Applications of Open Transit Data

Student Name: Kshitij Srivastava

IIT-D-MTech-CS-Mobile Computing-MT18099
June, 2020

Indraprastha Institute of Information Technology
New Delhi

Thesis Committee
Dr. Pravesh Biyani (Chair)
Dr. Madhu Errampalli
Dr. Vikram Goyal

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science in General Category

This work was partially funded by the Petroleum Conservation Research Association.

Keywords: open data, open transit data, bus breakdown, fuel consumption, clustering, thresholding, policy evaluation

Certificate

This is to certify that the thesis titled ”Applications of Open Transit Data” submitted by Kshitij Srivastava for the partial fulfilment of the requirements for the degree of Master of Technology in Computer Science Engineering is a record of the bonafide work carried out by him under my guidance and supervision at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Dr. Pravesh Biyani

Indraprastha Institute of Information Technology, New Delhi

Abstract

The concept of open data has become quite popular in recent times among governments and public-facing organisations promoting transparency and collaboration. In a similar expedition, the Open Transit Data Platform for Delhi was established, providing open access to data for buses in Delhi, for both static and real-time components. The work presented in this thesis primarily focusses on exploring the applications of this data for two scenarios: the first one being, identification of service breakdowns in buses. Congestion usually follows a breakdown. So once a breakdown is identified, apply a fuel consumption model to calculate fuel wastage due to the congestion that accompanies. We have identified two types of breakdown cases, viz. clustered - where GPS points form a cluster spread over a certain period and data loss - where the GPS trajectory of bus breaks for a significant amount of time. In our work, we figured the cost of fuel wasted by vehicles stuck in congestions because of such breakdowns to be around eight crores per year in Delhi alone.

The second application is in traffic policy evaluation with a particular case of the odd-even road rationing scheme in Delhi. Traffic speeds for major road sections and bus routes calculated using the open transit data for periods before and during the restriction phase showed that public transport buses during the phase moved as fast as private vehicles on regular days with a consistent decrement of 15-20% in travel times.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr Pravesh, for introducing me to such an exciting topic and providing his continual support and guidance. He has been there attending to all my doubts and correcting my mistakes.

I am thankful to Siddharth and Arav for their contributions to this work and Rajan for his helpful suggestions. This work would not have been possible without them.

I am also very much blessed to have some great friends, especially Manika, Vishal and Ashish, who made this journey more comfortable.

Lastly and most importantly, I am deeply grateful to my parents and sister for all the love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Open Data

Open data is the concept of making data available freely for everyone to use for anything they want. It has become prevalent in recent times among governments and public-facing organisations. There are many advantages to opening up governmental data. In addition to promoting transparency, open data promotes academic and industrial participation, it boosts the economy and helps in making better data-driven decisions.

1.2 Open Data Platforms

Open data platforms are used to share open data with interested users. These platforms are designed to smoothen the flow of data from creators to consumers. It is usually done via sharing data in the form of files of various formats such as text, comma-separated values (CSVs), excel sheets, among others through an online platform.

The Open Transit Data Platform is one such platform that shares the data for buses in Delhi (DIMTS), for both static and dynamic features. Static data includes trip schedules, route information, stop details while the dynamic information is about a vehicle's real-time location and its current trip details. Real-time data gets updated every 10 seconds. The static and real-time data is provided in the General Transit Feed Specification (GTFS) standard. It defines a common format for public transportation schedules and associated geographic information used worldwide.

All work done during this thesis is based on data from this platform.

1.3 Problem Statement

In this master thesis, the objective is to provide an understanding of how the Open Data, especially Open Transit Data, can help in solving problems by making transportation data freely available. We will be looking at the application of open transit data, first, in the detection of service gaps (breakdowns) in bus operations and its extension to estimate fuel costs wasted due to congestions because of these breakdowns. So, given the GPS data for a bus, the objective is to identify breakdowns when they occur in near real-time. Assuming congestion follows a bus breakdown, estimate fuel costs for the amount of fuel wasted by vehicles being stuck in such congestion.

Secondly, we will also look at the data's application in context to policy evaluation with the odd-even road rationing case in Delhi and how did it affect travel times during the period. The odd-even policy was in effect from 4th November to 15th November 2019 in an effort to reduce air pollution in Delhi. We will be looking at the policy's effect on travel times and travel speeds for major roads and routes across Delhi.

1.4 Background

Open Data Platforms have now been established in multiple countries across the world, providing free access to governmental data. These platforms promote transparency and help to build a sense of reliability towards the government. The sector which has gotten the most benefits out of such platforms is probably transportation. Transportation networks, because of their reach among the public interest the most from such advancements. For example, Transport for London (TfL), the transit agency for London oversees all transit movement across the city. They made their data open in the year 2010. Various estimates show that they provide an economic boost of approximately £130 million per annum in terms of development costs for apps, jobs provided along with saving passenger time as more and more reliable travel planning apps are launched.

Similar is the case for Singapore's Land Transport Authority (LTA), which made their transport data open in 2011. In an article by Singapore's Smart Nation initiative [4], they mention how public data has helped plan a better transport infrastructure. They said how the LTA had used open data to identify commuter hotspots along with vehicle sensors to facilitate transport planning to meet commuters' demands better. This resulted in a 92% reduction in the number of bus services with crowding issues despite a year-on-year increase in average daily bus ridership along with a decrement in average waiting time on popular service by about 3 to 7 minutes.

Even in India, states are moving forward with the idea of opening up their governmental data, especially transportation data. Multiple cities like Kochi [9], Hyderabad [10], Pune [11] have come forward with publishing their transit data on open platforms while some are in the process of opening up their data. In addition, to open platforms, there are also data stewardship [12] initiatives where an authority (trust) overlooks the process of delivery data to end consumers while taking care of terms of usage as set by the data provider.

1.5 Motivation and Aim

The future of cities will be built upon open data, and the sector which can benefit the most is transportation. Making transport data accessible for free use enables partners from the industry and academia to develop solutions for problems which otherwise would have been neglected simply because of the absence of data or it not being available freely. The latter is usually the case as most of the transit data is available with companies such as Google Maps, Here Maps, which provide the data for a fee. This restriction is a significant deterrent for many users. In 2018, IIIT Delhi worked with the Delhi Government to establish the Open Transit Data (OTD) Platform in Delhi. This platform aims to provide access to public transit data for buses in Delhi. At the time of writing this thesis, the platform offers static and real-time data for buses operated by the Delhi Integrated Multi-Modal Transit System (DIMTS) Limited, which is one of the two transit agencies operating public buses in Delhi. The other being Delhi Transport Corporation (DTC). The work presented in this master thesis revolves around solving problems using data from the OTD platform to provide an overview of the capabilities of such platforms. We have tried to pick problems to which a solution could have been produced using already existing infrastructure, but with no freely available data, it was not so straightforward. For instance, we have evaluated the effect of the odd-even road rationing policy on traffic movements, which is very much possible using third-party APIs from any of the map companies (Google, Bing, Here). Still, due to usage restrictions, users might be reluctant to approach the problem, or there won't be enough data points to support the study. As much as we have tried to solve multiple problems, our main objective through this work is to push forward the idea of using Open Data for building solutions and moving towards a sustainable future.

1.6 Thesis Structure

This thesis is divided into six chapters.

Chapter 2 discusses the bus breakdown and fuel consumption modelling problem, related work, the methodology employed and the results.

Chapter 3 discusses the policy evaluation aspect for the case of the Odd-Even road rationing scheme in Delhi.

Chapter 4 gives an overview of other relevant applications developed using the Open Transit Data Platform, Delhi.

Chapter 5 is a discussion of the results and insights of the overall work presented in this master thesis.

Chapter 6 contains conclusions drawn from the entire study and suggestions for further research.

Chapter 2

Bus Breakdown Detection - Fuel Consumption Modelling

2.1 Introduction

Public transport is said to be the backbone of a city and it is always appreciative for a city to have a dependable public transport system. Nowadays, we have metros and local trains but buses have always been the face of public transport. This makes it imperative to have a robust public bus network. In Delhi, buses have a daily ridership of around 42 lakhs, with a fleet size of approximately 5700 buses (including both DTC and cluster buses). However, the average fleet utilization is only 90%, leaving an odd 570 buses non-operational on any given day. This is the entire fleet size for smaller cities like Jaipur and Lucknow combined. As of June 2018, it was reported that at least 5500-6000 bus breakdowns happen every month, causing trouble to both the bus operators as well as commuters. These breakdowns cause major traffic blockages especially on arterial roads and there is little that can be done to prevent the losses incurred. According to a study by Central Road Research Institute (CRRI) on Losses of Petroleum Products at Traffic Intersections due to idling of vehicles at Delhi, it was found that in Delhi, with over 466 signalized intersections, 3,21,432 litres of Petrol and 1,01,312 litres of Diesel were being burnt every day due to the idling of vehicles. Converting these figures into monetary terms the total losses, at the current prevailing price of fuel, worked out to be Rs. 3.4 crore per day (Petrol at 80.43, Diesel at 80.53) for Delhi. E. RS 1241.42 crores per annum. Apart from the monetary loss there is productivity loss which is probably many folds this amount accompanied by increase in emissions causing pollution.

IIIT Delhi had previously developed an Open Transit Data Platform which provides the static and real-time data of Cluster Buses (DIMTS) running in Delhi. The data is obtained from the OTD Platform, Delhi. We have used this open data for extracting live geo-locations of buses throughout this project.

2.2 Related Work

As the use of GPS devices to track vehicles has increased, there have been studies to make use of this information for additional statistics about vehicle movement. A density-based clustering approach seems plausible to detect GPS stop gaps because of the accumulation of GPS points. GPS data is also accompanied by a temporal attribute signifying the timestamp of each data point. Birant et al. [8] proposed the ST-DBSCAN by modifying the DBSCAN algorithm to discover clusters on Spatio-temporal data such as GPS data points. Hwang et al. [5] presented an approach to detect stop episodes or stay points as they have described, in the GPS trajectory of an individual. The authors have presented a clustering approach based on DBSCAN, similar to ST-DBSCAN to detect GPS stop points. They also incorporated a gap treatment procedure to deal with data loss in case of GPS signal connection delay. Wang et al. [6] presented a statistical view of vehicle breakdown duration modelling and suggested that the duration for the breakdown of a vehicle depends on its type and the reporting mechanism. They showed that the breakdown duration followed a Weibull distribution. In a later paper, Wang et al. [7] showed the use of Fuzzy Logic and Artificial Neural Networks (ANN) for predicting breakdown duration. The authors presented that this approach outperformed all previous methods.

