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.

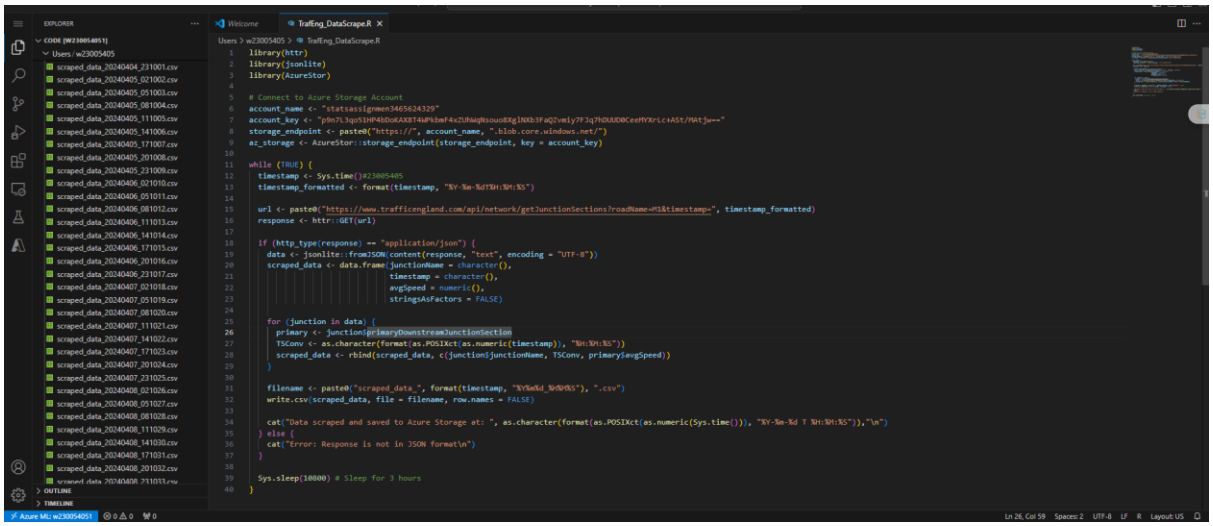
- **Temporal Sampling**: The data points are taken over different time points across multiple weeks at three-hour intervals.

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

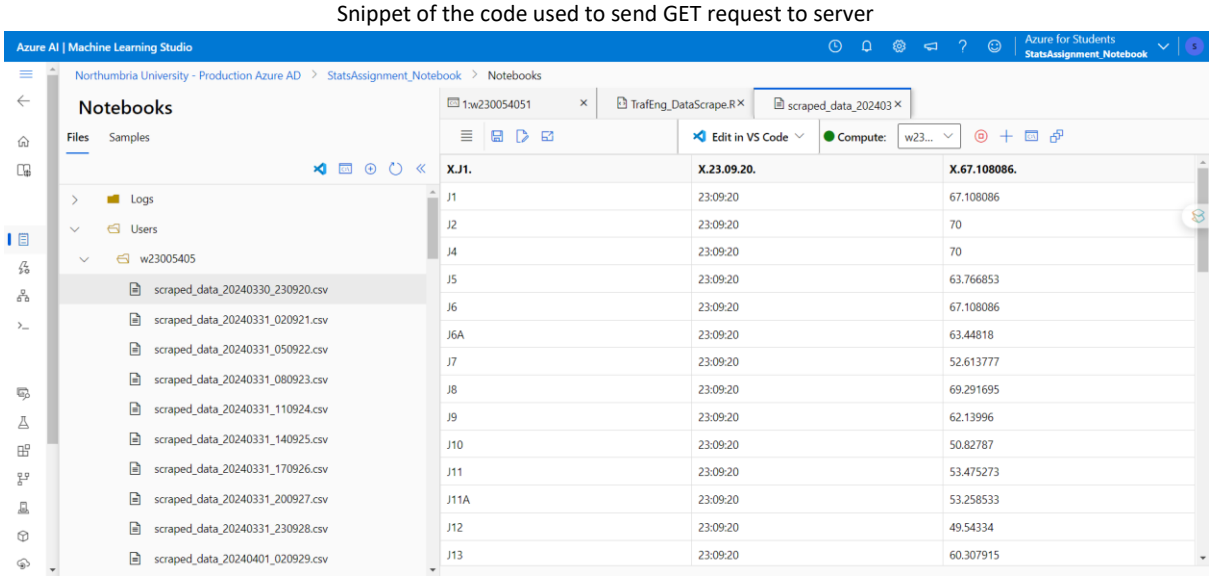
The dataset contains the average speed data for different junction sections of the M1 motorway. Each row corresponds to a specific junction and the average speed in miles per hour at a given time.

