**REAL-TIME DASHBOARD FOR ANALYZING THE DELAY IN DUBLIN BUS CAUSED BY LOCAL WEATHER**

**ABSTRACT**

Public transportation plays a crucial role in the daily lives of citizens across countries, facilitating their routine activities. Dublin Bus stands as a prominent public transport system within the city of Dublin, greatly facilitating the daily lives of its residents. However, occasional unforeseen delays in bus arrivals may inconvenience passengers, despite the overall utility the service provides. This study aims to analyze the delays experienced by Dublin buses on individual routes and explore any potential correlation with local weather conditions. It seeks to determine whether local weather factors contribute to these delays. The method involves developing a real-time dashboard to showcase delays, calculated using the NTA dataset, which is GTFS and GTFS-R data, in conjunction with local weather data. The dashboard provides valuable assistance to the public by enabling users to identify delays on specific Dublin bus routes at particular times or weather conditions. This functionality aids in planning journeys more effectively, especially during specific time frames or weather scenarios. The summary of the primary findings indicates that local weather significantly contributes to delays in Dublin buses.

**Keywords: Dublin Bus, Delay, Real-time dashboard**

1. **INTRODUCTION**

For any modern country, public transportation plays a crucial role in the daily lives of its citizens. A fast and reliable transportation network is vital for easy access across the country. Time is invaluable and cannot be regained once lost, making punctuality a key aspect of public transport services. One of the significant challenges in public transport is the city service, which is heavily utilized by most people for their daily activities such as commuting to work, school, and other destinations. The reliability and availability of city services, like buses, are crucial, especially in cities like Dublin, where the Dublin Bus system is the primary mode of transportation due to its widespread availability and dependability. Dublin Bus is an Irish state-owned bus operator providing services in Dublin(‘Dublin Bus’, 2024). Dublin bus is the largest bus operator in the city, it carried over 145 million passengers in the year of 2023(2024).

Despite being one of the most widely utilized public transportation systems in Ireland, Dublin Bus occasionally encounters unforeseen circumstances leading to significant delays. At times, passengers experience prolonged waiting periods before boarding a bus, and even upon boarding, journeys may take longer than expected to reach their destinations. Dublin Bus, affiliated with the National Transport Authority of Ireland, offers advanced services including travel schedules, estimated arrival times, travel predictions, vehicle positioning, and real-time updates to passengers. However, despite these provisions, passengers may occasionally experience delays in reaching their destinations within the estimated arrival times. Road transport, especially buses, is vulnerable to delays influenced by factors like traffic, weather conditions, and road infrastructure. Weather significantly affects journey durations. Adverse weather conditions not only affect the operational efficiency of buses but also alter road conditions, leading to increased travel times and extended stops at each station, consequently elongating transit durations. “Weather can affect the total trip duration by increasing the access time, transfer time and the normal trip duration and also by causing schedule disruptions”(Singhal, Kamga and Yazici, 2014). The aim is to examine the influence of weather on Dublin Bus, specifically assessing its impact on travel times and potential delays. This analysis seeks to understand how varying weather conditions contribute to fluctuations in the efficiency and punctuality of bus transportation services.

* 1. **Research Question**

Arrived at research question by initially selecting the broader topic of challenges in public transport in Ireland. Subsequently, honed in on the research problem, focusing specifically on the timeliness of public transport, particularly buses. The purpose statement outlined the objective of analyzing the causes of delays in scheduled arrival times. This process led to the final research question concerning the correlation between local weather conditions and the punctuality of public transport. And the research question is follows:

***“How much does the local weather affect the delay of Dublin buses?”***

Figure 1 Narrowed down to the Research Question

dfff

Problems in Public Transport in Ireland

Timeliness of Public Transport in Ireland, especially Dublin Bus

Analysis on the causes of delays in scheduled arrival time

Question?

* 1. **Research aim(s)**

The aim is to publish the extent to which weather conditions impact delays in public buses, especially Dublin buses, determining which weather conditions are most conducive to bus delays. The study also seeks to publish the average/current duration of delays associated with each distinct weather condition or distinct times of the day in general or from specific bus routes.