2.3 Methodology

This section describes the methodology followed for the detection of bus breakdowns and integrating this information with a fuel consumption model to estimate costs in terms of fuel wastage. We can divide the entire process into two sub-components:

- Breakdown detection
- Fuel consumption modelling

For breakdown detection, we start with the process of data collection for complaint records from bus depot. The next step is to digitise these complaint records to identify the types of breakdowns that happen. This is followed by defining algorithms to detect these breakdowns.

Once a breakdown has been identified, we move on to the fuel consumption modelling component. We started with measuring the traffic flow of a 3-lane under congested and non-congested conditions to measure respective travel speeds similar to before breakdown and after breakdown conditions. A speed based fuel consumption model is then employed to calculate fuel consumption difference between congested and non-congested states. This can be extrapolated to estimate total fuel wastage because of such breakdowns.

2.3.1 Data collection

The task of data collection required gathering actual complaint records from a bus depot. These complaint records helped in identifying which buses have broken down on any given day and the reason for it. This data is maintained in physical complaint registers, so the first step was to digitise these complaint registers manually. Optical Character Recognition (OCR) was not an option simply because of the illegibility of handwriting which for some cases was not clear even for a human observer. The only digital records maintained by any bus depot were the monthly statistics which included the total buses that broke down, total trips completed, to name a few. The following sections go about how the data was procured.

Complaint Records

DIMTS Kushak Nallah depot was chosen as the source of information/data for this project considering its larger operational fleet size of over 140 buses. The breakdown data was acquired from the depot in the month of August 2019 for the month of July 2019 in the form of complaint register records. These were scanned images of depot complaint records for the month of July (Figure 2.1). The features consisted of the "bus number", "complaint time", "place", "nature of the defect", "action time", "bus ok time" along with some other details.

The tasks that were performed on this data were as follows:

- Digitized complaint records for the first 30 days. All the scanned images were converted to a spreadsheet file. This file consisted of more than a thousand rows.
- Constructing maps for each complaint record.
- Observing the types of breakdown cases.
- Identifying breakdown cases algorithmically.

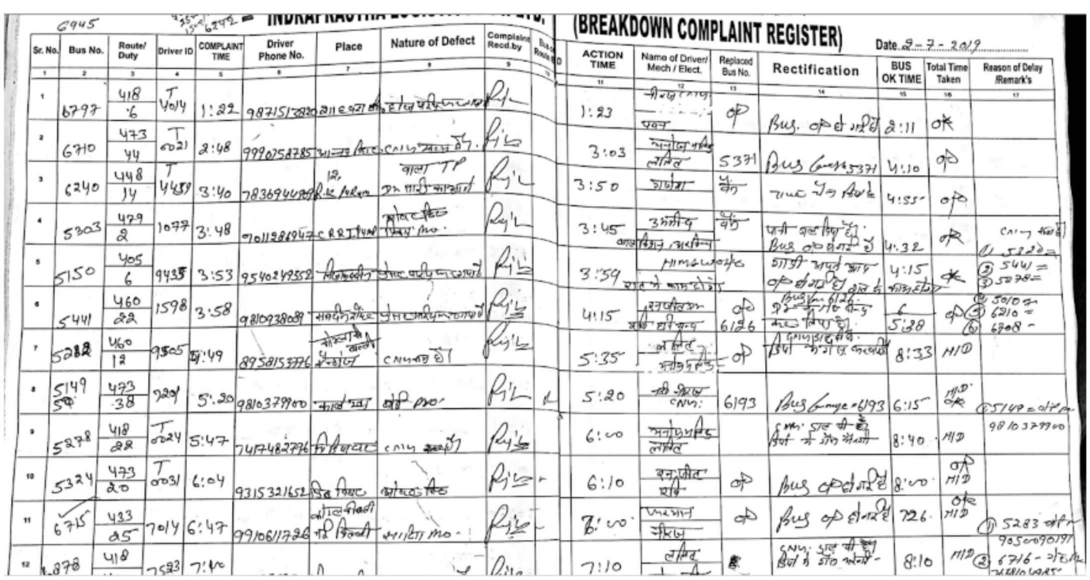


Figure 2.1: Depot Complaint Record Register

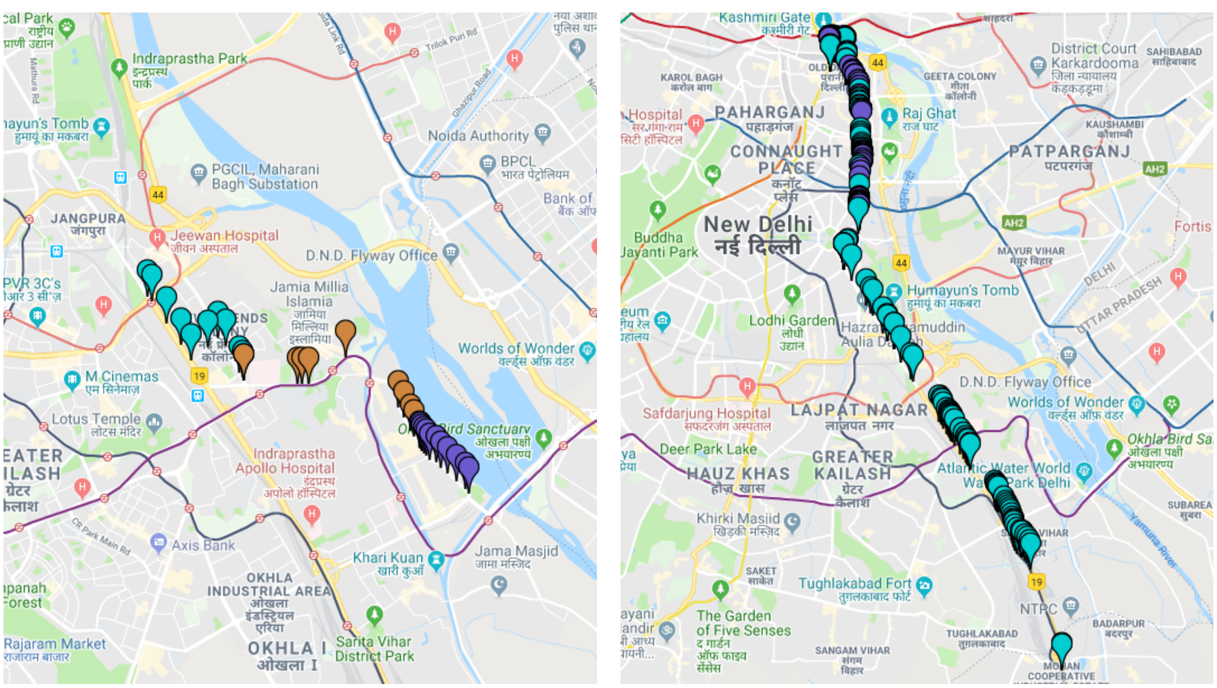
Data Cleaning

This digitized data as such was not fit to be used directly. They were a few features such as the “driver details”, “place”, “name of the mechanic” to list a few that were not required and hence, had to be put away with. From the above fields the ones of our interest were:

- Bus Number
- Complaint Time
- Action Time
- Bus ok Time
- Reason for Breakdown

Date	Bus No.	Complaint Time	Action Time	Bus Ok Time
01-07-2019	5315	05:38:00	06:00:00	7:10:00
02-07-2019	6708	06:40:00	07:00:00	8:00:00
03-07-2019	6708	09:40:00	11:05:00	11:05:00
04-07-2019	5312	09:06:00	09:50:00	11:05:00
05-07-2019	5312	09:08:00	09:10:00	10:50:00

Table 2.1: Sample of the digitized data



(a) (b)

Figure 2.2: Sample of GPS trajectories

2.3.2 Types Of Breakdown

The next step after cleaning and extracting the features from the data was to analyze it. For each day in July, all records were plotted on maps to see the similarities and variations in the GPS data obtained. There were mainly two cases of breakdown that arose after studying the graphs and the plots for the month. The first one was the formation of a cluster of GPS points around the breakdown region and The second case was the one wherein the bus lost its GPS signal minutes before having a breakdown until the entirety of the breakdown or even more.

These two cases are discussed briefly here and would be analyzed in detail in the further sections.

Cluster Cases

For the case of clustering, there was a continuous stream of GPS data that was coming after the bus had broken down, thereby maintaining connectivity with the server. The GPS points were plotted for a bus during the entirety of the breakdown. The inference drawn from the plot was that during the time of breakdown, a cluster of GPS points was formed around the point the bus had stopped.

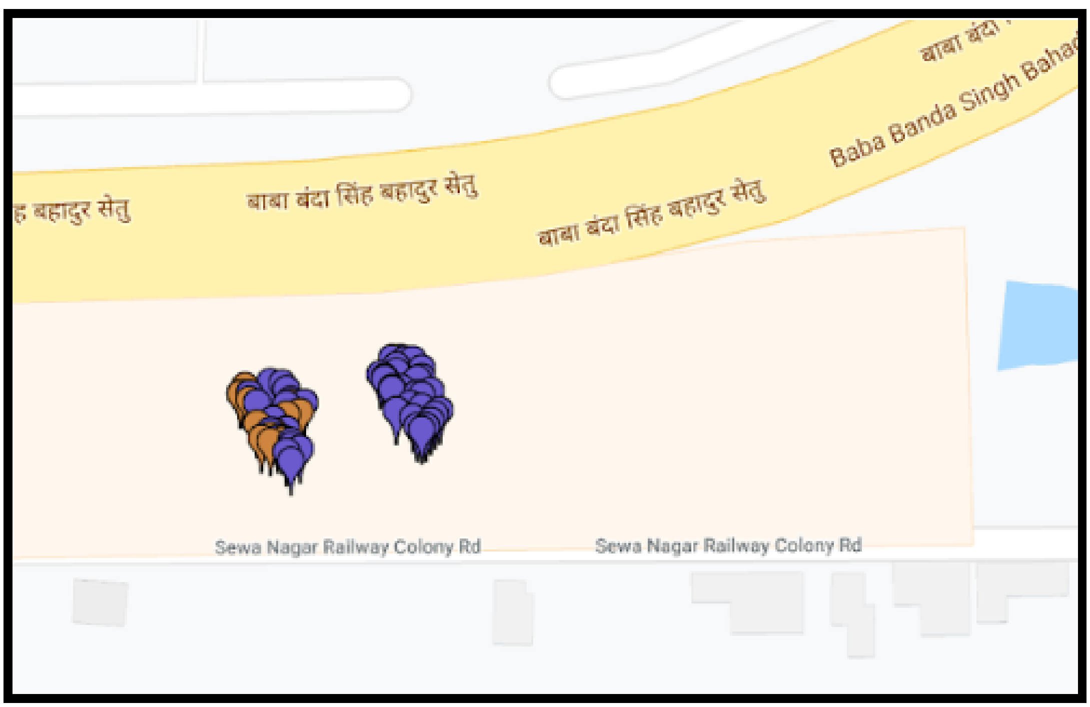


Figure 2.3: An example of cluster formation of GPS points

Data Loss Cases

For the case of Data Loss, the bus lost its GPS signal minutes before having a breakdown. By closely analyzing the cases of Data loss, it was observed that there could be two possible reasons for GPS data signal loss: Connectivity issue of the area in which the bus is moving When the bus breakdowns, the GPS device stopped working. Fig. 2.3.2 illustrates the loss of data between 10:15-11:04 AM