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.



```r
library(httr)
library(jsonlite)
library(AzureStor)

# Connect to Azure Storage Account
account_name <- "statsassignmen3465624329"
account_key <- "[redacted]"
storage_endpoint <- paste0("https://", account_name, ".blob.core.windows.net/")
az_storage <- AzureStor::storage_endpoint(storage_endpoint, key = account_key)

while (TRUE) {
  timestamp <- Sys.time()#23005405
  timestamp_formatted <- format(timestamp, "%Y-%m-%dT%H:%M:%S")

  url <- paste0("https://www.trafficengland.com/api/network/getJunctionSections?roadName=M1&timestamp=", timestamp_formatted)
  response <- httr::GET(url)

  if (http_type(response) == "application/json") {
    data <- jsonlite::fromJSON(content(response, "text", encoding = "UTF-8"))
    scraped_data <- data.frame(junctionName = character(),
                               timestamp = character(),
                               avgSpeed = numeric(),
                               stringsAsFactors = FALSE)

    for (junction in data) {
      primary <- junction$primaryDownstreamJunctionSection
      TSConv <- as.character(format(as.POSIXct(as.numeric(timestamp)), "%H:%M:%S"))
      scraped_data <- rbind(scraped_data, c(junction$junctionName, TSConv, primary$avgSpeed))
    }

    filename <- paste0("scraped_data_", format(timestamp, "%Y%m%d_%H%M%S"), ".csv")
    write.csv(scraped_data, file = filename, row.names = FALSE)

    cat("Data scraped and saved to Azure Storage at: ", as.character(format(as.POSIXct(as.numeric(Sys.time())), "%Y-%m-%d T %H:%M:%S")),"\n")
  } else {
    cat("Error: Response is not in JSON format\n")
  }

  Sys.sleep(10800) # Sleep for 3 hours
}
```

Snippet of the code used to send GET request to server



Azure AI | Machine Learning Studio

Northumbria University - Production Azure AD > StatsAssignment_Notebook > Notebooks

Notebooks

Files Samples

Logs
Users
w23005405
scraped_data_20240330_230920.csv
scraped_data_20240331_020921.csv
scraped_data_20240331_050922.csv
scraped_data_20240331_080923.csv
scraped_data_20240331_110924.csv
scraped_data_20240331_140925.csv
scraped_data_20240331_170926.csv
scraped_data_20240331_200927.csv
scraped_data_20240331_230928.csv
scraped_data_20240401_020929.csv

1:w230054051 | TrafEng_DataScrape.R | scraped_data_202403

Edit in VS Code | Compute: w23...

X.J1.	X.23.09.20.	X.67.108086.
J1	23:09:20	67.108086
J2	23:09:20	70
J4	23:09:20	70
J5	23:09:20	63.766853
J6	23:09:20	67.108086
J6A	23:09:20	63.44818
J7	23:09:20	52.613777
J8	23:09:20	69.291695
J9	23:09:20	62.13996
J10	23:09:20	50.82787
J11	23:09:20	53.475273
J11A	23:09:20	53.258533
J12	23:09:20	49.54334
J13	23:09:20	60.307915

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

```
Response [https://www.trafficengland.com/api/network/getJunctionSections?roadName=M1&timestamp=2024-03-29T18:34:13]
  Date: 2024-03-29 18:34
  Status: 200
  Content-Type: application/json;charset=UTF-8
  Size: 272 kB

Response [https://www.trafficengland.com/api/network/getJunctionSections?roadName=M1&timestamp=2024-03-29T18:34:14]
  Date: 2024-03-29 18:34
  Status: 200
  Content-Type: application/json;charset=UTF-8
  Size: 272 kB

Data scraped and saved to scraped_data_20240329_183414.csv
```

Response from the server

```
$`53:J47`$secondaryDownstreamJunctionSection$upStreamJunctionDestination
[1] "Garforth / A1(M) Interchange"

$`53:J47`$secondaryDownstreamJunctionSection$links
        id roadId length sectionId sectionOrder direction                                linkName roadName m6TollPromoLink speedLimit speed flow
1 200109149   2001   3245  50001027            0        SB M1 southbound between J47 and J46       M1           FALSE         70    NA   NA
2 200109375   2001   1628  50001027            1        SB          M1 southbound within J46       M1           FALSE         70    NA   NA
  journeyTime lastUpdated latitude longitude
1          NA          NA       NA        NA
2          NA          NA       NA        NA
```

The raw data received from the server.

