**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

| **Module Title:** | Research and Professional Ethics |
| --- | --- |
| **Assessment Title:** | Research Proposal |
| **Lecturer Name:** | Rory Byrne |
| **Student Full Name:** | [Paul Ryan](mailto:sbs23013@student.cct.ie) |
| **Student Number:** | sbs23013 |
| **Assessment Due Date:** | 12/11/2023 |
| **Date of Submission:** | 12/11/2023 |

**Declaration**

| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |
| --- |

# 

# 

[**Introduction 1**](#_egvx5ncufdbs)

[**Project 2**](#_al1ae9u8iztc)

[Title 2](#_h64clfgpvnm6)

[Topic Area 2](#_o06t427s6jnj)

[**Research Objectives: 3**](#_c1fjh1y6p6tm)

[Objective One 3](#_jlwqadakd12z)

[Objective Two 3](#_s49ycwr1obef)

[Objective Three 3](#_tav40xtyk7qr)

[**Literature Review 4**](#_bjr4u5rliok)

[Introduction 4](#_63jocaasny0m)

[Literature Review 4](#_omat7w4uy4pw)

[Public perception of Dublin Bus delays 4](#_h6aq21jx2hn2)

[Current issues facing public bus reliability 5](#_3thttjln72n2)

[The tracking and understanding delays and their severity. 6](#_7gv6v55mw3xy)

[Current strategies being implemented to mitigate delays. 7](#_iqpgj6gf2i07)

[Machine Learning and Big Data solutions designed for or adaptable to predicting traffic delays 7](#_hm7szxrkuiac)

[Conclusion 8](#_isomqkon9zkq)

[**Sampling Management Strategy 10**](#_m80qqfq02h3o)

[**Primary Research Methodology 11**](#_2avynvvobgxa)

[In-Depth interviews 11](#_xcwy8ee6uhoj)

[Validity Management 12](#_br1fjxynig09)

[**Ethical and Risk Considerations 13**](#_5qsrtvfpnt8c)

[Data Protection Design by Default 13](#_9kapkq9vqbyk)

[Pseudonymisation and Anonymisation 13](#_ufvqijr0lob5)

[Informed Consent to Data Processing 13](#_tj6lfpi57vha)

[Use of previously collected data (‘secondary use’) 14](#_v62o4gm06uj)

[Data Security 15](#_fr13xtwb6pr5)

[**Conclusion 16**](#_2b5lr115wsm2)

[**References 17**](#_y99nahsol9ws)

# 

# Introduction

The following is a research proposal designed to lay out a research problem and ensure that before commencing the project itself, certain criteria are met. These criteria include ensuring the research problem is significant and novel enough to warrant investigation, that the methodology being employed is both appropriate and can be carried out by the researcher, and that any results would be both an original and tangible contribution to the research field, as well as having the potential to generate further research. It is not designed to be a template which cannot be changed or deviated from, but rather a starting point for the development of the project. It will ensure that the premise and base is strong enough to start work, but will be subject to adaptation as the research progresses.

The title and topic area will be detailed, followed by the research objectives, these are the basis of the project and will guide all the other work done to support and answer these. A comprehensive literature review will look at the current state of the art in the topic area, what previous research has been done, as well as the selection and justification of further research. Sampling is required as part of the research being performed and so a strategy for this must be identified, with the appropriate population, methods and types being identified and explained. Primary research will be undertaken as part of the project, and the appropriate research methodologies will be identified and examples of how and why these have been chosen. As part of completing the project, ethical and risk considerations that may be encountered must be taken into account. In order to address these in an appropriate manner certain strategies will be identified and listed.

# Project

## *Title*

Modelling the influence of external factors on Dublin Bus scheduled reliability, to predict future delays.

## *Topic Area*

Dublin Bus served "a total of 227.7 million passengers ... on scheduled bus services" (CSO Ireland, 2020) in 2019. With this public transport serving such a large number of the population on a daily basis, and a number of reasons why choosing public transportation is preferable to private car use, reliability is a key factor for users.

Delays to scheduled times can cause serious disruption for users who depend on the services for transport to work, education, medical services, and other basic amenities. Dublin Bus provides real time updates for all of its routes, which allows for performance comparisons to the scheduled times. The severity of delays will vary and a number of factors can influence them. Large amounts of weather data is provided by Met Eireann, crash and traffic information is available historically from a number of public and private sources, and live event lists can be compiled for the time period being analysed. These factors will be quantified and used to create a model to predicate delays and their severity.

# Research Objectives:

## *Objective One*

Analyse the general trend of how real time information compares to scheduled times in order to provide a performance baseline and an understanding of what small, moderate and severe delays constitute.