Figure 2.4: Data loss between consecutive GPS points

Other Cases

Other cases included the breakdown of buses which do not impact the flow of traffic. Many of such cases included the buses undergoing repairs at the workshops and also those cases. The buses that were allowed to complete their respective trips even in case they had minor problems that could be solved en-route and issues that did not put the passenger's in harm's way, were also put under the list of breakdowns Some of those cases we found out were as follows:

- Door not working properly
- Glass Broken
- Conductor left
- Window was broken

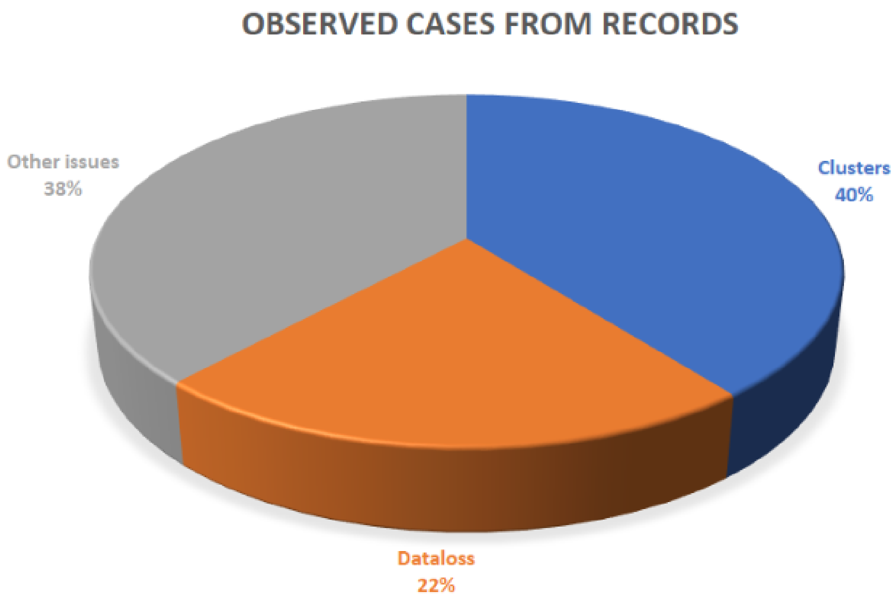


Figure 2.5: Distribution of breakdown cases

2.3.3 Cluster Case

Since there are GPS clusters (groups) being formed in Case 1 breakdowns, a clustering algorithm should be able to detect such clusters. However, the clustering algorithm needs to take into account not only the spatial feature but also the temporal nature of data. Using GPS positions, a breakdown is detected where a cluster of GPS pings gets formed. Here, a Spatio-temporal variant of density-based clustering (DB-SCAN) is used. The algorithm takes care of both spatial and temporal thresholds to define a cluster. Further, we will be looking at different GPS clusters for the month of July 2019.

Algorithm

The algorithm used for detecting clusters is a variant of DB-SCAN (density-based clustering), called ST-DBSCAN (Spatio-temporal DB-SCAN). It takes as parameters, three threshold values: Spatial threshold - region to consider around a data point for detecting neighbours Temporal threshold - time gap to consider for detecting neighbours Minimum neighbours - minimum neighbours required to consider a data point as 'core point'

There are three types of points detected by this algorithm:

- **Core points** - having minimum neighbours within the specified spatial and temporal thresholds.
- **Border points** - are not core-points but have at least one core point in the neighbourhood.

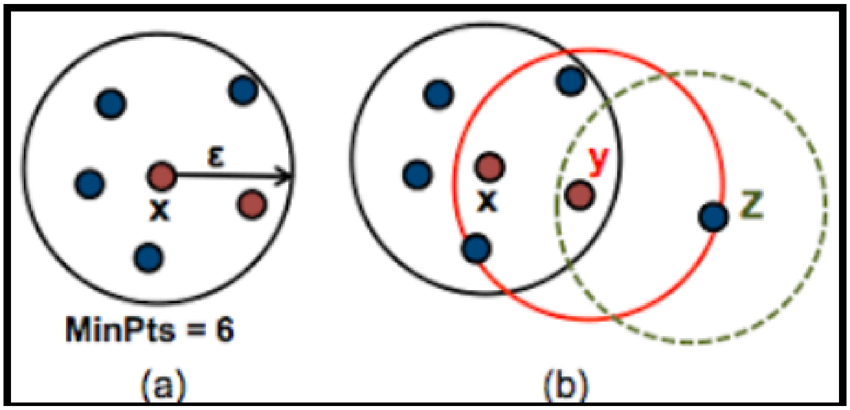


Figure 2.6: An example of the working of the DBSCAN algorithm

- **Noise/outliers** - don't have any core point in the neighbourhood.

There were many cases that arose where the detected clusters did not imply a breakdown. Such clusters were mainly found near the depot and the rest stops(stops permitted by the depot for resting of the bus drivers). To take care of these cases, any cluster happening within a specific radius of the depot or the rest stops were ignored.

Algorithm 1 ST-DBSCAN algorithm

Input: D, Eps1, Eps2, MinPts, $\Delta\epsilon$
Output: C = $(C_1, C_2, ..., C_n)$ (Set of clusters)

procedure STDBSCAN
  cluster label = 0;
  for i = 1 to n **do**
    if o_i *is not in a cluster* **then**
      X = Retrieve Neighbors(o_i, Eps1, Eps2)
    else
      cluster label = cluster label + 1
      for j = 1 to $|X|$ **do**
        Mark all objects in X with current cluster label
      Push(all objects in X)
      while not IsEmpty() **do**
        CurrentObj = Pop()
        Y = Retrieve Neighbors(CurrentObj, Eps1, Eps2)
        if $|Y|$ >= MinPts **then**
          for all objects o in Y **do**
            if o is not marked as noise or it is not in a cluster **then**
              if $|ClusterAvg() - o.Value| <= \Delta\epsilon$ **then**
                Mark o with current cluster label
                Push(o)

2.3.4 Data Loss Case

The bus breakdowns that resulted in the bus losing connectivity or experiencing an interruption in sending GPS pings for the duration of the breakdown were classified in this set of cases. The challenges associated with data loss were different and unlike that of the clustering case. This was challenging primarily because we were dealing only with GPS data. In the absence of any data, it used to become a herculean task to determine whether the bus had broken down or not. One of the significant hurdles faced was that not all places in Delhi transmit data with the same frequency. Many routes had stretches of road where the bus lost signals due to poor connectivity instead of a breakdown. A mapping-cum-thresholding algorithm was written from scratch to deal with the issue of network connectivity. This algorithm also helped in defining the irregularity quantitatively.

Network Grid

It was observed that GPS signals are not received uniformly throughout Delhi. There are areas with a low rate of GPS signal pings due to which the absence of data may not always imply breakdown, rather it may mean that it is in one of those low network areas. To achieve this the following steps were followed:

- The whole of Delhi was divided into a 200x200 grid.
- Each cell is assigned a value that was equal to the number of GPS points divided by the number of unique buses passing through that particular cell.
- Then the value of each cell was normalized using min-max normalization (to keep the values in each cell between 0 and 1) to maintain uniformity.

Each cell in the grid now had a value that could quantitatively define the connectivity for that particular cell. The average of all the values in the grid was taken and was kept as a thresholding parameter. Any cell value higher than this value was put in the category of a good network area and for the values below that parameter in the category of low network area. This resulted in buses having data loss cases in areas with excellent connectivity to be classified as breakdown cases with a high probability, whereas for buses having data loss in areas with mild connectivity, the thresholding parameters were relaxed and the state was left as ambiguous.

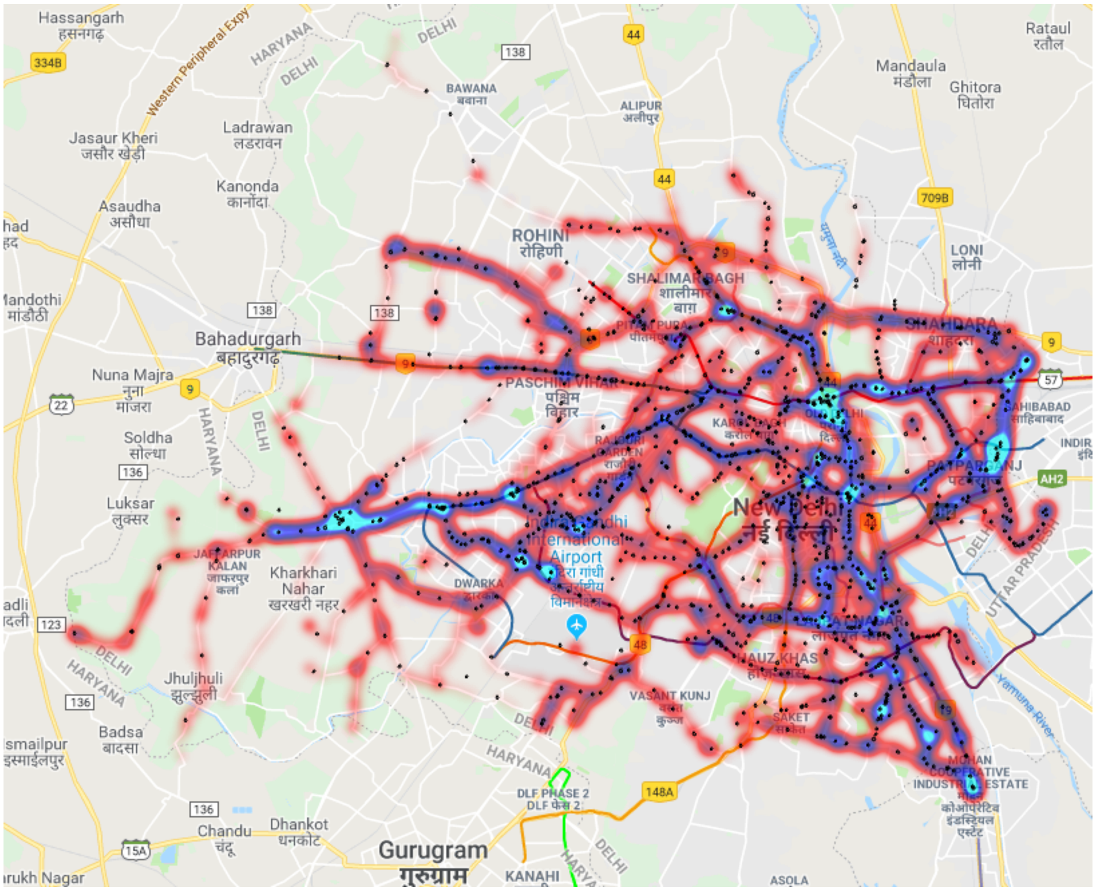


Figure 2.7: Network map of Delhi

Parameters for detecting breakdown in the case of data loss

- **Source/Destination:**
  It was observed that quite a few data loss cases were happening in the vicinity of the source or the destination of a particular route. On further analysis, it was understood that these were false positives and upon experimentation, it was decided to not consider those buses whose last data point was seen at a maximum distance of 300 meters from either the source or the destination.

