Kate Hamblin - Dissertation

by Kate Hamblin

UCL Bartlett School of Planning:

BPLN0008 Dissertation in City Planning / BPLN0039 Dissertation in Planning / BPLN0052 Major Research Project

To be completed by the student submitting the dissertation:

Candidate name:	Kate Hamblin
Programme name:	MSc Transport and City Planning
Time and date due in:	17:00 2nd September 2019
Supervisor name:	Dr. John Ward

To be completed by the School office:

Time and date actually submitted:	
Lateness penalty applied (if applicable)**:**	
Supervisor name:	
Second marker name:	
Third marker name (if applicable):	

UNIVERSITY COLLEGE LONDON FACULTY OF THE BUILT ENVIRONMENT
BARTLETT SCHOOL OF PLANNING

TO WHAT EXTENT CAN MEASURING BUS PERFORMANCE FROM THE CUSTOMER PERSPECTIVE INCREASE SUSTAINABLE MODE SHARE?

THE DEVELOPMENT OF TFL'S WEIGHTED BUS CUSTOMER JOURNEY TIME METRIC

(Anderson, 2018)

Kate Hamblin
BPLN0039: Dissertation in Planning
MSc Transport and City Planning
Supervisor: Dr. John Ward

Being a dissertation submitted to the faculty of The Built Environment as part of the requirements for the award of the MSc Transport and City Planning at University College London: I declare that this dissertation is entirely my own work and that ideas, data and images, as well as direct quotations, drawn from elsewhere are identified and referenced.

(signature)
(date)

Word Count (body): 9989
Word Count (appendices): 4956

Acknowledgements

I would firstly like to thank my supervisor Dr. John Ward for his continued support throughout this thesis. John, the guidance on how to turn this into a discursive topic for my thesis has been completely invalauble.

I would also like to thank my colleague, Nigel Kelt, for his assistance with both thinking through the logic and programming throughout this process. Without his VBA and Python expertise, it would not have been possible to process so much data in a short time period. To my other colleagues at Transport for London for their continued support throughout the production and delivery of this metric and for the access to the data which enabled the methodology to be developed.

To my friends and family for listening to the logic of measuring Bus Performance more times than they could count and to Mum, Dad, Natalie and Becky for their proof-reading.

Table of Contents

Table of Figures and Data Tables

Abstract

Increasing sustainable mode share (travel by walking, cycling or public transport) to 80% is the core objective of the current Mayor's Transport Strategy (MTS) (TfL, 2018). To achieve the 80% target, current car use must decline by 17%, meaning that by 2041, 10 million more daily journeys will need to be made by sustainable modes compared to 2015. London Buses are the largest and most flexible mode of public transport, with over 6 million daily journeys and 600 routes, it is required to adapt to ensure service is effectively provided to facilitate achieving these goals. However, since 2015, despite the best performance on record, passenger demand on buses has been declining, largely due to increases in overall passenger journey times.

Existing measures of bus performance heavily focus on monitoring that services maintain a consistent gap between services, without a customer perspective of overall travel times. There is a need to understand the bus customer perspective of bus performance and incorporate this into a holistic measure of the overall experience.

This thesis develops the methodology of a new bus performance measure. This moves away from measuring each element of the customer experience individually and instead incorporates average journey times, journey time reliability and customer perception into one overall quantitative metric. Average wait time, average in-vehicle time, interchange, crowding and buffer time (additional time accounting for variability) are captured from every customer origin and destination through to network level. Through utilising this metric TfL would maximise the effectiveness and targeting of interventions to ensure the customer experience is maximised. By providing an efficient and effective service, customers would be encouraged onto the bus network and measuring performance from the customer perspective would increase sustainable mode share.

1.0. Introduction

1.1. Background and Context

With the Mayor's Transport Strategy (MTS) outlining targets that by 2041, 80% of journeys in London are to be made by sustainable mode share (public transport, walking and cycling), ensuring the bus network meets the needs of its customers is essential (TfL, 2018a). Despite a period of extensive patronage growth on London's bus network between 2000 and 2012, bus passenger demand has been declining in the last few years (see Figure 1), largely due to increasing and unreliable journey times. Even with 6 million bus journeys made in London every day, sustainable mode share must increase. To achieve the 80% target, daily trips made by walking, cycling and public transport must increase by 10 million see Figure 2) (TfL, 2018a, 22). Transport for London (TfL) must make interventions to encourage Londoners out of their cars and to choose sustainable travel. With sustainable mode share at only 63% in 2015, there is a growing need to understand the factors that encourage customers to choose sustainable modes and so intervene to improve the customer experience and ensure progress towards the 80% target (TfL, 2018a).

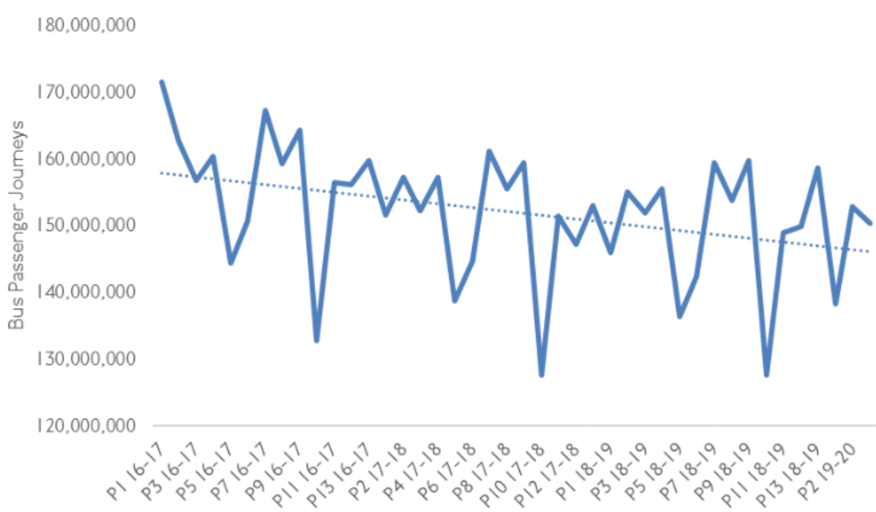


Figure 1 - London Bus Passenger Journeys by Period (4 weekly blocks) (source TfL, 2019d)

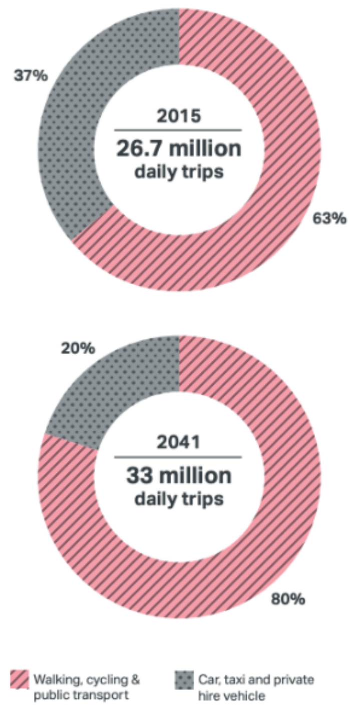


Figure 2 - MTS Sustainable Mode Share Targets (TfL, 2018a, 22)

As London's largest and most accessible public transport mode, with over 600 routes spanning across the city, buses provide the densest and widest reaching public transport option and are essential to increase sustainable mode share (see Figure 3) (Beyazit, 2011; TfL, 2019). As well as ensuring that services are provided to match demand with fully integrated land-use and transport planning, it is essential the bus network is both reliable and convenient for customers (Hall, 2013). The role of the bus is outlined in Policy 15 in the MTS:

> *"The Mayor, through TfL and the Boroughs, and working with stakeholders, will transform the quality of bus services so that they offer faster, more reliable, accessible, comfortable and convenient travel by public transport, while being integrated with, and complementing, the rail and Tube networks" (TfL, 2018a, 155).*

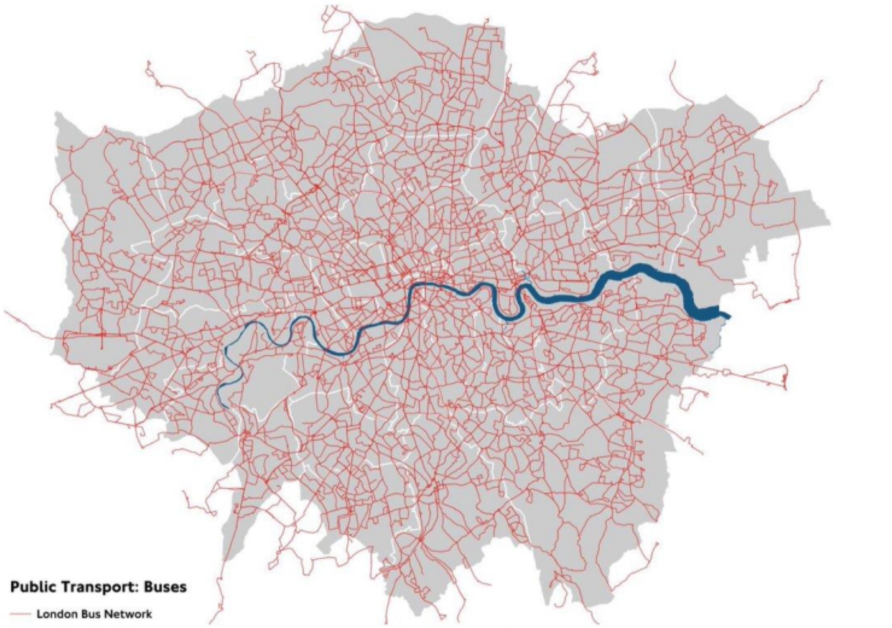


Figure 3 - The London Bus Network (source TfL, 2019e)

TfL requires a thorough understanding of the bus customer experience to ensure that services are planned and operated to focus on the customer. This requires bus performance measures to shift towards the customer perspective so they can identify where there are journey time, reliability and crowding problems and determine potential locations for improvement across the bus network. This would focus interventions where there is the largest benefit to customers (Bagherian et al, 2013; Gittens and Shalaby, 2015). However, current measures of public transport performance, particularly on buses, tend to be largely operationally focused. Current TfL metrics focus most on locations where there are the most buses operated, not where there are the most customers. Additionally, the two main metrics of bus performance (Excess Wait Time (EWT) and Bus Speeds) measure each journey component in isolation, with no view of the overall customer experience. There is a requirement for a new holistic bus performance metric, enabling TfL to plan, manage and intervene where it is needed most. TfL would gain a more detailed insight into the decline of bus patronage and be able to ensure reliable bus services effectively contribute towards the Mayor's 80% sustainable travel target.

1.2. Structure of Study

This study is split into a series of chapters. Chapter 2 evaluates the existing literature to understand the potential and need to measure bus performance from the passenger's perspective. Existing TfL performance measures will be discussed, exploring the possibilities for a new metric and identifying the key components of a customer's journey. Chapter 3 outlines the methodology, highlighting the available data sets for analysis and identifying data processing methods for manipulating large quantities of data from both automatic-vehicle location and smart-card data. Chapter 4 sets out the metric development, outlining the methods used to capture each element of the customer experience and defining the overall outputs. It demonstrates the technicalities and various levels of disaggregation. Chapter 5 analyses the preliminary metric outputs for validation. This assesses whether the metric achieves its purpose, looking at how it aligns with existing metrics and exploring how the spatial and temporal disaggregation effectively captures the holistic customer perspective. Chapter 6 covers potential policy implications, identifying existing policy gaps and how application could effectively contribute to TfL's strategic goals. The study concludes that BCJT is an effective measure of the customer perspective, providing a customer insight at both network and disaggregated levels. It demonstrates that the metric can be used to ensure that the bus network in London is planned, operated and managed to ensure the best overall customer experience for Londoners and encourage increased sustainable travel.

2.0. Literature Review

2.1. The Bus Customer Experience

To understand how bus service provision impacts the customer, the most important components of the bus customer experience must be identified. All transport providers aim to increase the attractiveness of their services and encourage new and existing passengers to meet ambitious sustainable mode-share targets (Diab et al, 2015; Jabareen, 2006). The drivers of passenger demand must be understood, as despite ongoing efforts to improve the London Bus Network, which saw unprecedented growth in both demand and coverage between 2000 and 2013, since 2015 passenger demand has been decreasing with an 8% passenger decline between 2015 and 2019 (SDG, 2010, TfL, 2019d).

A holistic view of each customer journey is needed to ensure that service provision is designed and managed systematically to benefit the customer (Carreira et al, 2013; Gentile et al, 2007). The customer experience involves any 'internal or subjective response that customers have to any direct or indirect contact with a company' (Meyer and Schwager, 2007, 118) and therefore measuring it requires an understanding of the range of experiences or interactions that all customers have over time rather than just on average. To understand how this can impact upon customer behaviour and demand, for example the choice a customer makes to take the bus, requires analysis to go beyond and include more complexities than just traditional service quality (Paulley et al, 2006). This means capturing both perceptual and emotional experiences alongside the operational service, for example how crowded the bus was as well as whether it arrived on-time (Carreira et al, 2013; Trompet et al, 2011).

Previous studies analysing the priorities of bus passengers have concluded that the most important customer concern is their ability to plan and thus the reliability and punctuality of service (SDG, 2010). A customer's perception is heavily dependent on whether a service is run frequently, on-time and without disruption whilst also being convenient and comfortable. Customers place high value on being able to consistently rely on how long it will take to reach their destination (Kwon et al, 2014).

Additionally, customers are not able to relate to a journey that they themselves have not experienced, for example the average customer experience is not something that the majority of passengers will get. The customer perspective is more focused on extremes rather than averages, with the ability to accurately estimate arrival times at destinations central to a customer's choice to take the bus (Gittens and Shalaby, 2015; Nam et al, 2005). Predictability and consistency are central components of the bus customer experience, with customers relying more on actual experience rather than scheduled information when planning journeys (El-Geneidy et al, 2011). Transit agencies must look beyond the ability to operate the schedule when trying to account for the customer perception of services. Incorporating a reflection of the overall range of the customer experience provides an understanding of the customer impact and the influence this has on passenger decision-making to use bus services.

With reliability and frequency identified as the most important components of the customer experience, the relationship between service provision, reliability and patronage needs to be understood, aligning service performance indicators to capture the passenger perspective (Bates et al, 2001; Diab et al, 2015). Capturing reliability is central to this requirement, however this must not be at the expense of capturing as many of the other aspects of the customer experience as possible, for example a customer's perception of different parts of the journey and comfort. An assessment is required to identify the technical possibility of using existing data sets held by transit organisations to fulfil this requirement (Peek and Van Hagen, 2002).

A major challenge is that public transport is not experience-centric. Paying to travel on a public transport service is unlikely to be a customer's primary objective, with travelling a means of reaching a destination (Carreira et al, 2013). As such, time spent waiting for and using public transport is not perceived by customer's to have been valuably spent (Larson, 1987; Leclerc et al, 1995). As Graham (1981, 336) outlined "time is money" and as a scarce resource it cannot be regained when it is lost. To maximise the customer experience, travel times must be both minimal and consistent. The ability to plan journey times is especially aversive to the value of time, demonstrating the importance of consistency on positively influencing a passenger's experience and increasing the customer's trust in a transport company (Abkowitz, 1978; Walker, 2012). This does not only apply to journeys which take longer than expected. Arriving early is also not necessarily a positive customer experience, as this is time that the customer has already spent and cannot be reused for a productive purpose (Gross, 1987). Effective customer measures must include a consistency measure and the customer's value of time.

2.2. Existing Measures of Bus Performance

This discussion highlights the drawbacks of existing bus performance measures, particularly the disconnect between operational and passenger perspectives. To capture the customer perspective each customer journey must be measured rather than the average operational service provision (Camus et al, 2005). It is this lack of passenger orientation that has been linked to a passenger's inability to relate to how bus service performance metrics reflect their experience (Trompet et al, 2011; Currie et al, 2012). Existing measures tend to be heavily based around averages and service deviation from the schedule rather than accounting for actual passenger journeys and how these are experienced and perceived by individual users (Diab et al, 2015; Gittens and Shalaby, 2015; Van Oort, 2014).

Traditionally, gaining an insight into the passenger perspective has involved the use of qualitative surveys, conversations and discourse analysis of social media (Diab et al, 2015). The TfL Bus Customer Satisfaction Score (CSS) is an example of this, using a sample of bus passengers to gain an understanding of how satisfied customers are with bus services. Whilst the CSS identifies five main drivers of customer satisfaction and aims to measure the overall customer experience (see Figure 4) (TfL, 2015), it is for a limited number of services and is not available daily. However, the findings from the CSS exemplify the importance of reliability to the customer experience, highlighting that a bus that gets a passenger to their destination on-time can be given a 10/10 score regardless of the other factors (TfL, 2014). An unreliable service consistently has a larger impact on the satisfaction scores than any other driver, therefore demonstrating the need for inclusion of reliability in any performance metric (TfL, 2015). However, alongside this, the CSS also highlights the importance of other qualitative factors of the customer experience, for example human factors like customer service. Whilst it would not be possible to cover all components of this with quantitative analysis, it is possible to use value of time weightings to account for customer perception, contributing to a measurement of stress, journey ease, human and personal comfort factors (Ryan, 1996). Measuring factors, for example vehicle temperature and driver behaviour, is not directly related to service performance and thus is not a necessary component for inclusion in a new service performance metric.