## *Objective Two*

Quantify the impact of external factors on delays to the regular bus schedules. External factors should include weather, road accident and construction data, and live events / attractions (concerts, sporting events, parades)

## *Objective Three*

Create a model which uses the external factors to predict the likelihood of delays which could be used to give advanced notice or create alternative options to alleviate the delays.

# Literature Review

## *Introduction*

This is a literature review designed to validate and inform the research problem and the project being proposed. For this literature review, the organising principles will be thematic. Therefore the sections will be divided into five different sub themes covering the problem question, topic area and research objectives. Throughout the literature review there have been twenty-two sources investigated with twelve included in this review. Out of the sources investigated, approximately eleven of the twelve were in support of the research objectives or had no position, with one indicating a different approach to the research problem.

The range of sources across sub themes included newspaper articles, journal articles, website and blog pages, peer-reviewed articles, and conference papers. Some sources were grey materials, where the source was created by someone else for another purpose, but is relevant to the topic being researched. These included certain newspaper articles and blog posts and did not make up more than 20% of the overall sources used, but were included due to the relevance to the research objectives.

## *Literature Review*

### *Public perception of Dublin Bus delays*

For citizens and regular public transport users in Ireland, it would seem based on anecdotal evidence that the performance of these transport services are not on par with other European standards, and even when adequate leave a lot to be desired, “it's not that the service is bad, it’s just that it could be so much better” (O’Connell, 2023). Some of this can be attributed to the impact of poor service on the user, even if statistically it is unusual, one bad experience can stay with a person, especially when they have no control or forewarning that there may be an issue. “For 4.4 million people, (93% of the population), the bus stop was the nearest public transport option” (Central Statistics Office, 2019), meaning that buses are generally first and foremost in the public mind when we think of public transportation. Stories such as “I had to walk home alone because the night bus didn’t show” (Wilson, 2023) are not uncommon to hear. So before looking at whether Dublin Bus does actually have an issue with delays, it can be said that there is a public perception of unreliability.

### *Current issues facing public bus reliability*

In the project being proposed, three factors which can influence the likelihood and severity of delays are being considered, weather, road incidents, and live events. All of these can contribute to congestion and impact the reliability of a scheduled bus service.

Taking the weather as the first factor, “adverse weather conditions have an impact on the level of service an operator provides. They also result in higher levels of congestion due to an increase of personal car usage.” (Hofmann and O’Mahony, 2005). This paper took the approach of looking at the effect of just weather conditions on the reliability of bus operators and showed that travel time was longer on rainy days than non-rainy days, leading to delays.

Another factor which will be looked at is road accidents and construction. These are very different in that the first is a random occurrence which we assume there is no forewarning of, whereas construction can be planned for along a specific timeline. The effect of construction on traffic flow has been looked at with a “moderate effect” being found in one specific study on the impact of an office building in Shanghai (Wang et al., 2019). Information of the impact of traffic accidents on traffic flow is harder to find, with a lot of the current information focused on the cause and outcomes of the accidents themselves.

Live events can include concerts or music festivals, sporting events, parades, or any other large organised gathering. These are almost always forecastable events, with notice being given well in advance. These events can have a variety of effects on traffic, depending on their size and duration. “Though traffic congestion accumulates leading up to the start of an event, its effects are much more pronounced at the end due to the spike in attendees leaving at the same time” (Micovison, 2019). This is an important point to bear in mind when mapping the event data, start and end times should be included as these are the points of highest impact.

### *The tracking and understanding delays and their severity.*

Currently Dublin Bus is continually measured against the standards set out in the Public Services Obligation (PSO) contract. This results in the publishing of performance reports, one such report being the ‘Dublin Bus Punctuality Report Route by Route’ (National Transport Authority, 2023) which records the punctuality and regularity of its buses. The measurements used break down between two separate groups, high frequency routes, which operate a frequency of 12 minutes or greater on a weekday, outside of the peak periods. As opposed to being measured against their scheduled times, they are measured using Excess Wait Time (EWT). This is calculated by subtracting the average planned waiting time from the average actual waiting time, and regularity standards are set out based on this. Low frequency routes are compared to their scheduled times.

### *Current strategies being implemented to mitigate delays.*

Technology is currently used to attempt to regulate and manage the flow of public transport. Transportation agencies rely on software for traffic planning, “it’s essential to plan services efficiently and be able to estimate expected demand” (Baeli, 2021) Dublin Bus states that part of their strategy is “embracing new technology that enhances our work process” (Dublin Bus, 2023) but there is not specifics on what such technology is.

