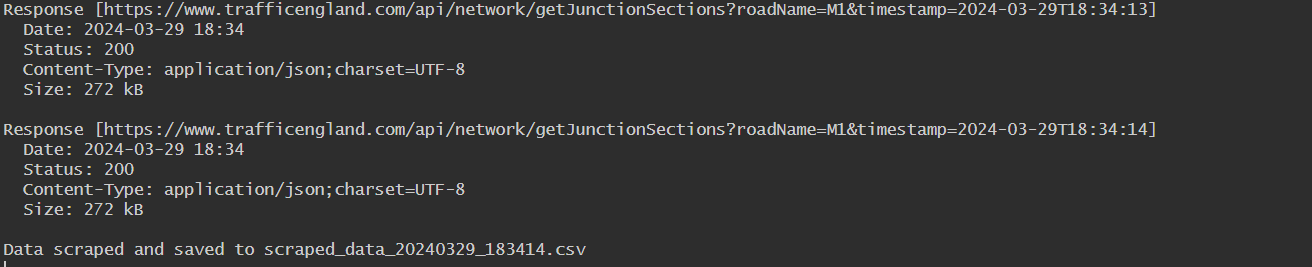
To collect this data, a data scraper program written in R language using the **httr** package sends a GET request to the Traffic England website at three-hour intervals and receives JSON data. The process is automate using an ML notebook on Azure Cloud which automatically stores the data in a cloud server using the **AzureStor** package. This data is processed in R using the **jsonlite** package and appended into a table.



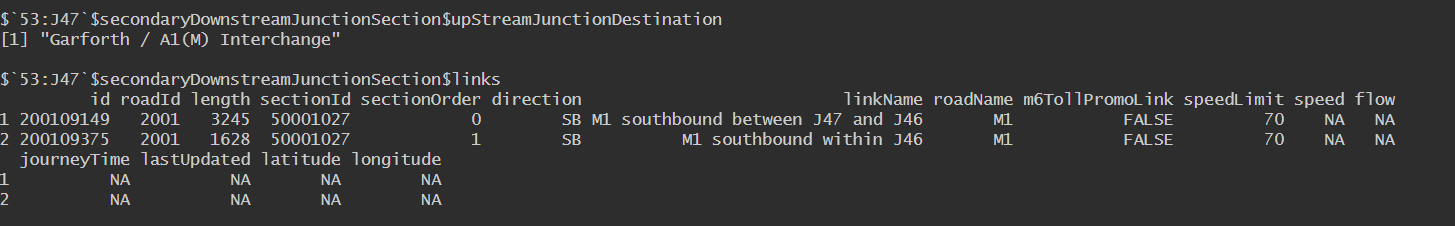
Snippet of the code used to send GET request to server



Sample of speed reading taken on 30-03-2024 at 23:09:20



Response from the server



The raw data received from the server.



Sample of processed response which is used as dataset for analysis.

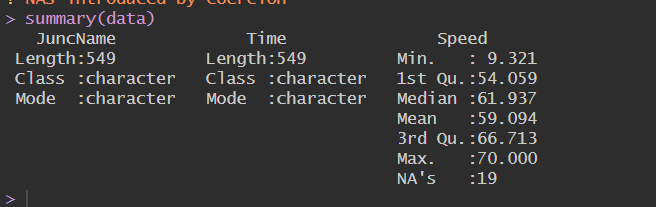
**3. Statistical Analysis**

**3.1 Descriptive Statistics**

Descriptive statistics include mean, median, and standard deviation for the average speeds in each direction (Northbound and Southbound) across different junction sections.

To calculate the statistical measure, we must first convert the average speed into a numeric value using,

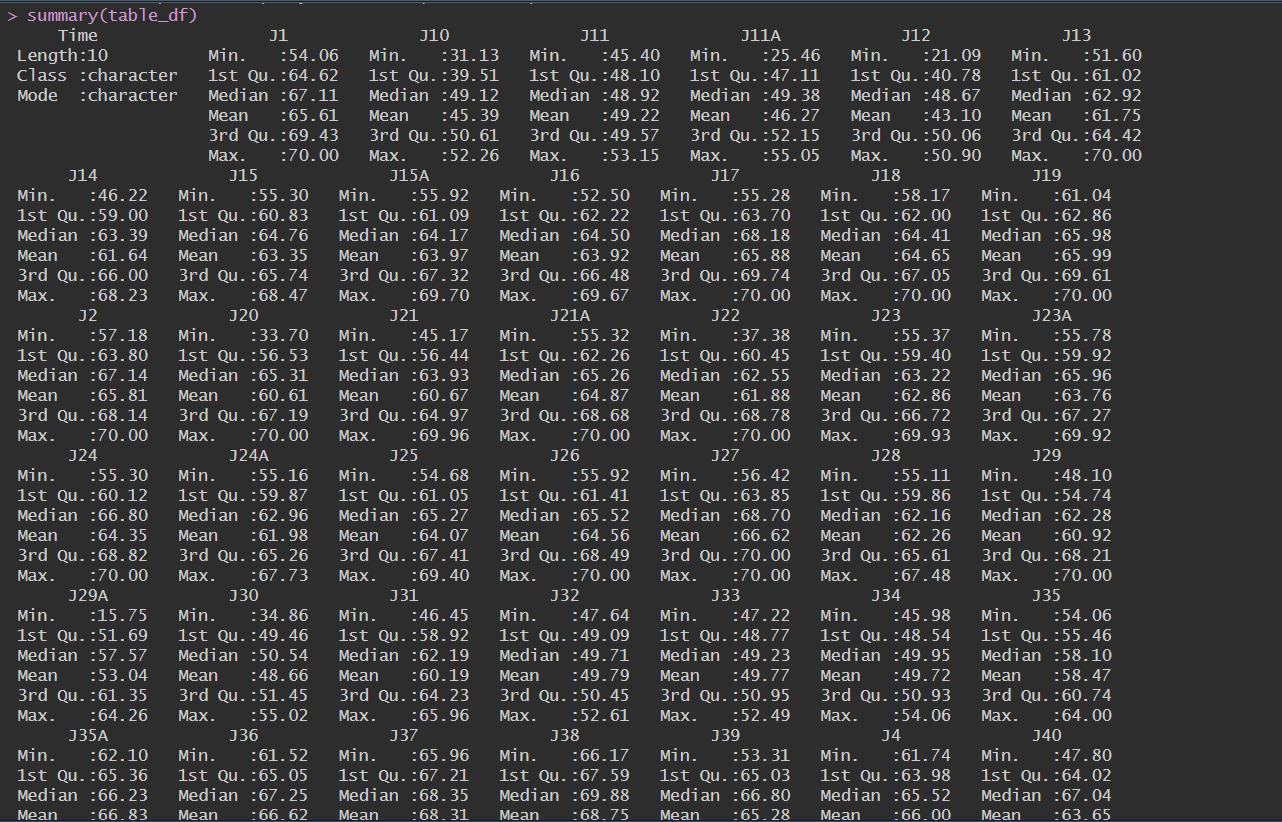
However, upon inspection of the data, we can find that there are 19 NA values present in the dataset which are added due to coercion from the as.numeric() function.