* 1. **Research objective(s)**

The National Transport Authority of Ireland provides a transport feed service containing essential information such as trip schedules and real-time updates, including vehicle positioning and arrival times. Utilizing this data along with local weather information from Dublin city allows us to conduct diverse analytical activities to achieve our research objectives. This involves publishing bus delays on specific days under a particular weather condition and identifying which weather conditions contribute most to delays. The objective also includes displaying the current delay, both in general and for specific routes of Dublin Bus.

* 1. **Research hypothesis**

Dublin Bus delays are more probable during days characterized by rain and strong winds. Adverse weather conditions, specifically rain and wind, tend to have a notable impact on the timeliness of bus services, increasing the likelihood of delays during such conditions.

1. **LITERATURE REVIEWS**

“Weather conditions are considered exogenous factors which indirectly influence the demand for transit”(Singhal, Kamga and Yazici, 2014). Bad weather conditions directly affect both the accessibility and the quality of transit services. The study done by Singhal, Kamga and Yazici in 2014 is analyzing the impact of weather on the Metropolitan Transportation Authority-New York City Transit subway ridership. They obtained two years of ridership data and weather information from the National Oceanic and Atmospheric Administration (NOAA) and the Weather Underground website. Subsequently, they developed a model using ordinary least squares regression (OLS).

“Weather conditions and built environment contribute 30.22 and 55.83% to ridership fluctuations, respectively”(Lin *et al.*, 2020). “Adverse weather, such as strong wind, high humidity, or heavy rainfall, has a more disruptive impact on leisure-related areas than on residence and office areas”(Lin *et al.*, 2020). The study of ‘Analyzing the relationship between weather, built environment, and public transport ridership’ done by Lin et al, gathered smart card data in Beijing, China, spanning from February 2018 to January 2019, including both bus and subway usage. The study considered various factors, including daily weather conditions (such as temperature, wind speed, humidity, rainfall, snowfall, and air quality) and built environment factors (like residential and office density, as well as accessibility of public transport infrastructure in a Traffic Analysis Zone). The Light Gradient Boosted Machine (LightGBM) algorithm was utilized to examine the influence of weather conditions (Lin *et al.*, 2020).

“Weather has a huge impact on many aspects of traveler’s travel decision, for example, departure time, route and mode choices”(Khattak and De Palma, 1997; Miranda-Moreno and Nosal, 2011). “All four weather variables, namely humidity, wind speed, rainfall and temperature are found to have statistically significant negative effects on bus ridership”(Li *et al.*, 2015). The study done by Li et al categorizes the ridership data from the smartcard according to route types (RTs) and seasons, and then analyzes the impact of different weather factors on various types of routes separately (Li et al., 2015).

“The effect of weather can be direct, by making people drive more slowly, or indirect, e.g. by people choosing different means of transportation or by increasing the amount of accidents which can then cause even more congestion” (Peltola, 2019). This study was carried out by recording bus link travel times between consecutive bus stops from live location data provided from Tampere regional buses during different weather circumstances (Peltola, 2019).

The research on forecasting irregularities in transit bus arrival times by Alam et al. involved predicting the likelihood of arrival time disruptions. They utilized GPS coordinates data provided by the Toronto Transit Commission (TTC) in conjunction with hourly weather data. Machine learning models, particularly Long Short-Term Memory Recurrent Neural Network (LSTM) models, were employed in this study (Alam *et al.*, 2021).

Turning to studies conducted on Dublin Bus, the research by Akhil Alfons and Kirthy Francis centers on constructing a system that gathers transit feed in real-time and utilizes this data to forecast delays (Kodiyan and Francis, 2020). Another study conducted by Pandurangi et al. focuses on predicting bus travel times through a Segmentation Approach. They developed a user-friendly application that employs machine learning techniques to estimate the travel time of buses and destinations (Pandurangi *et al.*, 2020). The research done by French and O’Mahony utilized Automatic Vehicle Location System (AVLS) data to explore the influence of adverse weather conditions on bus journey times. Rainfall, temperature, and wind speed were utilized as indicators to assess their impact. Additionally, that study aimed to determine if the impacts varied across three types of bus routes: those with bus lanes along the entire route, those with bus lanes along part of the route, and those where buses operate in mixed traffic (French and O’Mahony, 2021).