### *Machine Learning and Big Data solutions designed for or adaptable to predicting traffic delays*

Transit signal priority is a proposed method for improving transit system punctuality and efficiency by “developing the five following models: a bus status judgement model, a bus dwelling model, a speed guidance model, a step signal priority control model, and a speed adjustment model” (Wu et al., 2016). This however is a model which would take an effect on the surroundings, in order to alleviate delays, whereas the approach we are taking is to predict delays, not necessarily find a solution to them.

Other work has been done which is more in line with our approach, one such study used "a forward stepwise feature selection process…to select the variables with the highest predictive impact on the models” (Kinder, 2021). The kind of models which could be used range, although one paper suggests that “modelling future traffic conditions often relies heavily on complex…neural networks” (Zhu et al., 2023). This could be a good direction to go in considering the benefits of such models like “overlooking the inherent noise in the data” (Zhu et al., 2023), which is something likely to be present in the proposed project.

## *Conclusion*

The purpose of the literature review was to validate and inform the research problem, as well as guiding future research. After having completed the full review of all the sources listed, as well as examining a number of others for their relevance to the topic, the following can be surmised.

There is a definite public perception that the level of service provided by public transport in Ireland is not meeting the needs of the users, and that much could be done to improve this. Reliability is a key component of this and addressing it is important to encourage the use of public transportation in the future. There are factors which can impact the reliability of bus services, these include weather, traffic conditions such as accidents and construction, and live events. Some of these factors are more easily quantifiable and have significant research already undertaken, such as weather and construction. The higher occurrence of this type of material may be related to the amount of information publicly available. The methods used to measure the effects of each of these factors were looked at with best practices being noted for future use. Understanding how delays and reliably is recorded will be an important element of the project and this was looked at, with Dublin Bus providing a standard of measurement. The current strategy on mitigating delays is less clear in the case of Dublin Bus, with not a large amount of specific information available, but in general it is understood that companies are looking to technology to address reliability issues.

Finally the usage of big data processing and machine learning models as a solution to traffic prediction was looked at, with the current state of the art confirming that a neural network model is likely to be the most efficient option.

The research reviewed in this paper has some gaps that the proposal aims to address, primarily taking more than one factor into consideration when attempting to model the likelihood and severity of delays; this has been done previously but only taking weather as a factor. There also seems to be a lack of research and information around current methods being used to alleviate these issues, the in-depth interviews being proposed as part of this research should address some of that gap.

# Sampling Management Strategy

The sampling management strategy for this project will work to give a representative proportion of the relevant population, without having to conduct the primary research on all population cases (individual units).

In the problem statement of predicting future delays based on external factors, the population is any Dublin bus user or affiliated worker/representative.

In this case the sampling method will be non-probability, where some cases of the population will have no chance of selection, bias is inherent in this method. Self-selection, sometimes referred to as judgement sampling is the style that will be employed. The rationale behind choosing this style is that the primary research will be conducting in-depth interviews with a number of stakeholders in different areas of the problem statement. Self-selection allows the researcher to use their own bias to control the sample and select cases based on the knowledge they have of the problem statement, or how greatly they are impacted by it, or some combination of the two. It is important to be confident that the number of interviews conducted, and the stakeholder areas which they represent are a full representation of the entire population.

# Primary Research Methodology

## *In-Depth interviews*

Primary research will be undertaken as part of this project, this is research carried out by the researcher, specifically for the research problem and that is not previously available.

For this project it is intended to conduct in-depth interviews with stakeholders in the problem statement.

The sampling strategy for these interviews has been discussed and the groups being proposed are:

1. Daily Dublin Bus users

These are commuters who use Dublin bus as their primary method of transportation to either work or education, or as the primary transportation mode in general.

These users have no impact on the delays or the ability to mitigate these delays, but they are the most impacted group, and will be key to understanding the necessity of such research. The group size should be a minimum of three separate users.

1. Dublin Bus drivers

The drivers spend more time in buses, on the roads than any other stakeholder group, which puts them in a unique position to speak of the size and impact of this problem. Delays can be both in and outside of the drivers control, and can negatively impact them in a number of ways. Understanding the training, processes and realities of facing these issues as seen by the workers will enrich this side of the research. The group size should be a minimum of three drivers, ideally from separate depots.

1. Dublin Bus / National Transport Authority representatives.

Speaking to those whose responsibility it is to monitor and reduce these delays should give insight to the most current strategies employed and the official positions held on such issues. Conducting an interview with a Dublin Bus manager who is overseeing the current delays and working within the organisation to solve these issues could help influence the type of model being considered in the research problem. Interviewing an official from the National Transport Authority would gather the official position on how Dublin Bus as a whole addresses delays and modelling external factors into its strategies. The total group size would be two, one interview with each of the mentioned representatives.

