Using Smart Card Technologies to Measure Public Transport Performance: Data Capture and Analysis

Industrial Engineering

Pau Segarra Algueró
6/13/2013

Director: Luis Ferreira

Rapporteur: Mª Antònia Santos
Abstract

This project investigates a series of benefits arising from the use of smart card technologies apart from the ones they were initially conceived for. The purpose of this project is to use real smart card data from passengers provided by the transit authority in Brisbane, Australia. Two bus lines were studied, concerning some aspects of the travel time reliability and passenger demand. The aim was to show that smart card data could also be used for purposes other than the ones for which smart card ticketing systems were designed (mainly revenue collection).

The first part of the project intends to review the reason for the success of smart card ticketing systems over other traditional methods of fare collection. Then, focus is moved on to the stage of processing and analysis. In this stage performance data was extracted from passenger records in order to investigate the adequacy of smart card data to provide public transport performance indicators. This was done in two stages: the first one was focused on the method of data extraction, by designing a universal algorithm capable to extract useful data, not only for the scope of the current project but also that could be used on other possible future studies. Next, a data analysis phase was performed, in which the intention was to show examples of results and discern in what cases the method could be useful and the limitations concerning the approximations and length of the project.

Finally there are chapters dedicated to project planning, sustainability and cost breakdown, as well as conclusions and future similar studies that could be derived from this project.
Table of Contents

ABSTRACT ............................................................................................................. 3

TABLE OF CONTENTS ......................................................................................... 5

LIST OF FIGURES ................................................................................................. 7

LIST OF TABLES ..................................................................................................... 9

1. PREFACE ........................................................................................................... 11
   1.2. Origin of the project ...................................................................................... 11
   1.3. Motivation ..................................................................................................... 11
   1.4. Structure of the project ................................................................................ 12

2. INTRODUCTION ................................................................................................. 13
   2.1. Objectives of the project .............................................................................. 13
   2.2. Scope of the project ..................................................................................... 13

3. SMART CARD TICKETING SYSTEMS: INTRODUCTION AND LITERATURE REVIEW .................................................................................................................. 15
   3.1. The smart card .............................................................................................. 15
   3.2. Comparison of fare payment methods .......................................................... 16
      3.2.1. Technologies .......................................................................................... 19
      3.2.2. Privacy concerns ...................................................................................... 20
   3.3. The Go Card in Brisbane .............................................................................. 22
   3.4. Introduction and comparison with other smart card systems .................... 23
      3.4.1. Octopus in Hong Kong .......................................................................... 23
      3.4.2. Sydney and Melbourne .......................................................................... 24

4. BACKGROUND ON SMART CARD STUDIES ................................................... 27
   4.1. Travel Time .................................................................................................... 28
   4.2. Passenger Demand and Trip Planning ......................................................... 29
   4.3. Smart card studies ........................................................................................ 30
      4.3.1. Strategic-level studies .......................................................................... 30
      4.3.2. Tactical-level studies ............................................................................ 31
      4.3.3. Operational-level studies ...................................................................... 32
      4.3.4. Origin and Destination Inferring (O-D) and O-D Matrix Estimation .......... 32

5. CREATING A METHOD TO ANALYSE GO CARD DATA ......................... 35
   5.1. Introduction: Brisbane City .......................................................................... 35
5.2. The study ................................................................. 37
5.3. Data used for the study ........................................... 39
  5.3.1. Matlab® introduction ...................................... 39
  5.3.2. Initial considerations ...................................... 40
  5.3.3. Approximations ............................................ 41
  5.3.4. Flowchart of the process .................................. 43
5.4. Data Processing ................................................... 43
  5.4.1. Data cleaning and setting .................................. 44
  5.4.2. Data restructuration ....................................... 47
  5.4.3. Calculation of travel time and passenger demand indicators ...... 56
6. RESULTS AND ANALYSES ............................................. 61
  6.1. Travel time analysis ............................................ 61
    6.1.1. Comparing travel time statistics for different services and directions .... 62
    6.1.2. Travel time variation in different time periods ......................... 63
  6.2. Passenger Demand ............................................. 67
    6.2.1. Bus loading calculation ................................... 67
    6.2.2. Checking bus line sizing .................................. 69
7. SUSTAINABILITY OF THE PROJECT ................................. 71
  7.1. General Description of the Project and Consideration of Alternatives ...... 71
  7.2. Description of the Environment ................................ 71
  7.3. Identification and Assessment of Impacts on the Environment .................. 71
    7.3.1. Evaluation Criteria ....................................... 72
    7.3.2. Identification of Impacts. List of Activities ....................... 72
    7.3.3. Potential Environmental Impacts ............................ 73
    7.3.4. Assessment of Impacts ..................................... 74
  7.4. Planned Actions ................................................ 74
8. COST OF THE PROJECT ................................................. 75
  8.1. Cost breakdown ................................................ 75
  8.2. Total Cost .......................................................... 76
9. PLANNING .............................................................. 77
CONCLUSIONS ............................................................ 79
FUTURE STUDIES BASED ON SIMILAR SMART CARD DATA ............ 81
ACKNOWLEDGEMENTS .................................................... 83
REFERENCES .............................................................. 85
List of Figures

**Figure 1**: Sample picture of a smart card 16

**Figure 2**: Advantages and disadvantages of using the combi-card (sample of 160 passengers, in Tampere, Finland, May 2000) (Blythe P., 2004) 17

**Figure 3**: Using a Go Card 22

**Figure 4**: Brisbane and Barcelona extension comparison 35

**Figure 5**: Bus route layout for 109 and 412 bus services. 37

**Figure 6**: Brisbane CBD with Cultural Centre and Adelaide Street Stops. 38

**Figure 7**: UQ St Lucia Campus and main bus stops 39

**Figure 8**: Buses lining at Cultural Centre Busway Station on a Friday at 5:15PM (own capture) 41

**Figure 9**: Time calculations schematic diagram 42

**Figure 10**: Scheme of the data extracting process 43

**Figure 11**: Sample of the initial raw data. 45

**Figure 12**: Sample of some incomplete data that needed to be removed. 46

**Figure 13**: Example of classifying a cell string. 47

**Figure 14**: Example of the classification process 48

**Figure 15**: CS4 variable after first classifying set of steps 49

**Figure 16**: Example of a dendrogram and different cluster possibilities 50
Figure 17: Obtaining pairwise distances using pdist function in Matlab® 51

Figure 18: Classification example after clustering process 54

Figure 19: CS5 variable before joining same station records 55

Figure 20: CS6 variable after unifying same station records 56

Figure 21: Parameters calculated for travel time analysis. 58

Figure 22: Parameters calculated for Passenger Demand indicators 59

Figure 23: Sample of the final output data for 109 inbound bus service extracted to excel 60

Figure 24: Basic descriptive statistics for 109 outbound service during the month of March of 2013 63

Figure 25: 412 inbound travel time plots from Monday to Friday (April 8-12th 2013) 64

Figure 26: Amount of smart card users in 109 outbound bus departing at 8:30AM on 05/03/2013 70

Figure 27: Project task scheduling 77
List of Tables

Table 1: Advantages and disadvantages of the use of a smart card (Pelletier, Trépanier, & Morency, 2009) 19

Table 2: Characteristics of some lower level contactless smart card standards, (Trépanier, Barj, Dufour, & Poilpré, 2004), based on (McDonald, 2003) 20

Table 3: Principal sources of concern about smart cards (Pelletier, Trépanier, & Morency, 2009) 21

Table 4: Proportion of single paper tickets passengers compared to Smart Card users 66

Table 5: Boarding and alighting number of passengers in every stop for each bus studied 68

Table 6: Cumulative number of passengers as bus advances in every stop 68

Table 7: Costs associated with material resources 75

Table 8: Costs associated with human resources 76

Table 9: Total cost of the project 76
1. Preface

1.2. Origin of the project

This project was originated and planned amongst the context of high priority studies in the Transport Department in The University of Queensland. Several options were considered at the beginning of the stay under the supervision of Professor Luis Ferreira. Amongst the existing options, this present project looked interesting and with future applications on transit analysis. Due to the growing existence of smart card ticketing systems in cities around the world, this project seemed a practical way to introduce data extraction methods from smart card data from passengers, as well as to show some sorts of analyses that could be performed.

1.3. Motivation

Once understood the operation of smart card ticketing technologies, especially the system in Brisbane, the potential range of applications of smart card data to the monitoring of bus performance became clearer, and a latent idea of how useful could be to use smart card information to get performance indicators of public transport was appearing. The fact of using this existing smart card data, whose initial purpose was merely to collect fares in an efficient and quick way, to extract performance data and evaluate the public transport reliability and efficiency became fascinating. The correct consecution of this task would lead to a win-win situation for all parts involved in the environment of the public transport, such as the transit authority, the city, and obviously the passengers. The possible benefits of analysing the public transport system and finding solutions would have as a result a better functioning of the system (or indicators of performance showing the inefficiencies) with minimal cost for the authorities, because it uses an existing and successful infrastructure created for other purposes, thus the investment that should have to be spent on technologies or spendings would be minimal.
Another of the motivations at the time to plan this project was the possibility of not only using the skills and abilities already learnt before to undertake and successfully complete the project, but also to be able to learn some new skills to add to the personal background of the involved parts. Not only improving the knowledge in transport planning and monitoring performance, but to learn and use new methods of data processing and develop a step by step workplan to resolve the setbacks. The fact of having to process real smart card with Matlab® appeared to be challenging and motivating.

1.4. Structure of the project

This report is structured as follows: Firstly, a chapter is dedicated to provide an introduction to the study, including its objectives and scope. Then another chapter reviews what are smart card ticketing systems and what are the main advantages that have made these systems very popular around the world. The following part, indeed, introduces the concept of analysing bus performance and at the same time intends to provide a general view of previous studies using smart card data. Then, in another chapter there is explained the method that was developed to use smart card data from passengers the city of Brisbane to get performance data that could be used to monitor the performance of the buses. Next there is a section for some results obtained after analysing some samples of the smart card data. Finally, there are chapters for sustainability analyses, costs associated with the project and the conclusions derived from the realisation of the study. Also a brief guideline on possible similar future research based on smart card data is provided.
2. Introduction

2.1. Objectives of the project

This project is related to the public transit data analysis, and is centered in the use of smart card data to accomplish a certain number of goals.

First of all, as a secondary objective, it is intended to give a global vision of what is a smart card, exposing the advantages and disadvantages of this system and try to give an explanation why these systems are increasingly becoming more popular.

Subsequently, a method was developed to extract useful data from smart card records in the city of Brisbane, Australia. The objective was to extract performance data from smart card records in order to perform travel time and passenger demand analyses and try to discern how useful smart card data can be and which limitations it ballasts. So the method itself and the correct extraction of data and performance indicators was considered as a first objective, and the analysis of the results and limitations of this data extraction system was the second related goal.

Finally, another intention of this project was to draw a picture of some of the future research that could be done parting from similar sets of data as the one available in this project.

2.2. Scope of the project

As stated before, one of the main purposes of this project was to create the method to successfully extract smart card data for performance purposes. It departed from the idea of taking advantage of an existing infrastructure (in this case the smart card ticketing system in the city of Brisbane) to extract information that could help to improve the performance of the public transport network or give some hints of the inefficiencies. It was not intended to give exact or significant results of the performance of the public transport network and how could
it be improved in some way, but to show a possible method to extract the desired output data.

The study was centered in the city of Brisbane, and focused on the bus network, more specifically in two bus lines (number 109 and 412) that depart and arrive in the same places (from the center of the city to the bus stations in the University of Queensland) using different routes. Initially the study of the two bus lines was conceived to provide a possible comparison of both bus services. After analysing the results this option was finally discarded due to lack of substantial conclusions, though both bus lines were kept in the study and some analyses of each are shown in the results and analyses chapter. The inclusion or both lines is because the characteristics of the data in both bus lines is different, and the method to extract data was designed to fit in both cases. So, the fact of testing the Matlab® code in both bus services made the software more robust and universal. This universality feature was considered to be extremely important: The code and method designed in this project were conceived not only to meet the goals of this project, but also thinking on possible future studies based on similar sets of data, the scope of which lies beynd this project.

The data used for the study had been previously provided by Translink¹ to the Transport Engineering department of the Univerity of Queensland. This data was limited to two days of passenger records at the beginning stages of the process, so it was not possible to perform any deep analysis on this data, but only to check if the software worked well and was capable to extract the performance indicators it was designed for. Later on, Translink provided UQ with 6 months of Go Card data and the software could be then tested in a more demanding set of data and some results could be extracted and analysed.

¹Translink is the transit authority operating in the south-east of Queensland, including mainly the city of Brisbane and its greater area, but also to the north until the Sunshine Coast and has the lower limit in Coolangatta, in the Gold Coast.

3.1. The smart card

Smart Cards are being used throughout the world to collect fares for urban public transport systems. This chapter aims to introduce what a smart card is and why it can be used for. Also, it intends to give an answer to why smart card systems have been implanted in public transport, and what advantages and disadvantages the use of smart card ticketing technologies by a transit authority involve.