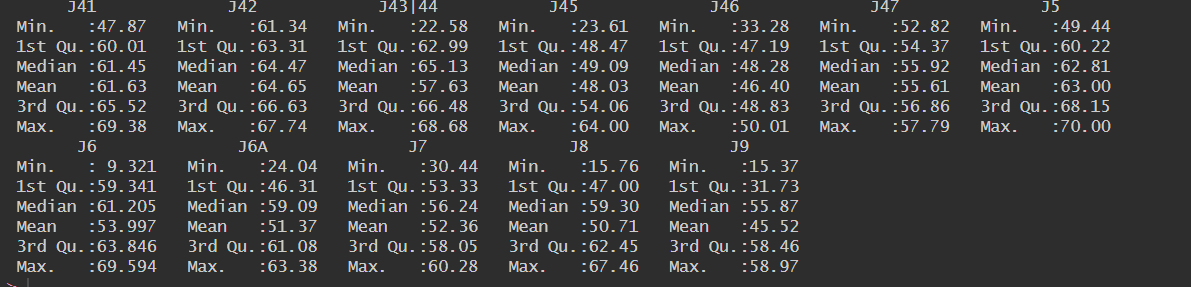
****

Summary of data

For further operations, the NA values will be removed using .

Now calculating the descriptive statistical measures using the function, we get the following values