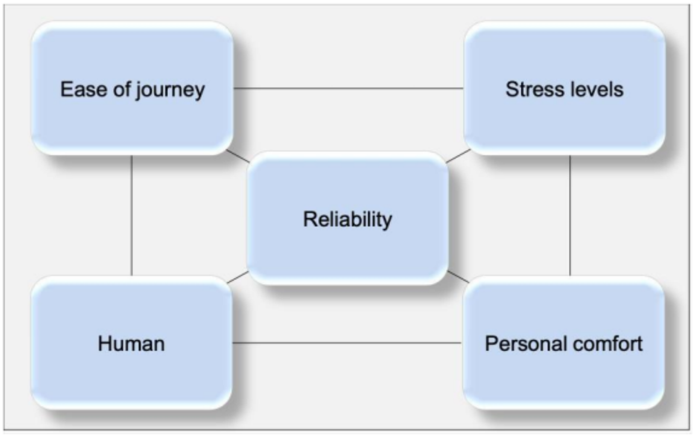


Figure 4 - Five Core Principles of the Bus Customer Experience (TfL, 2014)

However, despite the direction of the qualitative analysis to encompass the breadth of the customer experience, quantitative measures have continued to be dominated by the operational perspective. Transport for London, like most other bus companies around the world, lack a holistic customer experience measure, instead focusing on headway (regular service interval) adherence (EWT) and average bus speeds individually. Focusing on the averages rather than extremes does not capture the overall customer experience and lacks relativity from a passenger perspective (Schil, 2012).

EWT is measure of regular headway (gaps between services), looking at the deviation of the average actual headway from the scheduled headway and weighting by the number of observed buses. The output is a value in minutes which represents the average amount of additional time that customers have to spend waiting for services to arrive compared to what is defined in the schedule, due to buses not arriving at regular intervals (TfL, 2019b). The use of averages however and a focus primarily on the wait time element of the journey means that this does not capture the whole passenger experience. The focus of the measure is on operational ability to ensure that journeys arrive at regular intervals, with no measurement of the resulting impact this has on customer's overall travel times (Schil, 2012). This demonstrates a disconnect between the operational and the passenger perspective, as an operator actively trying to increase their service performance through lowering their EWT score, can have a detrimental impact on the overall customer experience through elongating customer travel times (Diab et al, 2015). For example, where there are uneven gaps between services, buses will often be held at bus stops to regain a regular interval. However, for passengers on-board the held bus this will increase their in-vehicle times and reduce their overall journey time reliability because of an inability for passengers to plan this kind of occurrence into their journey time (Fattouche, 2007).

TfL also uses Bus Speeds, measuring the average speed of buses across the network including dwell time (at stops etc.) (TfL, 2019c). Whilst this captures a different journey component to EWT, measuring in-vehicle times, there is no measure of how consistent the speed is and how much variability there is from the average. Whilst this enables TfL to monitor the impacts of changing network conditions to buses, for example showing where there has been speed deterioration over time, it is not relatable to the actual customer experience (El-Geneidy et al, 2011; Barron et al, 2013).

The disconnect between the operational and passenger perspective is highlighted further when considering the weighting mechanisms (Diab et al, 2015). Both bus speeds and EWT are weighted by service provision, i.e. scheduled mileage or the number of observed buses, rather than the number of passengers impacted. Neither metric is weighted to reflect the busiest points of the network, demonstrating the need for a new measure with the objective of aligning these perspectives and measure service performance with consideration of the actual customer experience (Ehrlich, 2010).

Furthermore, traditionally TfL has defined a passenger journey as a single boarding and alighting activity on one service, with no account of where a passenger may use two or more bus services to reach their destination (TfL, 2019b), demonstrating another missing component of measuring the overall customer experience (Goedkoop et al, 1999).

2.3. Opportunities for Capturing the Customer Experience

To holistically capture the customer experience, TfL must learn from its existing qualitative and quantitative performance measures and collate them. The lack of focus on extremes is a weakness of existing performance metrics and their applicability to the customer experience (Bagherian et al, 2016). Customers are more interested in actual rather than scheduled information with an acknowledgement that both schedules and averages are not relatable to the actual customer experience (see Figures 7 and 8) (Barron et al, 2013; El-Geneidy et al, 2011). Uniman et al (2010) highlight that customers account for the 'worst case' scenario when planning journeys. Predictability is a highly valued component of reliability perception (Bates et al, 2001). This is supported by the conclusion that a customer is likely to choose a journey which is longer yet consistent than faster and highly variable (Polus, 1978). Therefore, to effectively capture customer journey time, actual travel times and their variability for each origin and destination (start and end point of a customer journey) must be accounted for.

The availability of automatic-vehicle location (AVL) data enables recording of the range of travel times experienced by customers, with an actual time between each origin-destination pair, and therefore the calculation of journey time variability (Furth, 2000; Hollander and Buckmaster, 2009). Additionally, the increase in the availability of smart-card data means there is the opportunity to incorporate dynamic demand flows into bus performance measures (Pelletier et al, 2011; Nassir et al, 2015). This could weight measures to reflect where service performance is experience by the highest volume of passengers as well as the ability to track how passenger use bus services (Eboli and Mozzulla, 2012). There also provides a better understanding of travel patterns where there are interchanges between services and bus loadings.

2.4. Measuring the Customer Experience

Despite demonstrating that advancement in data technology has opened up more possibilities for performance management, the literature has shown that there is a gap in current bus performance management, with a focus on operations rather than the customer perspective (Carreira et al, 2013; Gentile et al, 2007). With reliability as the most important journey component of the customer experience this must involve not only a measurement of actual travel times but also of journey time variability, alongside an account for how journeys are perceived by the customer (Uniman, 2009; Leclerc et al, 1995, Ryan, 1996).

To be used as a measure of service performance, a new measure must be able to systematically utilise data for all bus services across the network, bringing together data sets which have not previously been used in combination. This will remove the disconnect between the operational and customer perspectives and incorporate as many of the features traditionally captured using qualitative measures as accurately as possible (Diab et al, 2015; Raoniar et al, 2015).

Available data sets enable the calculation of the following journey components:

1. Travel Time – the actual travel time between one point and another, including wait times, in-vehicle times and interchange time (Van Oort, 2014).
2. Journey Time Variability – the amount of time that a customer's journey can vary by – providing an indication of the amount of extra time a customer should allow as a buffer to secure on time arrival (Uniman et al, 2010).
3. Perception – the value of time of each of the journey components – i.e. the customer perception including wait time, interchange penalties and on-bus crowding (Ryan, 1996).

2.5. Research Objective

This project will aim to increase the understanding of bus performance from the customer perspective through formulating a methodology for a new TfL bus service performance metric. It will then complete a validation of outputs and potential policy applications to assess the effectiveness of combining operational and customer perspectives to measure bus performance.

2.6. Research aims:

1. To formulate an effective methodology to capture the customer experience using a quantitative metric.
2. To validate results from a new metric by demonstrating relationship with existing metrics and demonstrate the use cases of a more detailed insight into bus performance.
3. To identify potential applications and policy implications of a new bus performance metric.

3.0 Methodology

3.1. Formulation of Methods

The primary objective of this research was to fulfil three main purposes: the formulation of a new metric, its validation and its potential application for TfL. The majority of the methodology was based around the manipulation and development of existing data sets owned by TfL, with Visual Basic (VBA) and Python coding used to manipulate large quantities of data in combination and formulate a new measure of bus performance which represents both quantitative and qualitative components of the customer experience (Cresswell, 2009). For validation, the new metric has been tested against existing measures of bus performance, demonstrating the additional utility this measure can have at realising TfL's policies. Following this a set of potential policy implications for future applications of the metric have been identified.

3.2. Data Sets

This thesis was reliant upon data sets owned by TfL. The availability of GPS systems and smart card data facilitated the shift to focus on passenger flows in performance metrics (Bagherian et al, 2011). The dependency on manual survey analysis to understand either transit passenger usage or measure bus travel times has been removed with the availability of automatic vehicle location (AVL) data. This provided a comprehensive set of bus travel times and smart card data and enabled an understanding of all passenger movements (Mazloumi et al, 2009; Pelletier et al, 2011). Whilst existing bus performance metrics have made use of AVL data's ability to automatically report locational and travel time data, previously there has been little collaboration with smart-card data which enabled the shift from operational to customer perspective in performance management.

The development used the following datasets:

- London Reporting Database (LRD) (TfL, 2019g) - this database holds TfL's iBus system. Implemented in 2007, iBus is a GPS AVL system installed on all 8000+ TfL buses (Wong and Hounsell, 2010). Each bus has an on-board computer connected to a data server and feeds real-time databases. This informs network management and customer information such as bus stop countdown signs. This data could measure travel times for all origin-destination points, with automatic reporting of service headways, journey times and travel speeds. The potential for utilising this dataset for increasing TfL's understanding of the bus network has been demonstrated by other projects. For example, Hardy (2009) showed how bus priority location could be prioritised by identifying areas with regular service delays.

- ODX (TfL, 2019h)– this database collates journey information from all Oyster and contactless payment card transactions. Despite passengers not being required to tap out when they exit a bus service, ODX contains an algorithm to infer over 70% of customer destinations and rounding for where inference was not possible. The model records the behaviour of each oyster/contactless card, using regular travel patterns and the next card use location to infer the alighting point (Wang, 2010; Sanchez-Martinez, 2016). The result is a complete origin-destination matrix for the London bus network, enabling calculation of both crowding and interchange.

These databases are maintained and owned by TfL, meaning no data cleaning is necessary before use within metric development. My position as a permanent employee at TfL, resulted in fewer gatekeeper obstacles, which could 'grant or withhold access…for the purpose of research' as I already have access for the purpose of my work and agreement from TfL to utilise the data for the purpose of this study (Burgess, 1984 in Valentine, 1997, 148).

3.3. Data Handling

With over six million daily bus journeys in London, this metric required a large amount of data processing. To do this effectively it was built using VBA and Python coding. VBA enabled a user-friendly tool interface to be made using Microsoft Excel (which is available to all TfL employees), meaning that the metric can be readily accessed across the business (Walkenbach, 2010). Additionally, VBA has the ability to call Python, meaning that data processing was completed outside of Excel and then read in using VBA code. This enabled the extraction, processing and reporting of results within one Excel user interface (Sweigart, 2015).

The calculation process is outlined in Chapter 4 with all coding included in Appendix B.

3.4. Case Study

For metric validation, outputs have been tested in relation to results from existing bus performance metrics. This showed how BCJT results both correlate and add extra information to existing data sets. A regression analysis also demonstrated this relationship. Additionally, route 62 has been used to exemplify the potential of BCJT's spatial and temporal disaggregation, showing how the detail captures the customer perspective and thus provides insights to plan targeted interventions and effective contributions to TfL's strategic policies. This section of route was chosen as it had high variability in BCJT performance across different stops and times of travel.

3.5. Research Ethics

This research did not involve any ethical issues. The methodology relied on the use of TfL's "ODX" system which stores historic journey information for all smartcard users on the London Transport Network. To protect personal data, the lowest level of journey information disaggregation is hourly meaning no possible identification of individual customers. Access to this data was enabled through my position as a permanent employee at TfL. No publishing of personal data is included.

4.0 Building the Bus Customer Journey Time Metric

4.1. Metric Development Introduction

This chapter will outline the calculation process of using the identified data sets to develop a customer focused performance metric. With the literature identifying reliability and journey times as the most important aspects of a passenger journey, this section will focus on how available data can capture the customer experience into one overall service performance metric.

The components of customer journey time to be included in a new customer focused metric are:

- Wait Time
- In-Vehicle Time
- Interchange
- Crowding
- Buffer Time (journey time variability)

This is alongside ensuring that customer demand is accounted for throughout the process. To move away from a focus on service provision towards a focus on service demand, it is vital that weighting is given to the busiest points on the network. Through doing this, areas where demand is high but provision is low will become apparent, increasing the visibility of where there is potential to optimise service provision for customers. It will also highlight areas with capacity to reallocate resource. Additionally, a value of time will be allocated for each of the journey components, enabling assessment of how bus journeys are perceived by customers and providing an ability to balance and assess the impacts of improvements and interventions.

The experience of travelling on the London bus network can vary depending on the time of travel, with different times of day and days of week having different travel conditions. Therefore, in order to effectively capture each customer experience, there is a need to measure the bus customer experience at each point of the day. To do this effectively each hour on every day (from 5am to midnight) is measured individually with all averages being taken across an hour and then weighted by demand to scale to daily, weekly and periodic results (see section 4.7). Using an hour, ensures that variability is effectively measured at comparable times, accurately enabling the calculation of both the average travel time and a buffer time that customers would need to plan at the time of day they wish to travel at (see Figure 5).

Furthermore, for accuracy and validity of outputs, an hour ensures that at least 5 buses have passed through each origin/destination point, providing a balance between giving enough trips to calculate an average and buffer travel time and retaining connection to the specificity of customer journey planning. This also removes the impact of averaging across times/day types with different travel conditions which would smooth out the extremities that are experienced by large numbers of passengers at particular times.

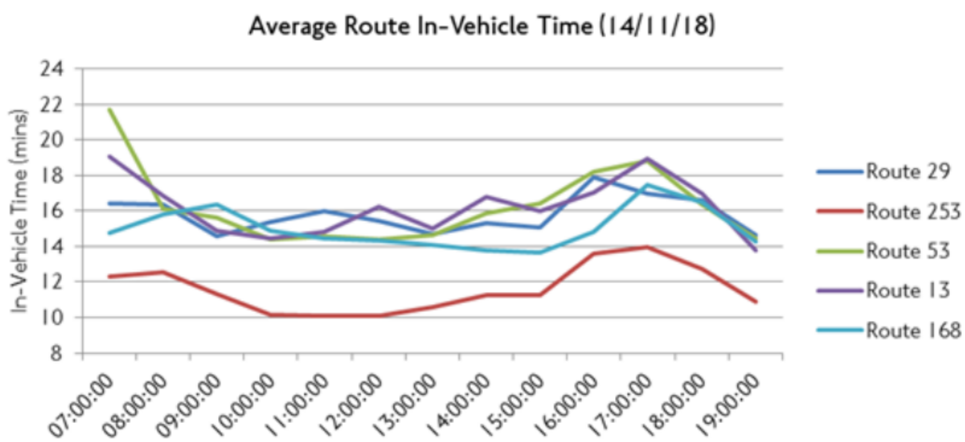


Figure 5 - Hourly In-Vehicle Time (14th November 2018) on a sample of 5 routes

4.2. Wait Time

Even though there is no data automatically recorded for wait times, there is a need to estimate as accurately as possible the amount of time a customer waits for a bus service. As the first customer interaction with the bus network, the amount of time spent waiting for services and the frequency of services have a large impact on a customer's decision to take the bus (Schil, 2012). The literature has shown that traditional measures of bus performance have had a large emphasis on wait times demonstrating the importance of their accurate inclusion in an overall measure of the customer experience. However, traditional measures have primarily focused on how wait times vary on average from the scheduled service, with little account for the actual total times customer spend waiting for bus services (see Figure 7). To be able to measure a bus customer journey time, there is a need to move away from schedule deviation as the main reference point and instead account for each customer experience of waiting at a bus stop.

iBus data from the LRD provides information on the arrival time of each service at each stop, enabling a measurement of an estimated wait time for every customer boarding every service across the network. By using half of the actual headway, the variation between services is captured, with the whole wait time accounted for rather than the average difference from the schedule. As there is no record of when passengers arrive at bus stops, this relies on the assumption that on average customers wait for half of the actual headway between services. For this reason, this methodology is only applicable for high frequency services (operating at 5 buses per hour or more) as this assumes that there is uniform arrival by passengers at stops throughout the whole of the gap between services. For example, for a service which operates 6 buses per hour this model would assume that on average customers wait 5 minutes for a service. However, on a bus which operates, 1 bus per hour, it would assume average wait time of 30 minutes – which is unrealistic as customers are likely to have service information. Whilst there are some limitations to this approach, for example it does not account for the use of real-time information applications, this methodology provides a mechanism for capturing the element of wait time for every operated service, rather than relying on sampling.