As an introduction, a smart card is basically a piece of plastic that has the size and dimensions of a credit-card, being very portable and durable (Lu, 2007). It has a microchip that performs all the necessary functions it was conceived for, mainly storing, processing and writing data. This capacity for the card for transferring data makes it very useful for a number of possibilities.

Smart cards can be used in a wide range of different sectors, but in this project only their use in public transport systems as a method of collecting fares is considered. Smart cards claim to revolutionize public transport users experience, and how transit systems operate. Smart cards are the future of high quality ‘integrated ticketing’ (Department of the Environment, Transport and the Regions, 1998). Smart card technology is not new, it was invented more than 30 years ago. However, at the beginning it was introduced to the market based on a push strategy instead of a pull strategy, which created several failures that discouraged the transport industry from investing in this kind of technology during the 90s. (Barry et al., 2002). However, over the time smart card ended taking off after the introduction of the Octopus card in Hong Kong and it was then in the 2000s when smart cards were increasingly implanted around the world.
3.2. Comparison of fare payment methods

The way in what fares are collected is one of the most important aspects in determining customers’ satisfaction in public transport. This section intends give arguments for or against the use of smart cards and try to discern, by comparing smart card use to other forms of payment for public transit (e.g. cash, prepayment tickets, monthly passes, etc.) whether they seem to increase the standard of performance, especially regarding passengers experience. This increase in the performance is mainly measured or analysed in two ways: customers’ satisfaction, and efficiency improvement respect traditional methods.

Conventional ways of fare collecting can be insecure and cause delays in boarding. This one of the main reasons why transit agencies began to show interest in smart card technologies, and many of them are now using it to replace the traditional magnetic cards or tickets as a viable payment option (Blythe P., 2004). Smart cards are perceived as a secure method of user validation and fare payment and offer a new ways to improve the collection and payment of fares (Blythe P., 2004). It also makes driver’s job easier, as he or she no longer has to collect the fare. Furthermore, the smart card improves the quality of passenger service.
of the data, gives transit a more modern look, and provides new opportunities for fare structuring (Dempsey, 2008).

Figure 2 shows a customer survey of a smart card type used in Finland to determine the opinion of the public towards this way of payment instad of the traditional ones. Characteristics of smart card ticketing systems are detailed below, based on Pelletier, Trépanier, & Morency in 2009, and using the criteria presented by Vuchic in 2005 for the evaluation of fare collection systems:

- **User friendly.** The smart card is a method for permanent fare payment which can be used during a number of years. This makes it a better system compared to other kind of cards, like magnetic cards, because they are not as durable, and of course regular tickets. Also, the smart card is contactless and does not require the user to insert the card in a reader, as it happens in magnetic cards (Blythe, 1988).

- **Bus delay.** Interactions between a smart card and a reader machine are relatively fast on boarding a bus. Even though in this case smart cards are not the best choice: the fastest way would be to show a pass (weekly, monthly…) to the driver. This is not usually an inconvenient in trains, where the readers are located at some point in the stations, instead of being installed inside the train.
• Easy payment tracking. Compared to magnetic cards and other forms of automated payment, smart card transactions are easy to treat to produce accurate financial reports for the transit authority.

• Cost of the equipment. The smart card requires costly equipment aboard the vehicle or at the stations. This can be considered as one of the major disadvantages of the system, but this equipment could be made more profitable by adding other functions, such as cash-counting devices, automated route display, and driver-related management roles.

• Security. Like magnetic cards, the smart card reduces fraud because it validates the right to travel each time a user boards a vehicle or enters a station. The correct reporting of tickets and cash payments is time-consuming and demands a large number of staff.

• Ability to use different fare structures and types. Smart cards can support different fare types at the same time, and the system can validate boarding and prioritize fares. Also, the fare structure can be easily modified by reprogramming the reading services. Complex fare structures with multiple zones are difficult to implement with traditional ticketing and monthly passes, because sometimes the system must validate at both entry and exit, which is incompatible with the drivers’ role.

Following there is Table 1 which is used in Pelletier, Trépanier, & Morency (2009) study in the conclusions section and contains advantages and disadvantages for general use of a smart card (collecting fares), but also regarding to analysis performance.
3.2.1. Technologies

A general picture is drawn here regarding technology issues involved to smart cards, as well as the standards used. Mainly, the difference between contact-based smart cards and contactless ones. Since the invention of the card in the 70s, the technology has significantly evolved and many features have been added to the first concept.

In the contact smart card, a chip is introduced in the plastic surface, not covered because it has to be in contact with the machine that reads it. Whereas in the contactless smart card the chip can be completely be inside the card because there is no need for contact so the chip can be protected from possible external damages. Contactless cards have also usually a small antenna installed inside to amplify the signal received (Pelletier, Trépanier, & Morency, 2009).

Table 1: Advantages and disadvantages of the use of a smart card. (Pelletier, Trépanier, & Morency, 2009)

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Minimizes user role in data collection previously obtained by survey</td>
<td>- Cannot provide information on trip purpose, or on user assessment of service</td>
</tr>
<tr>
<td>- Trip data combined with personal data improves data quality and increases</td>
<td>- Does not provide user’s ultimate destination, despite a method to deduce it</td>
</tr>
<tr>
<td>amount of statistics (Bagchi &amp; White, 2005).</td>
<td>- High research and development cost (Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>- Easier to examine travel behaviors by “reconstructing” user trips than</td>
<td>- Complexity of introducing new components and processes into systems</td>
</tr>
<tr>
<td>by studying existing data (Bagchi &amp; White, 2005).</td>
<td>(Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>- Improves feasibility and convenience of a variety of pricing options for</td>
<td>- High implementation cost (Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>road use, parking, and transit fares (Deakin &amp; Kim, 2001).</td>
<td>- High risk associated with investment (Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>- Provides universal payment method for a number of systems (Deakin &amp;</td>
<td>- Slow institutional change (Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>Kim, 2001), integration with other transportation fare collection</td>
<td>- Slow social acceptance (Deakin &amp; Kim, 2001).</td>
</tr>
<tr>
<td>activities (Cunningham, 1993).</td>
<td>- Need for service providers to undertake surveys to confirm analysis of use</td>
</tr>
<tr>
<td>- Reduces cost (McDonald, 2000).</td>
<td>and assumptions made (Bagchi &amp; White, 2005).</td>
</tr>
<tr>
<td>- Improves service (McDonald, 2000).</td>
<td>- No guarantee of profitability improvement (McDonald, 2000).</td>
</tr>
<tr>
<td>- Implementation of flexible and creative fare policies (McDonald, 2000).</td>
<td>- Success of implementation often depends on users (McDonald, 2000).</td>
</tr>
<tr>
<td>- Improvement in revenue management (McDonald, 2000).</td>
<td>- Market penetration needs to be sufficient to provide a representative</td>
</tr>
<tr>
<td>- Convenience and system utilization time improved for users (Bagchi &amp;</td>
<td>sample of the entire population (Utsunomiya, Attanucci &amp; Wilson, 2006).</td>
</tr>
<tr>
<td>White, 2005).</td>
<td>- The more complex the card, the less its reliability is guaranteed (Byrne,</td>
</tr>
<tr>
<td>- Improves user perception of public transport by using new technology</td>
<td>2004).</td>
</tr>
<tr>
<td>(Ibrahim, 2005).</td>
<td>- Potential for information-sharing offers innovative revenue-earning</td>
</tr>
<tr>
<td>- Improves payment mechanism and information flow (Byrne, 2004).</td>
<td>possibilities (Byrne, 2004).</td>
</tr>
<tr>
<td>- Smart card life span longer than that of traditional cards (Utsunomiya,</td>
<td>- Reduces user boarding time (Chira-Chavala &amp; Colman, 1996).</td>
</tr>
</tbody>
</table>
| - Reduces cost (McDonald, 2000).                                           | - Cannot provide information on trip purpose, or on user assessment of service (Bagchi & White, 2005).
3.2.1.1. Standards

Contact-based smart cards are usually covered by ISO/IEC7816, that defines the contact plate layout and usage, the electrical interfaces and the selection of applications (Hendry, 2007). For contactless smart card the standards are more diversified. In the following figure there is a summary of those standards and its characteristics. As international standardizaton is not widely accepted, and consequently several standards have been developed.

Table 2: Characteristics of some lower level contactless smart card standards, (Trépanier, Barj, Dufour, & Poilpré, 2004), based on (McDonald, 2003)

<table>
<thead>
<tr>
<th>Technology</th>
<th>Frequency (MHz)</th>
<th>Data transmission speed (kbits)</th>
<th>Activation distance</th>
<th>System</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO/IEC14443 (Type A or Bi)</td>
<td>13.56</td>
<td>106</td>
<td>10 cm</td>
<td>Open or closed</td>
<td>Transport, off-line purchasing, vending, and physical access control.</td>
</tr>
<tr>
<td>ISO/IEC15693 Vicinity card</td>
<td>13.56</td>
<td>26</td>
<td>until 1 m</td>
<td>Closed</td>
<td>Physical access control, ticketing, parking, and drive-thrus.</td>
</tr>
<tr>
<td>Felica</td>
<td>13.56</td>
<td>212</td>
<td>n/a</td>
<td>n/a</td>
<td>Transport, identification, and others</td>
</tr>
<tr>
<td>NFC (Near Field Communication) ISO/IEC18092 EX-PASS Proprietary Ultra-High-Frequency Technology</td>
<td>13.56</td>
<td>212</td>
<td>until 20 cm</td>
<td>Open</td>
<td>Payment.</td>
</tr>
<tr>
<td></td>
<td>902, 928 &amp; 5900</td>
<td>n/a</td>
<td>3 to 10 m</td>
<td>Closed</td>
<td>Highway toll booths and fast-food drive-thrus</td>
</tr>
</tbody>
</table>

3.2.2. Privacy concerns

This section is important for smart card users, furthermore it can set the willingness of the population towards acceptance of usage of smart card ticketing systems. There are some security concerns about smart cards if they are used for operations that user’s privacy or property could be violated. Example of that are the uses of smart cards for economic transactions or personal data. The use of smart cards in transport involves personal data use and storage, as well as the payment of fares. So smart card technology needs to possess a serial of characteristics to be considered as a viable way to collect the fares securely. For instance, the card has to be safe enough to ensure that the data cannot be modified or copied by an unauthorised person, either at the time to validate the ticket or by just being able to generate a copy of the card when possessing the original card. If security concerns are met, then it makes the card usable more than once, and it can hold
details of the card holder and also be used for other applications depending on its purposes (Blythe P., 2004)

It is sensible to think that a card that contains passenger personal information or money is in risk to be stolen or intercepted by third parties. But the concerns related to smart card usage are about the same as those for credit cards, cell phone communication and other tracking technologies (Clarke, 2001), and nowadays the use of these devices are widespread. This following table contains the main concerns passengers could have about smart cards, such as how can the smart card be linked to the personal data of the passenger\(^2\), and also who could have access to that information. In the ideal case only authorities should have access to the data, only being authorised to use it under special circumstances.

<table>
<thead>
<tr>
<th>Concerns</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of information gathered</td>
<td>The information can be linked to prepayment functionality, financial and trip data, personal information (passenger name, home and work address, age, and gender), biometric identification, and information on possible criminal activity of passengers.</td>
</tr>
<tr>
<td>Potential uses of the information gathered</td>
<td>The smart card can be used solely for fare payment, planning, and advertising purposes and to monitor the personal behavior of individuals in the network.</td>
</tr>
<tr>
<td>Who has access to the information?</td>
<td>Should the information be available only to authorized personnel of the transit agency, shared with other institutions, or be broadcast to the general public? There is a need to implement a strict system for accessing data (Schwartz, 2004).</td>
</tr>
<tr>
<td>Implications for personal privacy</td>
<td>Advances in technology have outpaced personal privacy law (Archer 2005).</td>
</tr>
<tr>
<td>Uses for the smart card</td>
<td>If the smart card is used for multiple purposes, there is risk of creating a central database that gathers all personal information.</td>
</tr>
</tbody>
</table>

\(^2\) Even if the smart card does not have any personal data payments using a credit card could be tracked and the user identified in some way.
3.3. The Go Card in Brisbane

The case of study in this project is the city of Brisbane and its public transport, centering the attention in Brisbane’s and South East of Queensland’s smart card, the Go Card. The transport operator in the city, Translink, introduced this smart card ticketing system, starting the project in July 2003. In July 2006, Go Card test stage started by Translink signing up around 1,000 volunteers to test the system in the Redcliffe Area. Finally, the Go Card was launched in Brisbane in February 2008. In 2003 Queensland government awarded the $134 million contract to design, build, operate and maintain the Go Card system to Cubic Corporation.

The Go Card is mainly centered in the Brisbane area and suburbs, but being extended to the south until the Gold Coast and to the north until the Sunshine Coast. In order to use the Go Card, the users must hold the card less than 10 cm away from the reader to "touch on" before starting a journey, and then they must do the same to touch off the service at the end of the journey. After touching off, the cost of each journey is subtracted from the Go Card balance. Go Cards can be topped up or recharged in machines located at the main stations, in specific shops, and through the internet.
To promote the use of the Go Card over traditional fare collection methods, Translink provides Go Card users with several advantages. Go Card fares are always cheaper than single paper tickets, but the amount of saving varies along the day. There is a period in the day considered off-peak where all trips are 20% cheaper compared to the peak period, while in peak hours the difference decreases. Also, students and some collectives are eligible to use a Concession Go Card, which benefits in a 50% reduction in all fares. Finally, all Go Card users who pay for 9 trips during the same week with a Go Card are free to travel without paying until the week ends, no matter the trip or destination (excepting some special cases like going to the airport from the city or vice versa).