## *Validity Management*

The two areas of validity management being specifically examined for this project are relevance and reliability,

The relevancy of the primary research is demonstrated by the sampling strategy, choosing the cases that have a high level of knowledge or experience with the problem statement.

A large amount of secondary research will be datasets containing information on bus times, weather, traffic incidents and live events. The first two elements have state organisations that publish live and historical data on these topics that can be considered reliable. Traffic incidents are reported on by government bodies but the scope and recency of such data will need to be explored. Live event data is provided by a number of outlets and may need to be collected into one usable source.

# Ethical and Risk Considerations

Ethical and risk considerations may be encountered when completing this project, these will be discussed and strategies to address them will be identified.

## *Data Protection Design by Default*

The first step in both mitigating issues completely or having strong procedures in place to address those that may appear will be to apply the ‘privacy by design’ concept as described by the European Commission (2021), which provides a framework to focus the, in this case project design, on a respect for data subjects’ fundamental rights.

This could involve using a measure such as the pseudonymisation and anonymisation of personal data, which will be discussed in more depth. Other measures include data minimisation and “arrangements that enable data subjects to exercise their fundamental rights (e.g. as regards direct access to their personal data and consent to its use or transfer)” (European Commission, 2021).

## *Pseudonymisation and Anonymisation*

Pseudonymisation and anonymisation of personal data can mitigate ethical concerns around the usage of personal data, making it so that data subject cannot be re-identified, is no longer personal data, and therefore is “outside the scope of data protection law” (European Commission, 2021).

It is anticipated that the data being collected as secondary research is aggregated statistics and not personally identifiable to one data subject.

The primary research involves in-depth interviews with selected groups. These subjects may be identified by job title if they are representative of Dublin bus or the NTA, the public transport group does not need to have their job title shared. For both the bus driver and transport user group anonymisation is possible with the sharing of only the group the interviewee belongs to. However true anonymisation may be difficult depending on the detail being shared as part of the interview itself.

## *Informed Consent to Data Processing*

“Informed consent is the cornerstone of research ethics” (European Commission, 2021). The procedure of obtaining informed consent is documented and the first step requires explaining the purpose of the research to the participants as well as what their participation will involve, and any associated risks. As in-depth interviews are being conducted, this procedure will have to be followed for all participants. After this information has been conveyed to and understood by the participants, the express permission to participate in the project may be obtained.

Records documenting the informed consent procedure will be maintained, including the information sheets and consent forms provided to research participants, and the acquisition of their consent to data processing. According the the paper on Ethics and Data protection, commissioned by the European Commission (2021), the definition of “informed” means that:

“[T]he data subject must be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language. As a minimum, this should include:

* the identity of the data controller and, where applicable, the contact details of the DPO;
* the specific purpose(s) of the processing for which the personal data will be used;
* the subject’s rights as guaranteed by the GDPR and the EU Charter of Fundamental Rights,
* in particular the right to withdraw consent or access their data, the procedures to follow
* should they wish to do so, and the right to lodge a complaint with a supervisory authority;
* information as to whether data will be shared with or transferred to third parties and for
* what purposes; and
* how long the data will be retained before they are destroyed.

The data subjects must also be made aware if data are to be used for any other purposes, shared with research partners or transferred to organisations outside the EU”

## *Use of previously collected data (‘secondary use’)*

The majority of the data that is anticipated to be used will be secondary sources of quantitative information provided by national organisations. This data is publically available, but all of the sources must still be documented with a clear indication that the data is open, publicly accessible and may be used for research purposes. That is to say that just because the data is accessible, does not relieve the researcher of the duty to ensure that the intended use of the data complies with and terms and conditions published by the data controller.

## Data Security

The collection of any personal data, regardless of the method, obligates the researcher both ethically and legally to ensure that the information of the participants is properly protected.

The General Data Protection Regulation (GDPR) requires “all data controllers and processors to implement appropriate technical and organisational measures to ensure a level of data security that is commensurate to the risks faced by the data subjects in the event of unauthorised access to, or disclosure, accidental deletion or destruction of, their data” (European Commission, 2021). In the case of the in-depth interviews being performed, both the participant information and the recording of the interview will be stored locally on the researcher's machine, which is password protected. The impact of a data breach as described above depends on the stakeholder group and the content of the interview. A bus driver giving a disparaging view of the organisation may feel more of a consequence than a commuter expressing their frustration. Given the localised storage of the information, the ability to edit personal information given as part of the interviews, and the unlikelihood of a concerted effort to reveal the identity of the participants, the risk of a data breach could be considered low at this point.