- **Network Grid:**
  To consider only those breakdown cases in good network areas it was required to find an optimum threshold for the network grid value in each cell (Data loss cases in low network areas are an ambiguity with equal chances of it being a breakdown and a false positive). Averaging out all the values in the grid, it was found that grids with a value of greater than 0.015 were considered to a reasonably good network and data loss cases in these grids ensured a breakdown with a high probability.

- **Rest stops:**
  Initially, there were many cases that showed data loss around Badarpur border. Upon enquiries with the depot officials, it was later understood to be a rest stop i.e. the stop

where drivers are allowed breaks and rest. Many a time these buses used to lose their GPS signal in such rest stops. Hence, any bus losing connectivity within 300 meters of the rest stops were not considered as breakdown cases.

- **Distance from depot:**
  The buses in and around the depots which had to undergo repairs had momentary connectivity and later lost the signal. This was detected by the algorithm and reported as false positives. To avoid such cases the buses within one kilometre of the depot were ignored.

2.3.5 Fuel Consumption Modelling

Fuel efficiency or Fuel Economy is the energy efficiency of a vehicle, expressed as the ratio of distance travelled per unit of fuel consumed in km/litre. Fuel efficiency depends on many parameters of a vehicle, including its engine parameters, aerodynamic drag, weight, and rolling resistance. Higher the value of fuel efficiency, the more economic a vehicle is (i.e., the more distance it can travel with a certain volume of fuel). Fuel efficiency also affects the emissions from the vehicles.

Factors Affecting Fuel Consumption

- **Traffic Congestion:**
  These include the flow of traffic, with a non-congested road average speed of the vehicle, will increase and hence fuel economy will increase.

- **Highway Network Related Factors:**
  These include geometric design features of the highway such as grade. The emission rate is very high at steep gradients, as the vehicle needs to put in more effort to maintain its speed. The highway network facilities such as signalized intersections, freeway ramps, toll booths, weaving sections, etc. also influence the vehicular emission rates.

- **Vehicle-Related Factors:**
  Vehicle-related factors include the engine sizes, horsepower and weight of the vehicle. Vehicles with large engine sizes emit more pollutants. Since larger sized engines are seen in vehicles with more horsepower and more weight, these factors also contribute to the emission rates. Another important factor is the age of the vehicle. Older vehicles have higher emission rates.

- **Other Factors:**

  - Ambient Temperature: Evaporative emissions are higher at high temperatures.
  - Urbanization: Congestion is higher in urban areas, and hence emissions are also higher.

Fuel Consumption Model

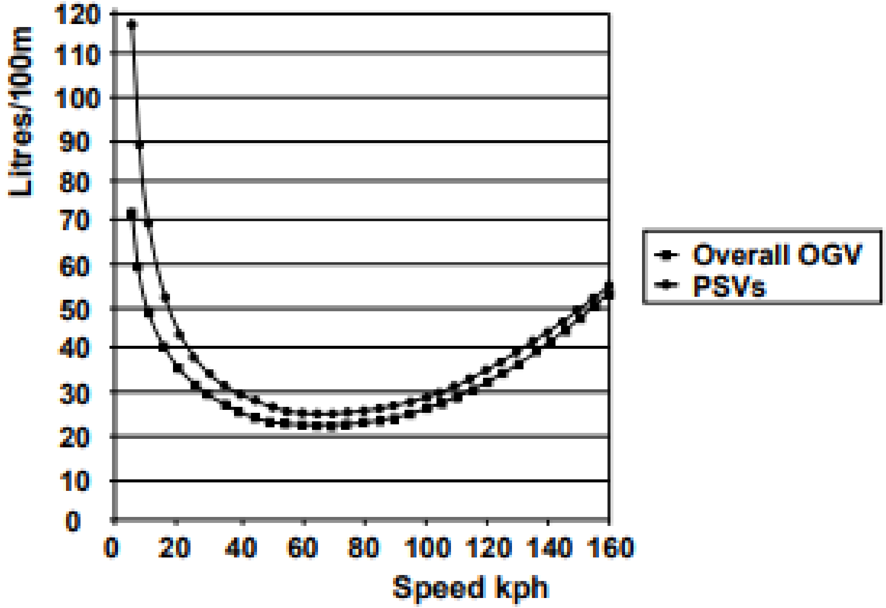


Figure 2.8: Fuel consumption as a function of speed

The Fuel Consumption is related to the average speed (or travel time) using the relation below

$$F = k_1 + k_2 T \tag{2.1}$$

$$F = k_1 + k_2 / v \tag{2.2}$$

where,
F is the Fuel consumed per vehicle per unit distance (litres/km)
T is the Travel time per unit distance, including stops and speed changes (minutes/km),
v is the Average speed measured over a distance including stops and speed changes
k1 is the parameter associated with fuel consumed to overcome rolling resistance, approximately proportional to vehicle weight (litres/Veh-km)
k2 is the Parameter approximately proportional to fuel consumption while idling (litres/hr).

2.4 Results

Cluster Case Results

From the detected cases, we found out the ones which are correct breakdowns and also false detected cases (i.e., the ones which were not breakdown/complaint cases but were detected as such by the algorithm).

Date	Complaint Clusters	Algorithm detected clusters	Clusters Matched
01-07-2019	9	15	9
02-07-2019	5	12	5
03-07-2019	10	15	10
04-07-2019	10	20	10
05-07-2019	11	21	11

Table 2.2: Cluster case results for Esp1=10, Esp2=900, minPts=90

Data loss Case Results

Date	Complaint cases	Algorithm detected cases	Data loss cases Matched
01-07-2019	9	29	8
02-07-2019	5	18	4
03-07-2019	9	30	6
04-07-2019	12	30	10
05-07-2019	18	26	6

Table 2.3: Data loss case results

Fuel Consumption Modelling

Table 2.4 shows fuel consumption calculations for an average car under congested and non-congested conditions.

Vehicle Type	K1	K2	Road Type	Fuel Consumption (Litre/Km)	Fuel Economy (Km/Litres)
Car	0.041	0.9	Congested	0.089	11.2
Car	0.041	0.9	Non Congested	0.061	16.23

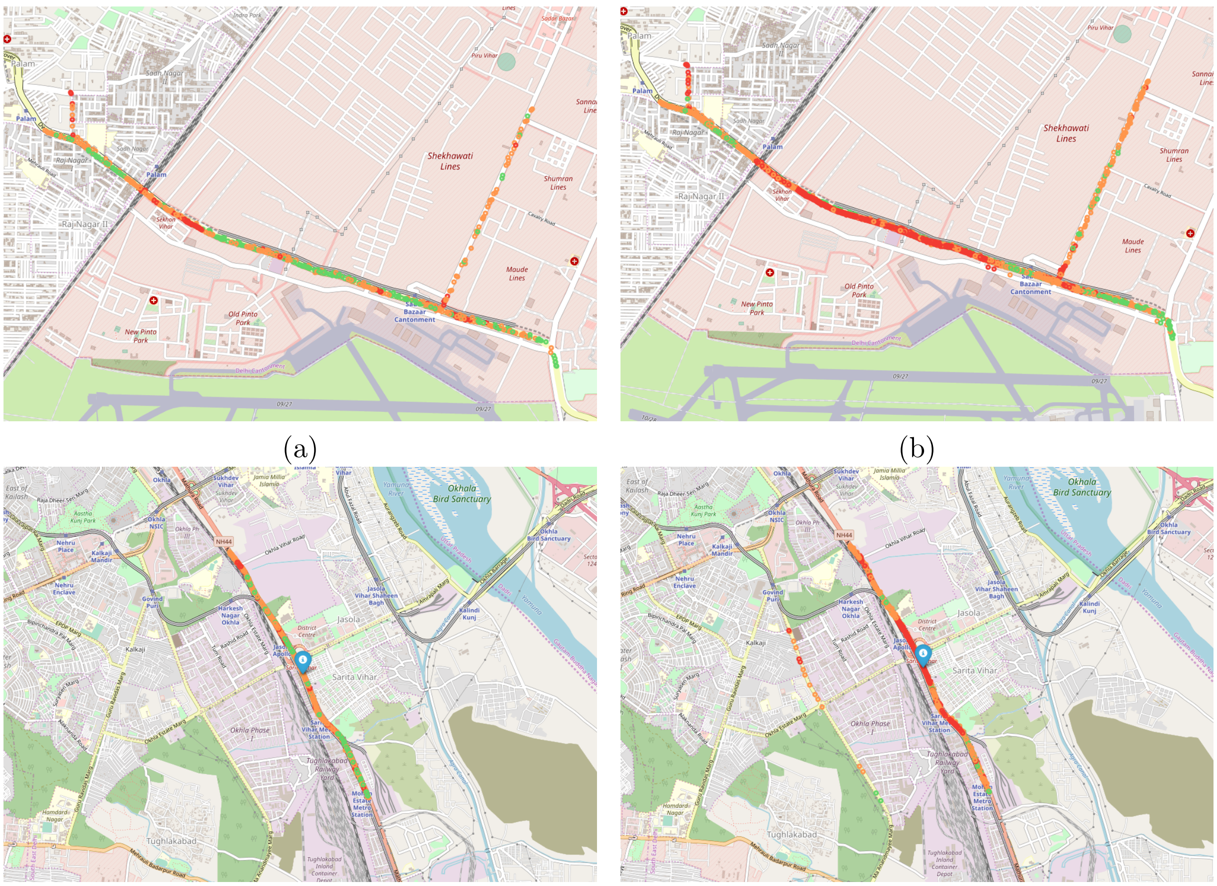
Table 2.4: Fuel Consumption Model Calculations

Monetary effects of Congestion

All the facts and estimates used in this study are for DIMTS (cluster) buses and using the data obtained from one depot (Kushak Nallah) but can be scaled reasonably to be applied to the entire Delhi. If we combine the results for both cases, i.e, cluster and data loss we observed around 4-5 congestion causing cases each day. Using the fuel consumption model, for a stretch of 1km, the difference in fuel consumption is 0.028 litres. The normal traffic flow for a 3 lane road during peak hours was observed to be around 100 vehicles per minute. If we take a congestion period of 10 minutes, there would be 1000 vehicles making a fuel consumption difference of 28 litres. There are 11 cluster depots and around 40 DTC depots leading up to at least 100 such cases each day for all depots would be a reasonable estimate. The estimate comes out to be around 8.22 crores/ year in fuel wasted by vehicles stuck in congestions due to a breakdown.