3.4. Introduction and comparison with other smart card systems

Other cities also use smart card for collecting the fares in their metropolitan areas. In this section some of the most relevant ones are reviewed and summarised, by introducing the Octopus and Upass card (first smart cards implemented in the world) and some comparison with other Australian cities that was performed empirically. The idea is to give some background to compare smart card systems and not to keep only Brisbane's Go Card functionality in mind.

3.4.1. Octopus in Hong Kong

Seoul and Hong Kong's smart cards, named Upass and Octopus were pioneers in the world at implementing smart card fare collecting systems in their respective cities. The first one was the Upass, with a first version of it used in June 1996, and it is issued by Seoul metropolitan bus operator Association and eB Card. The Octopus card, by its side, was launched in September 1997. The Octopus card is recognised internationally, winning the Chairman's Award of the World Information Technology and Services Alliance's 2006 Global IT Excellence Award for being the world's leading complex automatic fare collection and contactless smartcard payment system, and for its innovative use of technologies.
3.4.2. Sydney and Melbourne

The two biggest cities by population in Australia, Sydney and Melbourne, also use smart card fare payment systems as the one existing in Brisbane. The way the different smart card systems work are not exactly the same and present some differences amongst them. The most important ones are discussed here.

The first and most notorious one, is the area covered by the integrated fare collection zone. While the Go Card covers a zone from the Sunshine Coast to the Gold Coast (more than 200 km in straight line) and it is divided up into 23 zones, in Melbourne (with its smart card called Myki), the city is divided only in two zones\(^3\). There is a different fare that applies whether travelling inside the same zone or from a zone to the other. Even though in Melbourne passengers are supposed to touch on and off as in Brisbane, if travelling inside the same zone passengers are allowed to not touch off if they do not want to, and of course no fine is applied for not doing it\(^4\). This is in some way logical from the passenger’s point of view, because a lot of passengers travel inside zone 1 every day and that makes the process of alighting and boarding smoother, especially in rush or peak hours. Furthermore, zone 1 is fairly big (much more extensive than zone 1 in Brisbane) and the fare for travelling between stations in zone 1 is the same no matter the length or duration of the trip. The inconvenient of this procedure is that no alighting passengers data is kept on the transit authority database, which makes it not possible studies like the present one. Also, in systems like the one presented in Melbourne, an origin-destination (OD) algorithm (see section 4.3.4) needs to be

\(^3\) Zone 1: Melbourne city center and some inner suburbs, zone 2: outer suburbs of metropolitan Melbourne.

\(^4\) In Brisbane there is a fine applied to the user of the Go Card account if a user fails or decides not to touch off when alighting a vehicle. The amount varies depending on the transport mode and destination.
undertook first to estimate the OD matrix of the system before proceeding to possible studies using smart card data.

Sydney on the other hand, covers a big zone, in a similar way as Brisbane’s Go Card, comprising the city of Sydney and suburbs, as well as the following surroundings: the Blue Mountains, Central Coast, Hunter, Illawarra and Southern Highlands. Sydney’s smart card, Opal card, also allows the user to travel with different transport modes such as bus, train and ferry, and calculates the fare according to the distance travelled, according to certain predefined patterns. It is a method that works in a similar way to Brisbane’s, and also offers advantages to regular users of smart cards to promote opal card usage.

Also, another difference that characterizes Brisbane’s go card is the fact that it is refundable. When the user wants to start using a Go Card, pays a deposit of $5 which is returned when the user decides to stop using it. A Go card can only be returned if its balance is positive or equal to zero. The balance of money in the Go card account associated is returned too. In Melbourne, for example, the deposit for the Myki card is not refundable, but it is the balance in case it were positive.
4. Background on Smart Card Studies

In general terms, transit performance refers to how well a transit system achieves its intended goals and objectives. And measuring performance is an important task, because it provides information on how well the service is provided, while also helping to improve it given the case, by the interpretation of the indicators and performance metrics (Carrasco, 2012).

Currently, with the existence of new advances in technology some of the previous existing limitations to estimate transit performance have been overcome. Some examples of this are Automatic Vehicle Location (AVL) systems, that can track the position of the studied vehicle at any time, automatic passenger counting, and automatic fare collection systems. AVL systems often use a Global Positioning System (GPS) device located in the vehicle, which is a accurate method for obtaining its speed and position (as well as other aspects) but have the inconvenient that require an economic investment. This study pretends to offer a similar solution based on a similar concept of being able to locate (or approximate) the position of a vehicle at any moment, but parting from the point of an existing smart card ticketing system, thus using implemented systems instead of new ones. In the first part of this study, a general view of smart card ticketing systems has been shown. The intention is to expose what advantages can it have and why a transit authority could consider the option of implementing it, despite it is more expensive to deploy than traditional fare collecting methods. After that, and taking the example of the city of Brisbane (that already has the system working since beginnings of 2008) the purpose of the current chapter is to show some studies regarding smart card data. In further sections of the study, another method to estimate some aspects of transit performance is proposed. The main advantage is that the system is already existing and implemented, so opposite to other ways to get performance data, this method would not require additional economical investments more than the one done to launch the smart card system at the beginning of its conception.

Generally speaking, several types of analyses can be done with smart card data, including estimating origin-destination matrices, measuring passengers behavioral reactions to service changes, and evaluating service quality.
4.1. Travel time

In this section some indicators about travel time analysis are reviewed and presented. Even though in transit systems it can be difficult to define, reliability is usually measured by its consequences, such as the time lost, number of breakdowns, etc. Travel time reliability could be defined as consistency or dependability in travel times, measured from day to day for the same trip (Carrasco, 2012). Thus providing a reliable service is somehow related to keep buses on schedule and minimising time variability as much as possible.

An accurate travel time estimation can be useful to reduce transport costs, because it can lead to a better knowledge of the route characteristics and therefore it allows to find solutions, like for instance, avoiding congested sections (Lin, Zito, & A.P. Taylor, 2005). Travel times play a very important role nowadays and it is important to know what factors can have influence in its variation. Human, vehicles and facilities, as well as other factors that could have influence in some of these three elements, are the main components of the traffic environment (Lin, Zito, & A.P. Taylor, 2005). Different drivers, road conditions or vehicles, even in same time intervals or similar routes, can have quite different travel times (Li & McDonald, 2002). Other factors cited in previous studies that have influence on travel time include holiday and special incidents, signal delay, weather conditions, traffic operation and congested level.

Travel time variability has become more important as a performance measure because it reflects the reliability of a transportation system (Turochy & Smith, 2002). Some studies have found reliability of a transport system to be equal or more important than other aspects (Lam & Small, 2001). For example, a reduction in variability has been found to be at least as valuable as a reduction of the mean travel time (Bates, Polka, Jones, & Cook, 2001). Nevertheless, the amount of research into travel time reliability is quite small compared to the research and studies analysing travel times and measure behavioural reactions to the changes in the travel time reliability (Li R., 2006). Li’s thesis also states that the reason could be the limited availability of measured travel time data. This problem could be solved in the case of Brisbane, and the data used for this present study, because potentially all bus travel times could be extracted from the smart card data.
4.2. Passenger demand and trip planning

The demand of a bus service is an important aspect to consider, and it can be approached in many different manners. For each bus and stop the number of boarding and alighting passengers, to know how many passengers are in the bus at every moment. This can be used to know in what stops the passenger demand is higher and the patterns for every stop. Also to know if at any moment the passenger limit inside the bus is exceeded.

For example, using the example of Brisbane, a passenger who travelled every day from a determined residential suburb to UQ at the same time, could be assumed as a university employer, that has every day the same schedule to respect. While another passenger who went to the same suburb to UQ using a different day to day pattern could be a student, as lectures and tutorials do not have a fixed schedule every day of the week. Whereas another passenger who only went to UQ occasionally (perhaps to visit a friend or colleague) could be expected to be a leisure trip. At this point it is important to remark that this kind of study has not been the purpose of this project, trip and travel patterns, mainly due to the extension of a study of this characteristics. But it is mentioned in this section because this is a task that could be performed using a set of data similar to the one used for this project. The software designed and presented in further chapters of this project was conceived to extract data from passengers using smart cards in order to meet, between other options, the possibility to undertake these kind of studies.

The demand of the service consists in summing up the total number of passengers that board a bus service in all the stops. This concept offers an idea of the magnitude of a line, and to see how many people affect any change in this line. It is not the same a hypothetical bus route of 10 stops where, according to the characteristics of the passengers was used by 10 people from the beginning to end, to another line of 10 stops where 10 passengers boarded the bus in each stop and alighted it in the next stop. While the first bus would have a line demand of 10 passengers, the second bus would have a line demand of 100, having at any time both 10 passengers simultaneously inside the bus. While in the first example a possible inoperation of the bus line, totally or partially, would only affect to 10 passengers, in the second example could get to affect up to 100 customers. That is the reason why it is considered to be a important indicator of bus line 'importance', and also was decided to be included as one of the outputs of the smart card data extraction software.
4.3. Smart card studies

Transport planning purposes have not been treated directly in this project. However, at the time to extract smart card information they were considered to be a possible continuation of a project like the ones presented in this thesis. Thus, the design of this project kept in mind that the output data that it could be obtained had to be usable also for these type of studies. In addition, there are several kind of studies made using smart card data performed with transport planning purposes.

Studies using smart card data for transit planning purposes can be grouped in three different categories (Pelletier, Trépanier, & Morency, 2009). This section is based on the study undertook by Pelletier, Trépanier and Morency (2009), mainly in the way to classify studies based on smart card data. Studies related to long-term network planning, customer behaviour analysis and demand forecasting are clustered into strategic-level studies. Next proposed group are the tactical-level studies, which gather all studies related to schedule adjustment, longitudinal, and individual trip patterns. Not all of these studies are related directly to what is studied in this project, thus only the ones that are relevant are going to be analysed and discussed here in detail. And the third group is the one comprising operational-level studies, related to supply and demand indicators, and smart card system operations.

4.3.1. Strategic-level studies

In strategic-level studies the studies are centred basically in long-term planning. Most of the research is centred in user characterization and classification\(^5\). The use of smart card data over traditional fare-collecting systems is a great advantage in this strategic-level kind of studies. Because with the use of smart card data ticketing systems a large amount of data is

\(^5\) That does not imply the knowledge and usage of users' personal information. It is related to characterization and classification of user profiles.
collected, and that makes possible a better understanding of travel behaviour, because there are more observations than with any other mean of data collection (Agard, Morency, & Trépanier, 2006), (Bagchi & White, 2004). Also, for every smart card transaction several characteristics of the trip (such as the date, time, card number, service type, direction, boarding/alighting station etc.) are available to be studied, which makes it easier to calculate passenger statistics in a quite precise temporal and spatial scale. However, there is usually no information related to the passenger stored in the card, and if it is, it has to be treated with caution, not to vulnerate the legal rights of the people. So, as a result, the available data to study often lacks sociodemographic attributes, and studies may be complemented and enriched with other more traditional methods, for instance a household travel survey (Pelletier, Trépanier, & Morency, 2009)

4.3.2. Tactical-level studies

The tactical-level studies are not going to be approached in this project, mainly because the data available for the study had different characteristics and did not need to. This tactical-level approach is more oriented to systems that do not have certain information about the journeys and some methods for predicting or estimate this lack of data are necessary. For example, in a system where there was no alighting station passenger data, an algorithm to estimate the most probable alighting point (by using some approximation methods such as looking at the passengers next boarding point or by applying similarities to other trips made with the same card in the historical database) was performed (Trépanier, Chapleau, & Tranchant, 2007). Other remarkable tasks that can be done could include:

- Predicting the boarding stop (in case it is not known) by looking at the time of boarding compared to the bus schedule
- Analysing the boarding passengers variability on a determined route

Once algorithms have solved the problem of lacking data, and boarding and alighting point are available, further studies can be undertook on the data. Amongst other, some remarkable ones could be:

- Route load profile
- Detection of maximum boarding point and return runs
- Schedule coordination between bus and metro
4.3.3. Operational-level studies

On the third group, the operational-level studies, the goal is mainly to calculate precise performance indicators on a transit network (Pelletier, Trépanier, & Morency, 2009). Examples of this could be:

- Schedule adherence
- Vehicle-kilometers and person-kilometers for each individual run/route/day

Schedule adherence can be estimated by comparing all the boarding times at every stop with the theoretical route schedule. Vehicle-kilometers can be calculated by seeking the number of vehicles running at each time (for example knowing the real frequency of the service), and person-kilometers can be estimated by calculating the net amount of people being in the bus at each moment.

Also in this third group can be included studies that have used smart cards to detect irregularities or errors in the smart card payment itself. The simple presence and detection of errors in the data can quickly lead to identify the type of error committed. The most common ones are employee errors, fraud or detective equipment error, being the most common one the desynchronization between the onboard smart card reader and the planned routes.