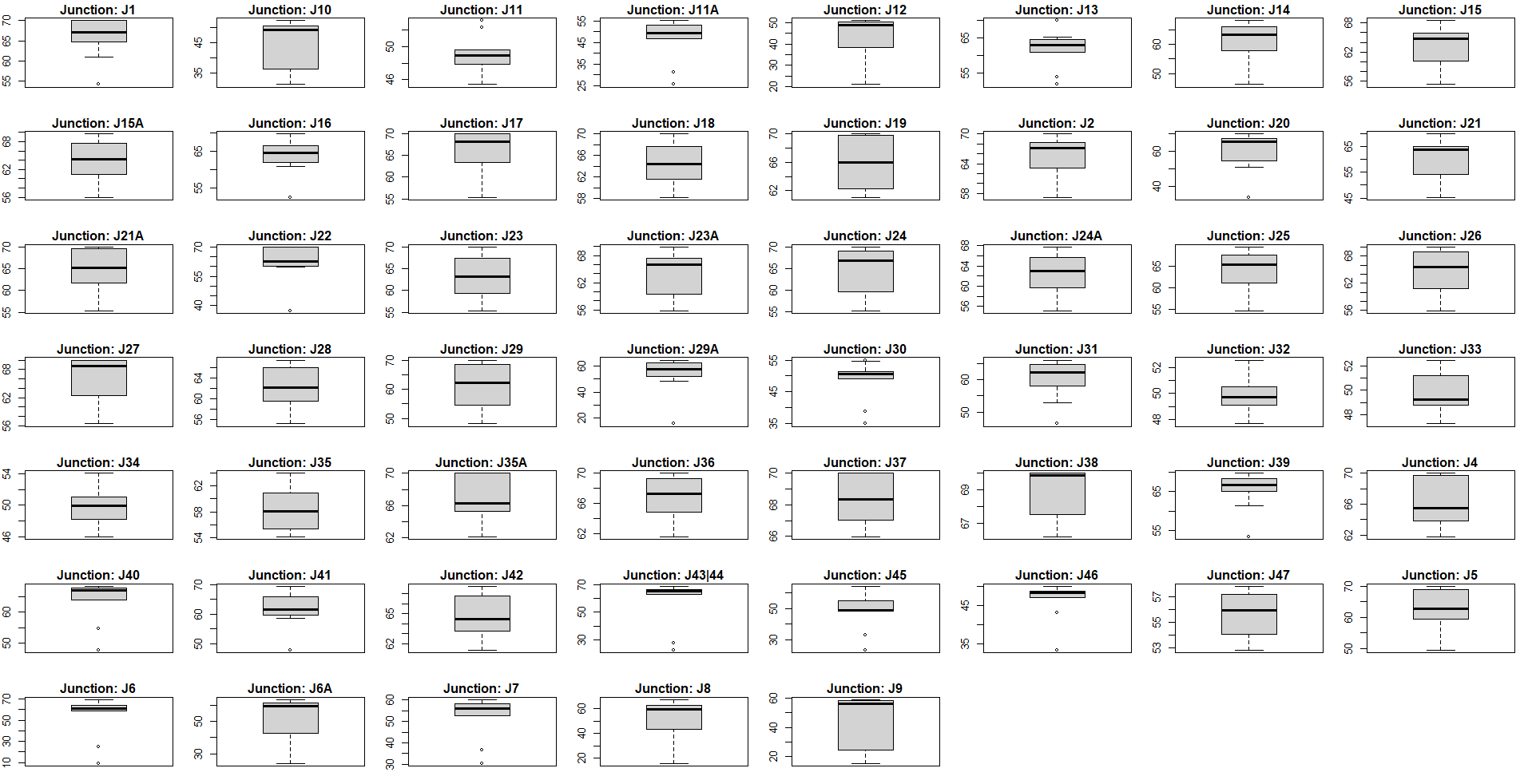


Summary of table

From the data, the following can be inferred,

1. **Central Tendency**: The values of mean and median are close to each other which indicates dataset is symmetrically distributed around central value.

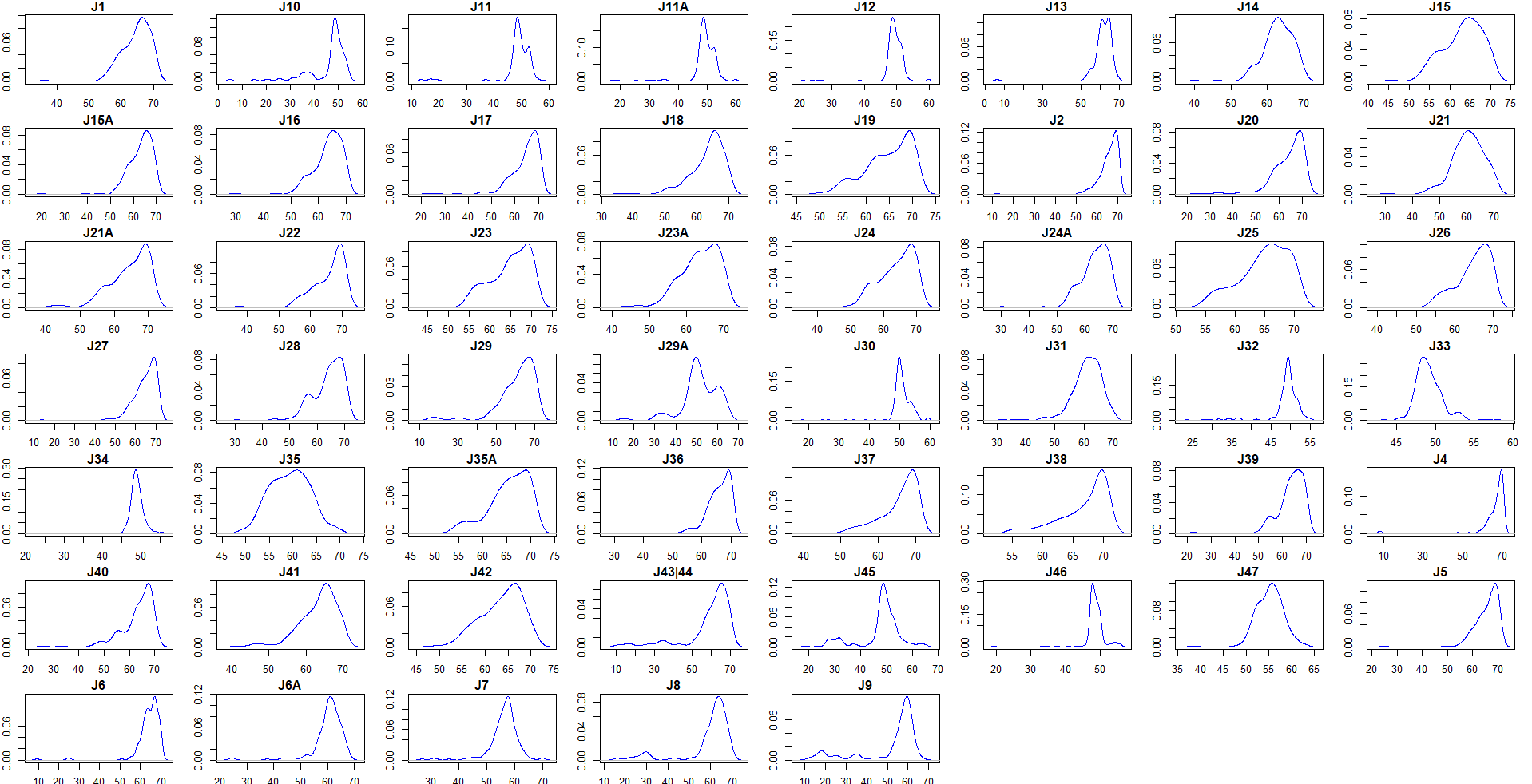
2. **Spread**: The data is moderately spread based on the range which indicates the minimum and maximum values (Inter Quartile Ranges) respectively.



Boxplots of data table

Additionally, we can see that there are outliers present which must be removed before further processing.