Additionally, it is well researched that wait time is perceived as longer than in-vehicle time for bus passengers with Quarmby (1967) concluding that walking and waiting times are worth between two and three times in-vehicle times. Additionally, in 2016 TfL commissioned research which concluded the value of time weighting for wait time was to be reduced from 2.5 to 2.0 for buses, to account for the impact of live bus arrival information (e.g. at bus stops) positively impacting the customer perception (Accent, 2016). Therefore, for this metric a value of time weighting of 2.0 is given to the wait time element.

$$Wait\ Time\ = (0.5 * Actual\ Headway) * 2$$

4.3. In-vehicle time

In-vehicle time is the simplest element of a customer experience to measure. iBus records the arrival and departure time of every service at every stop in London which thus accounts for both movement time and the dwell time spent at bus stops. From this the actual travel time between each origin-destination pair on every route can be accurately calculated to account for all customer travel times between every point on the network. This is calculated for each hour with the in-vehicle time as the average travel time between each origin-destination pair on the network and is then scaled up by demand geographically and temporally (see section 4.7).

$$In - Vehicle\ Time\ = Arrival\ time - Departure\ Time$$

4.4. Interchange

Existing measures of customer interchange (transferring from one bus to another) have continued to focus on surveys and models with little automated understanding of a customer's whole journey from origin to destination (Ehrlich, 2010). With the availability of smart-card data, the ability to track complete customer journeys across multiple services is now increasingly possible (Bagherian et al, 2013, Vanderwaart, 2016). Therefore, the inclusion of interchange into a bus network performance metric is essential.

However, with the previously explained methodology for wait times, the metric is already capturing an estimate of the wait time for each individual leg of customers' journey(s). It is not possible to use ODX data in isolation to calculate customer interchange time as this would mean that time spent waiting for any subsequent services would be double-counted (Nassir et al, 2015). It is necessary to calculate the amount of additional time customers spend interchanging between services. As such, the metric calculates the amount of time that it takes to walk from the point of alighting the first leg to boarding the second (see Figure 6).

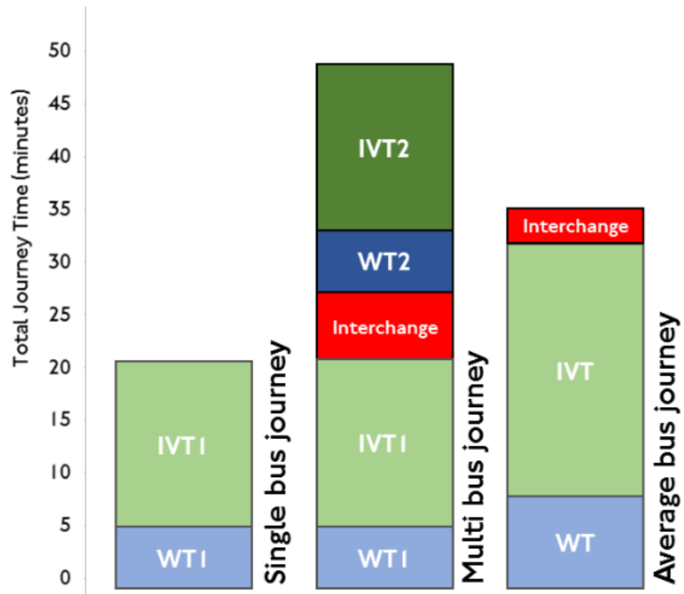


Figure 6 - Journey Leg Factor - scales up the average bus journey to be representative of the typical bus customer journey. (IVT = In-Vehicle Time, WT = Wait Time)

By extracting all bus stop information from iBus, the longitude latitude coordinates for each stop were fed through TfL's journey planner to calculate the average walk time between each possible interchange point on the network. This was completed for all stops with a different service pattern to ensure realism with where customers make interchanges in reality. Hourly ODX data was then used to identify the location and quantity of interchanges taking place on the network and weight the average walk times to generate an average hourly walk time where customer flows are accurately accounted for. This is then scaled to reflect the proportion of customer journeys which include an interchange, giving the hourly average interchange time on each bus journey in London.

A value of time weighting is then applied to ensure the metric captures the customer inconvenience of having to make multiple bus legs. Interchange has a value of time weighting of 2.0, doubling the walk time to account for customer perception (Quarmby, 1967). Additionally, an interchange penalty of 3.5 minutes is added and then scaled to reflect the proportion of customer journeys which encompass an interchange (Daly et al, 1973; Wardman and Hine, 2000, TfL, 2019i).

$$\textit{Network Interchange Time} = \frac{\left((\textit{Stop to Stop Interchange Distance}/\textit{Average Walk Speed}) * 2\right) * \textit{Number of Stop to Stop Interchangers}}{\textit{Total Network Demand}}$$

Additionally, to ensure the network metric is reflective of the complete bus customer experience, it is necessary to reflect the end-to-end customer journey, therefore measuring where journeys are comprised of multiple bus journey stages (Carreira et al, 2013). As ODX is able to infer where journey stages are made as part of a longer overall journey, it is possible to calculate the average number of legs comprising a bus journey in London during each hour. This is done by counting the total number of taps and comparing to the total number of complete journeys. The result is a journey leg factor – the average number of bus legs in each customer journey – and this is used to scale up the average bus customer journey time to better reflect the overall customer experience (see Figure 6).

4.5. Crowding

To ensure a new metric accounts for where the on-board experience is impacted by crowding, the methodology calculates how the value of time changes in line with crowding (Liu and Wen, 2016). Crowding on-board buses can have an impact on travel time for two main reasons, through reducing a passenger's ability to use travel time for other purposes, for example reading a book, and impacting on emotive factors such as personal stress, noise and security (O'Regan and Buckley, 2003). To generate a crowding factor for bus travel, the new metric utilises the crowding formula used by London Underground. This crowding factor inflates the value of average in-vehicle time with density, it is non-linear increasing as the square of density, reflecting how the experience changes for a customer as personal space decreases. The value of time begins to increase as the load exceeds seating capacity, with results ranging from 1 (where there is no standing) up to 2.6 (where crowding is at 90% crush standing capacity) (Crossley, 2003a; Crossley, 2003b).

The formula is as follows:

$$Crowding\ Factor = 1 + C0 + \left(C1 - CY\left(\frac{Seats}{Capacity}\right)\right) * (Pax - Seats)/(Capacity - Seats)$$

Where:

- C0 = 0.085
- C1 = 1.915
- CY = 1.03

C0, C1 and CY are defined by London Underground's crowding research and Seats, Capacity and Pax (vehicle load) are user input variables (Crossley, 2003a).

ODX data is used to calculate the load of each bus between each stop during each hour. By extracting the number of boarders and alighters at each stop, it is possible to calculate the departing load on each link. This information is then combined with the bus capacity information (extracted from iBus) in the above formula to provide a crowding factor for each link during each hour. This is then scaled by demand across each of the links and used to weight the hourly average in-vehicle time to give an additional average hourly crowded time for each route and the network. This is the only value of time weighting received by the in-vehicle time element of the metric.

$$Network\ \ or\ Route\ Crowding\ Factor = \frac{(Link\ Crowding\ Factor * Link\ Demand)}{Total\ Network\ or\ Route\ Demand}$$

$$Network\ or\ Route\ Crowded\ Time = Network\ or\ Route\ In - vehicle\ time * Network\ or\ Route\ Crowding\ Factor$$

4.6. Buffer Time (Journey Time Variability)

As Uniman et al (2010) highlighted it is not average travel times that customers have to plan for. To be able to make informed decisions about their travel choice, customers must account for the 'worst case scenario', adding additional time to account for possible journey delays to ensure that their travel choice does not make them late to their destination. Therefore, to effectively capture the customer experience it is important to capture the total spread of customer journey times (see Figures 7 and 8 below) (Barron et al, 2013).

The definition of the 'worst case scenario' is set at the 95th percentile. Uniman et al (2010) and Wood (2015) agree that this is an appropriate level as passengers find a once a month chance of late arrival as acceptable. This therefore removes any travel times that were impacted by a major event or disruptions and gives a realistic indication of the range of travel times that customers must allow for when planning their journeys. Therefore, as well as using iBus data to calculate the average hourly wait time and in-vehicle times, the new metric also calculates the 95th percentile wait and in-vehicle time in each hour for each origin-destination pair on the network. Buffer time is therefore defined as the difference between the 95th percentile travel time and the average travel time (see figures 7 and 8 below). This is calculated for each of the wait time, in-vehicle time and total journey time for each origin-destination pair on each route in each hour and is scaled by demand temporally and geographically (see section 4.7).

$$Buffer\ Time = 95th\ percentile\ travel\ time - mean\ travel\ time$$

This receives the same value of time weightings as the wait time and in-vehicle time respectively, i.e. wait time buffer receives a value of time weighting of 2 compared to in-vehicle time.

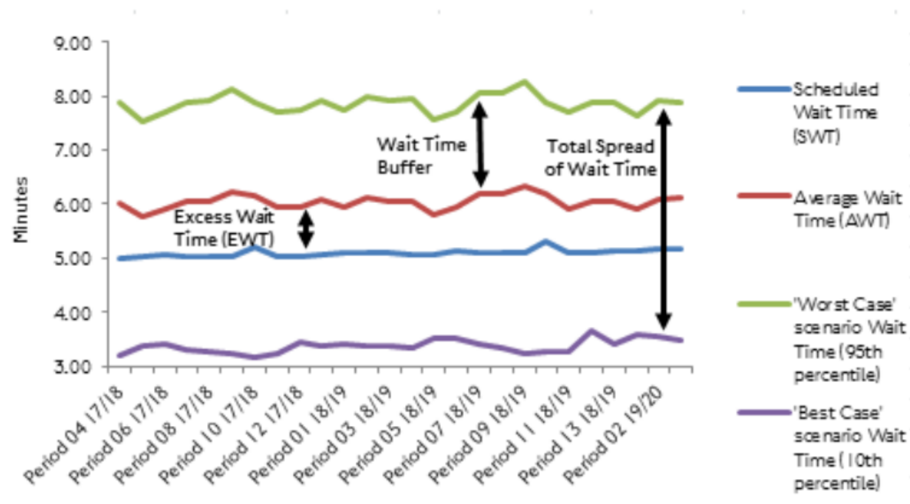


Figure 7 - Spread of Customer Wait Time

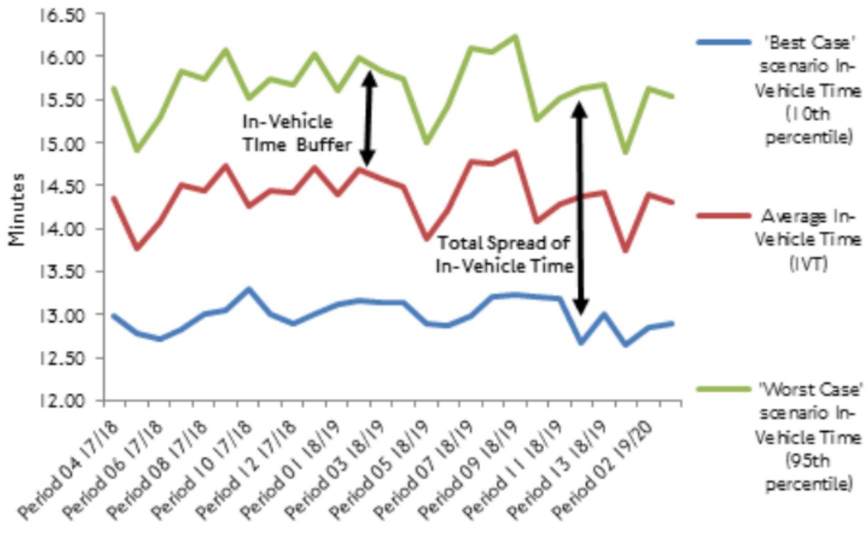


Figure 8 - Spread of Customer In-Vehicle Time

In addition to calculating buffer time to form a component of the overall metric, by capturing the spread of the total journey times experienced on each origin-destination pair, the metric also produces a normalised output. This compares the whole range of journey times, comparing the 'best' (10^{th} percentile) and 'worst' case (95^{th} percentile) scenarios to give a comparable output (the Excess Planning Time Index (EPTI)) which compares journey time variability over different routes, links and time bands (see formula below). This gives an indication of the proportion of the 'best' case scenario that customers must allow to give themselves 95% certainty that they will arrive on time. The larger the proportion the more variable the journey time (see Figures 7 and 8 for spread of total journey times).

$$EPTI = \frac{95th\,percentile\,travel\,time - 10th\,percentile\,travel\,time}{10th\,percentile\,travel\,time} + 1$$

4.7. Temporal and Geographical Aggregation by Demand

Each of the components in BCJT is calculated at the hourly level for each origin-destination. Demand data from the oyster card database is then used to scale this up both temporally and geographically to give results for the aggregated time periods, routes and the network. There are two main stages to this process:

1. Origin-destination to route level – this is done with an annual demand dataset recording the number of people travelling between each stop on each route across the network. This data is also used to scale temporal results for each origin-destination pair up to time period, daily, weekly, periodic and annual results.

2. Route to network level – total route demand is extracted for each hour on every day. This is used to aggregate results from route to network level and average across route level results temporally.

Demand weighting is given to the metric results to ensure the metric remains as accurately reflective of the actual customer experience as possible. Demand weighting ensures that the parts of the networks/times of day where there was the highest number of customers travelling receives the biggest contribution to the overall metric result, as this reflects the average result each customer experienced in reality, avoiding averaging out from area/times where demand is low but the network performs well.

$$Demand\ Weighted\ BCJT = \frac{(Route\ or\ Hour\ BCJT * Route\ or\ Hour Demand)}{Total\ Route\ or\ Hour\ Demand}$$

4.8. Metric Outputs

Once all of the components have been calculated these are combined to give one overall measure of the customer experience (bus customer journey time in “weighted” minutes). The scaled result across all the origin-destination pairs, hours, time bands, days and routes on the network gives one overall number which can be used to track the bus customer experience over time (see Figure 9). This can then be disaggregated to any level to enable detailed performance management of particular time periods, routes or links on the network as well as understand the breakdown of the different journey components (see Figure 10). This demonstrates the possibility of measuring service performance from the customer perspective and through using the BCJT alongside the EPTI measure, the normalised output can be used to compare locations across the network and ensure performance management is targeted and thus encourage sustainable mode shift.

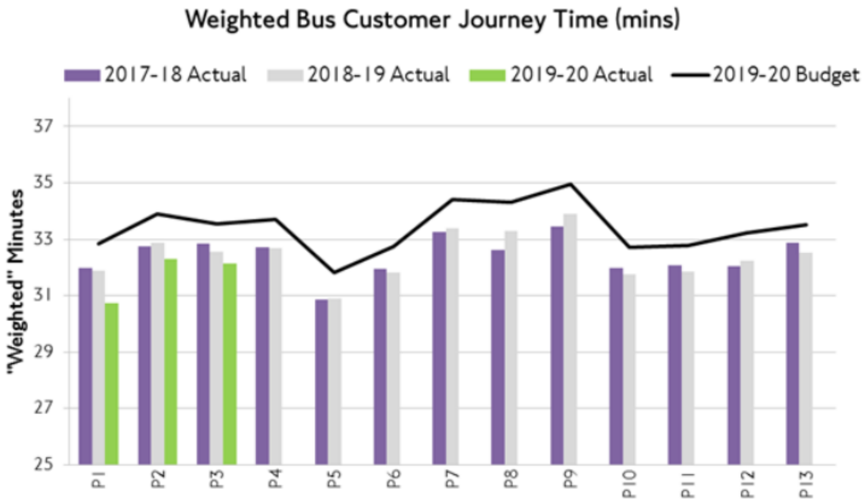


Figure 9 - Periodic Weighted Bus Customer Journey Time for the London Bus Network

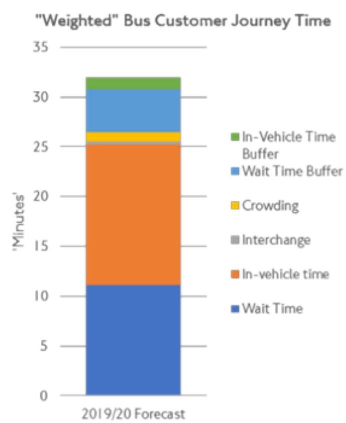


Figure 10 - Weighted Bus Customer Journey Time Components

5.0 Validating the Bus Customer Journey Time Metric

This chapter will compare BCJT results against the outputs from existing metrics to both validate the BCJT outputs and demonstrate the differences and enhanced potential for gaining a more holistic understanding of bus performance measured from the customer perspective. This chapter will be split into two main parts:

1. Validation with existing performance metrics – highlighting the utility of the additional components of BCJT. This will involve validation both in terms of correlation and also highlight parts of the customer experience previously missed by EWT and Bus Speeds.

2. Additional Insight of BCJT – the new metric gives a greater level of disaggregation, enabling more targeted interventions and offering insights into different components of the customer experience.

5.1. Validation of Bus Customer Journey Time with Existing Performance Metrics

BCJT is closely aligned and builds upon the outputs of existing performance metrics. One of the main advantages is that it balances both EWT and Bus Speeds into one overall measure, understanding how improvements/compromises to each impact the customer experience. Figure 11 demonstrates the relationship between BCJT, EWT and Bus Speeds. It shows that as EWT decreases (improves), there is a very strong positive correlation with BCJT (0.94). However, as the graph shows the reduction in BCJT is less than EWT, exemplifying both the influence of other journey components and the extent of the impact of EWT on the overall customer experience.