Effect of Congestion on Traffic speed

The changes in the average speed of vehicles on roads with a bus breakdown were quite noticeable. This idling of vehicles causes a significant reduction in the fuel economy of vehicles. Two sample cases of breakdown of bus affecting the speed has been shown below:



(a) (b)

(c) (d)

Figure 2.9: Effect of bus breakdown on traffic speeds. (a) shows the traffic speed of area before the breakdown of a bus - DL1PD3210 on 17th October, 2019 (4pm - 5pm). (b) shows the same area after the breakdown occured (after 5pm). (c) and (d) show a similar case for DL1PC6716.

Chapter 3

Odd-Even Road Rationing Policy Evaluation

3.1 Introduction

According to the Economic Survey of Delhi 2018-19, in the year 2017-18, there were 10.9 million registered vehicles in Delhi, with over 3 million four wheelers and 7 million two wheelers. The vehicle ownership per capita almost doubled from 317 in 2005 to 598 in 2017. This immense vehicle population has a great impact on the traffic flow around the city. Traffic congestions are a commonplace and make travelling within Delhi, a hassle. These vehicles also contribute towards pollution due to large vehicular emissions. The condition worsens around the time of Diwali festival. The combined effect of vehicular emissions, fireworks and crop burning in neighboring states end up making the city's air unsuitable for breathing. To curb the effect of vehicular emissions to some extent, the government of Delhi introduced the odd-even rule during the month of November. While the effects of the scheme on pollution are still debatable, there are visible effects on traffic congestion and that is what we are interested in.

There are studies which have analyzed of the scheme's effect on air pollution in great detail by collecting data from air monitoring devices and analyzing it. However, the studies focused on inspecting the effect of the odd even rule on traffic congestion have been supported by the traveller's perception. There have been surveys, interviews and manual data collection at major cross sections to estimate the policy's effect on traffic congestion. However, the problem with such studies is that it is always arguable that there were not enough data points to support the author's claims.
The work described here gives a view on the traffic restriction's effect on congestion over Delhi using the data acquired from the Open Transit Data (OTD) Platform. The approach presented in this paper is completely data driven with no human intervention at any stage for collecting data. The OTD platform makes it possible to perform a citywide as well as region wise analysis which is discussed in detail in upcoming sections.

3.1.1 Odd Even Traffic Restriction

In order to reduce the effect of vehicular emissions on the air quality of the city the government of Delhi announced the odd-even road rationing rule to be implemented from 4th November to 15th November 2019. The rule was first imposed in twice in 2016 and later in 2017.

According to the rule, vehicles with odd numbered license plates would be allowed to run on odd calendar dates and the ones with even numbered license plates on even calendar dates. The rule exempts two wheelers and electric vehicles along with women only vehicles with children upto 12 years of age. Unlike previous iterations, CNG vehicles were not exempted. In total, 29 vehicle categories had been exempted from following the rule. It is estimated that around 50% of private four wheelers had been affected. Assuming all registered vehicles ply on the roads of Delhi; at any given day there were 1.5 million lesser vehicles on roads during the period. The big amount of reduction in the number of on-road vehicles led to faster travel across the city. There were lesser congestions reported during peak hours. Through this study we aim to quantify these observations with the help of transit data acquired from the OTD platform.

3.2 Objective

The goal of this study is to assess the impact of the odd even road rationing policy on traffic congestion in a data driven manner. The odd even rule in Delhi is the first recorded instance of any such road rationing scheme to be implemented in India. There have been extensive studies done analyzing the scheme's effect on air quality using air monitoring devices which measure the levels of air pollutants. However, the studied effect on traffic congestion has been more supported by what people have perceived rather what has been measured. There have been claims about the policy benefiting the overall traffic flow but it has been hard to put a number to such claims. With Delhi's Open Transit Data Platform in place, we now have the data to perform a quantitative analysis of the scheme's effect on traffic congestion. This was the sole motivation behind this study. The odd even rule, adopted to curb rising air pollution levels in Delhi, is a big step towards road rationing and can benefit other congestion infested cities in India, not necessarily affected by air pollution.

While it has been difficult to state the policy's contribution towards reducing pollution levels, it's contribution towards smoothing traffic flow is well established by various studies worldwide.

3.3 Related Work

Road space rationing schemes, especially the odd even policy have been used at many places around the world, with minor variations. The studies are mostly concerned with the environmental aspect of the policy, since, such schemes are usually employed to curb air pollution. There have been some related studies however, highlighting the impact of such schemes on traffic congestion.

We will be looking at some of them in this section focusing on the data collection and analysis strategies.

Beijing, China 2008

The odd even rule was implemented in Beijing in 2008 during the summer Olympics. Li and Guo [2] analyzed the traffic volume using a total of 592 traffic detectors installed at major ring expressways and arterial roads. A reduction of 20-30% was reported in traffic volume with an overall increase in travel speed by 10-20% during the odd even period.

Jakarta, Indonesia 2016

Jakarta implemented the odd even rule on 30th August, 2016. However, unlike New Delhi, the rule was applicable only on limited roads segments in the Central Business District area. A 2% decrease in travel time and 2% increase in travel speed was reported [3]. According to a user survey done as part of the study, commuters reported that they did not experience any effect of the odd even rule.

The main difference between the above studies and ours is in terms of the data collection process, scalability and monetary cost of the process. The pre-existing Open Transit Data Platform was used, the data was freely available and no people were specially employed for data collection tasks. In addition to this, we had the flexibility to look for any particular area in Delhi for analysing the effect of the odd even rule.

3.4 Methodology

Data is served through an application program interface (API) at a 10 second update interval. The initial idea was to collect data right before and after the odd-even period and compare the two. However, it was a festive week due to Diwali. Since Diwali is a major festival, the roads are crowded with vehicles hopping marketplaces. This festive rush adds to the regular traffic and eventually slows it down.

To make sure that the effect of Diwali is minimal on traffic, a two week period before Diwali week is considered. However, it is still interesting to see how does the festival affect traffic flow. This led us to consider the three following phases for our study:

1. Pre-Diwali Phase: 7-16 October
2. Diwali Week Phase: 21-30 October
3. The Odd-Even Phase: 4-15 November

3.4.1 Data Storage

The OTD API serves the real-time data in the General Transit Feed Specification (GTFS). A format widely followed all around the world as a standard for transit data exchange. A proto buffer file containing the following fields is received every 10 seconds via the API:

- GPS coordinates (latitude, longitude)
- Timestamp
- Vehicle registration number
- Vehicle's route ID
- Vehicle's trip ID

The GPS coordinates represent the position of vehicle at the given timestamp. The route ID tells the vehicle's current route and the trip ID refers to which trip the vehicle is operating on. Up and Down directions are allotted different route IDs. Moving once from source to destination in one direction is counted as a trip. The span of a trip is anywhere around 1-2 hours having an average trip length of 10 kilometres. For each such trip, we consider the data points which satisfy the above proximity criterion.

It is possible for the data of an individual bus to not get updated with every proto buffer. One of the possible reasons for this is the unavailability of network signals for the vehicle's GPS device to transmit data. There is nothing that can be done about it as of now. However, this does not pose itself as a limitation as we are able to receive a good number of GPS points for each trip (around 300-400), making the effective update interval of around 30 seconds.

The data is stored in an SQLite database for each day using a script. For all the three phases combined, there are over 80 million unique data-points, spread across 533 routes.

3.4.2 Slot Division

Each hour is divided into 15 minute slots. This helps to divide individual trips into subsections, providing for a more granular analysis. So, we have a matrix in which the rows represent trips and columns represent slots of 15 minute interval from 8 am to 8 pm. Each cell of this matrix contains the average speed of vehicle for that particular slot corresponding to a unique trip. We create such matrices for each day. In the next section, we will see how these speeds are calculated for each cell.

3.4.3 Travel speed calculation

For each data point, we define a criterion on its proximity to a stop on route. We say a point satisfies the proximity criterion, if it is within 100 metres of a stop. Distance is calculated using the haversine distance formula, described below.

$$a = \sin^2(\psi B - \psi A/2) + \cos\psi A * \cos\psi B * \sin^2(\lambda B - \lambda A/2) \tag{3.1}$$

$$c = 2 * \arctan 2(a, \sqrt{(1-a)}) \tag{3.2}$$

$$d = R * c \tag{3.3}$$

where, ψ is the latitude angle difference
λ is the longitude angle difference in radians
R is the radius of the earth (6371 kilometres)

The points which do not satisfy the stop proximity criterion are ignored. For two consecutive points which do satisfy the proximity criterion, the difference between timestamps is the travel time. This travel time also includes the dwelling time and any time spent in congestion. Since, the points are in close proximity of stops ($\pm 100m$), we can consider the location of stops for distance calculation. So instead of directly calculating the distance between the points, we use inter stop distances. This makes for a better distance approximation. Haversine distance formula is used here too. The inter stop distances are static and thus can be pre calculated for faster computation. The distance and travel time calculated are used to find out the travel speed of the bus between two stops for a particular trip, using the following formula.

$$Speed = \frac{Distance\ between\ two\ consecutive\ stops}{Time\ taken\ to\ cover\ the\ distance} \tag{3.4}$$

Slots are assigned on the basis of the timestamp of the first data point by rounding them down to the closed quarter. If there are multiple pairs in a single slot, their speeds are averaged, resulting in a single slot speed. This process is performed for all the trips in a day (8 am to 8 pm). For convenience, we will refer to these speeds as "trip slot speeds".