	Time	J1	J10	J11	J11A	J12	J13	J14	J15	J15A	J16	J17	J18	J19	J2	J20	J21	J21A	J22
1	00:53:05	60.89437	50.64101	47.88293	55.04720	50.89556	53.85915	56.81609	55.30203	55.92341	61.93505	60.05813	61.50691	61.03927	62.03351	54.64643	54.05929	61.59046	60.13338
2	04:02:44	54.05929	49.15061	49.08832	49.66434	49.09259	64.13156	62.87649	64.23143	61.53472	52.49789	55.28039	58.17221	62.24690	57.17648	64.10394	63.67977	66.01476	59.71452
3	06:40:15	64.62260	48.96376	47.71825	47.80022	48.35747	70.00000	66.29621	60.13460	60.89437	60.92320	67.85774	63.46841	64.84217	67.05112	62.16618	69.96083	65.73963	61.38305
4	18:10:39	67.72946	50.51792	52.35986	52.81655	50.44482	64.52254	65.59020	62.91948	68.50822	66.50799	70.00000	70.00000	70.00000	70.00000	70.00000	64.18800	69.56987	70.00000
5	18:24:03	67.72946	34.17986	49.54502	50.13847	48.97457	51.60406	57.85424	58.75418	62.98100	63.50256	68.95055	63.50137	61.11395	69.35889	33.70133	45.16875	61.66517	63.44967
6	18:56:10	64.62260	31.13380	45.39745	46.88165	38.47668	61.41567	62.43404	65.35907	65.35523	63.06121	63.33656	65.31725	69.81314	63.04488	50.79162	48.21962	55.32215	37.37979
7	19:12:16	66.48672	36.35244	48.75896	31.10479	21.09109	65.27422	68.23135	68.46593	66.08785	69.23161	70.00000	70.00000	69.00954	67.23392	70.00000	65.07519	70.00000	70.00000
8	19:32:49	70.00000	52.25746	48.73689	25.46001	25.60705	62.20228	66.13165	67.22318	69.70482	69.67409	70.00000	65.38107	67.10809	68.19026	66.93692	63.57834	70.00000	70.00000
9	20:33:00	70.00000	49.08833	49.58237	49.08833	47.67922	63.64051	63.90641	65.28987	60.94051	65.50033	68.50076	67.60328	70.00000	66.05698	66.52095	64.66501	64.77202	65.10435
10	21:18:29	70.00000	51.63609	53.14591	54.69674	50.38794	60.89438	46.22176	65.86535	67.72947	66.41039	64.77253	61.51158	64.69727	67.97778	67.27926	68.13913	64.04864	61.65537

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,

$$data \ <- data \ \% > \% \ mutate\_at(vars(-JuncName, -Time), \sim as.numeric(as.character(.)))$$