# 

# Conclusion

The topic and research questions of this project revolve around Dublin Bus delays, how they are impacted by external factors, and how these can be modelled to create a forecast of such delays. The purpose of the proposal has been to demonstrate the validity of the research problem, and that the achievement of the research objectives is novel, impactful and feasible. In doing so there has been a comprehensive literature review. There has also been discussion and selection of the appropriate sampling methods. A non-probability method using self-selection was determined to be the best option for the project. This was also based on the primary research methodologies, which were discussed as being in-depth interviews with different groups of stakeholders in the problem area. Finally the ethical and risk considerations were discussed with the application to the current project, and research methods being examined. A range of issues were identified, including ensuring informed consent to data processing and data security. With the proposal organised and informed by best practices, it is possible to move towards the project with confidence in the foundations of the research problem and methodologies.

# References

Baeli, V. (2021). *Public transport disruptions: How to keep passengers’ trust*. [online] PTV Blog. Available at: https://blog.ptvgroup.com/en/city-and-mobility/public-transport-disruptions-how-to-keep-passengers-trust/ [Accessed 12 Nov. 2023].

Central Statistics Office (2019). *Public Transport Services - CSO - Central Statistics Office*. [online] www.cso.ie. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-mdsi/measuringdistancetoeverydayservicesinireland/publictransportservices/ [Accessed 12 Nov. 2023].

CSO Ireland (2020). *Public Transport - CSO - Central Statistics Office*. [online] www.cso.ie. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-tranom/transportomnibus2019/publictransport/ [Accessed 6 Nov. 2023].

Dublin Bus (2023). *Dublin Bus - Dublin Bus Strategy*. [online] Dublin Bus. Available at: https://www.dublinbus.ie/policys-and-procedures/dublin-bus-strategy [Accessed 12 Nov. 2023].

European Commission (2021). *Ethics and data protection*.

Hofmann, M.J. and O’Mahony, M.J. (2005). The impact of adverse weather conditions on urban bus performance measures. *IEEE Intelligent Transportation Systems*. [online] doi:https://doi.org/10.1109/itsc.2005.1520087.

Kinder, M. (2021). *Bus Punctuality Analysis & Prediction*. [online] Valkuren. Available at: https://www.valkuren.com/blog/2021/07/14/bus-punctuality/ [Accessed 12 Nov. 2023].

Micovison (2019). *How Live Events Affect Traffic | Miovision*. [online] miovision.com. Available at: https://miovision.com/blog/live-events-and-traffic [Accessed 12 Nov. 2023].

National Transport Authority (2023). *Dublin Bus Direct Award Contract Punctuality Report*. [online] *https://www.nationaltransport.ie/public-transport-services/public-transport-contracts/operator-performance/quarterly-performance-reports/bus/*. Available at: https://www.nationaltransport.ie/wp-content/uploads/2023/10/2021-to-2023-Dublin-Bus-Punctuality-Report-Route-by-Route.pdf [Accessed 12 Nov. 2023].

O’Connell, J. (2023). *Jennifer O’Connell: Irish public transport is a lesson in how to make easy things hard*. [online] The Irish Times. Available at: https://www.irishtimes.com/opinion/2023/05/14/jennifer-oconnell-irish-public-transport-is-a-lesson-in-how-to-make-easy-things-hard/ [Accessed 12 Nov. 2023].

Wang, Z., Bai, Y., Zhu, R., Wang, Y., Wu, B. and Wang, Y. (2019). Impact Analysis of Extra Traffic Induced by Project Construction during Planned Special Events. *Transportation Research Record*, 2673(7), pp.402–412. doi:https://doi.org/10.1177/0361198119840346.

Wilson, J. (2023). *‘I had to walk home alone because the night bus didn’t show’: Your public transport tales*. [online] The Irish Times. Available at: https://www.irishtimes.com/transport/2023/05/09/many-times-i-had-to-walk-alone-home-because-the-bus-didnt-show-your-public-transport-tales/ [Accessed 12 Nov. 2023].

Wu, Z., Tan, G., Shen, J. and Wang, C. (2016). A Schedule-based Strategy of transit signal priority and speed guidance in Connected Vehicle environment. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. [online] 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEE. doi:https://doi.org/10.1109/itsc.2016.7795945.

Zhu, Y., Ye, Y., Zhao, X. and James J.Q. Yu (2023). Enhancing Traffic Prediction with Learnable Filter Module. *arXiv (Cornell University)*. [online] doi:https://doi.org/10.48550/arxiv.2310.16063.