For calculating hourly speeds, the slots are clubbed into groups of four and then averaged. This results in 12 slots of 1 hour each.

3.4.4 Calculating Area Wise Speeds

For calculating area wise speed, we again take the help of stops. Given a stop, we can query the real-time database and find data points which are close to the stop. Since each data point has a trip ID and timestamp associated with it, these can be used to find the speed using trip slot speeds calculated earlier.

We find a stop in the area that we are interested in. Then find nearby stops which are within a 1 kilometre distance from this stop. For each of these stops, a bounding box is computed around each of these stops, with a diagonal of 2 km. Once, the bounding box is computed, we filter out the data points that fall within the bounding box. A slot is assigned to each of these points. Once the trip ID and slots for all such data points are known, we can query the respective "trip slot speeds" matrix to get travel speed.

3.5 Results

3.5.1 Travel Speeds

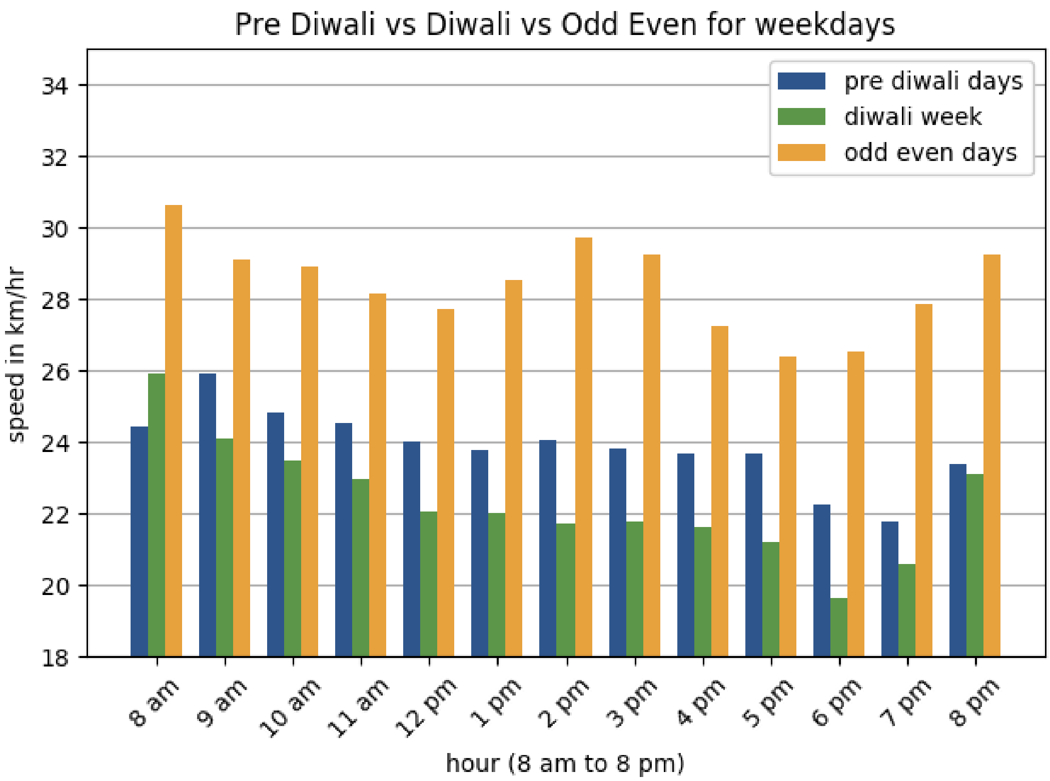


Figure 3.1: Hourly travel speeds across Delhi for the three phases

Fig. 3.1 shows a comparison between the travel speeds across Delhi for the three phases considered for weekdays (from Monday to Saturday). Speeds are averaged over all days in each phase for every hour. The speed for 8 am to 9 am is around 24 kmph, 26 kmph, 30 kmph for pre Diwali, Diwali and Odd Even phase. The speeds then go on to decrease until 12 pm. The differences remain quite similar among the phases, except for the festive week, which witnesses a significant decline. Overall, there is an increment of at least 20% for morning peak hours.

After the peak hours are over, there is an increase in speed seen for all three phases. However, it is quite significant for the odd even phase. For evening peak hours (5 pm to 8 pm), the difference is not much initially but increases later.

The pattern for travel speeds throughout the day is almost constant for all three phases; with the

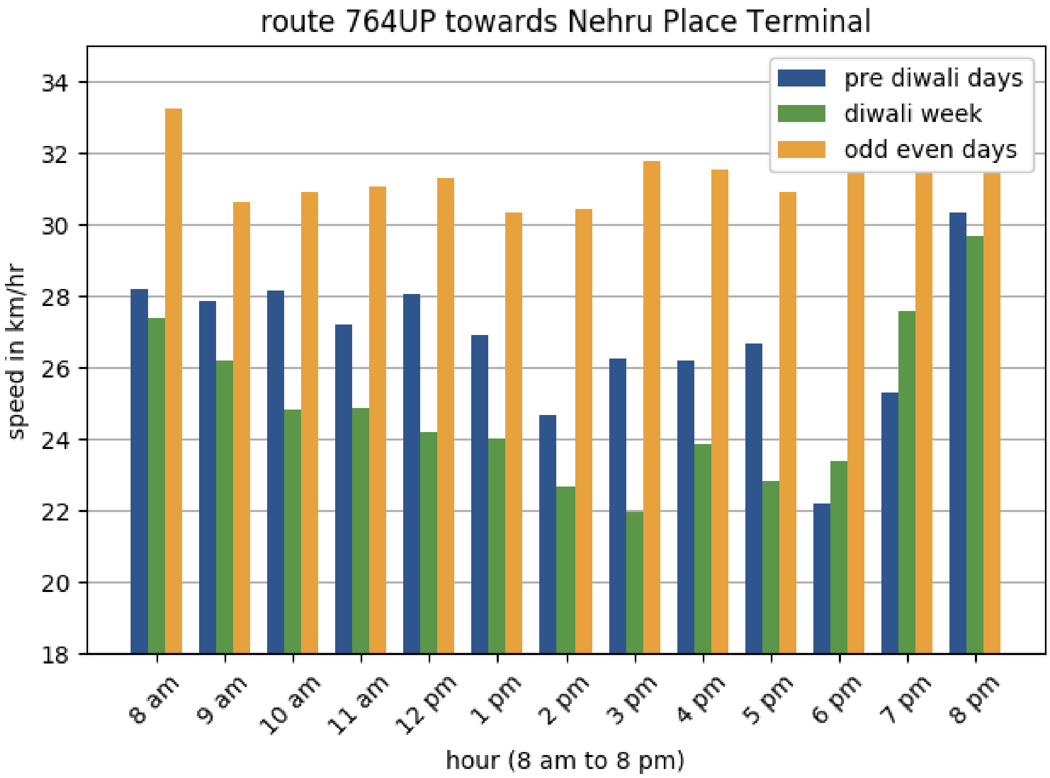


Figure 3.2: Route 764 UP

maximum speed being observed around 8 am. A little variation is seen during noon. The lowest is seen around 6 pm and it improves as the night sets in. The difference in the odd even phase still cannot be ignored. The following subsections discuss the rule's effect across routes and areas.

Across routes

Fig. 3.2 and 3.3 show the travel speed comparison for route 764 which operates between Nehru Place Terminal and Najafgarh. Unlike what we have seen earlier, there is a consistent increase in travel speed from 9 am to 11 am in both the directions. The area around Nehru Place being a well known corporate hub caters to a significant population of private vehicle owners. The increase that we see is most likely because a lot of these vehicles were removed from roads. Secondly, the area is also well linked via metro rail services, making it the preferred mode of transport during odd even phase.

Across areas

Fig. 3.4 shows the travel speeds around the Gupta Market area across the three phases. The figure shows the travel speeds around the Gupta Market area across the three phases. The area is a well known and busy marketplace. The results are quite similar to what we have seen earlier, with a slump seen during the Diwali phase, followed by an increment during the odd-even phase.