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(data)
   JuncName              Time                Speed
 Length:549         Length:549         Min.    : 9.321
 Class :character   Class :character   1st Qu.:54.059
 Mode  :character   Mode  :character   Median :61.937
                                       Mean   :59.094
                                       3rd Qu.:66.713
                                       Max.   :70.000
                                       NA's   :19
>
```

Summary of data

For further operations, the NA values will be removed using $data = na.omit(data)$.

Now calculating the descriptive statistical measures using the $summary()$ function, we get the following values

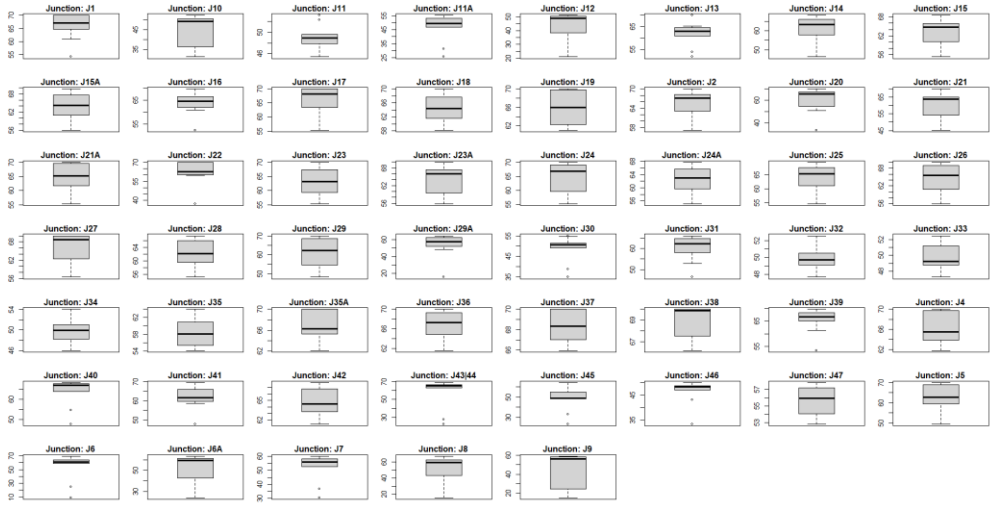
```
> summary(table_df)
     Time                J1              J10             J11             J11A            J12             J13
 Length:10          Min.   :54.06   Min.   :31.13   Min.   :45.40   Min.   :25.46   Min.   :21.09   Min.   :51.60
 Class :character   1st Qu.:64.62   1st Qu.:39.51   1st Qu.:48.10   1st Qu.:47.11   1st Qu.:40.78   1st Qu.:61.02
 Mode  :character   Median :67.11   Median :49.12   Median :48.92   Median :49.38   Median :48.67   Median :62.92
                    Mean   :65.61   Mean   :45.39   Mean   :49.22   Mean   :46.27   Mean   :43.10   Mean   :61.75
                    3rd Qu.:69.43   3rd Qu.:50.61   3rd Qu.:49.57   3rd Qu.:52.15   3rd Qu.:50.06   3rd Qu.:64.42
                    Max.   :70.00   Max.   :52.26   Max.   :53.15   Max.   :55.05   Max.   :50.90   Max.   :70.00