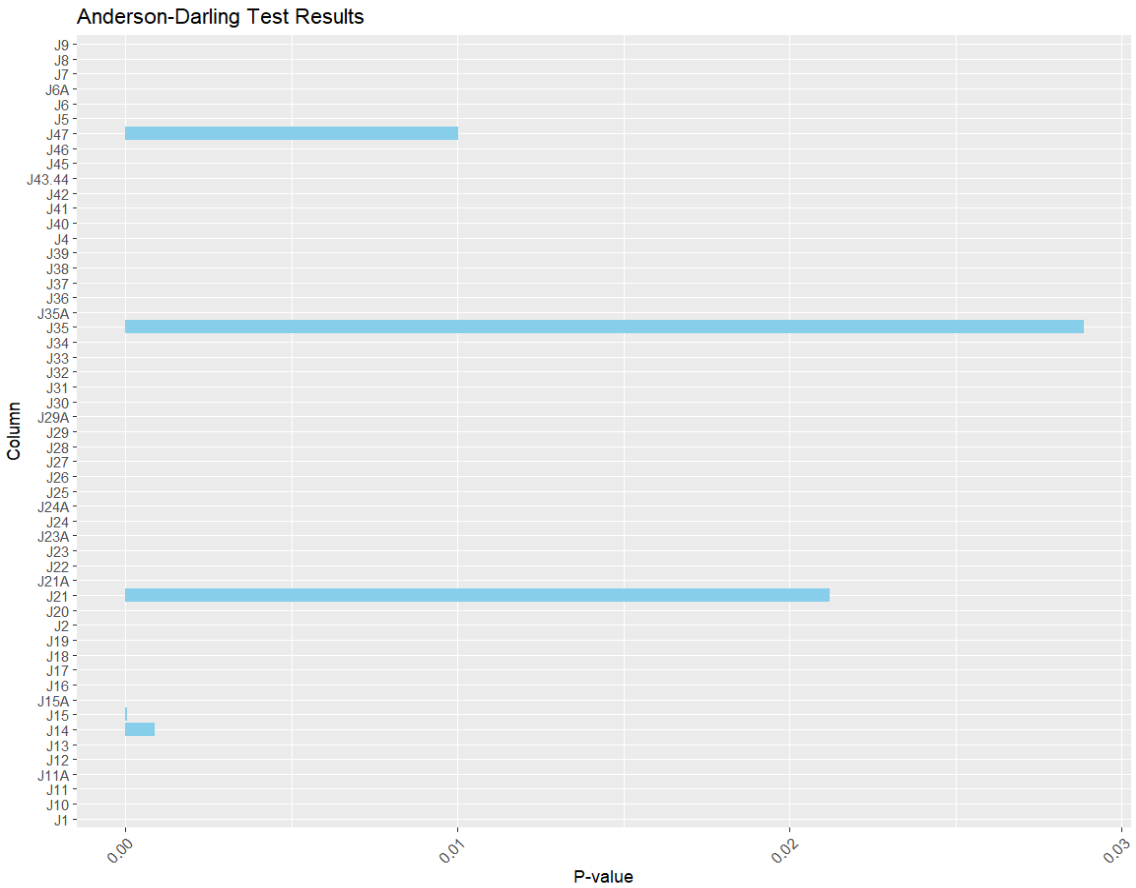
3. **Skewness**: The skewness comes out to -2.176 which indicates the dataset is negatively skewed. The same can be seen the density plot below,

****

Density plots for all junctions

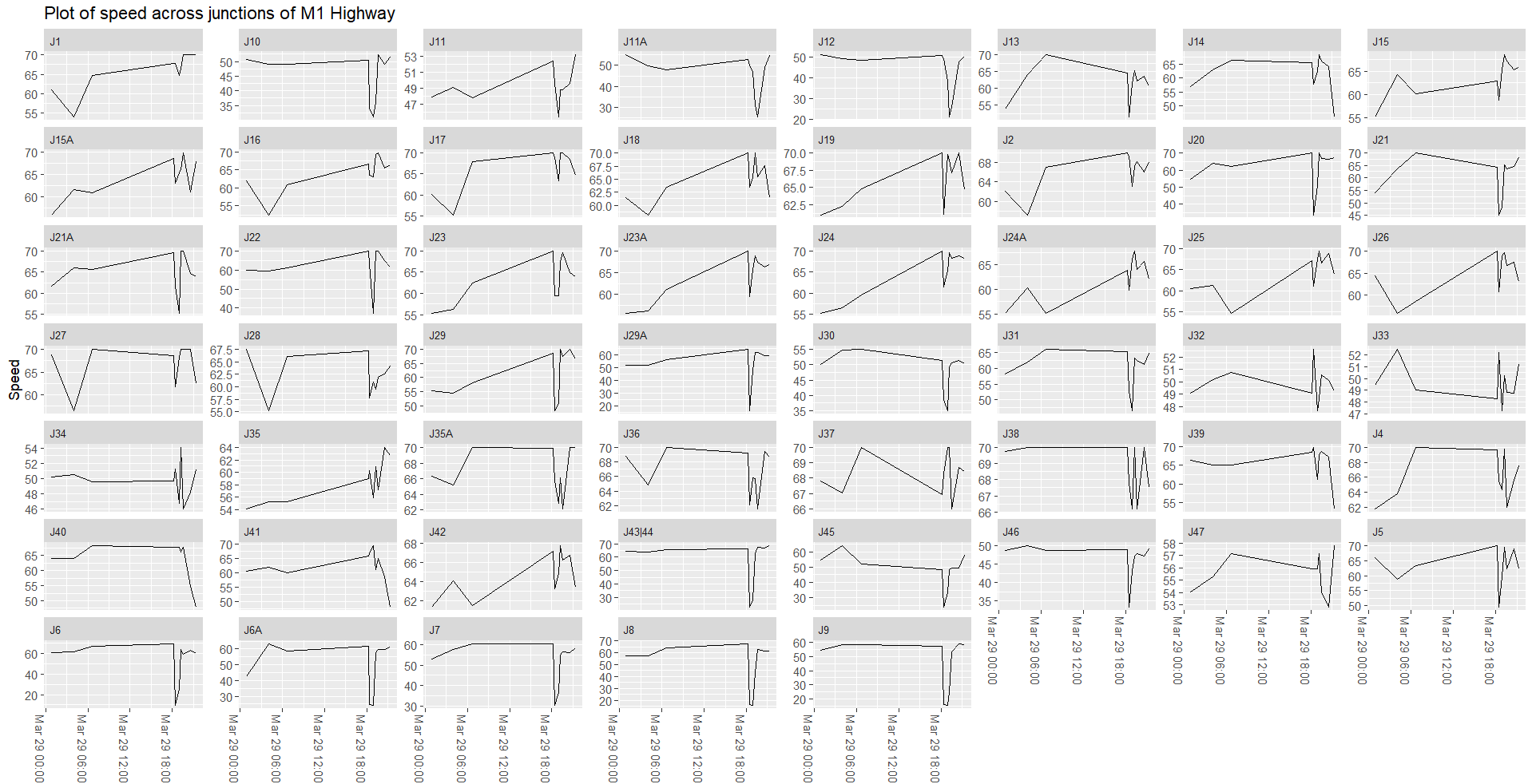
4**. Kurtosis:** The value is 13.23 which indicates a positive kurtosis which means the data has very sharp peaks when compared against a normal distribution. This is also called as leptokurtic distribution.