4.3.4. Origin and Destination inferring (O-D) and O-D matrix estimation

Several studies based on smart card data have been undertaken related to origin-destination matrixes. The purpose of this project was not related to this topic, mainly because for the structure of the data available it was not necessary. Usually these kinds of methods or algorithms are applied to systems where there is some part of the data related to passenger trips missing or not available. And then the objective is to estimate the origin and destination points or the O-D matrix of the system following a concrete procedure or method. With Brisbane’s Translink data, the real O-D matrix could be obtained because there is information regarding boardings and alightings. In this section a quick review of different methods for estimating O-D matrixes using smart card data are summarised.

One way to do it is by applying a trip-chaining method to infer bus passengers’ boarding and alighting locations, by linking journey stages and form complete journeys comparing
interchange times and the connecting bus route’s headway (Wang, 2010). For this study both boarding and alighting locations for passengers had to be estimated. The first step in this process consisted in inferring origins for bus passengers by matching smart card boarding transactions times with Automatic Vehicle Location (AVL) data. Then, the trip chaining methodology was used to infer the alighting location. Finally, results were compared with a origin and demand survey to validate the results.

Another option for undertaking a similar procedure consists in identifying each smart card separately to determine the sequence of trips made on each day, and see what patterns do the smart cards follow (Barry et al., 2002). This particular study started by sorting the smart cards data by their serial number and then extracting for each one the sequence of trips and stations used at the beginning of the trip. Some assumptions had to be made in this study to be able to complete the methodology. For example, the assumption that a high percentage of passengers end the day in the same stop they began their first trip of the day.

For estimating or inferring an O-D matrix, the process can start from a number of different passenger data collections. According to Cui’s study (in 2006), Automated Fare Collection (AFC) related to smart cards used in Chicago from Chicago Transit Authority (CTA), and Automatic Passenger Count (APC) were used to get data from passengers trips (Cui, 2006). Relating to the vehicle position, also Automatic Vehicle Location (AVL) data was used. The algorithm consisted in three steps: data preparation, estimation of single route O-D matrices for all the routes, and estimation a network level O-D matrix using transfer flow information.
5. Creating a Method to Analyse Go Card Data

5.1. Introduction: Brisbane City

Brisbane is a city located in the east coast of Australia. It is the capital of the state of Queensland being the third most populous city in Australia after Sydney and Melbourne. It has a population of 2.2 million people spread in its area of 5,950 km². This big area makes public transport system design a very important issue in order to ensure optimal service coverage for most of the population and suburbs. To have an idea of the size, the city of Barcelona covers an area of 102 km² whilst the whole Barcelona Greater Area covers 803 km².

It is important how is how the public transport is structured in the city of Brisbane. The city itself does not have a subway network, as other cities around the world have. To overcome this, it has to have other methods of providing the necessary and reliable public transport...
service to the users. The transport modes currently existing in the city of Brisbane are the bus, the train and the ferry (for mobility along/across the river). As commented before, there is a smart card fare collection system named the Go Card. All three means of transport are integrated within the Go Card fare zone. The region covered is divided by different sections or zones that cover a great extension, in fact between the Sunshine Coast and the Gold Coast (upper and lower limits of the coverage of the Go Card system) there are almost 200km by road, being Brisbane city located in between both boundary places. The fare is calculated considering the boarding and the alighting stations of each passenger, being the trip more expensive as more zones are travelled between these beginning and end stations. There are also available single tickets in the train stations and inside the bus, designed for people that do not possess a Go Card card, and there are also 3 and 5 day passes for visitors or tourists. Translink zones in the city of Brisbane cover from zone 1 in the CBD to zone 6 at the further suburbs of the city. In Coolangatta (southernmost point in the Gold Coast), the zone arrives until the number 18, and in Noosa (northernmost point in the Sunshine Coast) the zone gets up to 23.

As the city covers a big area and there is no subway system, a substitute was created to make public transport trips more efficient. The city has two busway routes, the south-east busway and the eastern busway, that are basically bus lanes completely closed to regular traffic, and allow buses to cover quite big distances in relatively small amounts of time. When busway enters the CBD the busway blends with the streets of the city and buses circulate through the roads mixed with the rest of the regular traffic.

---

CBD (a.k.a. Downtown) stands for Central Business District and it refers to the center of the city, often the commercial and geographic core.
5.2. The study

The research used Go Card data for two bus services, 109 and 412, that connect the city of Brisbane with The University of Queensland through different routes. While bus 412 is mixed with traffic, bus 109 uses the busway. These services were initially selected to compare the difference in time of both layouts, that use different types of ways to connect the same points. The standard deviation of both travel times can be an important issue to see too, due to traffic flow variations that should affect mixed with traffic sections more than the busway ones.

Figure 5: Bus route layout for 109 and 412 bus services.

In the previous figure it can be seen the map of a part of the city of Brisbane, showing the route followed by both of the bus services. While most of the route followed by 109 bus service takes place in the busway connecting the city and UQ in a fastest way (passing through South Brisbane, West End, Highgate Hill, Woolloongabba and Dutton Park suburbs), the bus service 412 uses the regular open-to-traffic road to link the city and UQ following the
river through its north bank, providing service to all the passengers who live in the suburbs of Milton, Auchenflower, Toowong and St. Lucia.

The following figures below show zoomed parts of the city where the most characteristic stops for this study are located. Figure 6 shows the city CBD with Adelaide Street Stop 16 (start/end stop for both 109 and 412 bus routes) and Cultural Centre Busway Station, a stop only for bus service 109 that sets the boundary between the busway section and the mixed with traffic one. Figure 7 indeed shows UQ St. Lucia campus, with the two bus stops that the campus has, UQ Lakes Busway Station and UQ Chancellors Place. In UQ Lakes arrive and depart all buses that travel via busway, including 109 service, while Chancellors Pl. is the final stop for all buses not travelling through the busway, like 412 bus service.

![Figure 6: Brisbane CBD with Cultural Centre and Adelaide Street Stops.](image-url)
5.3. Data used for the study

5.3.1. Matlab ® introduction

The main software used in this study was Matlab ® because of its performance at the time to perform calculations. Its basic data element is the matrix (in Matlab ®, even a simple integer is considered as a matrix of one row and one column), which allows to undertake a number of different operations. Matlab ® is a very powerful programming environment that allows to create programs and functions to perform repetitive tasks. This feature made it an excellent choice for the purposes of this project.

The main data structure used in this project to store and restructure all the smart card data in Matlab ® was the cell array, mainly in order to be able to process all types of data (numeric,
text, etc.) in the same variable. The best way to clarify what a cell array is to use the available definition by the help section in the Mathworks - Matlab® database:

“A cell array is a data type with indexed data containers called cells. Each cell can contain any type of data. Cell arrays commonly contain lists of text strings, combinations of text and numbers from spreadsheets or text files, or numeric arrays of different sizes.

There are two ways to refer to the elements of a cell array. Enclose indices in smooth parentheses, ( ), to refer to sets of cells — for example, to define a subset of the array. Enclose indices in curly braces, { }, to refer to the text, numbers, or other data within individual cells."

In this project the notation using curly braces\(^7\) \{{\}} is often used to refer to data in concrete positions of cells.

5.3.2. Initial considerations

This study is centered in using smart card data for monitoring performance purposes. Smart Card data was requested to Translink, the transit operator in Brisbane in order to undertake several studies in the Transport Department in UQ, such as this one. The available data during the period this project was being performed was enough for the objectives of this project, even though at the beginning only two days of data were available, which were used to build and test the program; while later on 6 months of data were received and adjustments could made to improve the roughest parts of the programmed code.

\(^7\) The creation and modification of cell arrays is one of the main points of the process. In Matlab®, the notation \{i,j\} is used to specify a position of a cell array, being \(i\) the row and \(j\) the column indicators. This notation is used throughout different sections in this project every time a position in a cell needs to be located/referred.
One of the goals of the project was not just the obtained output results, but to give an idea of some analyses that could be done with Go Card data from passenger's trips. The reliability of the results obtained in this project were not the main issue of concern, but it was to see that the studies undertaken in this project could be done and how could this be helpful to monitor and improve the reliability of the public transport. Being aware of the limitations of this method is also very important, as well as, under the framework of being affected by any possible limitation, try to discern how important it is in affectation of the result obtained.

5.3.3. Approximations

According to the characteristic of smart card and therefore the data available, some approximations had to be made at the time to calculate travel times with smart card data.

In Figure 9 there is a scheme of the different stages of a bus when arriving into a stop, waiting until all processes are finished (waiting at the stop, opening and closing doors, passengers boarding and alighting, etc.) and leaving the stop to reach the following one. The dwell time is defined as “the time in seconds that a transit vehicle is stopped for the purpose of serving passengers. It includes the total passenger service time plus the time needed to open and close the doors” (Highway Capacity Manual (HCM), 1985). The running time is the time between the bus departs from a stop and arrives to the following one.

![Figure 8: Buses lining at Cultural Centre Busway Station on a Friday at 5:15PM (own capture)](image-url)
As the only data available and considered in this study is smart card data, those times cannot be exactly known, but they can be approximated. For every bus in every stop there are several people boarding and alighting the bus, and this data is known and available. Thus from all times of passengers touching on and off available in a stop, the difference between the maximum and the minimum ones is considered to be the most similar to dwell time as it can be. This dwell time approximation is called "validation time", because it implies delay associated with passengers validating the ticket.

Stop delays, sometimes called idle times in this project, cannot be known by only considering smart card data with the layout and structure as the one available in this study. The reason why is similar to the case of the dwell time. Only passenger records are available in smart card data, so while the bus is in movement or somewhere in between two stops, there is no way to know in what situation is the bus, either travelling, stopped in a traffic light or traffic jam, or queuing to enter a bus stop if other buses arrived before. So the total travel time can be approximated by knowing the first passenger and the last one who board or alight the bus in the first and last stops of the route. And the difference between the total travel time and the sum of all the dwell times in all the stops would be the running times in the case the stop delays were zero. In this study, an approximation for the running time is performed, in which stop delays have been included in it due to the impossibility to split both by using only smart card data. It remains unknown whether that approximation can be a matter of concern. Nevertheless, in most stops, stop delays should not have to be a big issue, whilst in others can be troublesome (Figure 8). So, as a summary, in this project the travel time of a bus is divided into running time (time in which the bus is running, and contains a percentage of time belonging to stop delays) and validating time (best approximation found for the dwell time).
5.3.4. Flowchart of the process

The process that was undertook in this project to extract performance data from Go Card record is complex and consists of several steps that work one after the other in a chain process that leads to the final result. Figure 10 shows a simplified scheme of the methodology and steps in what the algorithm consists.

![Flowchart of the process](image)

**Figure 10:** Scheme of the data extracting process

5.4. Data processing

The data received from Translink could not be used in its original form, and needed to be processed in order to be suitable for the purposes of this project. The datasheet had a determined structure that needed to be modified. This step was important and was the one that has took more time to be completed.
5.4.1. Data cleaning and setting

The beginning point in this subsection was an Excel datasheet containing records of the trips of each passenger travelling in the bus services 109 and 412 for the days of data available. As showed in Figure 11, a sample of the initial raw data, each row represents a passenger trip, with several data related to this trip is stored in the columns. As not all of this data was necessary for the study, only the necessary columns of data were included in the further steps of data classification. The main modus operandi consisted in classifying the initial raw data according to a certain criterion with the aim of obtaining a datasheet that allowed having some results on monitoring the performance, as explained in more detail in section 5.4.1.2. As commented, only some necessary useful columns were to be used for this project after the cleaning stage. Those columns that were meaningful for the study are the following ones:

- Column 2: Date of the trip
- Column 3: Bus route (109 or 412)
- Column 4: Direction
- Column 5: Bus ID
- Columns 8 and 9: Boarding and Alighting date/time, respectively
- Columns 12 and 13: Boarding and Alighting stations, respectively

The structure of this data is the way it came from the transit authority in Brisbane, Translink, and it is initially not designed for monitoring performance purposes.
The first steps involved working on the raw excel sheet of data still without the need of disregarding the unuseful columns of data. Those columns that were not useful and were not included later on (chapter 5.4.2), in the stage involving the creation of new variables.

5.4.1.1. Cleaning

The first step was to clean the data, because not all the records for each passenger stored in the initial raw data were complete. Occasionally, some of the data were missing, and this could happen for several reasons. Usually those incorrect or missing records can be due to wrong lectures by the smart card reading machines in the buses or public transport stops. Another reason for a misleading record can happen when a passenger forgets to touch off when alighting the bus, so this bus trip record is incomplete and cannot be used in the project. Also, passengers using single paper tickets create incomplete records too, because the alighting record is not stored. The data at the end of this process looked like in Figure 11, with only the Go Card complete records included in the selection. In Figure 12 there is a sample of some of the incomplete records that needed to be removed at this stage.
As it can be seen, columns 9 and 13 (alighting time and station, respectively) are missing, so those records could not be included in the study. While records in rows 60-63 in this case belonged to regular adult Go Cards (column 10), rows 64-68 belong to single paper tickets purchased in the bus. It can be assumed that the records belonging to Go Card trips were not successfully recorded on the machine or the passengers forgot to touch off when alighting the bus.