      J14             J15             J15A            J16             J17             J18             J19
 Min.   :46.22   Min.   :55.30   Min.   :55.92   Min.   :52.50   Min.   :55.28   Min.   :58.17   Min.   :61.04
 1st Qu.:59.00   1st Qu.:60.83   1st Qu.:61.09   1st Qu.:62.22   1st Qu.:63.70   1st Qu.:62.00   1st Qu.:62.86
 Median :63.39   Median :64.76   Median :64.17   Median :64.50   Median :68.18   Median :64.41   Median :65.98
 Mean   :61.64   Mean   :63.35   Mean   :63.97   Mean   :63.92   Mean   :65.88   Mean   :64.65   Mean   :65.99
 3rd Qu.:66.00   3rd Qu.:65.74   3rd Qu.:67.32   3rd Qu.:66.48   3rd Qu.:69.74   3rd Qu.:67.05   3rd Qu.:69.61
 Max.   :68.23   Max.   :68.47   Max.   :69.70   Max.   :69.67   Max.   :70.00   Max.   :70.00   Max.   :70.00
      J2              J20             J21             J21A            J22             J23             J23A
 Min.   :57.18   Min.   :33.70   Min.   :45.17   Min.   :55.32   Min.   :37.38   Min.   :55.37   Min.   :55.78
 1st Qu.:63.80   1st Qu.:56.53   1st Qu.:56.44   1st Qu.:62.26   1st Qu.:60.45   1st Qu.:59.40   1st Qu.:59.92
 Median :67.14   Median :65.31   Median :63.93   Median :65.26   Median :62.55   Median :63.22   Median :65.96
 Mean   :65.81   Mean   :60.61   Mean   :60.67   Mean   :64.87   Mean   :61.88   Mean   :62.86   Mean   :63.76
 3rd Qu.:68.14   3rd Qu.:67.19   3rd Qu.:64.97   3rd Qu.:68.68   3rd Qu.:68.78   3rd Qu.:66.72   3rd Qu.:67.27
 Max.   :70.00   Max.   :70.00   Max.   :69.96   Max.   :70.00   Max.   :70.00   Max.   :69.93   Max.   :69.92
      J24             J24A            J25             J26             J27             J28             J29