5. **Anderson-Darling Test:** Before starting any analysis, the dataset will be assumed to be normally distributed. However, this is not always the case and the distribution must be verified using the Anderson-Darling Test (Chanoknath Sutanapong and Louangrath, 2018).



Anderson-Darling test across different junction

From the test, we get the p-value as 0.011 which is well below the above the significance value of 0.05. Hence, the null hypothesis cannot be rejected as evidence is not available to state that the dataset is distributed normally. This is further augmented by skewness and the density plot graphs.

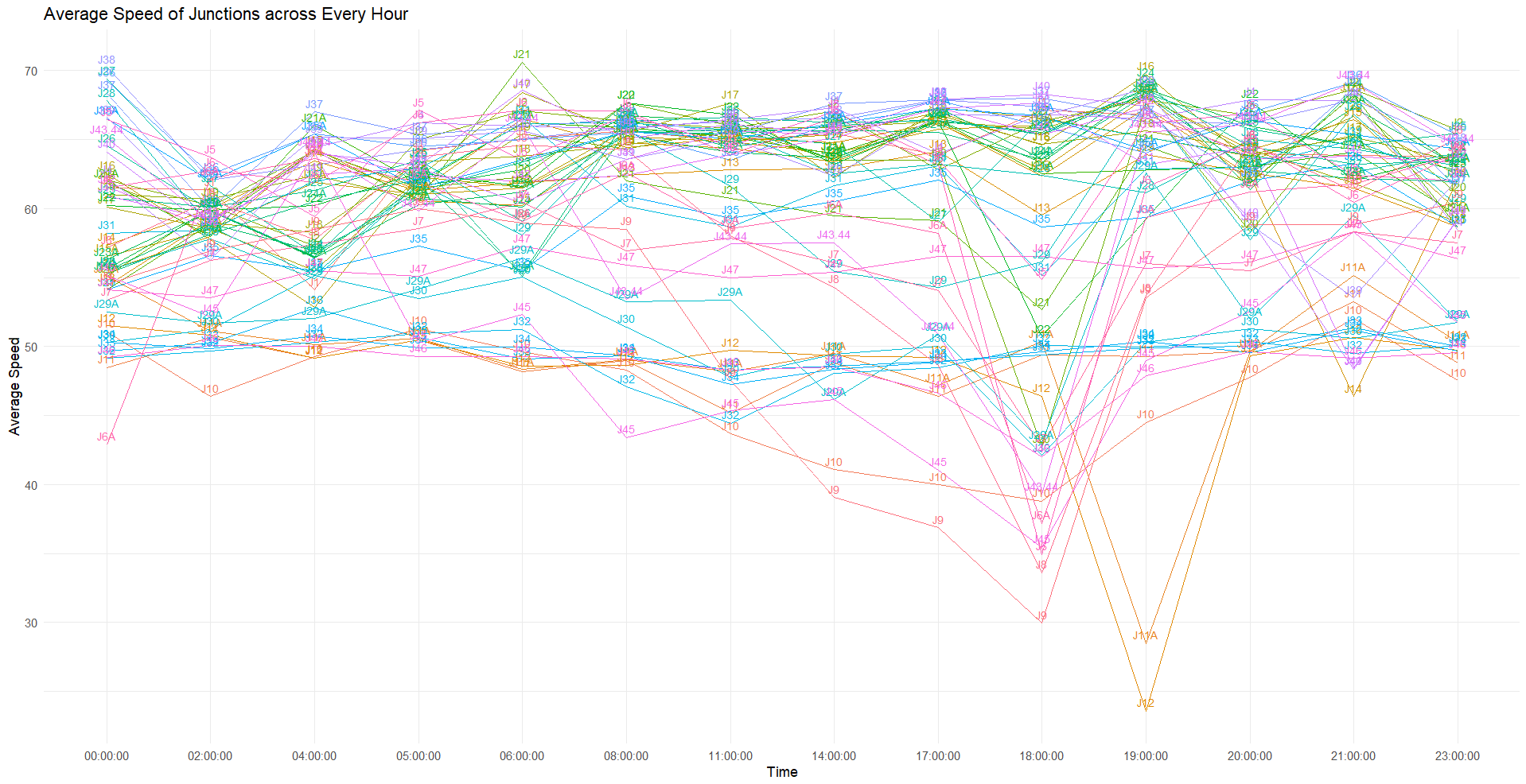