After completing this step, the following stage consisted in classifying the data to start the process of extracting only the necessary information from the initial raw data. Just before starting with this process, all the dates and times in the datasheet were transformed into numbers using the Matlab® function `Datenum`, to make it easier to work and perform calculations with them. When the calculation process was finished, they were converted back into date/time format (using Matlab® `Datestr` function).

5.4.1.2. Classifying

Even though all stages were necessary for the correct consecution of the study, this stage of data classification, together with next section of joining the data, was one of the most important and time-consuming parts of all the program-building stage of this project. This is why the concept of classification is now explained. The classifying process takes the initial data structure (after the cleaning stage) and modifies it to create new datasheets with a new chosen structure, more suitable for the studies and analyses carried on later. It is a key process, not only for the procedures performed, but also because the way data was chosen and repositioned depended on what was decided to be studied. So the entire classification and processing stage was conceived to fit the purposes of the study. That means that all the steps in this stage were designed and oriented to a specific output.

![Figure 12: Sample of some incomplete data that needed to be removed.](image)
The concept of classification is simple, and it is made through a Matlab ® subprogram (named classify.m). That collects the data of a same kind together in a new datasheet (stored in a new variable, by subdividing it in groups according to a chosen criteria (generally a column of the cell). Usually the new variable is a cell containing more cells inside it, containing each one of them one of the groups resulting of the classification. An example of what the process does is shown in the figure below.

![Figure 13: Example of classifying a cell string.](image)

The example case is simple because data is classified only according to one criteria. When the data needs to be classified according more than one criteria, the classification becomes slightly more complex. This means that the classifying step needs to be undertook several times in a row to get the final result. In the case of this project, in the second and successive steps of the classification stage, an index variable needed to be created in order to scan the subordinate cells inside the main cell to reclassify the data again according to the next criteria.

### 5.4.2. Data restructuration

In this point, data was ready for being restructured according to the process shown in the flowchart (see chapter 5.3.4). It is important to remind that at the beginning of the process,
each row is a record for a passenger trip. By having in mind the goal of the data restructuration it is easier to understand the process. This objective is to obtain a list of data similar to the one used in the input, but having buses information instead of passengers' records in the rows. This is easier to understand by seeing an example of the final output data in Figure 23.

![Figure 14: Example of the classification process](image)

The main action undertook in this part is the classification of data under a specified criteria. Several different steps were required (combined with other calculations explained along the section) until the data took the desired structure. The different variables generated are named from CS1 to CS7 (CS stands for Classifying Step). Usually each new variable is generated after a classification step, hence the choice of the name.

Thus, the process started by splitting the data according to the bus service (109 or 412). So after this step variable CS1 was formed by two cells, containing each the entire records for each bus service studied. After this Classification step, the next two were similar, because they imply data classification too, but under different criteria. After the second step, data was classified by direction (inbound/outbound directions are split), and after the third step data was split by Bus ID. Also, as in this point the data was still passenger’s data, so CS3 variable was split into boarding (columns 8-12) and alighting (columns 9-13) records (stored respectively in CS3Boas and CS3Alis). So basically every passenger record was split into two. This task was performed because in the next step, stored in variable CS4, these
boarding and alighting records were joined in a different structure. Then, in CS4v2, the data was sorted in ascending order. In Figure 14 there is an example of the process of classification. First level of CS4 variable contains the groups of records for each day. The second level, which is inside the classification of every day, there is the inbound and outbound division. The third level contains all the buses in a concrete day and direction. Finally, the fourth level contains the actual data that matches the specified criteria.

At this point, the generated data was of all the trips each bus does in a day. Some buses travel several times in a day the same route, so the different trips needed to be identified and split in order to have each bus trip alone to be able to study it. But sometimes a bus can do several trips (for example) in a day in a same route, let us say for instance the 109, while other buses may do only one trip because they run in other services during their day (such as the 412, 169, 150, etc.). In other words, the number of trips a bus does a day in a given situation is unknown. For the purposes of the project, data of each bus trip needed to be alone. So the next section explains the process followed to split different bus trips performed by a same bus during the day, having in mind that this number is different in each case.

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>170</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.5829</td>
<td>1</td>
<td>University o. . .</td>
<td>'on'</td>
</tr>
<tr>
<td>171</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4773</td>
<td>1</td>
<td>St Lucie We. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>172</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4838</td>
<td>1</td>
<td>St Lucia Sou. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>173</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4821</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>174</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4820</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>175</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4821</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>176</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4819</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>177</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4821</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>178</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4821</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>179</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4822</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>180</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4785</td>
<td>1</td>
<td>St Lucie Fer. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>181</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4820</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>182</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4820</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
<tr>
<td>183</td>
<td>735334</td>
<td>412</td>
<td>Outbound</td>
<td>0.4821</td>
<td>1</td>
<td>University o. . .</td>
<td>'off'</td>
</tr>
</tbody>
</table>

Figure 15: CS4 variable after first classifying set of steps

5.4.2.1. Iterative method find the optimal number of clusters in each case

The previous chapter contains a set of steps necessary to clean and classify the data. The current chapter, in addition, intends to explain a method used to perform a slightly more
complex process of differentiating groups of data. In fact, for the realisation of this project several methods were considered and some of them were substantially easier than this one. The method chosen and shown is more universal, and pretends to lay the foundations for other similar studies to use a similar method. On the contrary, the use of a more simple method could imply the possibility of not being able to be used in other cases.

This chosen method is based on hierarchical clustering. In data mining (computational process based on discovering patterns in large data sets), hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. This hierarchy allows the data to be clustered in different groups depending on the reference value or criteria available.

The results of hierarchical clustering are usually presented in a dendrogram, which is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering (Wikipedia, 22). Once this tree diagram is drawn, it is easier to see the different groups in which data can be clustered. As it can be seen in the example in Figure 16, depending at the point in the Y axis in which the cut is done, the number of clusters set is different.

The functions that have been used are existing functions in Matlab® and are the following ones:

- The function cluster: This function was used to perform the task of grouping the data into smaller groups. This function allows several ways of usage, being the chosen one in this project the following one:

![Figure 16: Example of a dendrogram and different cluster possibilities](image)
a. \( T=\text{cluster}(Z, 'maxclust', n) \): constructs a maximum of \( n \) clusters using the 'distance' criterion. \textit{cluster} finds the smallest height at which a horizontal cut through the tree leaves \( n \) or fewer clusters.

- \textit{Pdist} vector: The vector of minimum distances was calculated to perform an iterative algorithm to calculate the minimum number of clusters. The vector \textit{pdist} is equivalent to the upper side of the paired distance matrix. Given a pairwise – distance matrix example, "\( M \)"; the way to obtain the vector "\( V \)" using \textit{pdist} function would be the following:

\[
M = \begin{pmatrix}
0 & 1 & 3 & 2 & 1 & 0 \\
1 & 0 & 4 & 1 & 0 & 1 \\
3 & 4 & 0 & 5 & 4 & 3 \\
2 & 1 & 5 & 0 & 1 & 2 \\
1 & 0 & 4 & 1 & 0 & 1 \\
0 & 1 & 3 & 2 & 1 & 0 \\
\end{pmatrix}
\]

\( \rightarrow \ V = (1, 3, 2, 1, 0, 4, 1, 0, 1, 5, 4, 3, 1, 2, 1) \)

**Figure 17:** Obtaining pairwise distances using \textit{pdist} function in Matlab ®

5.4.2.1.1 Setting the criterion to decide

Having introduced the theoretical basis and the elements and functions used in the Matlab ® to build this algorithm, its operation is explained below:

The initial data from which the process starts is like the one shown in the Figure 17 above. It is the result of the classification 5.4.2, where data is classified by Bus ID. Here, data is kept inside a cell variable (often inside other cells) where there is information regarding several stations of a same bus, direction and day. As stated before, the number of trips a bus does in a day in the same bus route is unknown. Sometimes just one trip is done in a day, while at other times the bus can do three or four trips. The objective is to perform a hierarchical cluster of each of these bus groups (by calculating the distances between all pairs of stations involved using function \textit{pdist}), and to create a dendrogram with the different possibilities of aggrupation. Then, a criterion is needed to choose one of those aggrupation as the correct one.

The division is automatically calculated by Matlab ® by giving it a boundary value from which consider whether the distance of two points can belong to a same bus trip or not (in...
other words, bigger than the time spent between two consecutive stops of the same trips and smaller than two stops belonging to different trips). This boundary value needs to be reliable and not commit the mistake of splitting data of the same bus trip or keeping different bus data in the same group. It is calculated to be the double of the mean of one trip (from the first stop to the last) plus a time spent on an end station.

For the study of this topic, there are several cases in which a bus could be involved. They have been classified into two groups:

- The less favorable case would be the one in which the same bus travels in one direction, when it reaches the destination returns in the opposite direction, and when back to the beginning station repeats the route (e.g. 109 Inbound – Outbound – Inbound, etc.).
- On the opposite, a case in which a bus reaches its destination station and then it changes the route (e.g. 109 Outbound bus reaches UQ Lakes and then becomes 169 Outbound) would not be an inconvenient because it would take much longer than case 1 to repeat the trip.

According to the worst of the cases presented above, the minimum time that would have to pass until a same bus could repeat the exact same trip is modelled by the equation 5.1:

\[ t_{\text{min}} = 2 \cdot t_t + 2 \cdot d_t \] (5.1)

The mean travel time for a trip of a 109 and 412 bus has been considered to be about 16 and 20min respectively\(^9\). Also, the idle time spent at an ending/beginning stop has been

\(^9\)Bulk approximation empirically calculated in order to have a first idea of the means of the bus routes and move forward in the process. After getting the results and having statistics, the choice is validated as correct.
rounded to 5 min. That makes a \( t_{min} \) time of 42 min for the 109 bus service and 50 min for 412.

5.4.2.1.2 Operation

Once all elements involving this method are introduced, the operation is quite simple. Every time a cell with an unknown number of buses is accessed, the first thing consists in getting the distances between all pairs of point, using \( pdist \) function as described before. The design of the algorithm was more complex than understanding it, because this \( pdist \) distances vector needs to be accessed and every value compared to the value \( t_{min} \). The tricky point is the following one: The correct number of clusters was found to be (using the example matrix in Figure 17) equal to the dimension of the matrix (6 in the example case) less the number of rows in which there is at least one zero over the main diagonal\(^{10} \) (2 in the example). In the algorithm, all values in the matrix below \( t_{min} \) are treated as zeros, as they are considered to belong to the same bus trip, thus no “distance between them”. So an iterative algorithm was used to loop through the vector array \( pdist \), in order to count the number of zeros having in mind that the loop had to start every time a row started (see divisions made in vector \( pdist \) in Figure 17). In Figure 18 there can be seen how it acts after all the process and the classification is completed. Matlab \( ^{10} \) creates in column 9 an ID value that is shared between all the records of a same group. So they can be split to have all records of a same bus located together using the function Classify.m shown in the Annex.

\(^{10} \) Not studied in profundity due to lack of time, but looks like it could be related to linear dependence between vectors in the base of the subspace that forms the matrix of distances.
5.4.2.1.3 Another example that would work in this study

Just as a comment, another example of an easier procedure that would have fit the case scenario in this project is briefly presented here. Starting from the same point, it would consist in comparing two consecutive bus times at stations at a time (as it turns up data is sorted in ascending at this point) and seeing whether the euclidean distance between them is bigger than $t_{min}$, which would mean that they belong to different bus trips. Then, every time a different bus trip is found, a counter would keep the record of it. Then, the number of clusters in which divide the set of data would equal to the value in the counter plus one.

![Figure 18: Classification example after clustering process](image)

Then, after this clustering process was completed, next step consisted in classifying the bus records according to these clusters formed in the previous stage. Therefore, the number of divisions created to the data each time was variable, depending on the number of clusters in each case. If a bus had 3 trips it was divided into 3 different groups, while if another bus had only one, no division was made.
5.4.2.2. Joining same station records to have bus trips

The next step consisted on creating a new variable in which storing the data in a different format to have information about each trip of each bus, parting from the classification made after clustering the data into groups. In Figure 19 there is a sample of the data before joining same station records.

As it can be seen in this figure, CS5 variable stores in each of its subdivisions all the lists of records of a same bus, bus route, day, direction and clustered group. In other words, it is the last step of all the classifications performed, and contains records for every record in each stop, with information about the time the record was taken and also if it belongs to a boarding or an alighting. So the idea was to modify the way data is structured to create a new variable, CS6. Each subdivision of this variable CS6 represents a trip of a bus in a specific day, route and direction. So Matlab® automatically calculates this same operation for each bus in routes 109 and 412. Each row is for each stop in which the bus stops, and it mainly contains the following information:

- Arriving/departing time at each station (columns 4-6). This is an approximation (see 5.3.3) as, according to the objectives and sources of information available in the project, the data regarding to a bus arriving or departing a station is not known. One of the goals of the project is to use smart card data to show how to monitor the performance and extract indicators of bus trips, so here these arriving/departing
times are approximated respectively to the first and last record available in that stop of a passenger using a Go Card.