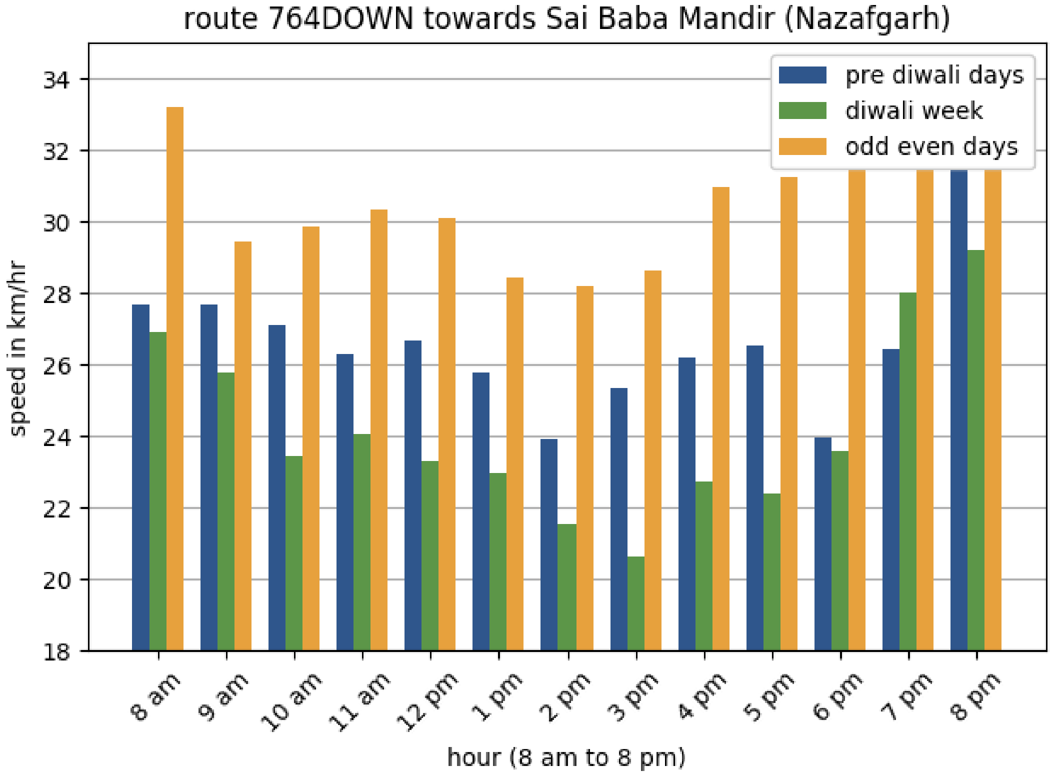


Figure 3.3: Route 764 DOWN

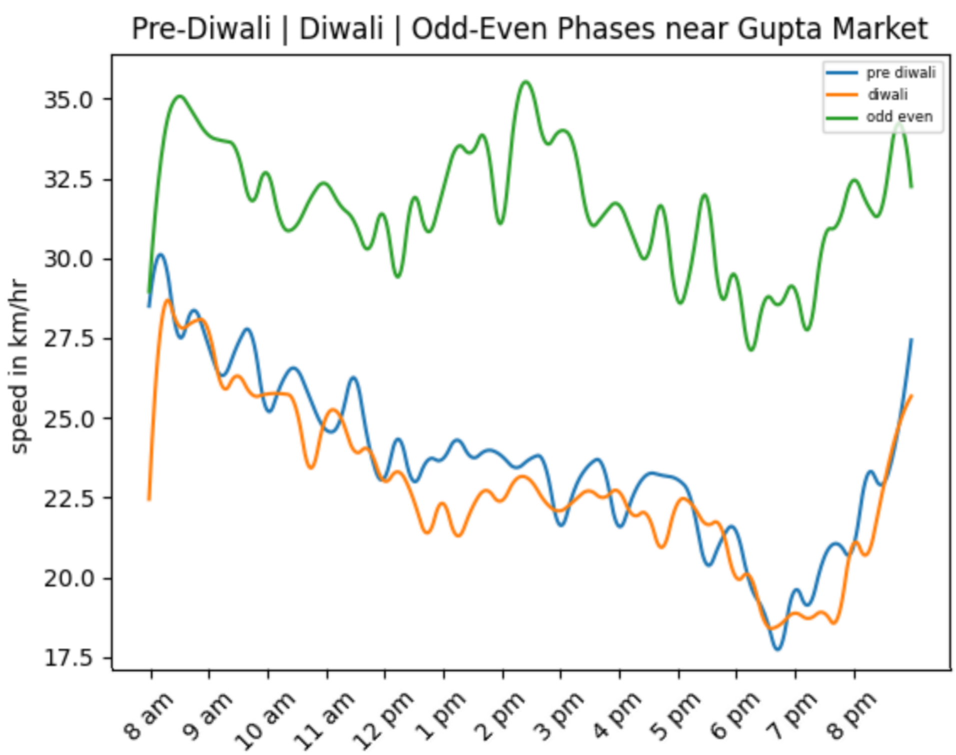


Figure 3.4: Travel speeds around the Gupta Market are across the three phases

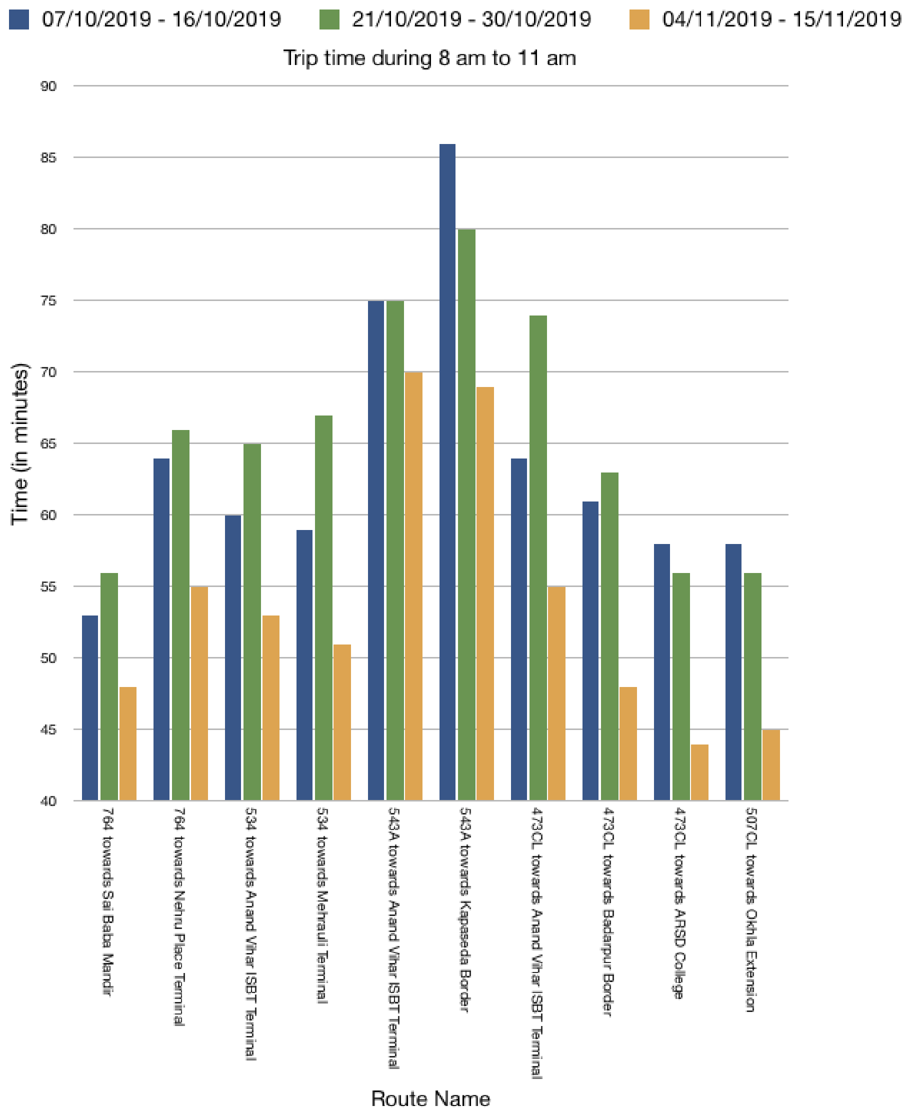


Figure 3.5: Travel times across different routes during morning peak hours

The increase in speed is pretty consistent for the odd-even period with trends similar to seen in earlier sections. A peak difference of about 10km/hr can be seen around 2 p.m. which is almost 50% of pre-diwali phase.

3.5.2 Travel Time

A more comprehensive difference is visible when we look at travel times. Fig. 3.5 shows the variation in trip travel times for five major routes during morning peak hours. Both up and down directions are shown in the plot. Route 764 is seen to have a reduction of around 10 minutes in both the directions. The biggest difference is seen for route 543A moving towards Kapashera Border with the trip time reducing by 15 minutes in one direction. If we look at percentage reduction in travel times however, the difference remains pretty much the same at around 15% to 20%.

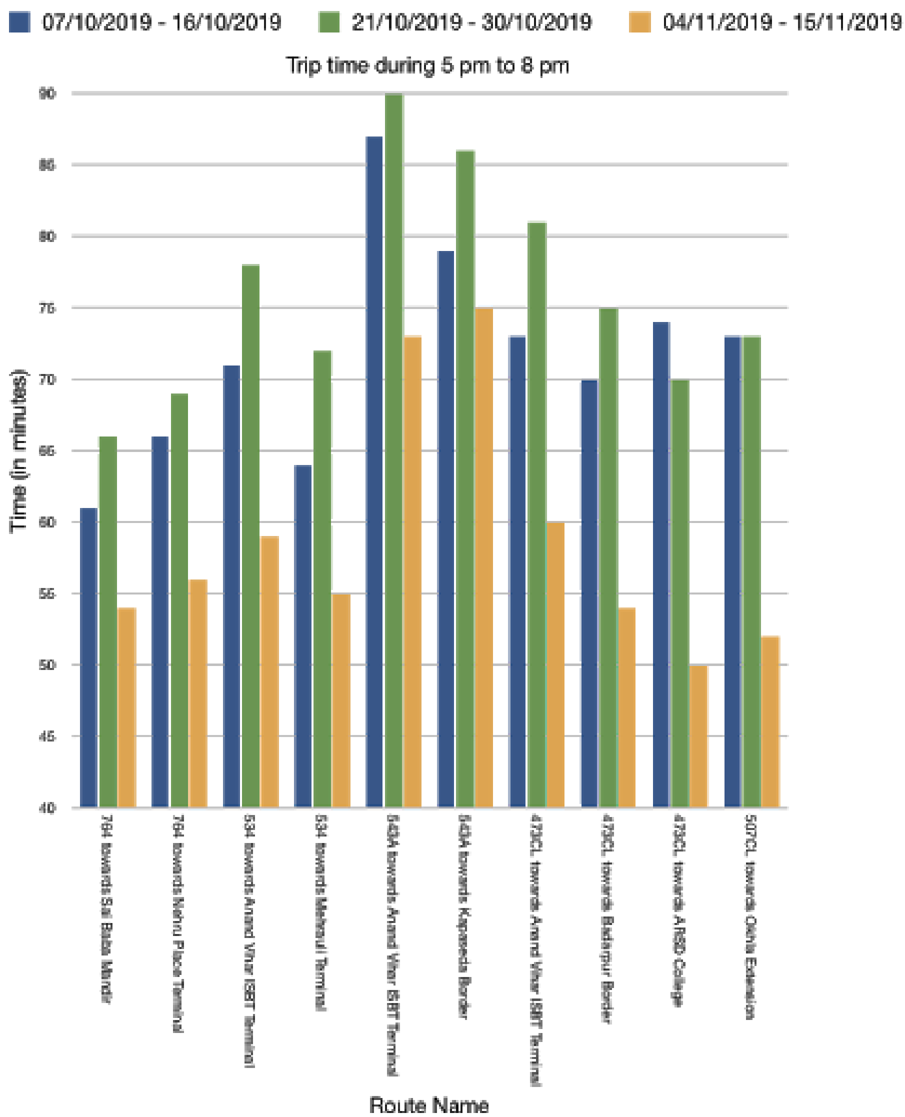


Figure 3.6: Travel times across different routes during evening peak hours

In Fig. 3.6, we see the travel time during evening peak hours. A similar change is visible for route 764 as we saw for morning peak hours. However, for route 543A, the other direction towards Anand Vihar ISBT Terminal takes more time. This is possibly because more traffic moves towards Anand Vihar in the evening in comparison to morning.