****

Line graph of average speed across both directions

**3.2 Visual Analysis**

**3.2.1. Temporal Analysis**

**3.2.1.1 Average Speed of junction through a day**

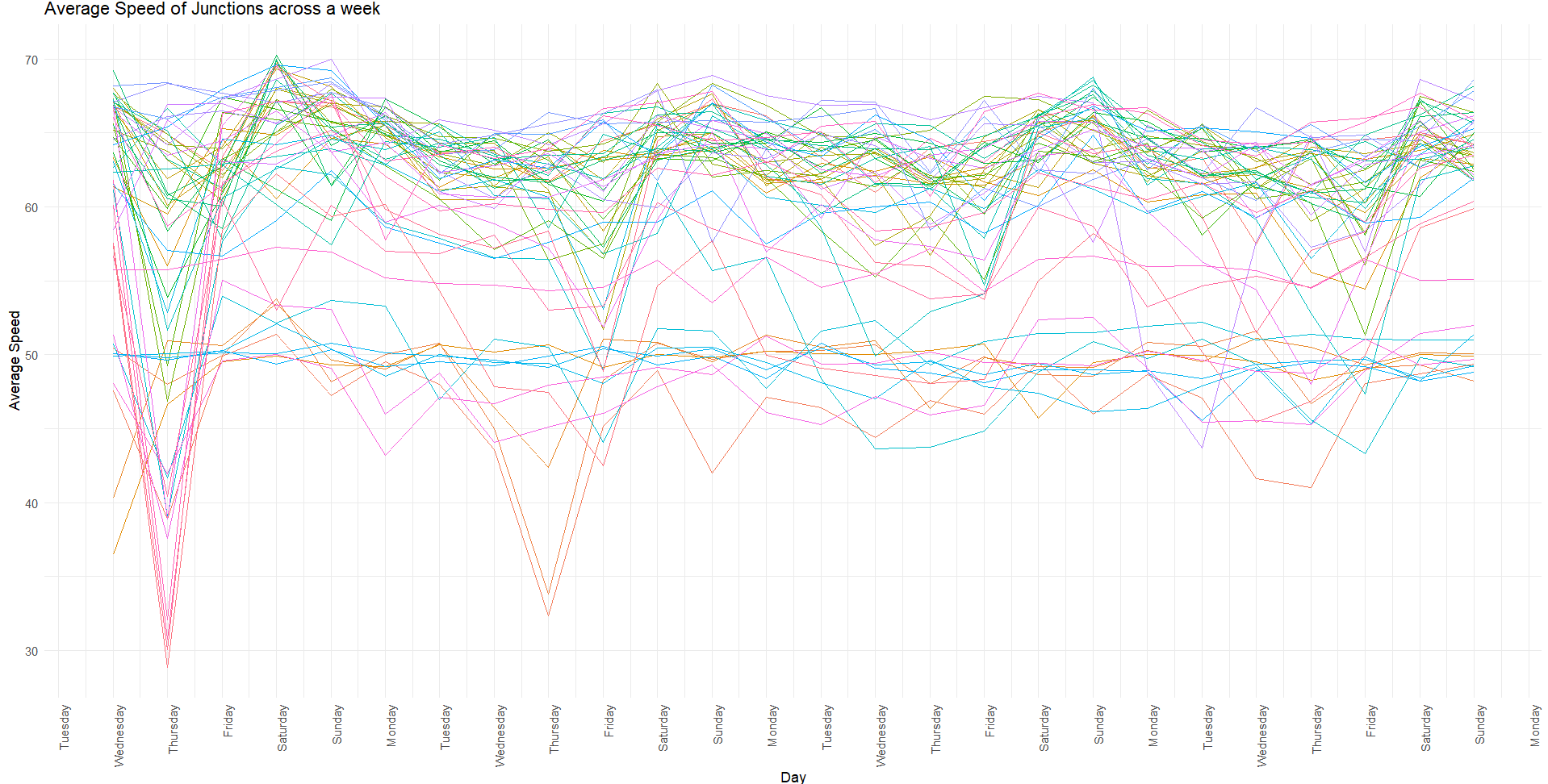
****

Average Speed of junction through a day

From the graph, we can clearly see a steady reduction in traffic speed across the entire day with the lowest speed during evenings around 18:00-20:00. This is due to rush hour on the M1 and heavy traffic on Friars Wash Interchange (J9), Chalton Interchange (J11A) and Toddington Interchange (J12).

This is not an isolated case as the graph tracks average speed across multiple weeks at the time intervals specified on the X-axis.