 Min.   :55.30   Min.   :55.16   Min.   :54.68   Min.   :55.92   Min.   :56.42   Min.   :55.11   Min.   :48.10
 1st Qu.:60.12   1st Qu.:59.87   1st Qu.:61.05   1st Qu.:61.41   1st Qu.:63.85   1st Qu.:59.86   1st Qu.:54.74
 Median :66.80   Median :62.96   Median :65.27   Median :65.52   Median :68.70   Median :62.16   Median :62.28
 Mean   :64.35   Mean   :61.98   Mean   :64.07   Mean   :64.56   Mean   :66.62   Mean   :62.26   Mean   :60.92
 3rd Qu.:68.82   3rd Qu.:65.26   3rd Qu.:67.41   3rd Qu.:68.49   3rd Qu.:70.00   3rd Qu.:65.61   3rd Qu.:68.21
 Max.   :70.00   Max.   :67.73   Max.   :69.40   Max.   :70.00   Max.   :70.00   Max.   :67.48   Max.   :70.00
      J29A            J30             J31             J32             J33             J34             J35
 Min.   :15.75   Min.   :34.86   Min.   :46.45   Min.   :47.64   Min.   :47.22   Min.   :45.98   Min.   :54.06
 1st Qu.:51.69   1st Qu.:49.46   1st Qu.:58.92   1st Qu.:49.09   1st Qu.:48.77   1st Qu.:48.54   1st Qu.:55.46
 Median :57.57   Median :50.54   Median :62.19   Median :49.71   Median :49.23   Median :49.95   Median :58.10
 Mean   :53.04   Mean   :48.66   Mean   :60.19   Mean   :49.79   Mean   :49.77   Mean   :49.72   Mean   :58.47
 3rd Qu.:61.35   3rd Qu.:51.45   3rd Qu.:64.23   3rd Qu.:50.45   3rd Qu.:50.95   3rd Qu.:50.93   3rd Qu.:60.74
 Max.   :64.26   Max.   :55.02   Max.   :65.96   Max.   :52.61   Max.   :52.49   Max.   :54.06   Max.   :64.00
      J35A            J36             J37             J38             J39             J4              J40