- Number of passengers boarding and alighting the bus at each stop (columns 7 and 9). This operation is performed by accumulating the values of boarding and alighting people from the previous step and then calculating the difference between both.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7407</td>
<td>Adelaide St ...</td>
<td>0.7467</td>
<td>29</td>
<td>'on'</td>
<td>0</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>2</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7517</td>
<td>Cribb Street...</td>
<td>0.7517</td>
<td>0</td>
<td>'on'</td>
<td>1</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>3</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7528</td>
<td>Auchenflower...</td>
<td>0.7532</td>
<td>0</td>
<td>'on'</td>
<td>1</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>4</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7550</td>
<td>Toowong Rd...</td>
<td>0.7551</td>
<td>5</td>
<td>'on'</td>
<td>4</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>5</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7552</td>
<td>Galley Road...</td>
<td>0.7562</td>
<td>1</td>
<td>'on'</td>
<td>4</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>6</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7557</td>
<td>St Lucia We...</td>
<td>0.7569</td>
<td>0</td>
<td>'on'</td>
<td>4</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>7</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7577</td>
<td>St Lucia Fer...</td>
<td>0.7578</td>
<td>0</td>
<td>'on'</td>
<td>5</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>8</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7591</td>
<td>Coldridge St...</td>
<td>0.7592</td>
<td>0</td>
<td>'on'</td>
<td>8</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>9</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7633</td>
<td>University o...</td>
<td>0.7634</td>
<td>1</td>
<td>'on'</td>
<td>4</td>
<td>'off'</td>
<td>2:04</td>
</tr>
<tr>
<td>10</td>
<td>735334</td>
<td>412</td>
<td>'Outbound'</td>
<td>0.7649</td>
<td>St Lucia Sou...</td>
<td>0.7651</td>
<td>0</td>
<td>'on'</td>
<td>5</td>
<td>'off'</td>
<td>2:04</td>
</tr>
</tbody>
</table>

Figure 20: CS6 variable after unifying same station records

### 5.4.3. Calculation of travel time and passenger demand indicators

Following a similar process than in the last step but starting from the data in the last figure, the data was at this point prepared to calculate the performance indicators chased in all the process. After the process, some of the performance indicators for every bus that were considered important, were collected together in a list and extracted from Matlab © to an excel datasheet. The process itself starts from CS6 variable, as the sample shown in Figure 20, having each bus data identified, organised and ready to perform calculations on them.

#### 5.4.3.1. Parameters for travel time analysis

For the travel time analysis, this is what was mainly calculated:

- Calculation of the Travel times for each bus (or section of the route, in the case of the 109 bus line)
- Calculation of the validation time (=dwell time in chapter 5.3.3) and running times to see their proportion and impact on the total travel time.
In each cell, which belongs to a particular trip of a bus, several calculations were undertook and placed in other blank positions of the same cell\textsuperscript{11}. The same process was automatically done for all the buses available after the previous data processing stage, in both 109 and 412 services. Figure 21 shows a sample of the final result of the whole process for a bus trip belonging to a 109 bus service in outbound direction. 109 service is shown here instead of the 412 is because the 109 is more complex to calculate, while 412 is simpler because it only contains a part of the calculations 109 route has.

Regarding the calculations on travel times, the process was done very similarly for both bus services 109 and 412 but with slight differences. For bus route 412 only travel time between the city and UQ stops was considered, because the route is mixed with traffic at all time. The city station for bus 412 is \textit{Adelaide St Stop 16}, and the UQ stop is the one located in \textit{Chancellors Pl, University of Queensland – Zone D}. However, bus route 109 has two main sections in its route that were decided to be split for a better understanding of the characteristics of the travel time in this route. As commented before, the section of the route between \textit{Adelaide St – Stop 16}, and \textit{Cultural Centre – Busway Station} is mixed with traffic as the bus route 412 is, and the remaining section, between \textit{Cultural Centre – Busway Station} and \textit{UQ Lakes Busway Station} takes place in the busway. As traffic and route characteristics for these two sections are significantly different, travel time for 109 service was calculated for each section, having also the total travel time calculated.

As it can be seen in Figure 21, column 4 contains data for the first passenger who touches on/off in the bus in the station considered (each row belongs to a concrete station). As commented before (chapter 5.3.3: Approximations), this was considered as the arriving time of the bus at the stop. Column 6 contains the data for the last passenger touching on/off in the station, so it was approximated as the time the bus departs the stop. From this column, the

\textsuperscript{11} All calculations undertook in Matlab \textsuperscript{\textregistered} using date/time strings are done first converting the strings into numbers and then reversing the process to obtain the result in a string.
departure times for the stops of Adelaide St, Cultural Centre and UQ Lakes busway station\textsuperscript{12} were kept in positions \{1,12\}, \{1,13\} and \{1,14\}. This allowed to calculate the travel times in the different sections considered for the study. In the example shown in Figure 21, by substracting the cells in positions \{1,13\} and \{1,12\} the travel time between Cultural Centre and Adelaide St was obtained, and stored in the position \{2,12\}. Same for the remaining section (Cultural centre to UQ Lakes), substracting positions \{1,14\} and \{1,13\} and storing the result in the position \{2,13\}. The total travel time (located in the position \{2,14\}) could be obtained either by summing up the travel times of the sections or by substracting end and beginning stations stop times.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
3 & 6 & 5 & 4 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\
\hline
1 & ... & 9:39:30 AM & Adelaide.. & ... & ... & ... & ... & 9:39:20 AM & 9:43:31 AM & 9:53:58 AM & 00:01:12 & 00:02:51 & 00:11:47 \\
2 & ... & 9:43:12 AM & Cultural .. & ... & ... & ... & ... & 00:04:11 & 00:10:27 & 00:14:38 & 2.199e-04 & | & | \\
3 & ... & 9:45:31 AM & South B.. & ... & ... & ... & ... & [ ] & [ ] & 0 & | & | \\
4 & ... & 9:50:58 AM & Dutton P.. & ... & ... & ... & ... & [ ] & [ ] & 1.7961e-04 & | & | \\
5 & ... & 9:53:26 AM & UQ Lake.. & ... & ... & ... & ... & [ ] & [ ] & 3.7073e-04 & | & | \\
\hline
\end{tabular}
\caption{Parameters calculated for travel time analysis.}
\end{table}

The second stage of this process, validating time was calculated, and used to obtain the approximation of the running time as well. Column 15 stores the result for the time the bus is stopped at each station. As each row of the cell is a record for a station, the formula \(i,15\) = \(i,6\) - \(i,4\)\textsuperscript{13} stores in column 15 this time in which a bus is stopped at each station. Summing up the results in the whole column 15 the total validating time for the bus in the trip is

\textsuperscript{12} Stations that set the limit between sections studied. Cultural center splits busway and mixed with traffic sections of the route. For 412 service, the departure times considered are Adelaide St stop and UQ Chancellors Pl stop.

\textsuperscript{13} Being \(i\) an index variable that scans the rows of the cell, in the example shown in Figure 21 it would be from the first to the fifth row.
obtained, and stored in position \(1,16\). The difference between the total travel time and the validating time is the approximated running time (position \(1,17\)).

### 5.4.3.2. Parameters for passenger demand analysis

Following a similar procedure as shown in the previous chapter 5.4.3.1: Parameters for Travel Time analysis, this section is about how parameters related to the passenger demand and occupancy of the bus were extracted and calculated from the smart card passenger data. Note that these parameters were located inside each cell as the travel time parameters, right next to them. So these parameters are the ones that can be seen in the example Figure 22 in columns 18 and 19.

For the demand of the line, defined as the total number of passengers that board the bus in all stops (from the beginning to the end), the calculation is just the result of summing up column 7, that represents the number of boarding passengers at each stop (row). The result was saved in column 19 row 1. Also, the maximum number of passengers (more loaded track of the trip) is calculated (column 19 row 3: it can be seen in Figure 22 how it coincides with the actual maximum of passengers in column 18 row 4, 6 passengers in this case in the station of Matter Hill). Finally, a message that claims "limit exceeded" appears in column 19 row 2 if the bus occupancy at some point exceeds 62 passengers, which is the maximum. In the rest of the cases, the message is "limit not exceeded".

Also, the bus occupancy was kept in column 18, and represents the exact number of passengers that remain inside the bus after every stop. It is the cumulative net balance of boarding and alighting passengers after each stop, which was useful for further sections in this project. This feature was automatically calculated for every bus, but not extracted to the excel datasheet list of performance indicators.
In Figure 23 there can be seen a sample of the final output data for 109 inbound bus service extracted to excel.

In Figure 23 there can be seen a sample of the final list of data extracted for bus service 109. Each row contains the chosen inforamation for every bus trip. To obtain this list the procedure was based in creating an iterative loop to scan every subdivision of the cell CS7 (every subdivision is a bus trip), and extract from each one the selected performance indicators. As a comment, two more indicators were created on the excel, as they were considered they could be useful in some cases do perform statistical studies:

- High occupancy: Indicator that showed if a particular bus had at any moment of the trip an occupancy equal of bigger than 20 passengers (customizable feature).
- Condensed trip: In the case a bus had at least a minimum occupancy of 10 passengers, this indicator checks if the maximum number of passengers that are inside the bus simultaneously is at least the 80% (again, customizable feature) of the total number of passengers that use that bus. That indicator could give an idea about the distribution of passengers at stops (if the passengers that use the bus line are spread along the different stations, or on the contrary, most of them travel between similar sections of the route).
6. Results and Analyses

The previous chapter contains the set of steps necessary to have the data processed and ready for the analyses. In this section are exposed a number of analyses that can be done with the data processed. Thus the aim of this section is to provide a possible roadmap that could be useful in further analyses undertaken on similar samples of smart card data, but not to present any final result on Brisbane bus reliability.

In the field of travel analysis there are a number of actions that can be done, and it is the purpose of the analysis that will decide the best option.

6.1. Travel time analysis

This section is related to the travel time analysis and its characteristics. The main intention is to investigate the travel time data and extract as much information as possible from the travel times of the buses to possible service alterations.

One of the main achievements of the previous process on travel time was to get the travel times for each of the bus services and directions, and also the travel times for each of the sections in which some of the lines were decided to be split. This data break down allows study of the different sections of the routes and provides data for a better understanding of the patterns of the travel time and makes it easier to find possible variables that influence in its variation.

Let us remember the travel times data available for each bus that was obtained from the raw data received from Translink. For both bus services and directions, 109 and 412 both inbound and outbound, all the total travel times between the two ends of the line that were
interesting for the study\textsuperscript{14} (located in the city CBD and in The University of Queensland), as well as the division of the travel time between the running and the validating time. Also, for the specific case of the 109 service, another division has been made, by splitting the travel times in the busway and mixed with traffic.

### 6.1.1. Comparing travel time statistics for different services and directions

The display of descriptive statistics can give some important information to begin. Minitab has been used to extract descriptive statistics for the services 109 and 412 in inbound and outbound directions. In inbound services Sect1 represents the Travel time in the busway section, while in outbound direction, as the direction is the opposite, the busway section is Sect2. The months chosen for the study were March and April of 2013. The whole month of March for the study of 109 and one week for 412, according to the data extracted for travel time and passenger demand studies in the following sections.

The main observable results in terms of basic statistics are the following (109 Outbound basic statistics shown in Figure 24, see Chapter B in the Annexes for the rest of descriptive statistics):

- Travel time in 109 Outbound was found to be substantially higher than 109 Inbound. 412 had a similar mean for both Inbound and Outbound services.
- Bus route 412 had a lower standard deviation compared to 109 in all comparable aspects (travel, validating and running times, nº of passengers, etc.). That was surprising because 412 runs mixed with traffic all the time, so traffic flow variations should have more influence in the total travel time than the 109.
- Number of passengers using 109 bus line is higher than the 412.

\textsuperscript{14}Line 412 does not end or start in UQ Chancellors Pl, but this the stop is considered in this study as the theoric end of line for the 412 service, because it allows the comparison between 109 and 412 services. Mainly due to same start/finish locations and very similar distances run.
There was a remarkable difference between The sections 1 and 2 in the 109 bus routes. While in Inbound direction the short section mean (Adelaide Street – Cultural Centre) was much smaller than the busway section (166,3 versus 830,98 seconds), in Outbound direction the difference between both decreases substantially (468,7 versus 690,4 seconds).