Chapter 4

Other Open Transit Data Applications

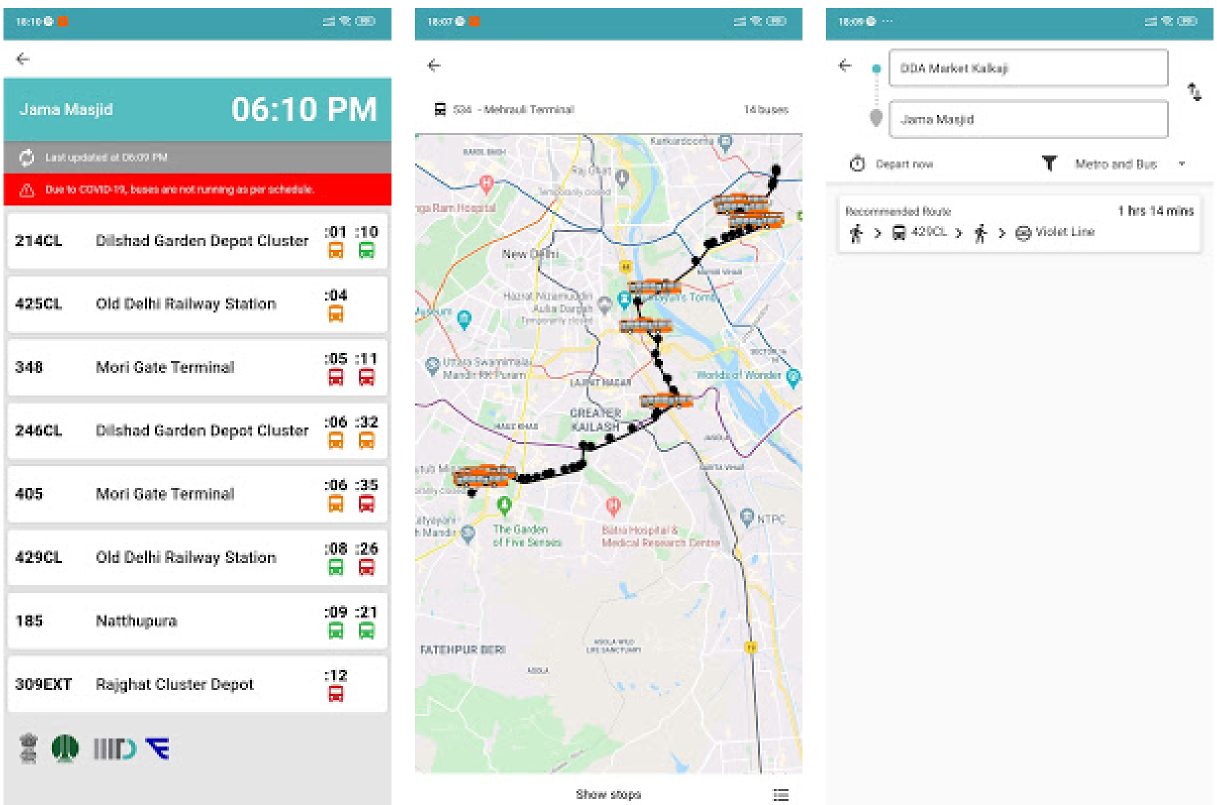
Apart from the applications discussed above, some other applications have also been developed based on the Open Transit Data which are worth mentioning. The novelty of these applications is in their ability to be adapted to any GTFS transit feed with minor adjustments.

4.1 Passenger Information System

A plug-n-play Passenger Information System (PIS) which can be installed at any bus stop to show information about upcoming buses at that stop including their ETA, route names and the type of bus (AC or Non-AC). The system only requires a display unit and a Raspberry Pi computer to load a web page and it is good to go. A snapshot of the PIS can be seen in Figure 4.1.

4.2 Chartr Mobile Application

A one-stop transit application where a user can search for routes, look at real-time location of buses on any particular route and even plan multi-modal trips within the city using public transport (bus and metro).



(a) (b) (c)

Figure 4.1: Application Snapshots

Chapter 5

Insights and Discussions

5.1 Breakdown detection

1. In the presented work, all analysis has been performed on the GPS data of public transport buses. A clustering and thresholding technique is used to identify breakdowns. The shortcoming of relying solely on GPS data for such an analysis is that it might not be possible to accurately identify the reason for such discontinuation in GPS data points. It could be because of a traffic jam or a road accident or an actual breakdown. Such false positives can surely be minimised by selecting suitable clustering parameters; however, to be sure about a breakdown, more data concerning the vehicle's health is imperative. Generally, vehicles have temperature, infrared, vibration and sound sensors, which act as vehicle health indicators. Access to this sensor data in real-time can significantly improve breakdown detection.

2. There was a lot of manual work involved initially in the digitisation of complaint records since all of the breakdown data is maintained manually in registers. In addition to this, there are no defined standards for filling entries in these complaint registers. Hence, it is highly dependent on who is the responsible individual filling out these complaint entries. Manually digitising complaint records is possible for maybe one depot but it is certainly not a scalable solution. Depot complains records data is significant because it helps identify breakdown cases among other service disruptions. An intermediate daily digitisation step would help a lot in conducting further studies on this problem, especially in Delhi.

3. Use of such a breakdown/ service-disruption detection tool can help transit agencies to manage these scenarios, should they arise. Once a breakdown is detected by the algorithm, transit operators can enquire the reason for it and make adequate arrangements. Similarly, the tool can be used by traffic management personnel to know about probable congestion events beforehand.

5.2 Odd Even Policy Evaluation

1. In general, the odd even scheme led to an increment in average travel speed all over Delhi with variations throughout the day. These variations were in sync with the pre odd even phases, i.e., they increased and decreased the same way just on a different scale.

2. Delhi has over 10.9 million vehicles on road, with almost 3.3 million private cars and jeeps and the rest being other passenger vehicles. During the odd even phase private four wheeler vehicles were banned on alternate days based on their number plates. This meant that on any given day during the odd even period, around 1.6 million vehicles were taken off the roads of Delhi, amounting to a reduction of 16% in terms of percentage reduction in vehicle population. In our study, we found a similar increase in terms of travel speeds (20%). Li and Guo [2] in their Beijing study suggested a non linear relationship between the two, based on the traveller's preference to use roads on which the rule was applied.

3. In general, travel times during the evening hours are more than travel times during morning hours. This suggests a greater traffic volume during evening as compared to morning. A possible reason could be that even people who did not take their vehicles out through out the day, got them out during the evening hours.

4. Usually it is observed that the travel time is different between up and down directions. It also varies with time of the day. Like, for route 543A travel time towards Kapashera border is higher during evening hours as compared to morning hours. This is possibly because of traffic volume being higher in one direction during the morning hours and it being higher for the other direction during evening hours. Directions with higher traffic volumes would have lower travel speeds and hence higher travel time. This information can be used to estimate traffic flow variations along a particular route/area. For instance, in route 543A, travel time is higher towards Anand Vihar ISBT Terminal during morning hours suggesting more people go towards Anand Vihar probably for work. While travel time is higher towards Kapashera Border during evening hours, suggesting a higher traffic volume in this direction, most probably because of people coming back to home after work.

Chapter 6

Conclusion and Future Work

6.1 Thesis Conclusion

In this thesis, we showcase different areas where Open Transit Data (OTD) or Open Data, in general, can help in solving a problem and can also provide a data-driven perspective of the situation. We propose to use OTD for the detection of service gaps (breakdowns) in buses. Once these are identified and we have an estimate of traffic flow affected because of the breakdown, we have calculated the cost of fuel wasted because of such breakdowns in Delhi for a year which came out to around eight crores.

The second application that we propose is to use OTD for traffic policy evaluation. We have used the odd-even road rationing policy for this purpose and shown how travel times and speeds were affected during the period. This was done across multiple routes and major cross-sections of Delhi. We have shown that there was an increment of 15-20% in travel speeds of buses, putting them close to average speeds of private vehicles on regular days.

6.2 Future Work

In addition to the work discussed in this thesis, Open Transit Data can also be used for a host of other applications. One of the most promising future extension of this data is in building a no interaction e-ticketing system where the passenger can buy a ticket for his/ her journey by just using a mobile application/ interface. Another problem that we are very interested in solving is of bus bunching, where a group of two or more buses, running along the same route, which were scheduled to be evenly spaced, but instead run in the same place at the same time. This can be solved by preparing a suitable trip schedule and applying bus bunching mitigation techniques, like stop skipping whenever it happens.

Bibliography

[1] Birant, D. and Kut, A. (2007). St-dbscan: An algorithm for clustering spatial–temporal data. Data Knowledge Engineering, 60(1):208 – 221. Intelligent Data Mining.

[2] Li, R., Guo, M., Effects of odd-even traffic restriction on travel speed and traffic volume: evidence from Beijing Olympic Games, Journal of Traffic and Transportation Engineering (English Edition) (2016), doi: 10.1016/j.jtte.2016.01.002.

[3] Nafila, O., Marseveen, M.V. (2018). ROAD SPACE RATIONING TO REDUCE TRAFFIC CONGESTION: AN EVALUATION OF THE ODD- EVEN SCHEME IN JAKARTA, INDONESIA.

[4] "Open Data Analytics for Urban Transportation" 7 Apr. 2020, https://www.smartnation.gov.sg/what-is-smart-nation/initiatives/Transport/open-data-and-analytics-for-urban-transportation-1

[5] Hwang S., Evans C., Hanke T. (2017) Detecting Stop Episodes from GPS Trajectories with Gaps. In: Thakuriah P., Tilahun N., Zellner M. (eds) Seeing Cities Through Big Data. Springer Geography. Springer, Cham

[6] W. Wang, H. Chen and M. C. Bell, "A study of characteristics of motorway vehicle breakdown duration," Eleventh International Conference on Road Transport Information and Control, 2002. (Conf. Publ. No. 486), London, UK, 2002, pp. 151-155, doi: 10.1049/cp:20020223.

[7] Wang, W Chen, H Bell, Margaret. (2005). Vehicle Breakdown Duration Modelling. Journal of Transportation and Statistics. 8.

[8] Birant, D. and Kut, A. (2007). St-dbscan: An algorithm for clustering spatial–temporal data. Data Knowledge Engineering, 60(1):208 – 221. Intelligent Data Mining.

[9] https://kochimetro.org/open-data/

[10] https://data.telangana.gov.in/story/gtfs-hyderabad-public-transport-systems

[11] https://smartcities.data.gov.in/resources/public-transport-mode-sharepune-2015-2018

[12] https://datatrust.in/