 Min.   :62.10   Min.   :61.52   Min.   :65.96   Min.   :66.17   Min.   :53.31   Min.   :61.74   Min.   :47.80
 1st Qu.:65.36   1st Qu.:65.05   1st Qu.:67.21   1st Qu.:67.59   1st Qu.:65.03   1st Qu.:63.98   1st Qu.:64.02
 Median :66.23   Median :67.25   Median :68.35   Median :69.88   Median :66.80   Median :65.52   Median :67.04
 Mean   :66.83   Mean   :66.62   Mean   :68.31   Mean   :68.75   Mean   :65.28   Mean   :66.00   Mean   :63.65
      J41             J42             J43|44          J45             J46             J47             J5
 Min.   :47.87   Min.   :61.34   Min.   :22.58   Min.   :23.61   Min.   :33.28   Min.   :52.82   Min.   :49.44
 1st Qu.:60.01   1st Qu.:63.31   1st Qu.:62.99   1st Qu.:48.47   1st Qu.:47.19   1st Qu.:54.37   1st Qu.:60.22
 Median :61.45   Median :64.47   Median :65.13   Median :49.09   Median :48.28   Median :55.92   Median :62.81
 Mean   :61.63   Mean   :64.65   Mean   :57.63   Mean   :48.03   Mean   :46.40   Mean   :55.61   Mean   :63.00
 3rd Qu.:65.52   3rd Qu.:66.63   3rd Qu.:66.48   3rd Qu.:54.06   3rd Qu.:48.83   3rd Qu.:56.86   3rd Qu.:68.15
 Max.   :69.38   Max.   :67.74   Max.   :68.68   Max.   :64.00   Max.   :50.01   Max.   :57.79   Max.   :70.00
      J6              J6A             J7              J8              J9
 Min.   : 9.321   Min.   :24.04   Min.   :30.44   Min.   :15.76   Min.   :15.37
 1st Qu.:59.341   1st Qu.:46.31   1st Qu.:53.33   1st Qu.:47.00   1st Qu.:31.73
 Median :61.205   Median :59.09   Median :56.24   Median :59.30   Median :55.87
 Mean   :53.997   Mean   :51.37   Mean   :52.36   Mean   :50.71   Mean   :45.52