BCJT and Bus Speeds have a strong negative correlation, meaning that as bus speeds increase (improves), BCJT decreases (also improves). The correlation between them is -0.84. This means that BCJT is impacted by both of the existing performance metrics and as such the influence of an improvement of EWT on BCJT will be minimised if such an improvement has a detrimental impact on bus speed. For example, in P3 19-20 (as shown in Table 1 and Figure 12), the average EWT has decreased (improved) by 0.26% compared to P2 19-20 whereas BCJT has only decreased (improved) by 0.10% and this is due to the 0.07% decrease (worsening) of Bus Speed.

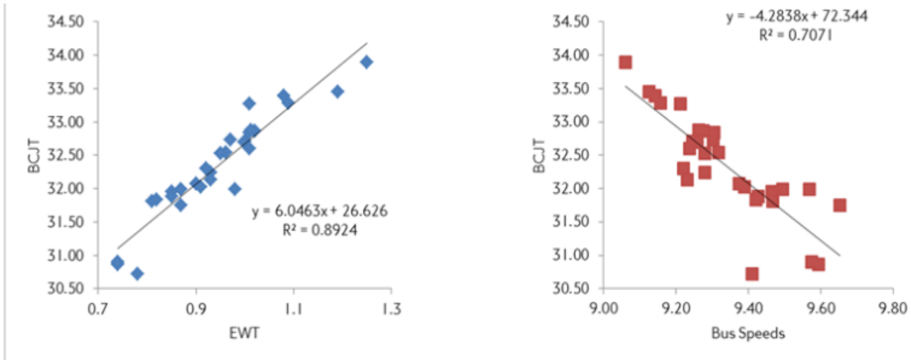


Figure 11 - Scatter Graphs to Demonstrate Relationship between BCJT and EWT and BCJT and Bus Speeds

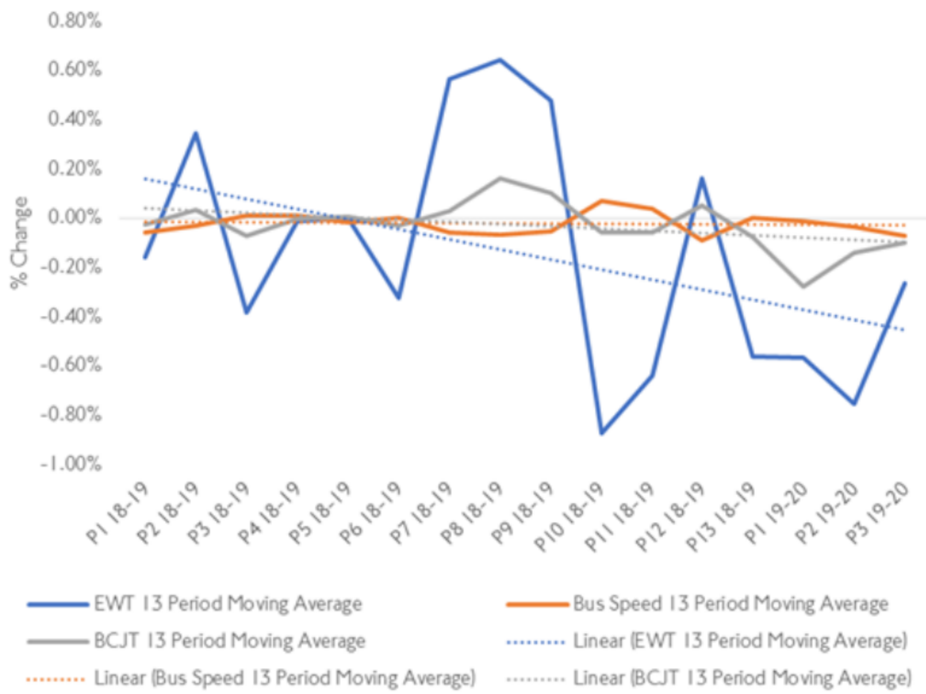


Figure 12 - BCJT compared to Bus Speed and EWT Change in 13-period Moving Average

Table 1- Performance Metric Periodic Results and % Change in 13 Period Moving Average

Financial Period (4 weeks)	EWT	Bus Speed	BCJT	EWT 13 Period Moving Average (% change)	Bus Speed 13 Period Moving Average (% change)	BCJT 13 Period Moving Average (% change)
P1 17-18	0.87	9.50	31.99			
P2 17-18	0.97	9.30	32.73			
P3 17-18	1.01	9.30	32.84			
P4 17-18	1.00	9.25	32.71			
P5 17-18	0.74	9.59	30.87			
P6 17-18	0.85	9.47	31.95			
P7 17-18	1.01	9.21	33.27			
P8 17-18	1.01	9.24	32.60			
P9 17-18	1.19	9.13	33.46			
P10 17-18	0.98	9.57	31.99			
P11 17-18	0.90	9.38	32.07			
P12 17-18	0.91	9.39	32.03			
P13 17-18	1.02	9.28	32.86			
P1 18-19	0.85	9.43	31.88	-0.16%	-0.06%	-0.03%
P2 18-19	1.01	9.26	32.88	0.35%	-0.03%	0.04%
P3 18-19	0.96	9.32	32.54	-0.38%	0.01%	-0.07%
P4 18-19	1.00	9.26	32.69	-0.01%	0.01%	0.00%
P5 18-19	0.74	9.58	30.90	0.00%	-0.01%	0.01%
P6 18-19	0.81	9.47	31.81	-0.32%	0.00%	-0.03%
P7 18-19	1.08	9.14	33.39	0.56%	-0.06%	0.03%
P8 18-19	1.09	9.16	33.29	0.64%	-0.07%	0.16%
P9 18-19	1.25	9.06	33.89	0.48%	-0.05%	0.10%
P10 18-19	0.87	9.65	31.75	-0.87%	0.07%	-0.06%
P11 18-19	0.82	9.42	31.83	-0.64%	0.04%	-0.06%
P12 18-19	0.93	9.28	32.24	0.16%	-0.09%	0.05%
P13 18-19	0.95	9.28	32.53	-0.56%	0.00%	-0.08%
P1 19-20	0.78	9.41	30.72	-0.57%	-0.01%	-0.28%
P2 19-20	0.92	9.22	32.30	-0.76%	-0.04%	-0.14%
P3 19-20	0.93	9.23	32.13	-0.26%	-0.07%	-0.10%

Despite the strong parallel with existing performance metrics however, BCJT incorporates more than EWT and Bus Speed, also accounting for other elements of the customer experience. The following regression analysis confirms this (see Table 2). Whilst this model validates the strong interaction between BCJT, EWT and Bus Speeds, demonstrating its accuracy in weighting and balancing the results of both, the adjusted R^2 value 0.9 shows that 10% of BCJT is not explained by EWT and Bus Speeds. This reflects the additional impact of journey time variability, interchange, crowding, customer perception and demand weighting on the overall BCJT result. The p-value of 0.06 for bus speeds also confirms that this model is not statistically significant as it is over the 0.05, 95% significant level (Bross, 1971). This again demonstrates the importance of other journey components on the overall customer experience (as shown in the metric output breakdown in Figure 10).

Table 2 - Multiple Regression Analysis of BCJT with EWT and Bus Speeds

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.95							
R Square	0.91							
Adjusted R Square	0.90							
Standard Error	0.24							
Observations	29.00							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2.00	14.92	7.46	125.47	0.00			
Residual	26.00	1.55	0.06					
Total	28.00	16.47						
BCJT	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	37.35	5.51	6.78	0.00	26.02	48.68	26.02	48.68
EWT	4.98	0.67	7.42	0.00	3.60	6.35	3.60	6.35
Bus Speed	-1.04	0.53	-1.95	0.06	-2.14	0.06	-2.14	0.06

5.2. Additional Insights of BCJT

Furthermore, BCJT provides a more disaggregated insight into bus performance than previously available. This means that users are able to access data with increased spatial and temporal accuracy with data available for each of the journey components on each stop-to-stop movement in each hour on a daily basis. For example, Table 3 shows an example extract for 14:00 on each day type between 3 stops on route 62.

The variability in the results show that on Sunday there were significant reliability issues between stop 19300 and 19298, with average in-vehicle time up to 6.45 minutes compared to 2.23 minutes on school days and 4.63 on Saturday. This was alongside an increased in-vehicle buffer. Looking at this in context (Figure 13) this movement involves a left turn onto Lodge Avenue, and therefore this could highlight a problem with the signal timings on Sundays for example. By having the data at this level, it is possible to identify exactly where and when reliability issues arise and by following these results and looking for trends, TfL would be able to use this information to make targeted interventions across the network exemplifying the effectiveness of measuring the customer perspective of bus performance.

Table 3 - BCJT Select Link Analysis for Route 62 (Marks Gate to Barking) from Stamford Road Health Centre (20392) to Maplestead Road (19296) - 2-3pm 12/11/18-18/11/18

DAY TYPE	HOUR	FROM	TO	LOAD	BOARDERS	EPTI AWT	EPTI IVT	EPTI TJT	Average Wait Time	Average In Vehicle Time	Crowding	Wait Time Buffer	In-Vehicle Time Buffer
Saturday	14:00	20392	19300	106	8	2.54	1.78	2.30	5.76	1.52	0.00	1.52	0.81
Saturday	14:00	19300	19298	100	8	1.86	1.76	1.81	5.53	4.63	0.00	1.84	1.59
Saturday	14:00	19298	19296	68	8	2.47	1.49	2.18	4.87	1.50	0.00	2.27	0.30
Sunday	14:00	20392	19300	45	8	2.06	1.28	1.94	10.03	1.43	0.00	3.62	0.13
Sunday	14:00	19300	19298	49	5	1.93	6.65	2.72	7.27	6.45	0.00	6.44	3.10
Sunday	14:00	19298	19296	44	2	2.18	1.94	2.15	7.92	1.52	0.00	5.73	0.20
School Day	14:00	20392	19300	401	40	3.26	1.60	2.71	6.17	1.61	0.00	1.75	0.31
School Day	14:00	19300	19298	399	25	2.88	1.87	2.50	5.16	2.23	0.00	2.78	0.90
School Day	14:00	19298	19296	376	31	3.10	1.84	2.79	5.81	1.21	0.00	2.44	0.42

Figure 13 - Route 62 (Marks Gate to Barking Gascoigne Estate). Zoomed area shows stops included in Table 3 (TfL, 2019f).

6.0 Policy Implications of the Bus Customer Journey Time Metric

The previous two chapters have demonstrated the capability and validated the effectiveness of a new bus performance metric. There are several potential policy implications and applications of BCJT due to its customer perspective (Paulley et al, 2006). Trompet et al (2011) highlight the limitations of the current service performance metrics, demonstrating how they are unrelatable to the actual customer experience. This chapter will explore how BCJT could be effectively incorporated into bus service planning and performance management.

6.1 Network Performance Tracking – Trend Analysis

Until 2019, EWT was the bus performance measure included on the TfL scorecard. In 2018/19, bus performance was the best it had ever been with EWT achieving an all-time low score of 0.95 minutes (TfL, 2019b). However, at the same time the qualitative CSS for reliability actually dropped 2%– from 84% to 82%, and the increase in passenger demand historically aligned with increased reliability was not realised. Therefore by switching to use the BCJT for tracking overall network performance, TfL would have a much broader understanding of the actual customer experience and be able to understand how each component of a bus customer journey is impacting the customer, not just the average EWT element and ensure an increase in EWT was actually improving the customer experience (Diab et al, 2015). BCJT is now included on the TfL Scorecard, providing a mechanism for ensuring that all customer journeys are accounted for. As a result, service improvements and bus priority interventions are targeted where customers will be benefitted the most (Carreira et al, 2013).

6.2 Bus Priority Interventions

In the 2019/20 TfL budget there is £177 million assigned to 'Healthy Street' projects targeted at 'recognising the value of increasing walking, cycling and public transport', this includes £15 million for Bus Priority interventions (TfL, 2018b, 132). The spatial and temporal disaggregation of BCJT demonstrates the potential to inform prioritisation of this spending to ensure interventions are delivered where they are needed most.

BCJT enables identification of required bus priority at both route and stop levels. Furthermore, the metric can also identify which component of the customer's journey is performing poorly and thus inform the type of intervention required. For example, if the metric identifies a junction which is experiencing highly variable customer journey times, it could indicate the need for a signal intervention to reduce passenger journey time variability (e.g. between stops 19300 and 19298 in section 5.2). This would be supported further if multiple routes using the same junction were all experiencing high levels of variability. If average in-vehicle times were higher than expected in a particular location it could suggest a need for bus priority intervention, such as a bus lane. Prioritisation is also then supported by the demand weighting used within the metric, as the number of passengers who would benefit from a bus priority intervention can be quantified and used as part of the business case.

This is a significant advance on the type of information used for planning bus priority interventions. Previously interventions have been focused where average speeds are lowest. However as shown, capturing the average speed does not reflect the actual customer experience, whereas BCJT provides the ability to include journey time variability and crowding into bus priority interventions (Barron et al, 2013).

6.3 Strategic Bus Service Planning

BCJT can be applied to ensure bus service planning is effectively contributing towards achieving TfL's strategic goals (TfL, 2018a). To increase bus patronage and thus increase sustainable mode share towards the 80% MTS target, it is essential that the bus service provided operates effectively and efficiently, whereby customers are informed and have the ability to rely on the service (Kwon et al, 2014).

The BCJT metric not only highlights to planners and performance managers where there are existing potential problems on the bus network, but also has the ability to identify potential service changes to enhance the customer experience further. For example, through being able to disaggregate to link level, the metric shows where there are particular pinch points on certain routes. Routings can then be compared with alternatives, raising the possibility of changing route alignments to faster and more reliable roads where possible. Additionally, BCJT contributes towards preventing route alignments from being changed to locations where there would be a negative impact on the customer experience, for example it would provide justification for not moving to another road where routes are currently experiencing high journey time variability, therefore improving the overall customer experience (Carreira et al, 2013).

Through using this approach, service planners would make more customer focused planning decisions, ensuring that service interventions are made with every customer experience in mind (Uniman, 2009). Considering every customer journey when analysing potential changes on the bus network, moves away from using averages to assess where an intervention is required. This demonstrates the importance of measuring performance from the customer perspective as it will positively influence a customer to choose to take the bus journey and subsequently contribute towards achieving the MTS sustainable mode share targets (Gittens and Shalaby, 2015).

This is particularly important in the current context of limited financial resources. For example, TfL is currently undertaking several strategic cuts to services in inner London to help release resources and redistribute services where they are needed most (TfL, 2018a). The TfL 2019/20-2023/24 Business Plan defines a net reduction in bus operated kilometres over the next 5 years with total service levels falling by 2% from 482 million to 473 million operated km per year (TfL, 2018c). BCJT will ensure that service changes minimise the customer impact, identifying the best locations for freeing up resource (e.g. where crowding low and service supply exceeds passenger demand) and tracking service change implementation to ensure there is no negative impact on the overall customer experience.

7.0. Conclusion

7.1. Summary of Findings

This thesis has demonstrated how a new bus performance metric can focus on the customer's perspective and has shown the potential impact of BCJT on how the bus network is managed and planned. The previous three chapters demonstrate the methodology, validation and potential implementation of this metric, showing how TfL's existing quantitative data sets can be manipulated and analysed to create an effective measure of the holistic customer experience (Carreira et al, 2013). BCJT provides a new mechanism for balancing existing performance insights into one overall customer experience measure with the ability to disaggregate temporally, spatially and into each of the journey components.

It became clear throughout the literature review that existing performance metrics tend to focus on averages and a primarily operational perspective with weightings based on service volumes rather than passenger demand (Uniman et al, 2010, Nam et al, 2005). This thesis has shown it is possible to switch to the customer's perspective, where weightings are given based on customer perception and passenger demand. The methodology developed in Chapter 3 shows how the iBus and ODX datasets can be used in combination to capture every bus customer journey and each of its components into one overall measure. The introduction of interchange, crowding and buffer time (both wait and in-vehicle time) is a new addition compared to existing bus performance metrics which have been based on average wait and in-vehicle times, demonstrating the extent of the customer's journey which was not being captured previously. Analysis of the CSS score identified that reliability was the most important aspect of a customer's experience, with Chapter 4 showing how 95% of customer's journey times are captured in BCJT.

There are many potential applications and uses of the new bus performance metric, with Chapter 6 outlining the business and policy applications of BCJT in service planning and performance management. BCJT creates a far better understanding of how bus passengers experience the bus network and enables TfL to make decisions with the customer in mind (Bagherian et al, 2013). The findings have shown that if this is done effectively, TfL can work to ensure that any service changes on the network minimise negative impact on customers and that performance interventions are targeted where there is the most customer benefit. With ambitious goals set out to increase sustainable travel in London, and the bus network being the most effective way to reach those not already travelling sustainably, this project has demonstrated the impact of measuring performance from the customer perspective, ensuring that work is done to help encourage users on to the bus network.