**Results for: 109Outbound**

**Descriptive Statistics: Sect1; Sect2; Total TT; Running _Tim; Dwell Time; ...**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sect1</td>
<td>732</td>
<td>121</td>
<td>468,71</td>
<td>4,35</td>
<td>144,66</td>
<td>28,00</td>
<td>365,00</td>
<td>450,50</td>
<td>505,00</td>
<td>677,00</td>
</tr>
<tr>
<td>Sect2</td>
<td>505</td>
<td>52</td>
<td>490,41</td>
<td>4,37</td>
<td>95,47</td>
<td>474,00</td>
<td>602,00</td>
<td>677,00</td>
<td>721,00</td>
<td>1032,00</td>
</tr>
<tr>
<td>Total TT</td>
<td>781</td>
<td>75</td>
<td>1158,3</td>
<td>6,43</td>
<td>179,6</td>
<td>72,00</td>
<td>1032,00</td>
<td>1146,00</td>
<td>1146,00</td>
<td>1146,00</td>
</tr>
<tr>
<td>Running _Tim</td>
<td>781</td>
<td>75</td>
<td>943,73</td>
<td>5,43</td>
<td>151,78</td>
<td>252,00</td>
<td>842,50</td>
<td>923,00</td>
<td>923,00</td>
<td>923,00</td>
</tr>
<tr>
<td>Dwell Time</td>
<td>857</td>
<td>0</td>
<td>205,30</td>
<td>4,59</td>
<td>134,30</td>
<td>0,00</td>
<td>102,00</td>
<td>183,00</td>
<td>183,00</td>
<td>183,00</td>
</tr>
<tr>
<td>Trip_Demand</td>
<td>857</td>
<td>0</td>
<td>29,390</td>
<td>0,673</td>
<td>19,692</td>
<td>2,000</td>
<td>15,000</td>
<td>24,000</td>
<td>24,000</td>
<td>24,000</td>
</tr>
<tr>
<td>Max _passengers</td>
<td>857</td>
<td>0</td>
<td>26,914</td>
<td>0,632</td>
<td>18,513</td>
<td>0,000</td>
<td>13,000</td>
<td>22,000</td>
<td>22,000</td>
<td>22,000</td>
</tr>
<tr>
<td>Day week</td>
<td>857</td>
<td>0</td>
<td>4,0455</td>
<td>0,0505</td>
<td>1,4777</td>
<td>1,0000</td>
<td>3,0000</td>
<td>4,0000</td>
<td>4,0000</td>
<td>4,0000</td>
</tr>
</tbody>
</table>

**Figure 24:** Basic descriptive statistics for 109 outbound service during the month of March of 2013

### 6.1.2. Travel time variation in different time periods

In this section some aspects regarding travel time are presented. The program built in Section 5 of the project was designed to extract different travel time output data. After considering several options, only the running/validating time has been considered for the study.
Figure 25: 412 inbound travel time plots from Monday to Friday (April 8-12th 2013)
Bus service 412 was studied in this case, in inbound direction. Chapter 4.1 can be useful to understand the analyses undertook in this section. Firstly, a week was chosen randomly to be studied amongst all the available options. The first complete week in the month of April was the selected one (April 8-12th, 2013). Considering the bus service and direction, some possible results for the study performed for this section were considered beforehand. Especially the fact that travel time could be affected by variations in running time, as 412 bus service runs all the time mixed with traffic. However, after watching the results some other conclusions were considered to be mentioned.

Watching the Monday-Friday variation in Figure 25 the main appreciations would be:

- Running and validation times do not seem to explain the travel time variations as good as it was expected, mainly due to the limitations of the approximations performed (see chapter 5.3.3: Approximations).
- There are cases in which it seems that running and validation times have some of the idle times mixed, assigned sometimes to one or other, what makes inverse tendencies in the plots in short periods of times (one increases while the other decreases in the same trip). These idle times can sometimes be important Figure 8.
- Also, the effect of circumstances that are external to the considerations of the study can have more importance compared to what was initially expected. For example, passengers not using smart cards can cause delay at stops while purchasing the ticket to the bus driver, which would be seen in the validation time series plot.

Especially concerning the last conclusion, a hypothesis was made. The hypothesis consisted in considering peak periods as studiable using the method developed throughout this project, while non-peak periods were considered to be treated cautiously for this regular possible presence of single paper ticket users (passengers not using smart cards). The hypothesis was based in the possible fact that people that travel in peak hours travel regularly so know the route and system, while sometimes out of that periods some passengers that are not regular public transport users can be found. If this hypothesis is found to be consistent, peak periods could be considered to be studied with this method.

First thing to do was to find the mean of passengers along the day. In order to do that, the same week considered in Figure 25 was used to find the following information:
Table 4: Proportion of single paper tickets passengers compared to Smart Card users

<table>
<thead>
<tr>
<th>Time in the day</th>
<th>Mean smart card passengers (pass/bus)</th>
<th>Total number of buses</th>
<th>Non smart card users (all week)</th>
<th>Non smart card users per bus</th>
<th>% Smart card users</th>
<th>% Single paper users</th>
<th>% of increase respect smaller</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 to 7</td>
<td>35.83</td>
<td>6</td>
<td>9</td>
<td>1.50</td>
<td>95.58%</td>
<td>4.02%</td>
<td>87.96%</td>
</tr>
<tr>
<td>7 to 10</td>
<td>34.15</td>
<td>25</td>
<td>49</td>
<td>1.95</td>
<td>94.57%</td>
<td>5.43%</td>
<td><strong>153.83%</strong></td>
</tr>
<tr>
<td>10 to 14</td>
<td>30.52</td>
<td>51</td>
<td>60</td>
<td>1.18</td>
<td>98.29%</td>
<td>3.71%</td>
<td>73.62%</td>
</tr>
<tr>
<td>14 to 17</td>
<td>34.18</td>
<td>75</td>
<td>56</td>
<td>0.75</td>
<td>97.86%</td>
<td><strong>2.14%</strong></td>
<td>0.00%</td>
</tr>
<tr>
<td>17 to 23</td>
<td>30.32</td>
<td>34</td>
<td>67</td>
<td>1.97</td>
<td>93.90%</td>
<td>6.10%</td>
<td><strong>185.46%</strong></td>
</tr>
</tbody>
</table>

According to the results obtained, it looks like only the afternoon peak period may be considered to be correct under the hypothesis scenario, while not the morning peak. Because apparently, in the morning peak there is a similar (even higher) amount of single paper users (in percentage, compared to smart card users in each case). Between all percentages of single paper passengers, a comparison between them was performed taking as a reference the smaller one, the afternoon peak period. The difference between both peak periods is considerable, being the morning one up to 153.83% bigger.
6.2. Passenger demand

In this section of the project some analyses on passenger demand are presented using the output data generated in the data process section. Some results are shown, with example purposes, using 109 outbound bus data for this section. As commented in the goals section, the intention is to see how smart card data can be a useful tool to study and keep a monitoring of the bus performance. It is important to remember at this point that what it is being calculated is not exactly the real situation. Not all the passengers that appear in the calculations of this section are the actual number that travelled that day in 109 outbound service, because the method in this project uses only smart card records. Single paper ticket passengers remain beyond the scope of the current project. Nevertheless, in the time period studied in this section this is not considered as an inconvenient because the percentage of passengers using single paper tickets in peak periods is, in proportion, almost negligible compared to smart card users.

6.2.1. Bus loading calculation

This section parts from data of each bus in a particular day (Figure 22). A specific day was chosen to perform this analysis, by looking for the day with more available records (by calculating the statistical mode of all the days), and finally March 5th of 2013 was the selected one. Only bus data in the peak period in the morning (7AM to 10AM) was considered, mainly because the need of a homogeneous period of time to study in terms of bus frequency. In peak period bus frequency is bigger and there is more sample available for the study. Another reason for choosing peak period is because it allows sizing the bus line in the worst case (more passengers expected to be using the bus) Outbound direction was considered to be more useful in this example, because more passengers were bound to be travelling away

15 This same study could be extended to any other period of time or different day. The purpose in this section is to show how it would be done for a particular case.
from the city of Brisbane in the morning. 23 buses were found in the morning peak hour, and they were extracted into an excel datasheet. Next, the following table was generated:

Table 5: Boarding and alighting number of passengers in every stop for each bus studied

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BOARDINGS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adelaide St</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>28</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>12</td>
<td>33</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>34</td>
<td>10</td>
<td>229</td>
<td></td>
</tr>
<tr>
<td>Cultural Ctr</td>
<td>15</td>
<td>6</td>
<td>1</td>
<td>23</td>
<td>27</td>
<td>16</td>
<td>35</td>
<td>26</td>
<td>20</td>
<td>12</td>
<td>4</td>
<td>6</td>
<td>27</td>
<td>3</td>
<td>4</td>
<td>18</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>42</td>
<td>11</td>
<td>339</td>
</tr>
<tr>
<td>South Bank</td>
<td>7</td>
<td>11</td>
<td>0</td>
<td>30</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>20</td>
<td>33</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>4</td>
<td>200</td>
</tr>
<tr>
<td>Mater Hill</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>PAH Station</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boggo Rd</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>21</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>3</td>
<td>146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UQ Lakes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ALIGHTINGS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adelaide St</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cultural Ctr</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>South Bank</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Mater Hill</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>PAH Station</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Boggo Rd</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>UQ Lakes</td>
<td>17</td>
<td>17</td>
<td>2</td>
<td>72</td>
<td>48</td>
<td>29</td>
<td>37</td>
<td>65</td>
<td>90</td>
<td>15</td>
<td>15</td>
<td>38</td>
<td>64</td>
<td>19</td>
<td>13</td>
<td>68</td>
<td>49</td>
<td>63</td>
<td>16</td>
<td>0</td>
<td>21</td>
<td>77</td>
<td>27</td>
<td>862</td>
</tr>
</tbody>
</table>

Table 4 contains all the boarding and alighting passengers in each stop for all the 23 buses running between 7AM and 10AM. The sum of all boardings and alightings for all the period was calculated in the “Sum” column. Using the information in this table, next Table 5 was build, calculating the cumulative number of passengers in each stop.

Table 6: Cumulative number of passengers as bus advances in every stop

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CUMULATIVE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adelaide St</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>26</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>28</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>12</td>
<td>33</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>34</td>
<td>10</td>
<td>235</td>
</tr>
<tr>
<td>Cultural Ctr</td>
<td>16</td>
<td>17</td>
<td>2</td>
<td>49</td>
<td>27</td>
<td>19</td>
<td>35</td>
<td>54</td>
<td>23</td>
<td>17</td>
<td>14</td>
<td>17</td>
<td>34</td>
<td>15</td>
<td>11</td>
<td>30</td>
<td>45</td>
<td>18</td>
<td>14</td>
<td>10</td>
<td>14</td>
<td>76</td>
<td>21</td>
<td>578</td>
</tr>
<tr>
<td>South Bank</td>
<td>21</td>
<td>20</td>
<td>2</td>
<td>78</td>
<td>33</td>
<td>33</td>
<td>35</td>
<td>71</td>
<td>55</td>
<td>16</td>
<td>16</td>
<td>25</td>
<td>46</td>
<td>21</td>
<td>11</td>
<td>40</td>
<td>49</td>
<td>34</td>
<td>16</td>
<td>12</td>
<td>20</td>
<td>74</td>
<td>25</td>
<td>753</td>
</tr>
<tr>
<td>Mater Hill</td>
<td>18</td>
<td>18</td>
<td>2</td>
<td>77</td>
<td>40</td>
<td>31</td>
<td>41</td>
<td>71</td>
<td>68</td>
<td>17</td>
<td>16</td>
<td>25</td>
<td>52</td>
<td>20</td>
<td>13</td>
<td>40</td>
<td>49</td>
<td>33</td>
<td>16</td>
<td>12</td>
<td>20</td>
<td>74</td>
<td>26</td>
<td>779</td>
</tr>
<tr>
<td>PAH Station</td>
<td>13</td>
<td>17</td>
<td>1</td>
<td>69</td>
<td>40</td>
<td>29</td>
<td>37</td>
<td>65</td>
<td>75</td>
<td>17</td>
<td>16</td>
<td>23</td>
<td>44</td>
<td>20</td>
<td>13</td>
<td>41</td>
<td>49</td>
<td>28</td>
<td>16</td>
<td>11</td>
<td>21</td>
<td>69</td>
<td>24</td>
<td>738</td>
</tr>
<tr>
<td>Boggo Rd</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>72</td>
<td>48</td>
<td>29</td>
<td>37</td>
<td>65</td>
<td>90</td>
<td>15</td>
<td>14</td>
<td>38</td>
<td>64</td>
<td>19</td>
<td>13</td>
<td>68</td>
<td>49</td>
<td>63</td>
<td>16</td>
<td>0</td>
<td>21</td>
<td>77</td>
<td>27</td>
<td>844</td>
</tr>
<tr>
<td>UQ Lakes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

As commented, this table contains for each of these 23 buses, the amount of passengers in the bus after every stop. As it can be seen in Table 6, in each bus the tendency is a gradual increase of passengers after every stop, until UQ Lakes, the last stop, where all passengers alight the bus. In this table, all the values that exceed 62 passengers are highlighted in red, because that is the maximum number of passengers allowed simultaneously in a bus. In the last column there is the sum of all columns stop by stop. That shows the most loaded track, which is used for dimensioning the line. In this case, the most loaded track is the last one,
between Boggo Road and UQ Lakes stations, with a total number of passengers of 844 during this morning peak hour studied.

6.2.2. Checking bus line sizing

With all the information extracted and calculated in the previous section, some performance indicators of the bus line were obtained and used to check the correct sizing of the bus line. The starting point was the most loaded track of 844 passengers, because is the one that sets all the line.