 3rd Qu.:63.846   3rd Qu.:61.08   3rd Qu.:58.05   3rd Qu.:62.45   3rd Qu.:58.46
 Max.   :69.594   Max.   :63.38   Max.   :60.28   Max.   :67.46   Max.   :58.97
```

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.

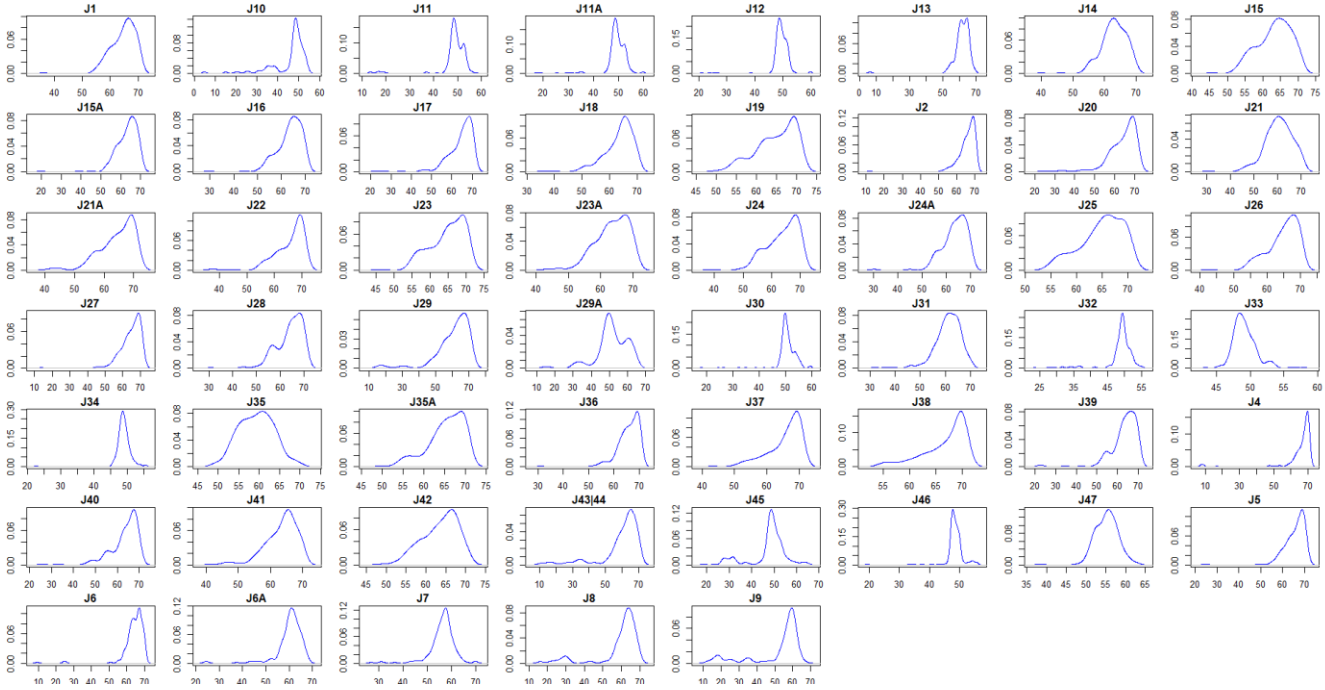
2. **Spread**: The data is moderately spread based on the range which indicates the minimum and maximum values (Inter Quartile Ranges) respectively.



Boxplots of data table

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

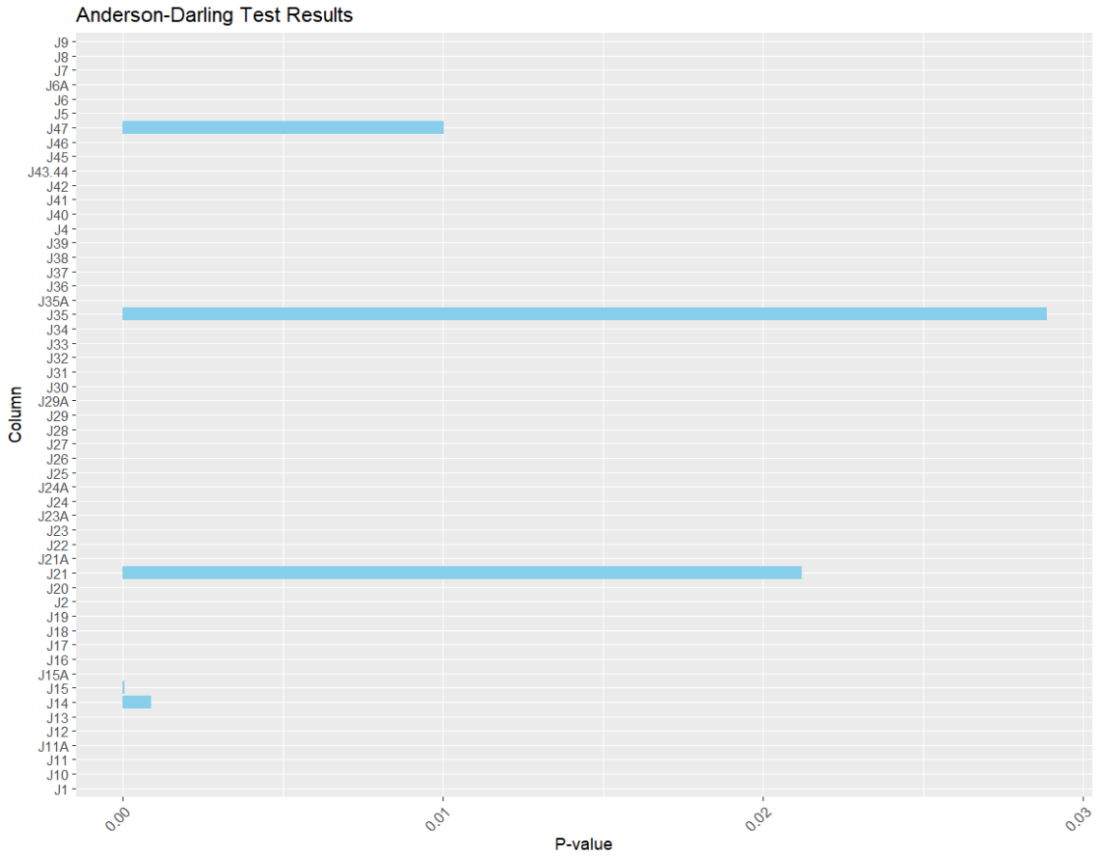
3. **Skewness**: The skewness comes out to -2.176 which indicates the dataset is negatively skewed. The same can be seen the density plot below,



Density plots for all junctions

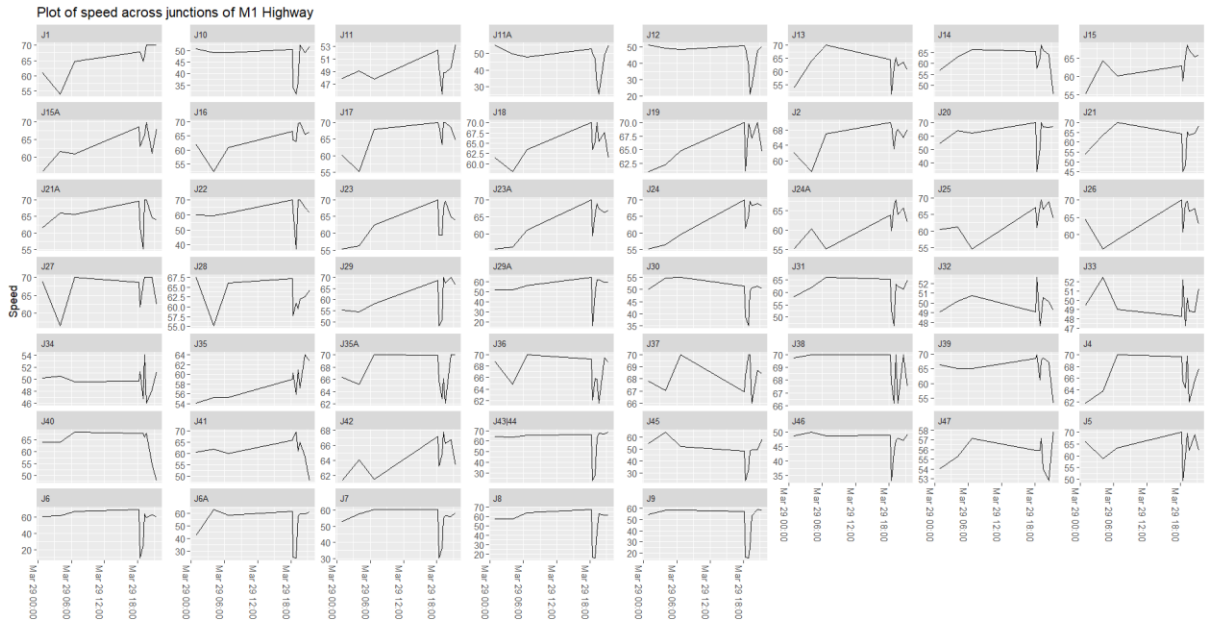
4**. Kurtosis:** The value is 13.23 which indicates a positive kurtosis which means the data has very sharp peaks when compared against a normal distribution. This is also called as leptokurtic distribution.

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.

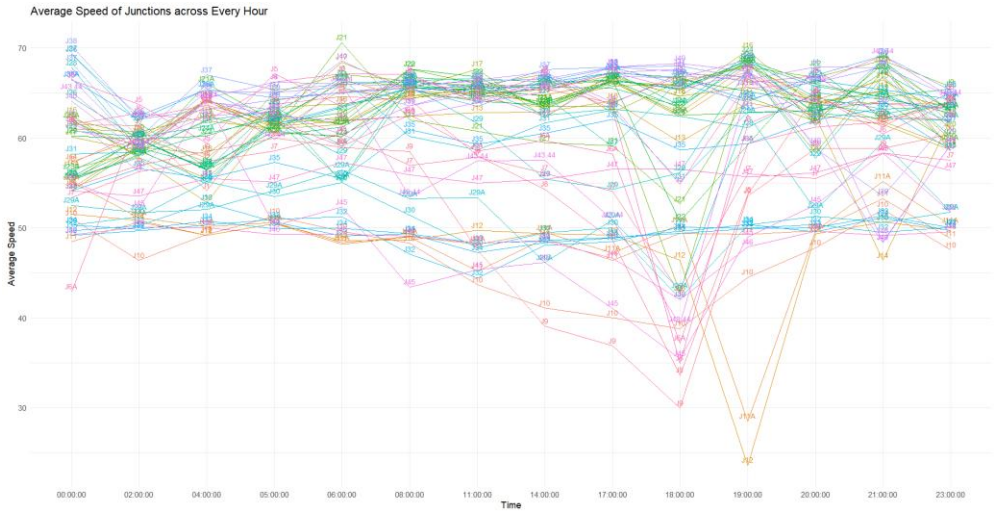


Line graph of average speed across both directions

3.2 Visual Analysis

3.2.1. Temporal Analysis

3.2.1.1 Average Speed of junction through a day

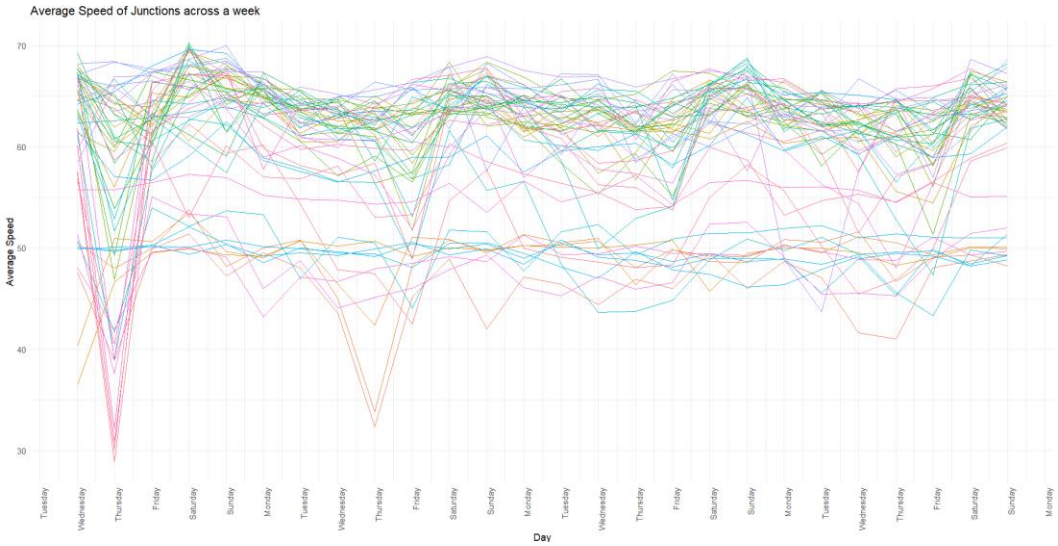


Average Speed of junction through a day

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



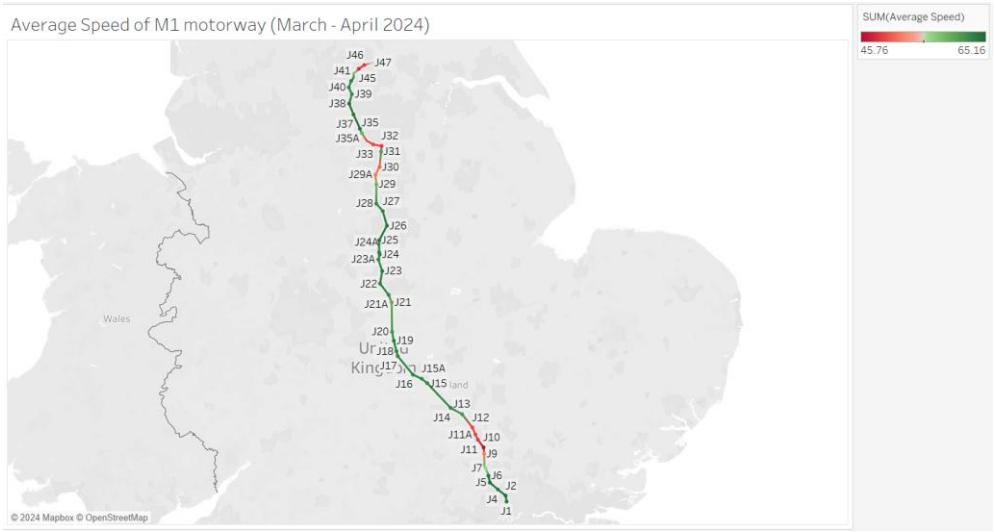
Average Speed of junction across a week

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

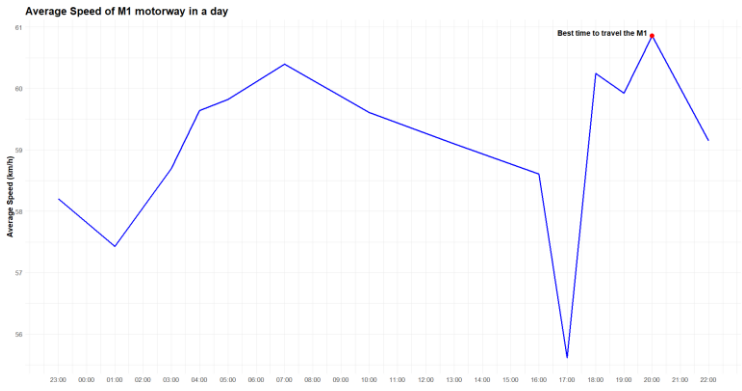


Speed map of the M1 motorway

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.