BCJT captures as much of the overall customer experience as possible quantitatively with value of time weightings accounting for the qualitative customer perception of a bus journey. Whilst the metric does not capture every component of a customer's perception, such as temperature, through the crowding component it accounts for the unpleasant experience of a customer who is unable to get a seat when on board a bus. Additional weighting is given for the more inconvenient components of a customer's journey, for example wait time and interchange. This reflects that these parts of the journey feel longer to a customer as they are components not undertaken when travelling by car (Quarmby, 1967). Through doing this BCJT becomes more relatable to the customer experience, moving away from focusing on average and scheduled journey time, as these are not something regularly experienced by the customer and do not account for the holistic customer experience (El-Geneidy et al, 2011; Barron et al, 2013).

A greater focus on the customer experience mean that TfL has now included BCJT on both its TfL and Surface scorecards (used to track overall TfL performance), replacing the existing bus performance measures. This has demonstrated the effectiveness of this study with the outputs showing that the customer perspective can be effectively captured within bus performance through using the journey component breakdown (as seen in Chapter 4). Through ongoing tracking of BCJT it will be possible for TfL to assess the impacts of large-scale network changes, such as the Central London changes, to move towards driving an overall improvement in the bus customer experience, delivering a more efficient, reliable and attractive bus network for customers, to achieve the goals set out in the MTS (TfL, 2018a).

7.2. Suggestions for Further Study

This study fills a literature gap by demonstrating the breadth of factors that can be measured quantitatively and systematically on the London bus network to capture the customer perspective. By using Python coding, this thesis manipulated large quantities of data to effectively measure every customer experience. Whilst this is now ready to be used for performance and planning purposes within TfL, it would be useful to assess the ability to use this methodology to revolutionise real-time customer information. This would inform customers of the time they should plan for their journeys and the experience they should expect to receive (e.g. crowding levels) (Kwon et al, 2014). Further study could explore how this data could be extracted and manipulated in real-time to better inform customers and help encourage more users to the bus network. Additionally, this metric uses the logic from existing London Underground performance metrics, particularly for the crowding and interchange components. It would be beneficial to undertake some bus-specific research to confirm the suitability of factors used within the crowding formula (section 4.5) and the interchange penalty (section 4.4) for bus users.

Additionally, whilst this study confirms the potential use of BCJT, for it to become embedded as part of business as usual practices at TfL, collaboration is required across the Bus Planning, Network Management and Performance teams. By including BCJT on the TfL and Surface scorecards, support from directors is being provided to promote it. However, further work is required to provide results in formats that are fit for purpose and easy to use with a clear and consistent understanding of metric outputs so that benefits can be maximised.

Carreira, R., Patrico, L, Natal Jorge, R., Magee, C. and Q. Hommes (2013) 'Towards a holistic approach to the travel experience: A qualitative study of bus transportation', *Transport Policy,* 25, 233-243.

Creswell, J. (2009) *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.), Thousand Oaks, CA: Sage.

Crossley, A (2003a) *CP Project Note,* Internal Project Note, Transport for London <16th August 2019>.

Crossley, A (2003b) *Degrees of Train Crowding Project Note,* Internal Project Note, Transport for London <16th August 2019>.

Currie, G., Douglas, N. and I. Kearns (2012) *An Assessment of Alternative Bus Reliability Indicators*, Presented at Australian Transport Research Forum, Perth, Australia.

Daly, A.J., Philips, G.W., Rogers, K. and Smith, P.J. (1973) Planning Urban Bus Services. Local Government Operational Research Unit. Daly, A.J., Philips, G.W., Rogers, K. and Smith, P.J. (1973) Planning Urban Bus Services. Local Government Operational Research Unit.

Diab, E., Badami, M and A. El-Geneidy (2015) 'Bus Transit Reliability and Involvement Strategies: Integrating the Perspectives of Passengers and Transit Agencies in North America', *Transport Reviews*, 35, 292-328.

Eboli, L. and G. Mazzulla (2012) 'Performance Indicators for an Objective Measure of Public Transport Service Quality', *European Transport*, 51, 1-21.

Ehrlich, J. (2010) *Applications of Automatic Vehicle Location Systems towards Improving Service Reliability and Operations Planning in London,* Masters Thesis, Cambridge MA: Massachusetts Institute of Technology.

El-Geneidy, A., Horning,J. and K. Krizek (2011) 'Analysing Transit Service Reliability using Deltailed Data from Automatic Vehicular Locator Systems', *Journal of Advanced Transportation*, 45, 66-79.

Fattouche, G., (2007) *Improving High-Frequency Bus Service Reliability through Better Scheduling,* Masters Thesis: Department of Civil and Environmetal Engineering, Massachusetts Institute of Technology.

Furth, P. (2000) 'TCRP Synthesis 34: Data Analysis for Bus Planning and Monitoring. Transit Cooperative Research Program', *Transportation Research Board*, Washington, D.C.

Gentile, C., Spiller, N., G. Noci, (2007) 'How to Sustain the Customer Experience:: An Overview of Experience Components that Co-create Value With the Customer', *European Management Journal,* 25, 395–410.

Gittens, A. and A. Shalaby (2015) 'Evaluation of Bus Reliability Measures and Development of a New Composite Indicator', *Transportation Research Record: Journal of the Transportation Research Board*, 2533, 91-99.

Goedkoop, M., Van Halen, C. and H. Te Riele (1999) *Product service systems, ecological and economic basics: Amersfoort The Netherlands*, The Hague: Ministry of Housing, Spatial Planning and the Environment, Communications Directorate.

Graham, R. (1981) 'The Role of Perception of Time in Consumer Research', *Journal of Consumer Research*, 7, 335-342.

Gross, B. (1987) 'Time Scarcity: Interdisciplinary Perspectives and Implications for Consumer Behaviour' in Jagdish, N. and E. Hirschman (eds) *Research in Consumer Behaviour*, Greenwich, CT: JAI, pp. 1-54.

Hall, P. (2013) *Good Cities, Better Lives*, London: Routledge.

Hardy, N. (2009) *iBUS Benefits Realisation Workstream: method & progress to date*, Paper presented to the 16th World Congress on Intelligent Transport Systems, Stockholm.

Hollander, Y. and M. Buckmaster (2009) *The Reliability Component in Project Evaluation*, Paper presented at Transportation Research Board.

Jabareen, R. (2006). Sustainable urban forms: Their typologies, models, and concepts. *Journal of Planning Education and Research*, *26*(1), 38–52.

Kwon, Y., Kyun-Kim, C., Kim, T., Hagen, J., Barone, R. and D. Joaquin (2014) 'Improving the Customer Experience' (WWW) London: Transport Leadership Summit [https://transitleadership.org/docs/TLS-WP-Improving-the-Customer-Experience.pdf] (accessed 17th April 2019).

Larson, R. (1987) ‘Perspectives on Queues: Social Justice and the Psychology of Queuing’, *Operations Research*, 35, 895-905.

Leclerc, F., Schmitt, B. and L. Dube (1995) ‘Waiting time and decision making: Is time like money?’, *Journal of Consumer Research*, 22, 110–119.

Liu, J. and H. Wen (2016) ‘Public Transport Crowding Valuation: Evidence from College Students in Guangzhou’, *Journal of Public Transportation,* 19, 78-97.

Mazloumi, E., Currie, G. and G. Rose (2009) ‘Using GPS Data to Gain Insight into Publuc Transport Travel Time Variability’, *Journal of Transportation Engineering*, 136, 623-631.

Meyer, C. and A. Schwager, (2007) ‘Understanding Customer Experience’, *Harvard Business Review*, February, 117–126.

Nam, D., Park, D. and A. Khamkongkhun, (2005) ‘Estimation of value of travel time reliability’, *Journal of Advanced Transportation*, 39, 39–61.

Nassir, N., Hickman, M. and Z. Ma (2015) ‘Activity detection and transfer identification for public transit fare card data’, *Transportation*, 42, 683-705.

O’Regan, B and F. Buckley (2003) ‘The Psychological Effects of Commuting in Dublin’, *Centre for Research in Management Learning and Development Working Paper Series*.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M.,Shires, J. and P. White (2006) ‘The demand for public transport: The effects of fares, quality of service, income and car ownership’, *Transport Policy* 13,295–306.

Peek, G. and M. Van Hagen (2002) ‘Creating synergy in and around stations: Three strategies for adding value’, *Transportation Research Record: Journal of the Transportation Research Board*, 1793, 1–6.

Pelletier, M., Trepanier, M. and C. Morency (2011) ‘Smart card data uses in public transit: A literature review’, *Transportation Research Part C: Emerging Technologies*, 19, 557-568.

Polus, A. (1978) ‘Modelling and Measurements of Bus Service Reliability’, *Transportation Research*, 12, 253-256.

Quarmby, D.A. (1967) Choice of Travel Mode for the Journey to Work. Journal of Transport Economics and Policy 1, pp.273-314.

Raonir, R., Rao, M. and V. Senathipathi (2012) 'Public Transport Performance Evaluation Techniques – A Review', *Indian Road Transport Journal*, 3, 1-8.

Ryan, A. (1996) *The Value of Time,* London: London Transport Marketing.

Schil, M. (2012) *Measuring Journey Time Reliability in London Using Automated Data Collection Systems,* Masters Thesis, Cambridge MA: Massachusetts Institute of Technology.

Steer Davies Gleave (SDG) (2010) 'Bus passengers' priorities for improvements in London' (WWW) London: London Travel Watch [https://www.londontravelwatch.org.uk/documents/get_lob?id=2157&field=file] (accessed 17th June 2019).

Sweigart, A. (2015) *Automate the Boring Stuff with Python: Practical Programming for Total Beginners,* San Francisco: William Pollock.

TfL (2014) 'Exploring the Bus CSS Metrics' (WWW) London: Transport for London [http://content.tfl.gov.uk/exploring-the-bus-css-metrics.pdf] (accessed 17th June 2019).

TfL (2015) 'Bus Customer Satisfaction' (WWW) London: Transport for London [http://content.tfl.gov.uk/stp-20150715-part-1-item10-bus-customer-satisfaction.pdf] (accessed 17th June 2019).

TfL (2017) Business Case Development Manual, London: Transport for London.

TfL (2018a) 'Mayor's Transport Strategy' London: Mayor of London.

TfL(2018b) 'TfL Budget 2019/20' (WWW) London: Transport for London [http://content.tfl.gov.uk/tfl-budget-2019-20.pdf] (accessed 17th June 2019)

TfL (2018c) 'TfL Business Plan 2019/20 – 2023/4' (WWW) London: Transport for London [http://content.tfl.gov.uk/tfl-business-plan-2019-24.pdf] (accessed 17th June 2019).

TfL (2019a) ‘Annual Performance Summary’ (WWW) London: Transport for London [http://content.tfl.gov.uk/2017-18-annual-performance-summary.pdf] (accessed 17th June 2019).

TfL (2019b) ‘Bus Routes & Borough Reports’ (WWW) London: Transport for London [https://tfl.gov.uk/forms/14144.aspx] (accessed 17th June 2019).

TfL (2019c) ‘Buses Performance Data’ (WWW) London: Transport for London [https://tfl.gov.uk/corporate/publications-and-reports/buses-performance-data] (accessed 17th June 2019).

TfL (2019d) *Analysis Packs*, Electronic Dataset, Transport for London <7th August 2019>.

TfL (2019e) *London Bus Network,* Map, Transport for London <7th August 2019>.

TfL (2019f) *Pinpoint,* Map, Transport for London <7th August 2019>.

TfL (2019g) *London Reporting Database*, Electronic Dataset, Transport for London <1st July 2019>.

TfL (2019h) *ODX,* Electronic Dataset, Transport for London <10th July 2019>.

Trompet, M., Liu, X. and D. Graham (2011) ‘Development of Key Performance Indicators to Compare Regularity of Service Between Urban Bus Operators’, *Transportation Research Record: Journal of Transportation Research Board*, 2216, 33-41.

Uniman, D. (2009) *Service Reliability Measurement using AFC Smart Card Data - A Framework for the London Underground,* Master’s Thesis, Massachusetts Institute of Technology.

Uniman, D., Attanucci, J, Mishalani, R. and N. Wilson (2010) ‘Service Reliability Measurement using Automated Fare Card Data’, *Transportation Research Record: Journal of the Transportation Research Board*, 2143, 92-99.

Valentine, G. (1997) 'Tell me about ...: using interviews as a research methodology', in Flowerdew, R. and D. Martin (eds) Methods in Human Geography, London: Longman, pp.110-126.

Vanderwaart, C. (2016) *Planning transit networks with origin, destination and interchange inference,* Master’s Thesis, Massachusetts Institute of Technology.

Van Oort, N. (2014) 'Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations', *Research in Transportation Economics*, 48, 92-100.

Walkenbach, J. (2010) *Excel 2010 Power Programming with VBA*, London: Wiley & Sons.

Walker, J. (2012) *Human Transit: How Clearer Thinking About Public Transit Can Enrich Our Communities and Our Lives,* Washington DC: Island Press.

Wang, W. (2010) *Bus passenger origin-destination estimation and travel behaviour using automated data collection systems in London, UK,* Master's Thesis, Massachusetts Institute of Technology.

Wardman, M. and Hine, J. (2000) Costs of Interchange: A Review of the Literature. Working Paper. Institute of Transport Studies, University of Leeds, Leeds, UK. Working Paper 546.

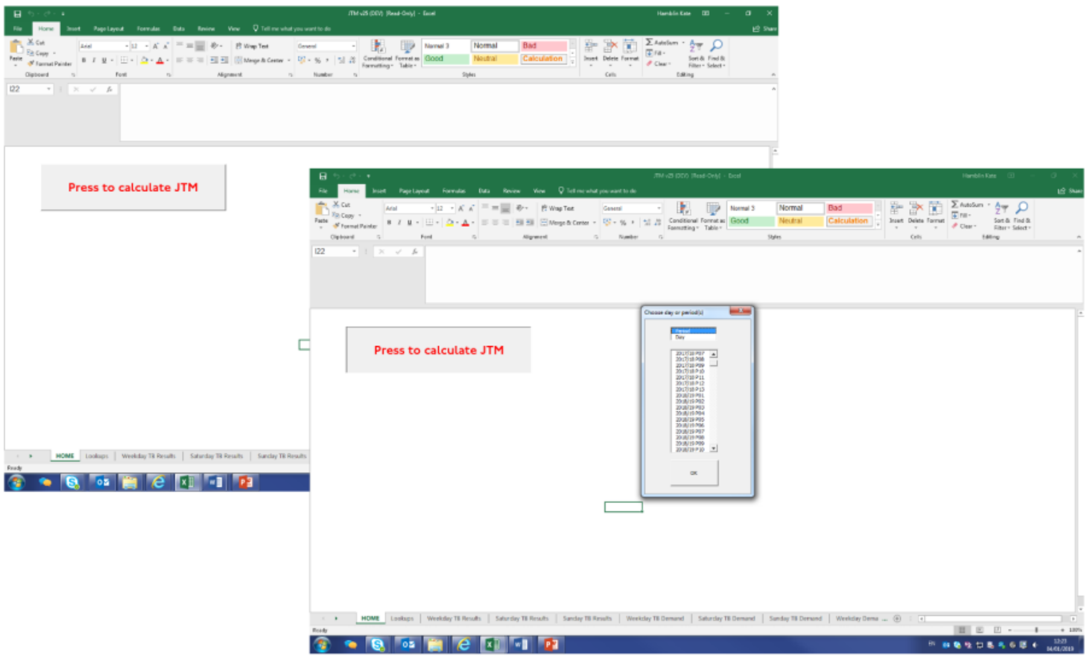
Wong, A. and N. Hounsell (2010) 'Using the iBus System to Provide Improved Public Transport Information and Applications for London', *Transportation Research Group*, 12 WCTR, Lisbon, Portugal.

Wood, D. (2015) *A framework for measuring passenger-experienced transit reliability using automated data,* Master's Thesis, Massachusetts Institute of Technology.

9.0. Appendices

Appendix A: BCJT User Interface

The code will run BCJT for user selected date range and produce all the outputs automatically at route and network level.