**3.2.1.2 Average Speed of junction across a week**

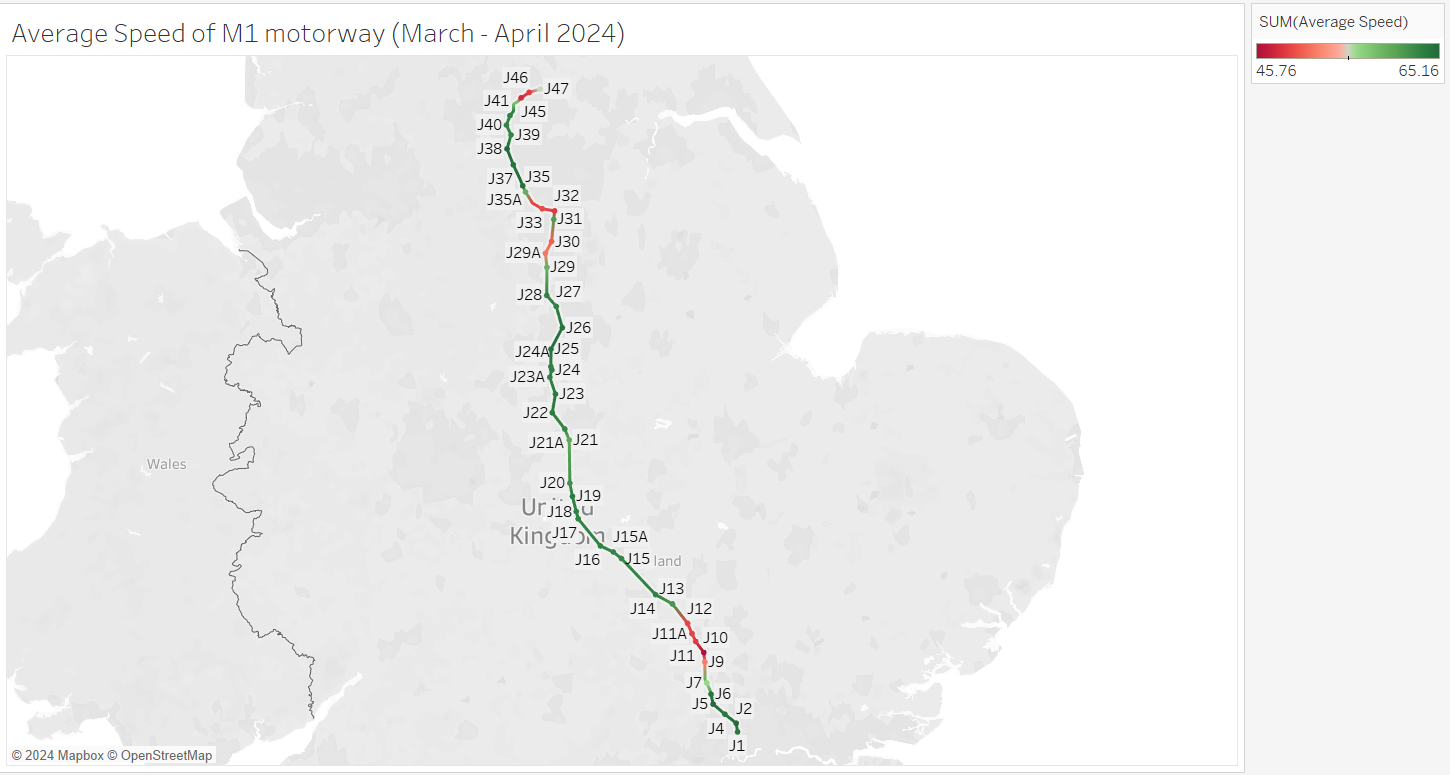
****

Average Speed of junction across a week

The M1 motorway has varying speeds across different time points in a week. Using this data, we can deduce that a significant drop in speed can be seen towards the weekends starting Thursday with some exceptions. This is possible due to increased influx of tourists and heavy long-haul trucks which need to deliver supplies over the weekend. Another reason is possible road works which significantly affect traffic movement.

However, the graph also shows this is not a prolonged problem as traffic speed usually resets towards its normal state around Saturday-Sunday.