- **Passengers per hour**: \( \frac{844 \text{ passengers}}{3 \text{ hours}} = 281.33 \text{ pass/h} \)
- **Frequency**: \( \frac{180 \text{ min}}{23 \text{ buses}} = 7.83 \text{ min} = 7 \text{ min } 50s \)
- **Waiting time**: \( t_w = \frac{7.83 \text{ min}}{2} = 3.92 \text{ min} = 3 \text{ min } 54s \)
- **Mean passengers per bus**: \( \frac{844}{23} = 36.7 \text{ pass/bus} \)

According to the result, the conclusion is that the bus line is well sized, because it is designed to absorb the existing demand of passengers during the peak period. However, despite the correct sizing, it does not mean that a better solution could be found for this case. According to Table 6, there are 5 buses in which the maximum demand of 62 passengers is exceeded. There are two main solutions considered in order to solve this:

- The first one would imply an increase in the frequency of the buses. On one hand that would solve the problem, but on the other hand that would make the rest of the buses worse sized, because they would travel emptier. It is important not to forget that the mean of passengers for this time period is 36.7 passengers per bus, which is slightly over the half of maximum occupancy (31 passengers).
- The second and more viable option is to identify at which times tend to be those limit exceeded buses and use double-length buses to cover those services. In the example those buses are the ones departing from the first stop (Adelaide Street station) at 7:09, 7:19, 7:25, 8:09, 8:25, 8:30 and 9:17AM.
In this last Figure 26 there is a graphic that shows the bus occupancy after every stop. The worst case in Table 6 has been chosen in order to show a bus exceeding the limit of 62 passengers. It corresponds to the bus number 9 (which departs Adelaide St station at 8:30AM), because it carries up to 90 passengers in its maximum loading track, between Boggo Rd and UQ Lakes stations.

![Bar chart showing bus occupancy](image)

**Figure 26:** Amount of smart card users in 109 outbound bus departing at 8:30AM on 05/03/2013
7. Sustainability of the Project

Based on the guidelines set out in the Royal Legislative Decree 1/2008, 11th of January, by which the text *Texto refundido de la Ley de Evaluación de Impacto Ambiental de proyectos* is presented, the study of sustainability of this project is performed.

7.1. General description of the project and consideration of alternatives

The realization of this project basically consists in processing data using computer software and analyzing it. Thus, the study of sustainability of this project is conditioned exclusively for this task. Being this project mainly theoretical, it is considered that there are no alternative solutions to its realization.

7.2. Description of the environment

This project was carried out entirely in the Intelligent Transport Laboratory facilities, located in the Transport Engineering Department in Hawkeen Engineering building of The University of Queensland, and it lasted about 8 months. It is believed that its implementation has not changed substantially the levels of environmental quality or the elements that compose it.

7.3. Identification and assessment of impacts on the environment

This section discusses the various potential impacts that the project may have on the environment and may be due to:

- The existence of the project
- The use of natural resources
• The emission of pollutants (in the atmosphere), the formation of harmful substances or waste treatment.

7.3.1. Evaluation criteria

First, a number of technical definitions essential for the assessment of environmental impacts considered significant are presented. These terms refer to different types of effects that can lead the impacts of a project are:

Depending on the type of effect:

• Positive effect: The effect admitted as such by both the scientific community and the general population, in the context of a complete analysis of the costs and benefits of generic external contingencies contemplated action.
• Negative effect: Effect that results in the loss of natural value, aesthetic and cultural landscape, ecological productivity, or damages arising from increased erosion, pollution and other environmental hazards in disagreement with the organic structure-geographical character and personality of a given location.

Depending on the impact of the effect:

• Direct effect: The effect that has an immediate impact on some aspect of the environment.
• Indirect or secondary effect: The one that has direct impact respect to the relationship between environmental sectors.

7.3.2. Identification of impacts. List of activities

At the time to do a study of sustainability of a project, there are some major steps or activities that are to be considered, such as the construction, implementation and decommissioning activities. In this case, however, it only makes sense to analyze the stage of the activity as no construction is realized.
The resources that were needed for this activity are mainly power and office material consumption such as paper, printer cartridges, pens, etc.

7.3.3. Potential environmental impacts

This section analyzes the potential impacts to the environment and society that may have caused the completion of this project. Specifically, there are three possible types of impact:

- **Impact on the existence of the project.** It is believed that if the current project or similar further studies carried on similar basis achieve their goals totally or partially, it could derive to a positive effect on environment. The increase in public transport performance and reliability would carry a reduction of travel times by decreasing the negative contributing factors. That would mean also a reduction in bus fuel consumption. Furthermore, if the public transport became more reliable and well sized, more passengers would be inclined to use it, thus also a certain decrease in the number of private vehicles used could be expected, meaning that less gasoline expelled to the atmosphere.

- **Impact on resource use.** There is a direct impact on resource use basically office equipment as a result of the use of computer equipment, office fitting, etc. Waste management is performed with the aims to recycle as much as it is possible. All used resources were classified and disposed on their correspondent recycle bins, such as paper, plastic, etc.

- **Impact on emissions.** It can also be considered the existence of an indirect impact arising from the consumption of electricity resulting from the emission of combustion gases at power plants. The main pollutant CO₂ released is responsible for the greenhouse effect. Considering that the project was developed around 8 months and the use of the computer is considered to be around 8 hours during working days, the energy consumption derived from computer usage would be around 1280 hours. The energy consumption of a computer varies substantially, but an average value considering all peripherals could be around 250W. So under these circumstances the energy consumed during the project is 320 kWh. Finally, from the data of CO₂ released per kWh
by the Spanish Electric Network, emissions of 300 g CO$_2$/kWh are approximated. If, on the edge of this approach, it is considered that the unit of energy produced is equivalent to the unit of energy consumed and, therefore, there is no loss in transport, the mass of CO$_2$ released to the atmosphere by the realization of this project is 96 kg of CO$_2$.

7.3.4. Assessment of impacts

All impacts of this study can be assessed as compatible impacts; therefore, the overall impact associated with the project is compatible too, given the planned corrective measures listed below.

7.4. Planned actions

Corrective measures planned to minimize environmental impacts were focused primarily on the environmental management of waste (paper, printer ink cartridges, etc.). Also, the use of paper documents were intended to be minimized, by the use of documents in electronic format and also using double-sided and black and white page printing.
8. **Cost of the Project**

Below is a list of expenses incurred in the course of this project. The different costs are broken down into two categories: the cost of material resources consumed and the cost of human resources.

### 8.1. Cost breakdown

Material resources consumed during the realization of this project are mainly related to the consumption of office supplies, electricity and water. It should be noted that all activity has been developed in the Transport Engineering Laboratory in the University of Queensland facilities and therefore, the computer equipment used was already existing in the institution. Consequently, the depreciation of computer equipment owned by UQ is considered regarding its costs. Concerning the software used, it is considered that the cost of licensing the operating system - Microsoft Windows ® 7 Professional - and the office suite - Microsoft ® Office 2010 - are included in the acquisition cost of the computer. However, the specific software license that was used – Minitab ® and Matlab ® R2012b – had a cost considered as zero since UPC and UQ benefit from academic licenses.

The costs of resources used materials are presented in Table 7:

<table>
<thead>
<tr>
<th>Concept</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amortization of computer equipment</td>
<td>200 €</td>
</tr>
<tr>
<td>Office supplies</td>
<td>30 €</td>
</tr>
<tr>
<td>General Supplies</td>
<td>40 €</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>270 €</strong></td>
</tr>
</tbody>
</table>

In this section costs associated with human resources, there are considered basically the amounts time of the person in charge of the project, as well as other people who have had an active and have made respectively on the direction and advice.
Thus, it is considered a cost of 20 € / hour for a junior engineer who has done the project, devoting 40 hours a week for slightly over 8 months. Also included in the costs the dedication of two doctors in Transport Engineering and a PhD Candidate.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Hourly Cost (€/H)</th>
<th>Hours Spent</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior engineer</td>
<td>20</td>
<td>1.320</td>
<td>26.400 €</td>
</tr>
<tr>
<td>Doctor in Transport Eng.</td>
<td>80</td>
<td>260</td>
<td>20.800 €</td>
</tr>
<tr>
<td>PhD candidate</td>
<td>40</td>
<td>420</td>
<td>16.800 €</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>64.000 €</strong></td>
</tr>
</tbody>
</table>

Below there is the total amount of the expenses resulting from the completion of the project (Table 9).

<table>
<thead>
<tr>
<th>Concept</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material resources</td>
<td>270 €</td>
</tr>
<tr>
<td>Human resources</td>
<td>64.000 €</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td><strong>64.270 €</strong></td>
</tr>
<tr>
<td>Tax (IVA) (21%)</td>
<td>13.496,7€</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>77.766,7 €</strong></td>
</tr>
</tbody>
</table>

Thus, the final cost of the project is 77.766,7 €.
9. Planning

Figure 27 shows the schedule of activities followed in conducting this project. This project lasted approximately 8 months.

The first two months were invested in getting introduced to smart card systems and some studies on public transport performance using smart card data. Then, once enough knowledge was acquired, it was the moment to start with Matlab® code programming introduction and the design of what the software should be able to show as an output result.

The following stage was the most time consuming, and involved the programming of all the code, as well as testing it to make sure the correct functioning under different scenarios. In the last phases of this process more literature review was undertook to ensure the correct consecution of the program built according to what was being intended to design.

Once the data extraction method was finished, it was time to generate output data and decide which results to show, and analyze the results. And the last part involved the writing of the conclusions, as well as the revision of the correct matching of the different sections in the project.
Conclusions

After all the studies and processes undertook in this project, this section has the aim of gather the conclusions found during the consecution of the study.

Smart card ticketing technologies are a more convenient way of collecting fares compared to other, traditional methods. The advantages could be mainly summarised in a more dynamic passenger movement when entering and alighting vehicles, also by reducing the amount of work the driver of a bus has when having to sell tickets and having to drive the vehicle at the same time. And especially, the simplicity of usage for the passenger, who does not need to worry about different kinds of tickets and fare zones, because it is just as simple as having a valid smart card and topping it up with money.

As for the main goal of the project, using smart card data collected from passengers to extract performance data for understanding and analysing the public transport system, the main conclusions are:

- The approximation made of validation time and running time seemed to be mixing the content between both in certain cases. The fact that is unknown if a bus is stopped queuing in a station with the doors closed (usually waiting for other buses to leave) can cause distortion in travel times measured.
- Single paper ticket users were kept apart from the study due to the impossibility to treat their data in a succesful way as smart card users, but this was not considered to be a serious trouble because they are a minority. It has been seen, though, that in some cases, a few single paper ticket users can cause considerable delay in a stop.
- The hypothesis of single paper ticket users travelling mainly on non peak periods seems that could be only partially true. So peak hours cannot be fully considered to be “free of single paper ticket users”.
- When checked bus line sizing, the result was that the line was properly sized though the bus loadings were not uniformly distributed along the period studied, causing several buses to exceed their limit of passengers while others travelled semi-empty. In this framework, increasing/diminishing the passing frequency of buses was not a proper solution, so the incusion of double-lenght buses for some services is conceived as the best solution to consider.
Future studies based on similar smart card data

At the end of this project, after having developed an algorithm that extracts performance data from an initial set of smart card data from passengers trips, and after having seen the different kind of studies that could be done with the data, it is moment to review future research that could be done with smart card data.

- **Operational studies:** The first possibility is the direct continuation of what has been shown in this study. Results and analyses proposed in this project could be done with reliability in certain conditions (in the framework of the limitations expressed). Analysing travel times and passenger demand patterns as in this project would be a main doable task, as well as keep developing and researching on the software proposed to reduce its limitations and make it reliable under a wider range of possible scenarios. The combination of the method proposed in this project together with a GPS system and/or any other bus position tracker (AVL system) could help in calibrating different scenarios and situations.

- **Modelling:** Using Go Card data could be also useful for modelling the travel times. In the case of Brisbane there is a bus that connects the city of Logan with Brisbane, and this bus service that uses the Go Card smart card ticketing system also has a GPS device that provides the position of the bus at every moment. By having a large amount of Go Card records, a mathematical model could be designed, and trying to infer different variables that can have influence on the travel time, such as the time in the day, the weather, the direction, the number of passengers in the bus, etc. The results of the model could be calibrated with the GPS system to maximise the accuracy of the model, which could be then applied to other bus lines.

- **Transport planning:** The fact that Go Card data contains a lot of information such as boarding and alighting stations and times, bus identifier, Go Card identifier, bus service, direction and day, etc. is a great advantage for transport planning purposes. The movements of the passengers are easy to identify, and using Go card data passenger trip patterns could be studied and analysed, in order to know how people travel through the city in different times and days, being theoretically a reliable method (even though single paper ticket users would be beyond the study).
Acknowledgements

I would like to express my very great appreciation to Professor Luis Ferreira and Mª Antònia Santos from UQ and UPC-ETSEIB respectively for their valuable and constructive suggestions during the planning and development of this research work. Their willingness to give their time so generously has been very much appreciated.

I am also grateful to PhD candidate Zheng-Liang Ma in The University of Queensland who without his help and guidance on Matlab ® programming this project would not have been completed with the same degree of success.

Finally, I wish to thank my parents and family for their tremendous contributions and support both morally and financially towards the completion of this project.
References


Barry et al. (2002). Origin and Destination Estimation in New York City with Automated Fare System Data.


Lam, T., & Small, K. (2001). *The value of time and reliability: measurement from a value pricing experiment*. Transportation Research Part E.