Appendix B: VBA and Python BCJT Extract

1. Metric Control VBA Extract

```
Sub JTMControl()
Dim ary
Dim intCol
Dim rngTarget As Range
Dim x As Integer
Dim sLine As String
Dim dDate As Date
Dim i As Long
Dim aPeriods
Dim sThisPeriod As String
Dim sPrevPeriod As String
Dim dPrevPeriodStart As Date
Dim dPrevPeriodEnd As Date
Dim dP09Start As Date
Dim dP09End As Date
Dim sP09 As String
Dim aRoutes
Dim aTimebands
Dim aTermDates
Dim aBankHols
Dim aBusCapacities
Dim iStart As Integer
Dim sString As String
Dim iFields As Integer
Dim iTry As Integer
```

```
Dim bIgnoreODX As Boolean
Dim aDayTypes

If dStartDate = 0 Then End

sPythonPath = "C:\Users\" & Environ("UserName") &
"\AppData\Local\Microsoft\AppV\Client\Integration\2552BB88-E62D-4806-A8DE-
6B2E698727DA\Root\Python.exe"


sPyFilePath = sFolder & "\JTM_tmp" & iRandom & ".py"
sTimebandFilePath = sFolder & "\Timebands_tmp" & iRandom & ".CSV"
sLrdFilePath = sFolder & "\LRD_tmp" & iRandom & ".CSV"
sDayTypeFilePath = sFolder & "\DayTypes_tmp" & iRandom & ".CSV"
sRouteStopFilePath = sFolder & "\RouteStops_tmp" & iRandom & ".CSV"
sBusCapacityFilePath = sFolder & "\BusCapacities_tmp" & iRandom & ".CSV"

aBusCapacities = Range("BusCapacities")
Open sBusCapacityFilePath For Output As #1
For i = 1 To UBound(aBusCapacities)
    aBusCapacities(i, 1) & "," & aBusCapacities(i, 2) & "," & aBusCapacities(i, 3) & ","
& aBusCapacities(i, 4)
Next i
Close #1

aTimebands = Range("Timebands")
Open sTimebandFilePath For Output As #1
aTimebands(1, 1) & "," & aTimebands(1, 2) & "," & aTimebands(1, 3) & "," & aTimebands(1,
4)
For i = 2 To UBound(aTimebands)
    aTimebands(i, 1) & ",1900-01-01 " & Format(aTimebands(i, 2), "hh:mm:ss") & "," &
aTimebands(i, 3) & "," & aTimebands(i, 4)
Next i
Close #1

aRoutes = Range("Routes")
sRoutes = ""
For i = 1 To UBound(aRoutes)
    sRoutes = sRoutes & "'" & aRoutes(i, 1) & "'" & IIf(i < UBound(aRoutes), ",", "")
Next i

aDayTypes = Range("DayTypes")
Open sDayTypeFilePath For Output As #1
"DATE,DAY_TYPE"
For i = 2 To UBound(aDayTypes)
    Format(aDayTypes(i, 1), "dd/mm/yyyy") & "," & aDayTypes(i, 2)
Next i
Close #1

connStr = "Provider=MSDAORA;Data
Source=(DESCRIPTION=(ADDRESS_LIST=(ADDRESS=(PROTOCOL=TCP)(HOST=pdc2ora012.onelondon.tfl.l
ocal)(PORT=1521)))(CONNECT_DATA=(SERVER=DEDICATED)(SID=lrd10))); User
Id=LRD;Password=CHANBAR_2002"

'determine last period and last Period 9
aPeriods = Range("Periods")
For i = 1 To UBound(aPeriods)
    If dStartDate >= aPeriods(i, 2) And dStartDate <= aPeriods(i, 3) Then
        dPrevPeriodStart = aPeriods(i - 1, 2)
        dPrevPeriodEnd = aPeriods(i - 1, 3)
        sPrevPeriod = aPeriods(i - 1, 1)
        sThisPeriod = aPeriods(i, 1)
        For j = 1 To 13
            If Right(aPeriods(i - j, 1), 2) = "09" Then
                sP09 = aPeriods(i - j, 1)
                dP09Start = aPeriods(i - j, 2)
                dP09End = aPeriods(i - j, 3)
                Exit For
            End If
        Next j
        Exit For
    End If
Next i

'create last Autumn (Period 9) demand if not exist
iTry = 0
sDemandFile = sFolder & "\Period 9 Demand\" & Replace(Left(sP09, 7), "/", "_") & " P09
Demand.CSV"
Do While Dir(sDemandFile) = "" And iTry <= 3
    Application.StatusBar = "Creating " & Replace(Left(sP09, 7), "/", "_") & " P09
Demand.CSV"
    Call PyDemand(dP09Start - #1/1/1980#, dP09End - #1/1/1980#, sP09)
    Call ShellAndWait(sPythonPath & " """ & sPyFilePath & """)
    iTry = iTry + 1
Loop

'create last Autumn (Period 9) Link Demand
```

```
iTry = 0
sLinkDemandFile = sFolder & "\Period 9 Demand\" & Replace(Left(sP09, 7), "/", "_") & "
P09 Link Demand.CSV"
Do While Dir(sLinkDemandFile) = "" And iTry <= 3

    Call LRD_Extract(MySql2(dP09End - 1), sRouteStopFilePath, False)

    Call PyLinkDemand
    Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
    iTry = iTry + 1
Loop

'create last period's 10th percentile if not exist
s10thPercFile = sFolder & "\10th Perc\" & Replace(sPrevPeriod, "/", "_") & " 10th
Perc.CSV"
Do While Dir(s10thPercFile) = ""
    Application.StatusBar = "Creating " & Replace(sPrevPeriod, "/", "_") & " 10th
Perc.CSV"
    Call DeleteFiles
    Erase iNDayTypes

    For dDate = dPrevPeriodEnd To dPrevPeriodEnd - 365 Step -1

        For i = 1 To UBound(aDayTypes)
            If aDayTypes(i, 1) = dDate Then
                sDayType = aDayTypes(i, 2)
                Select Case sDayType
                    Case "Weekday": iNDayTypes(1) = iNDayTypes(1) + 1
                    Case "Saturday": iNDayTypes(2) = iNDayTypes(2) + 1
                    Case "Sunday": iNDayTypes(3) = iNDayTypes(3) + 1
                    Case "School Holiday": iNDayTypes(4) = iNDayTypes(4) + 1
                End Select
                Exit For
            End If
        Next i

        If sDayType = "Weekday" And iNDayTypes(1) > 20 And dDate < dPrevPeriodStart Then
GoTo NextDate
        If sDayType = "Saturday" And iNDayTypes(2) > 4 And dDate < dPrevPeriodStart Then
GoTo NextDate
        If sDayType = "Sunday" And iNDayTypes(3) > 4 And dDate < dPrevPeriodStart Then
GoTo NextDate
        If sDayType = "School Holiday" And iNDayTypes(4) > 5 And dDate < dPrevPeriodStart
Then GoTo NextDate
        If sDayType = "Bank Holiday" Then GoTo NextDate

        iStart = 1 - iNRoutesPerRun
        Do
            iStart = iStart + iNRoutesPerRun
            aRoutes = Range("Routes")
            sRoutes = ""
            For i = iStart To Application.Min(UBound(aRoutes), iStart + (iNRoutesPerRun -
1))
                sRoutes = sRoutes & "'" & aRoutes(i, 1) & "'" & ","
            Next i
            sRoutes = Left(sRoutes, Len(sRoutes) - 1)

            Call LRD_Extract(MySql(dDate), sFolder & "\" & Format(dDate, "dd-Mmm-YYYY") &
"_tmp" & iRandom & ".CSV", True)

        Loop While i < UBound(aRoutes)
NextDate:

    Next dDate

    Call Py10thPerc
    Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
    Call DeleteFiles

Loop

For dDate = dStartDate To dEndDate

    If Month(dDate) = 12 And Day(dDate) = 25 Then GoTo NextDay

    sBankHoliday10thPercFile = ""

'if Bank Holiday then create 10th perc from last year's same Bank Holiday
    For k = 1 To UBound(aDayTypes)
        If aDayTypes(k, 1) = dDate And aDayTypes(k, 2) = "Bank Holiday" Then
            sBankHoliday10thPercFile = sFolder & "\Bank Holiday 10th Perc\" &
Format(dDate, "yyyy") - 1 & " " & aDayTypes(k, 3) & ".CSV"
            Do While Dir(sBankHoliday10thPercFile) = ""
                Call DeleteFiles
                For j = k - 1 To 1 Step -1
                    If aDayTypes(j, 3) = aDayTypes(k, 3) Then
```

```
                        iStart = 1 - iNRoutesPerRun
                        Do
                            iStart = iStart + iNRoutesPerRun
                            aRoutes = Range("Routes")
                            sRoutes = ""
                            For i = iStart To Application.Min(UBound(aRoutes), iStart +
(iNRoutesPerRun - 1))
                                sRoutes = sRoutes & "'" & aRoutes(i, 1) & "'" & ","
                            Next i
                            sRoutes = Left(sRoutes, Len(sRoutes) - 1)

                            Call LRD_Extract(MySql(aDayTypes(j, 1)), sFolder & "\" &
Format(aDayTypes(j, 1), "dd-Mmm-YYYY") & "_tmp" & iRandom & ".CSV", True)

                        Loop While i < UBound(aRoutes)

                        s10thPercFile = sBankHoliday10thPercFile
                        Call Py10thPerc
                        Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
                        Call DeleteFiles
                        s10thPercFile = sFolder & "\10th Perc\" & Replace(sPrevPeriod,
"/", "_") & " 10th Perc.CSV"
                    End If
                Next j
            Loop
        End If
    Next k

'Day Type
    For i = 1 To UBound(aDayTypes)
        If aDayTypes(i, 1) = dDate Then
            sDayType = aDayTypes(i, 2)
            bIgnoreODX = aDayTypes(i, 5) = "YES"
            Exit For
        End If
    Next i

'Hourly Demand
    If bIgnoreODX = False Then
        sHourlyDemandFile = sFolder & "\Hourly Demand\" & Replace(sThisPeriod, "/", "_")
& "\" & Format(dDate, "yyyy-mm-dd") & " Demand.CSV"
        iTry = 0
        Do While Dir(sHourlyDemandFile) = "" And dDate < Int(Now()) And iTry <= 3
            Application.StatusBar = "Creating " & Format(dDate, "yyyy-mm-dd") & "
Demand.CSV"
            Call PyHourlyDemand(dDate - #1/1/1980#)
            Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
            iTry = iTry + 1
        Loop
    End If

'Timeband Demand
    If bIgnoreODX = False Then
        sTimebandDemandFile = sFolder & "\Timeband Demand\" & Replace(sThisPeriod, "/",
"_") & "\" & Format(dDate, "yyyy-mm-dd") & " Demand.CSV"
        Do While Dir(sTimebandDemandFile) = "" And dDate < Int(Now())
            Application.StatusBar = "Creating " & Format(dDate, "yyyy-mm-dd") & "
Demand.CSV"
            Call PyTimebandDemand(dDate)
            Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
        Loop
    End If

'Hourly Interchange Times & JLF
    If bIgnoreODX = False Then
        iTry = 0
        sHourlyInterchangeResultFile = sFolder & "\Hourly Interchange Results\" &
Replace(sThisPeriod, "/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " Interchange.CSV"
        sHourlyJLFResultFile = sFolder & "\Hourly JLF Results\" & Replace(sThisPeriod,
"/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " JLF.CSV"
        sTimebandInterchangeResultFile = sFolder & "\Timeband Interchange Results\" &
Replace(sThisPeriod, "/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " Interchange.CSV"
        sTimebandJLFResultFile = sFolder & "\Timeband JLF Results\" &
Replace(sThisPeriod, "/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " JLF.CSV"
        Do While Dir(sHourlyInterchangeResultFile) = "" Or Dir(sHourlyJLFResultFile) = ""
And iTry <= 3
            Application.StatusBar = Replace(sThisPeriod, "/", "_") & "\" & Format(dDate,
"yyyy-mm-dd") & " Interchange.CSV"
            Call PyInterchange(dDate - #1/1/1980#)
            Call ShellAndWait(sPythonPath & " """ & sPyFilePath & "")
            iTry = iTry + 1
        Loop
    End If

' Crowding
    If bIgnoreODX = False Then
        iTry = 0
```

```
        sHourlyCrowdingResultFile = sFolder & "\Hourly Crowding Results\" &
Replace(sThisPeriod, "/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " Crowding.CSV"
        sTimebandCrowdingResultFile = sFolder & "\Timeband Crowding Results\" &
Replace(sThisPeriod, "/", "_") & "\" & Format(dDate, "yyyy-mm-dd") & " Crowding.CSV"
        Do While Dir(sHourlyCrowdingResultFile) = "" And iTry <= 3
            Application.StatusBar = Replace(sThisPeriod, "/", "_") & "\" & Format(dDate,
"yyyy-mm-dd") & " Crowding.CSV"
            Call PyCrowding(dDate - #1/1/1980#)
            Call ShellAndWait(sPythonPath & " """ & sPyFilePath & """)
            iTry = iTry + 1
        Loop
    End If

'Hourly JTM
    sHourlyResultFile = sFolder & "\Hour Results\" & Replace(sThisPeriod, "/", "_") & "\"
& Format(dDate, "yyyy-mm-dd") & " JTM.CSV"
    Do While Dir(sHourlyResultFile) = "" And dDate < Int(Now())
        Application.StatusBar = Replace(sThisPeriod, "/", "_") & "\" & Format(dDate,
"yyyy-mm-dd") & " JTM.CSV"

        iStart = 1 - iNRoutesPerRun
        Do
            iStart = iStart + iNRoutesPerRun
            aRoutes = Range("Routes")
            sRoutes = ""
            For i = iStart To Application.Min(UBound(aRoutes), iStart + (iNRoutesPerRun -
1))
                sRoutes = sRoutes & "'" & aRoutes(i, 1) & "'" & ","
            Next i
            sRoutes = Left(sRoutes, Len(sRoutes) - 1)

            Call LRD_Extract(MySql(dDate), sLrdFilePath, False)

            If sBankHoliday10thPercFile <> "" Then s10thPercFile =
sBankHoliday10thPercFile
            Call PyHourlyJTM
            s10thPercFile = sFolder & "\10th Perc\" & Replace(sPrevPeriod, "/", "_") & "
10th Perc.CSV"
            Call ShellAndWait(sPythonPath & " """ & sPyFilePath & """)
        Loop While i < UBound(aRoutes)

    Loop

'Timeband JTM
    sTimebandResultFile = sFolder & "\Timeband Results\" & Replace(sThisPeriod, "/", "_")
& "\" & Format(dDate, "yyyy-mm-dd") & " JTM.CSV"
    Do While Dir(sTimebandResultFile) = "" And dDate < Int(Now()) And
Dir(sHourlyDemandFile) <> "" And Dir(sHourlyResultFile) <> ""
        Application.StatusBar = "Creating " & Replace(sThisPeriod, "/", "_") & "\" &
Format(dDate, "yyyy-mm-dd") & " JTM.CSV"
        Call PyTimebandJTM
        Call ShellAndWait(sPythonPath & " """ & sPyFilePath & """)
    Loop

NextDay:

Next dDate

If Dir(sLrdFilePath) <> "" Then Kill sLrdFilePath
If Dir(sPyFilePath) <> "" Then Kill sPyFilePath
If Dir(sTimebandFilePath) <> "" Then Kill sTimebandFilePath
If Dir(sDayTypeFilePath) <> "" Then Kill sDayTypeFilePath
If Dir(sRouteStopFilePath) <> "" Then Kill sRouteStopFilePath
If Dir(sRouteStopFilePath) <> "" Then Kill sRouteStopFilePath

Close #1

End Sub
```