**3.2.2. Geographical Analysis**

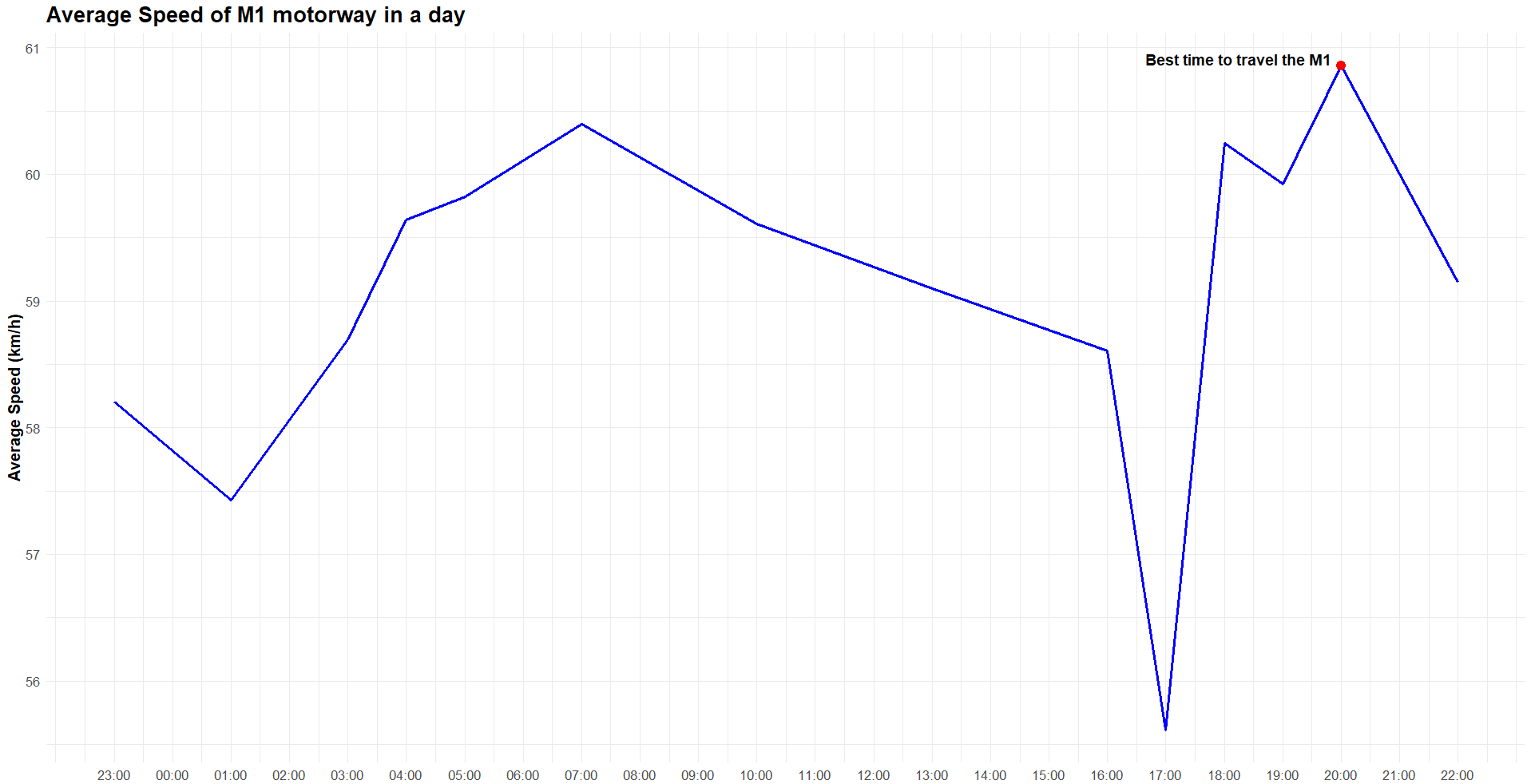


Speed map of the M1 motorway

The above image shows a map of the M1 highway different colours representing the average speed at every junction using data collected from March to April 2024. We can clearly infer those junctions are affected from traffic jams either due to ongoing road works or due to rush hour.

We can see a drop in average speed around junctions J33 and J11A which is consistent with the graphical visualization shown in the above section.

**3.2.3 Travel Analysis**



Average speeds along M1 throughout a day

Given that the company is a major manufacturer, we will assume they use Heavy Transport Vehicles greater than 7.5 tons to move goods around the country. Given that these vehicles cannot travel at high speeds, and any stall in traffic will result in delays and loss to company, calculations suggest the best time for travelling is 21:00 BST on any given day.

**4. Results and recommendations**

The analysis reveals significant changes in the average speeds across consecutive junctions of the M1 highway. Some sections of the M1 exhibit higher average speeds, while others have lower speeds, potentially due to factors such as congestion, road works, and time of day.

To reduce excess costs, the company must take the following recommendations into consideration:

1. Try to avoid high density junctions such as J11A and J33 if possible and use alternative routes.

2. If in case it is not possible to avoid these junctions, another suggestion would be to make deliveries during the weekdays and try to avoid Thursday to Saturday.

3. Delivery trucks must also not travel during rush hour that is from 16:00 to 20:00.

**References**

Chanoknath Sutanapong and Louangrath, P. (2018). Descriptive and Inferential Statistics. Zenodo (CERN European Organization for Nuclear Research). [online] doi:https://doi.org/10.5281/zenodo.1320727.

Chou, Y.-T., Hsia, S.-Y. and Lan, C.-H. (2017). A hybrid approach on multi-objective route planning and assignment optimization for urban lorry transportation. [online] doi:https://doi.org/10.1109/icasi.2017.7988625.

Donders, A.R.T., van der Heijden, G.J.M.G., Stijnen, T. and Moons, K.G.M. (2006). Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology, 59(10), pp.1087–1091. doi:https://doi.org/10.1016/j.jclinepi.2006.01.014.

‌Janus Christian Jakobsen, Gluud, C., Jørn Wetterslev and Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, [online] 17(1). doi:https://doi.org/10.1186/s12874-017-0442-1.

Lopez, V. and Whitehead, D. (2013). Sampling data and data collection in qualitative research. [online] ResearchGate. Available at: https://www.researchgate.net/publication/255950308\_Sampling\_data\_and\_data\_collection\_in\_qualitative\_research [Accessed 16 Mar. 2024].

Stopher, P., Greaves, S. and Fitzgerald, C. (n.d.). Developing and Deploying a New Wearable GPS Device for Transport Applications Developing and Deploying a New Wearable GPS Device for Transport Applications. [online] Available at: https://australasiantransportresearchforum.org.au/wp-content/uploads/2022/03/2005\_Stopher\_Greaves\_Fitzgerald.pdf [Accessed 16 Mar. 2024].

Trafficengland.com. (2024). Welcome to National Highways. [online] Available at: https://www.trafficengland.com/traffic-report [Accessed 29 Mar. 2024].

Zhou, Z., Zhou, X., Lu, Y., Yan, H., Guo, B. and Wang, S. (2021). Multi-Source Data-Driven Route Prediction for Instant Delivery. 2021 17th International Conference on Mobility, Sensing and Networking (MSN). [online] doi:https://doi.org/10.1109/msn53354.2021.00064.