2. Crowding Python

```
"import pandas as pd"
"import odbc"
"import numpy as np"

"connectstr = 'DRIVER=SQL
Server;SERVER=10.107.24.63\INS005;UID=EXCEL_USER;PWD=EXCEL_USER;APP=2007 Microsoft Office
system;WSID=PDC2CXP017;DATABASE=MASTER'"
"sql = ""SELECT route, direction, tripnumber, stopcode, observedarrivaltime as
arrivetime, observeddeparturetime as departtime, inferredalighted, AllBoardings,
ScaledAlighted FROM pdwlink.odx_data.dbo.BusEventsFact WHERE DayKey = " & iDate & " AND
route IN(" & sRoutes & ") AND observedarrivaltime IS NOT NULL AND observeddeparturetime
IS NOT NULL"""
"connect = odbc.odbc(connectstr)"
"odx = pd.read_sql_query(sql, connect)"

"odx['arrivetime'] =  pd.to_datetime(odx['arrivetime'], errors='coerce')"
```

```
"odx['departtime'] =  pd.to_datetime(odx['departtime'], errors='coerce')"
"odx = odx.fillna(0)"
"odx = odx.sort_values(['route', 'direction', 'tripnumber', 'departtime'],
ascending=[True, True, True, True])"

"odx['route'] = odx['route'].astype(str)"
"odx['route'] = odx['route'].str.replace('[^A-Z0-9]', '')"

"odx['IVT'] = odx['arrivetime'].shift(-1) - odx['departtime'].where((odx['route'].shift(-
1) == odx['route']) & (odx['direction'].shift(-1) == odx['direction'])  &
(odx['tripnumber'].shift(-1) == odx['tripnumber']))"
"odx['netboarders'] = odx['AllBoardings'] - odx['inferredalighted'] -
odx['ScaledAlighted']"
"odx['load'] = odx.groupby(['route','direction','tripnumber'])['netboarders'].cumsum()"
"odx = odx.dropna()"

"# read in bus types by route"
"GarageRoutes = pd.read_excel(r'" & sFolder & "\Crowding\Garages_&_Routes.xls', 'Sheet')"
"GarageRoutes = GarageRoutes[['Route','Vehicle Type and Qty']]"
"GarageRoutes['BusType'], GarageRoutes['Qty'] = GarageRoutes['Vehicle Type and
Qty'].str.split(' ', 1).str"
"GarageRoutes['Qty'] = GarageRoutes['Qty'].str.extract('([0-9]\w{0,})').astype(float)"
"GarageRoutes = GarageRoutes.fillna(0)"
"GarageRoutes = pd.pivot_table(GarageRoutes,index=['Route', 'BusType'],
values=['Qty'],aggfunc=[max])"
"GarageRoutes.reset_index(inplace=True)"
"GarageRoutes.columns = GarageRoutes.columns.droplevel(level=1)"
"GarageRoutes = GarageRoutes.loc[GarageRoutes.groupby(['Route'])['max'].idxmax()]"

"# read bus capacities and merge with GarageRoutes"
"BusCapacities = pd.read_csv(r'" & sBusCapacityFilePath & "', low_memory=False)"
"GarageRoutes = pd.merge(GarageRoutes, BusCapacities, left_on=['BusType'],
right_on=['Vehicle Type'],how='left') "
"GarageRoutes = GarageRoutes[['Route','BusType','Capacity','Seats','Pax Upscale']]"

"table = pd.merge(odx, GarageRoutes, left_on=['route'], right_on=['Route'],how='inner') "
"table['load'] = table['load'] * (1 + table['Pax Upscale'])"
"table['Hour']= pd.to_datetime(table['departtime'].dropna().dt.hour.astype(np.int64),
format='%H', errors='coerce' )"
"timebands = pd.read_csv(r'" & sTimebandFilePath & "')"
"timebands['Hour'] =  pd.to_datetime(timebands['Hour'], errors='coerce')"

If sDayType = "Weekday" Or sDayType = "School Holiday" Or sDayType = "Bank Holiday" Then
    "table['DAY_TYPE'] = 'Weekday'"
Else
    "table['DAY_TYPE'] = '" & sDayType & "'"
End If

"table = pd.merge(table, timebands, left_on=['DAY_TYPE','Hour'],right_on=['Day','Hour'],
how='inner')"
"table.reset_index(inplace=True)"


"C0=0.085"
"C1=1.915"
"CY=1.03"

"table['IVT'] = table['IVT'].dt.total_seconds().astype(np.int64)"
"table['CrowdingFactor'] = np.minimum(2.5, np.maximum(1, 1 + C0 + (C1 - CY *
(table['Seats'] / table['Capacity'])) * (table['load'] - table['Seats']) /
(table['Capacity'] - table['Seats'])))"
"table['CrowdingFactor(W)'] = table['CrowdingFactor'] * table['load'] * table['IVT']"
"table['Weight'] = table['load'] * table['IVT'] "

"table2 = pd.pivot_table(table,index=['Route', 'Hour'],
values=['CrowdingFactor(W)','Weight'],aggfunc=sum)"
"table2.reset_index(inplace=True)"
"table2['CrowdingFactor'] = np.minimum(2.5, np.maximum(1, table2['CrowdingFactor(W)'] /
table2['Weight']))"
"table2 = table2.fillna(1)"
"table2 = table2[['Route','Hour','CrowdingFactor']]"
"table2.to_csv(r'" & sHourlyCrowdingResultFile & "', mode='w', float_format='%.5f',
index=False)"

"table2 = pd.pivot_table(table,index=['Route', 'Timeband'],
values=['CrowdingFactor(W)','Weight'],aggfunc=sum)"
"table2.reset_index(inplace=True)"
"table2['CrowdingFactor'] = np.minimum(2.5, np.maximum(1, table2['CrowdingFactor(W)'] /
table2['Weight']))"
"table2 = table2.fillna(1)"
"table2 = table2[['Route','Timeband','CrowdingFactor']]"
"table2.to_csv(r'" & sTimebandCrowdingResultFile & "', mode='w', float_format='%.5f',
index=False)"
```

3. Interchange Python

```
"import odbc"
"import pandas as pd"
"import numpy as np"

"connectstr = 'DRIVER=SQL
Server;SERVER=10.107.24.63\INS005;UID=EXCEL_USER;PWD=EXCEL_USER;APP=2007 Microsoft Office
system;WSID=PDC2CXP017;DATABASE=MASTER'"
"connect = odbc.odbc(connectstr)"
"sql = ""SELECT fromnodeid, tonodeid, fromtime, totime, DATEPART(hh, fromtime) As HOUR,
prestigeid, linkedjourneyindex, linkedstageindex FROM pdwlink.odx_data.dbo.odx_output
WHERE DayKey = " & iDate & " and subsystemid = 1 AND routeid IN(" & sRoutes & ")"""
"odx = pd.read_sql_query(sql, connect)"
"odx = odx.sort_values(['prestigeid', 'linkedjourneyindex', 'linkedstageindex'],
ascending=[True, True, True])"
"daytype= 'Weekday'"
"odx['DAY_TYPE'] = daytype"
"odx['HOUR']= pd.to_datetime(odx['HOUR'], format='%H', errors='coerce' )"

"# create timebands"
"timebands = pd.read_csv(r'" & sTimebandFilePath & "')"
"timebands['Hour'] =  pd.to_datetime(timebands['Hour'], errors='coerce')"

If sDayType = "Weekday" Or sDayType = "School Holiday" Or sDayType = "Bank Holiday" Then
    "odx['DAY_TYPE'] = 'Weekday'"
Else
    "odx['DAY_TYPE'] = '" & sDayType & "'"
End If

"odx = pd.merge(odx, timebands, left_on=['DAY_TYPE','HOUR'],right_on=['Day','Hour'],
how='inner')"
"odx.reset_index(inplace=True)"

"# total hourly legs"
"hourlylegs = pd.pivot_table(odx,index=['HOUR'],
values=['prestigeid','fromtime'],aggfunc=len)"
"hourlylegs.reset_index(inplace=True)"
"hourlylegs.rename(columns={'prestigeid': 'legs'}, inplace=True)"
"hourlylegs['HOUR2'] = hourlylegs['HOUR'].dt.hour"

"# total timeband legs"
"timebandlegs = pd.pivot_table(odx,index=['Timeband'],
values=['prestigeid','fromtime'],aggfunc=len)"
"timebandlegs.reset_index(inplace=True)"
"timebandlegs.rename(columns={'prestigeid': 'legs'}, inplace=True)"

"#read in journey planner walktimes"
"odx['nodeid2'] = odx['fromnodeid'].shift(-1).where((odx['linkedjourneyindex'].shift(-1)
== odx['linkedjourneyindex']) & (odx['linkedstageindex'].shift(-1) ==
odx['linkedstageindex'] + 1) & (odx['prestigeid'].shift(-1) == odx['prestigeid']))"
"walktimes = pd.read_csv(r'" & sFolder & "\Interchange\Walktimes.CSV', low_memory=False)"
"odx = pd.merge(odx, walktimes, left_on=['tonodeid', 'nodeid2'], right_on=['From Stop',
'To Stop'],how='left')"
"odx = pd.pivot_table(odx,index=['prestigeid', 'DAY_TYPE', 'HOUR', 'Timeband',
'linkedjourneyindex'], values=['fromtime', 'Walktime'],aggfunc={'fromtime':np.min,
'Walktime':np.sum})"
"odx.reset_index(inplace=True)"

"# total hourly journeys"
"hourlyjourneys = pd.pivot_table(odx,index=['HOUR'],
values=['prestigeid','linkedjourneyindex'],aggfunc=len)"
"hourlyjourneys.reset_index(inplace=True)"
"hourlyjourneys.rename(columns={'prestigeid': 'journeys'}, inplace=True)"
"hourlyjourneys['HOUR2'] = hourlyjourneys['HOUR'].dt.hour"

"# total timeband journeys"
"timebandjourneys = pd.pivot_table(odx,index=['Timeband'],
values=['prestigeid','linkedjourneyindex'],aggfunc=len)"
"timebandjourneys.reset_index(inplace=True)"
"timebandjourneys.rename(columns={'prestigeid': 'journeys'}, inplace=True)"

"# calculate hourly journey leg factor "
"HourlyJourneyLegFactor = pd.merge(hourlyjourneys, hourlylegs, on=['HOUR2'],
how='inner')"
"HourlyJourneyLegFactor['JLF'] =
HourlyJourneyLegFactor['legs']/HourlyJourneyLegFactor['journeys']"
"HourlyJourneyLegFactor.rename(columns={'HOUR_x': 'Hour'}, inplace=True)"
"HourlyJourneyLegFactor.to_csv(r'" & sHourlyJLFResultFile & "', mode='w', index=False,
float_format='%.5f',sep=',', header=True, columns=['Hour','JLF'])"

"# calculate timeband journey leg factor"
"TimebandJourneyLegFactor = pd.merge(timebandjourneys, timebandlegs, on=['Timeband'],
how='inner')"
"TimebandJourneyLegFactor['JLF'] =
TimebandJourneyLegFactor['legs']/TimebandJourneyLegFactor['journeys']"
"TimebandJourneyLegFactor.to_csv(r'" & sTimebandJLFResultFile & "', mode='w',
index=False, float_format='%.5f',sep=',', header=True, columns=['Timeband','JLF'])"
```

```
"# average hourly walktime for ALL interchanges (ie unweighted timeband interchange
time)"
"hourly = odx.fillna(0)"
"hourly = pd.pivot_table(hourly,index=['HOUR'],
values=['Walktime'],aggfunc={'Walktime':np.mean})"
"hourly.reset_index(inplace=True)"
"hourly.to_csv(r'" & sHourlyInterchangeResultFile & "', mode='w', index=False,
float_format='%.0f',sep=',', header=True, columns=['HOUR','Walktime'])"

"# average timeband walktime for ALL interchanges (ie unweighted timeband interchange
time)"
"timeband = odx.fillna(0)"
"timeband = pd.pivot_table(timeband,index=['Timeband'],
values=['Walktime'],aggfunc={'Walktime':np.mean})"
"timeband.reset_index(inplace=True)"
"timeband.to_csv(r'" & sTimebandInterchangeResultFile & "', mode='w', index=False,
float_format='%.0f',sep=',', header=True, columns=['Timeband','Walktime'])"

"#walktime must be  x by JLF-1 i.e. average number of interchanges made"
```

4. Hourly Metric Results Python

```
"# functions"
"def Perc_10th(g):"
 return np.percentile(g, 10)"
"def Perc_95th(g):      "
  return np.percentile(g, 95)"
"def mydate(row):"
 return dt.datetime(1980,1,1,0,0) + dt.timedelta(days=row['daykey'])"

"import pandas as pd"
"import numpy as np"
"import datetime as dt"

"# read Period 9 ODX demand"
"dtypes={'ROUTE': object, 'DIRECTION': int,'DAY_TYPE': object,'HOUR': object,
'fromnodeid': object, 'tonodeid': object, 'TAPS': int}"
"odx = pd.read_csv(r'" & sDemandFile & "',dtype=dtypes,low_memory=False)"
"odx['HOUR'] =  pd.to_datetime(odx['HOUR'], format='%Y-%m-%d %H:%M:%S', errors='coerce'
)"

"# read LRD file"
"names=['SERVICEDAY','ROUTE','DIRECTION','TRIPNR','STOPSEQUENCE','STOP_NAME','STOP_NUMBER
','OBSERVED_ARRIVAL_TIME','OBSERVED_DEPARTURE_TIME','OBSERVED_HEADWAY']"
"dtypes={'SERVICEDAY': object, 'ROUTE': object, 'DIRECTION': int,'TRIPNR': int,
'STOPSEQUENCE': int, 'STOP_NAME': object, 'STOP_NUMBER': object, 'OBSERVED_ARRIVAL_TIME':
object, 'OBSERVED_DEPARTURE_TIME': object, 'OBSERVED_HEADWAY': int}"
"iBus1 = pd.read_csv(r'" & sLrdFilePath &
"',sep=';',header=None,names=names,dtype=dtypes,low_memory=False)"

"# convert to datetime"
"iBus1['SERVICEDAY'] =  pd.to_datetime(iBus1['SERVICEDAY'], format='%d/%m/%Y',
errors='coerce' )"
"iBus1['OBSERVED_ARRIVAL_TIME'] =  pd.to_datetime(iBus1['OBSERVED_ARRIVAL_TIME'],
format='%d/%m/%Y %H:%M:%S', errors='coerce' )"
"iBus1['OBSERVED_DEPARTURE_TIME'] =  pd.to_datetime(iBus1['OBSERVED_DEPARTURE_TIME'],
format='%d/%m/%Y %H:%M:%S', errors='coerce' )"

"# Day Type"

If sDayType = "Weekday" Or sDayType = "School Holiday" Or sDayType = "Bank Holiday" Then
    "iBus1['DAY_TYPE'] = 'Weekday'"
Else
    "iBus1['DAY_TYPE'] = '" & sDayType & "'"
End If


"# HOUR bin"
"iBus1 = iBus1.dropna(subset=['OBSERVED_ARRIVAL_TIME']) "
"iBus1 = iBus1.dropna(subset=['OBSERVED_DEPARTURE_TIME'])   "
"iBus1['HOUR']= pd.to_datetime(iBus1['OBSERVED_DEPARTURE_TIME'].dt.hour.astype(np.int64),
format='%H', errors='coerce' )"

"# duplicate and merge to create O to D table where O<D"
"iBus2 = iBus1"
"LeftCols = ['SERVICEDAY','DAY_TYPE','ROUTE', 'HOUR',
'TRIPNR','STOPSEQUENCE','STOP_NUMBER','OBSERVED_HEADWAY','OBSERVED_DEPARTURE_TIME']"
"RightCols = ['SERVICEDAY','ROUTE',
'TRIPNR','STOPSEQUENCE','STOP_NUMBER','OBSERVED_ARRIVAL_TIME']"
"iBus = pd.merge(iBus1[LeftCols],iBus2[RightCols], on=['SERVICEDAY','ROUTE','TRIPNR'])"
"iBus = iBus[(iBus['STOPSEQUENCE_x'] < iBus['STOPSEQUENCE_y'])]"
"iBus = iBus[(iBus['OBSERVED_DEPARTURE_TIME'] < iBus['OBSERVED_ARRIVAL_TIME'])]"

"# clear memory"
"del iBus1"
```

```
"del iBus2"

"# create measures"
"iBus['ACT_IVT'] = (iBus['OBSERVED_ARRIVAL_TIME']  -
iBus['OBSERVED_DEPARTURE_TIME']).dt.total_seconds()"
"iBus['ACT_AWT'] = iBus['OBSERVED_HEADWAY'] / 2"
"iBus['ACT_TJT'] = iBus['ACT_AWT'] + iBus['ACT_IVT']"

"# clean dataframe"
"iBus = iBus[iBus['OBSERVED_HEADWAY'] != 0]"

"LeftJoinCols = ['ROUTE','STOP_NUMBER_x','STOP_NUMBER_y','DAY_TYPE','HOUR']"
"RightJoinCols = ['ROUTE','fromnodeid','tonodeid','DAY_TYPE','HOUR']"
"iBus = pd.merge(iBus, odx, how='left', left_on=LeftJoinCols,
right_on=RightJoinCols).dropna()"

"del odx"

"iBus['DAY_TYPE'] = '" & sDayType & "'"

"# read 10th perc"
"routes = iBus['ROUTE'].unique()"
"routes=pd.DataFrame(routes, columns=['ROUTE'])"

"dtypes={'ROUTE': object, 'STOP_NUMBER_x': object, 'STOP_NUMBER_y': object, 'DAY_TYPE':
object, 'HOUR':object, 'Perc_10th': object, 'Perc_10th': object, 'Perc_10th': object}"
"chunksize = 100000"
"chunks = []"
"for chunk in pd.read_csv(r'"; s10thPercFile &
"',skiprows=[1,2],dtype=dtypes,low_memory=False,error_bad_lines=False,
chunksize=chunksize):"
 chunk = pd.merge(chunk, routes, on=['ROUTE'])"
 chunks.append(chunk)"
"period = pd.concat(chunks, axis=0)"

"period = period.rename(columns={period.columns[5]: 'Perc_10thACT_AWT'})"
"period = period.rename(columns={period.columns[6]: 'Perc_10thACT_IVT'})"
"period = period.rename(columns={period.columns[7]: 'Perc_10thACT_TJT'})"

"period = period.dropna()"
"period['Perc_10thACT_AWT'] = period['Perc_10thACT_AWT'].astype('float')"
"period['Perc_10thACT_IVT'] = period['Perc_10thACT_IVT'].astype('float')"
"period['Perc_10thACT_TJT'] = period['Perc_10thACT_TJT'].astype('float')"
"period['HOUR'] =  pd.to_datetime(period['HOUR'], format='%Y-%m-%d %H:%M:%S')"

"#create day results"
"day = pd.pivot_table(iBus.dropna(),index=['SERVICEDAY', 'ROUTE', 'STOP_NUMBER_x',
'STOP_NUMBER_y', 'DAY_TYPE', 'HOUR', 'TAPS'],
values=['ACT_AWT','ACT_IVT','ACT_TJT'],aggfunc=[np.mean,Perc_95th])"
"day.reset_index(inplace=True)"
"del iBus"

"# adjust AWT/IVT 10th percentiles so the they sum to the TJT 10th percentile"""
"day['Total_95th_Perc'] = day[('Perc_95th','ACT_AWT')] + day[('Perc_95th','ACT_IVT')]"
"day[('Perc_95th','ACT_AWT')] = day[('Perc_95th','ACT_AWT')] *
day[('Perc_95th','ACT_TJT')] / day['Total_95th_Perc']"
"day[('Perc_95th','ACT_IVT')] = day[('Perc_95th','ACT_IVT')] *
day[('Perc_95th','ACT_TJT')] / day['Total_95th_Perc']"
"del day['Total_95th_Perc']"

"# merge period 10th Percentile into day"
"LeftJoinCols = [('ROUTE',''), ('STOP_NUMBER_x',''), ('STOP_NUMBER_y',''),
('DAY_TYPE',''), ('HOUR','')]"
"RightJoinCols = ['ROUTE', 'STOP_NUMBER_x', 'STOP_NUMBER_y', 'DAY_TYPE', 'HOUR']"
"day = pd.merge(day,
period[[('Perc_10thACT_AWT'),('Perc_10thACT_IVT'),('Perc_10thACT_TJT')]+RightJoinCols],
how='left', left_on=LeftJoinCols, right_on=RightJoinCols)"
"day.reset_index(inplace=True)"
"del period"

"# create demand by OD, Hour, day type to get hourly TJT stats"
"DayHourDemand =
pd.pivot_table(day.dropna(),index=[('DAY_TYPE',''),('HOUR',''),('ROUTE','')],
values=[('TAPS','')],aggfunc=[np.sum])"
"DayHourDemand.reset_index(inplace=True)"
"DayHourDemand.columns = DayHourDemand.columns.get_level_values(0)"
"DayHourDemand = DayHourDemand.rename(columns={('DAY_TYPE',''): 'DAY_TYPE'})"
"DayHourDemand = DayHourDemand.rename(columns={('HOUR',''): 'HOUR'})"
"DayHourDemand = DayHourDemand.rename(columns={('ROUTE',''): 'ROUTE'})"
"DayHourDemand = DayHourDemand.rename(columns={'sum': 'TAPS'})"

"# missing Schedule TJT / Fixed 10th and 95th TJT / Fixed 10th"
"day['EPTI_AWT'] = 1 + ((day['Perc_95th','ACT_AWT'] - day['Perc_10thACT_AWT']) /
day['Perc_10thACT_AWT'])"
"day['EPTI_IVT'] = 1 + ((day['Perc_95th','ACT_IVT'] - day['Perc_10thACT_IVT']) /
day['Perc_10thACT_IVT'])"
```

```
"day['EPTI_TJT'] = 1 + ((day['Perc_95th','ACT_TJT'] - day['Perc_10thACT_TJT']) /
day['Perc_10thACT_TJT'])"

"Cols = [('mean','ACT_AWT'), ('mean','ACT_IVT'),
('mean','ACT_TJT'),('Perc_95th','ACT_AWT'), ('Perc_95th','ACT_IVT'),
('Perc_95th','ACT_TJT'),('Perc_10thACT_AWT'), ('Perc_10thACT_IVT'), ('Perc_10thACT_TJT'),
('EPTI_AWT'), ('EPTI_IVT'), ('EPTI_TJT')]"
"day[Cols] = day[Cols].multiply(day[('TAPS','')], axis='index')"
"day.drop('HOUR', axis=1, inplace=True)"
"day.drop('DAY_TYPE', axis=1, inplace=True)"
"day.drop('ROUTE', axis=1, inplace=True)"
"day.rename(columns=''.join, inplace=True)"

"day = pd.pivot_table(day.dropna(),index=['ROUTE','DAY_TYPE','HOUR'],
values=['meanACT_AWT','meanACT_IVT','meanACT_TJT','Perc_95thACT_AWT','Perc_95thACT_IVT','
Perc_95thACT_TJT','Perc_10thACT_AWT','Perc_10thACT_IVT','Perc_10thACT_TJT','EPTI_AWT','EP
TI_IVT','EPTI_TJT'],aggfunc=[sum])"
"day.reset_index(inplace=True)"
"day.columns = day.columns.get_level_values(0)"
"day.columns.values[3] = 'EPTI_AWT'"
"day.columns.values[4] = 'EPTI_IVT'"
"day.columns.values[5] = 'EPTI_TJT'"
"day.columns.values[6] = 'AWT_Perc_10th'"
"day.columns.values[7] = 'IVT_Perc_10th'"
"day.columns.values[8] = 'TJT_Perc_10th'"
"day.columns.values[9] = 'AWT_Perc_95th'"
"day.columns.values[10] = 'IVT_Perc_95th'"
"day.columns.values[11] = 'TJT_Perc_95th'"
"day.columns.values[12] = 'ACT_AWT'"
"day.columns.values[13] = 'ACT_IVT'"
"day.columns.values[14] = 'ACT_TJT'"

"#Weight"
"LeftJoinCols = ['DAY_TYPE', 'HOUR', 'ROUTE']"
"RightJoinCols = ['DAY_TYPE', 'HOUR','ROUTE']"
"weighted = pd.merge(day, DayHourDemand[['TAPS']+RightJoinCols], how='left',
left_on=LeftJoinCols, right_on=RightJoinCols)"
"weighted.reset_index(inplace=True)"

"Cols =
['EPTI_AWT','EPTI_IVT','EPTI_TJT','AWT_Perc_10th','IVT_Perc_10th','TJT_Perc_10th','AWT_Pe
rc_95th','IVT_Perc_95th','TJT_Perc_95th','ACT_AWT','ACT_IVT','ACT_TJT']"
"weighted[Cols] = weighted[Cols].divide(weighted['TAPS'], axis='index')"
"weighted.drop('index', axis=1, inplace=True)"
"weighted.drop('TAPS', axis=1, inplace=True)"
"weighted.to_csv(r'" & sHourlyResultFile & "', mode='a', index=False,
float_format='%.5f')"
```

5. Python Demand

```
"import odbc"
"import pandas as pd"

"connectstr = 'DRIVER=SQL
Server;SERVER=10.107.24.63\INS005;UID=EXCEL_USER;PWD=EXCEL_USER;APP=2007 Microsoft Office
system;WSID=PDC2CXP017;DATABASE=MASTER'"
"connect = odbc.odbc(connectstr)"
"sql = ""SELECT routeid AS ROUTE, DIRECTIONOFTRAVEL AS DIRECTION, CHOOSE(DATEPART(DW,
dateAdd(day,daykey,'1980-01-
01')),'Sunday','Weekday','Weekday','Weekday','Weekday','Weekday','Saturday') AS DAY_TYPE,
DATEPART(hh, fromtime) AS HOUR, fromnodeid, tonodeid, COUNT(*) TAPS FROM
pdwlink.odx_data.dbo.odx_output WHERE DayKey BETWEEN " & iP09StartDate & " AND " &
iP09EndDate & " AND routeid IN(" & sRoutes & ") AND tonodeid <> '' GROUP BY
CHOOSE(DATEPART(DW, dateAdd(day,daykey,'1980-01-
01')),'Sunday','Weekday','Weekday','Weekday','Weekday','Weekday','Saturday'), routeid,
DIRECTIONOFTRAVEL, DATEPART(hh, fromtime), fromnodeid, tonodeid"""
"odx = pd.read_sql_query(sql, connect)"
"odx = odx[odx['ROUTE'] != '']"
"odx['HOUR']= pd.to_datetime(odx['HOUR'], format='%H', errors='coerce' )"
"odx.to_csv(r'" & sDemandFile & "', mode='w', index=False)"

"sql = ""SELECT routeid AS ROUTE, DATEPART(hh, fromtime) AS HOUR, COUNT(*) TAPS FROM
pdwlink.odx_data.dbo.odx_output WHERE DayKey = " & iDate & " AND routeid IN(" & sRoutes &
") GROUP BY routeid, DATEPART(hh, fromtime)"""
"odx = pd.read_sql_query(sql, connect)"
"odx = odx[odx['ROUTE'] != '']"
"odx['HOUR']= pd.to_datetime(odx['HOUR'], format='%H', errors='coerce' )"
"odx = odx.dropna()"
"odx.to_csv(r'" & sHourlyDemandFile & "', mode='w', index=False)"
```

Appendix C: Risk Assessment

RISK ASSESSMENT FORM
FIELD / LOCATION WORK

The Approved Code of Practice - Management of Fieldwork should be referred to when completing this form
http://www.ucl.ac.uk/estates/safetynet/guidance/fieldwork/acop.pdf

DEPARTMENT/SECTION: BARTLETT SCHOOL OF PLANNING
LOCATION(S): TRANSPORT FOR LONDON
PERSONS COVERED BY THE RISK ASSESSMENT: Kate Hamblin

BRIEF DESCRIPTION OF FIELDWORK : Data analysis with Transport for London.

Consider, in turn, each hazard (white on black). If **NO** hazard exists select **NO** and move to next hazard section.
If a hazard does exist select **YES** and assess the risks that could arise from that hazard in the risk assessment box.
Where risks are identified that are not adequately controlled they must be brought to the attention of your Departmental Management who should put temporary control measures in place or stop the work. Detail such risks in the final section.

ENVIRONMENT	**The environment always represents a safety hazard. Use space below to identify and assess any risks associated with this hazard**
e.g. location, climate, terrain, neighbourhood, in outside organizations, pollution, animals.	Examples of risk: adverse weather, illness, hypothermia, assault, getting lost. Is the risk high / medium / low ? Low

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**
☐	work abroad incorporates Foreign Office advice
X	participants have been trained and given all necessary information
☐	only accredited centres are used for rural field work
☐	participants will wear appropriate clothing and footwear for the specified environment
☐	trained leaders accompany the trip
☐	refuge is available
☐	work in outside organisations is subject to their having satisfactory H&S procedures in place
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

EMERGENCIES	Where emergencies may arise use space below to identify and assess any risks
e.g. fire, accidents	Examples of risk: loss of property, loss of life

Low risk – office environment with established emergency protocol.

CONTROL MEASURES	Indicate which procedures are in place to control the identified risk

☐	participants have registered with LOCATE at http://www.fco.gov.uk/en/travel-and-living-abroad/
☐	fire fighting equipment is carried on the trip and participants know how to use it
X	contact numbers for emergency services are known to all participants
X	participants have means of contacting emergency services
X	participants have been trained and given all necessary information
X	a plan for rescue has been formulated, all parties understand the procedure
☐	the plan for rescue /emergency has a reciprocal element
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

EQUIPMENT	Is equipment used?	NO	If 'No' move to next hazard If 'Yes' use space below to identify and assess any Risks
e.g. clothing, outboard motors.	Examples of risk: inappropriate, failure, insufficient training to use or repair, injury. Is the risk high / medium / low ?		

CONTROL MEASURES	Indicate which procedures are in place to control the identified risk

☐	the departmental written Arrangement for equipment is followed
☐	participants have been provided with any necessary equipment appropriate for the work
☐	all equipment has been inspected, before issue, by a competent person

☐	all users have been advised of correct use
☐	special equipment is only issued to persons trained in its use by a competent person
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

LONE WORKING	**Is lone working a possibility?**	**NO**	**If 'No' move to next hazard** **If 'Yes' use space below to identify and assess any risks**
e.g. alone or in isolation lone interviews.	Examples of risk: difficult to summon help. Is the risk high / medium / low?		

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**
☐	the departmental written Arrangement for lone/out of hours working for field work is followed
☐	lone or isolated working is not allowed
☐	location, route and expected time of return of lone workers is logged daily before work commences
☐	all workers have the means of raising an alarm in the event of an emergency, e.g. phone, flare, whistle
☐	all workers are fully familiar with emergency procedures
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

ILL HEALTH	The possibility of ill health always represents a safety hazard. Use space below to identify and assess any risks associated with this Hazard.
e.g. accident, illness, personal attack, special personal considerations or vulnerabilities.	Examples of risk: injury, asthma, allergies. Is the risk high / medium / low? Low – normal work place

CONTROL MEASURES	Indicate which procedures are in place to control the identified risk
X	an appropriate number of trained first-aiders and first aid kits are present on the field trip
☐	all participants have had the necessary inoculations/ carry appropriate prophylactics
☐	participants have been advised of the physical demands of the trip and are deemed to be physically suited
☐	participants have been adequate advice on harmful plants, animals and substances they may encounter
☐	participants who require medication have advised the leader of this and carry sufficient medication for their needs
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

TRANSPORT	Will transport be required	NO		Move to next hazard
		YES	X	Use space below to identify and assess any risks
e.g. hired vehicles	Examples of risk: accidents arising from lack of maintenance, suitability or training Is the risk high / medium / low? Low – London Transport Network			

CONTROL MEASURES	Indicate which procedures are in place to control the identified risk
X	only public transport will be used
☐	the vehicle will be hired from a reputable supplier
☐	transport must be properly maintained in compliance with relevant national regulations
☐	drivers comply with UCL Policy on Drivers http://www.ucl.ac.uk/hr/docs/college_drivers.php
☐	drivers have been trained and hold the appropriate licence

☐	there will be more than one driver to prevent driver/operator fatigue, and there will be adequate rest periods
☐	sufficient spare parts carried to meet foreseeable emergencies
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

DEALING WITH THE PUBLIC	**Will people be dealing with public**	**NO**	**If 'No' move to next hazard** **If 'Yes' use space below to identify and assess any risks**
e.g. interviews, observing	Examples of risk: personal attack, causing offence, being misinterpreted. Is the risk high / medium / low?		

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**

☐	all participants are trained in interviewing techniques
☐	interviews are contracted out to a third party
☐	advice and support from local groups has been sought
☐	participants do not wear clothes that might cause offence or attract unwanted attention
☐	interviews are conducted at neutral locations or where neither party could be at risk
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

WORKING ON OR NEAR WATER	**Will people work on or near water?**	**NO**	**If 'No' move to next hazard** **If 'Yes' use space below to identify and assess any risks**
e.g. rivers, marshland, sea.	Examples of risk: drowning, malaria, hepatitis A, parasites. Is the risk high / medium / low?		

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**

☐	lone working on or near water will not be allowed
☐	coastguard information is understood; all work takes place outside those times when tides could prove a threat
☐	all participants are competent swimmers

☐	participants always wear adequate protective equipment, e.g. buoyancy aids, wellingtons
☐	boat is operated by a competent person
☐	all boats are equipped with an alternative means of propulsion e.g. oars
☐	participants have received any appropriate inoculations
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

MANUAL HANDLING (MH)	**Do MH activities take place?**	**NO**	**If 'No' move to next hazard** **If 'Yes' use space below to identify and assess any Risks**
e.g. lifting, carrying, moving large or heavy equipment, physical unsuitability for the task.	Examples of risk: strain, cuts, broken bones. Is the risk high / medium / low?		

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**

☐	the departmental written Arrangement for MH is followed
☐	the supervisor has attended a MH risk assessment course
☐	all tasks are within reasonable limits, persons physically unsuited to the MH task are prohibited from such activities
☐	all persons performing MH tasks are adequately trained
☐	equipment components will be assembled on site
☐	any MH task outside the competence of staff will be done by contractors
☐	OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

SUBSTANCES	**Will participants work with substances**	**NO**	**If ‘No’ move to next hazard** **If ‘Yes’ use space below to identify and assess any risks**
e.g. plants, chemical, biohazard, waste	Examples of risk: ill health - poisoning, infection, illness, burns, cuts. Is the risk high / medium / low?		

CONTROL MEASURES	**Indicate which procedures are in place to control the identified risk**

- ☐ the departmental written Arrangements for dealing with hazardous substances and waste are followed
- ☐ all participants are given information, training and protective equipment for hazardous substances they may encounter
- ☐ participants who have allergies have advised the leader of this and carry sufficient medication for their needs
- ☐ waste is disposed of in a responsible manner
- ☐ suitable containers are provided for hazardous waste
- ☐ OTHER CONTROL MEASURES: please specify any other control measures you have implemented:

OTHER HAZARDS	**Have you identified any other hazards?**	**NO**	**If ‘No’ move to next section** **If ‘Yes’ use space below to identify and assess any risks**
i.e. any other hazards must be noted and assessed here.	Hazard: Risk: is the risk		

CONTROL MEASURES	**Give details of control measures in place to control the identified risks**

Have you identified any risks that are not adequately controlled?	**NO**	☒	**Move to Declaration**
	YES	☐	**Use space below to identify the risk and what action was taken**

Is this project subject to the UCL requirements on the ethics of Non-NHS Human Research?	NO
If yes, please state your Project ID Number	
For more information, please refer to: http://ethics.grad.ucl.ac.uk/	

DECLARATION	The work will be reassessed whenever there is a significant change and at least annually. Those participating in the work have read the assessment.
	Select the appropriate statement:
X	I the undersigned have assessed the activity and associated risks and declare that there is no significant residual Risk
X	I the undersigned have assessed the activity and associated risks and declare that the risk will be controlled by the method(s) listed above

NAME OF SUPERVISOR: John Ward

**** SUPERVISOR APPROVAL TO BE CONFIRMED VIA E-MAIL ****