The rationale is to enhance the public transit planning through real-time analysis of Dublin bus delays and weather. From previous research and studies, the problem becomes evident. Various analytical methods have been employed in previous studies to address the challenges related to timelines faced on public transport like Dublin Bus. However, this research introduces a novel approach by providing a real-time dashboard displaying delay information alongside corresponding weather data. This dashboard enables individuals to access current delay information as well as historical delays, which can vary due to external factors. The significance of a real-time dashboard lies in its ability to assist the public in planning their journeys more effectively, offering additional insights into potential delays they may encounter during their travels.

The research is conducted by collecting and analyzing data using the General Transit Feed Specification (GTFS). The General Transit Feed Specification (GTFS) is an open standard used by most of the public transport providers to publish relevant information about transit systems to riders (‘GTFS: Making Public Transit Data Universally Accessible’, no date). Since there is no existing dataset available for this study, data needs to be recorded in the desired format. This involves utilizing GTFS schedule and real-time data to capture information in a specific format. By integrating this data with weather data, the goal is to establish a robust real-time data analytical model. The objectives include investigating the impact of weather conditions on delays and determining which weather conditions are more conducive to public transport delays, all accessible through the dashboard.

1. **METHODOLOGY**

For any dashboard, the primary focus is on data flow. Data must seamlessly transition from its source to the backend services and ultimately display in the specific format on the dashboard. The key consideration is ensuring this process occurs smoothly, given that the dashboard operates in real-time. Dublin Bus collaborates with the National Transport Authority, disseminating schedule and trip updates to passengers in GTFS format. GTFS, or General Transit Feed Specification, is the standard format for this data. The methodology involves utilizing both GTFS and GTFS-R data along with the open weather data from Dublin city to compile the information for the delay dashboard.

* 1. **GTFS**

The General Transit Feed Specification (GTFS) is an open standard employed to disseminate pertinent details regarding transit systems to passengers. This standard enables public transit agencies to publish their transit data in a format that can be easily utilized by a wide array of software applications. Presently, the GTFS data format is adopted by numerous public transport providers. GTFS consists of two primary components: GTFS Schedule and GTFS Realtime. GTFS Schedule encompasses information pertaining to routes, schedules, fares, and geographic transit particulars, all presented in simple text files. This user-friendly format facilitates simple creation and maintenance without the need for complex or proprietary software. On the other hand, GTFS Realtime includes updates on trips, vehicle positions, and service alerts. It utilizes Protocol Buffers, which serve as a language- and platform-neutral mechanism for serializing structured data (‘GTFS: Making Public Transit Data Universally Accessible’, no date).

* 1. **DATA COLLECTION**

The data collection process entails gathering GTFS static data and creating a code module responsible for extracting delays from all major routes and inserting them into the database. The database used is the GTFS schedule, which primarily consists of static transit information. This data is structured into several text files (.txt) contained within a single ZIP file. Each file within the archive describes a specific aspect of transit information, such as stops, routes, trips, and fares (‘GTFS: Making Public Transit Data Universally Accessible’, no date). The GTFS static data is accessible for download from the National Transport Authority's developer portal and can be imported into a database for further analysis. GTFS Realtime is a feed specification enabling public transportation agencies to deliver real-time updates about their fleet to application developers. It serves as an extension to GTFS (General Transit Feed Specification), an open data format for public transportation schedules and related geographic data. GTFS Realtime prioritizes ease of implementation, seamless interoperability with GTFS, and a strong emphasis on providing passenger information (‘GTFS: Making Public Transit Data Universally Accessible’, no date). The feeds are delivered through HTTP and are updated regularly. Among the plethora of information provided in GTFS-R, the focus is on collecting trip updates. This category primarily includes details about delays, cancellations, and route modifications, all of which are crucial for monitoring the real-time status of transit operations. The TripUpdates services are invoked at regular intervals to calculate delay parameters, and the resulting data can be pushed into the database. This database can be either the same as the one containing the static GTFS data or a separate one.

One challenge encountered when reading GTFS static data is the unpredictable nature of changes resulting from updated route and timetable information provided by operators. These changes, originating from the operator or agency level, can occur unexpectedly. The solution recommended by the NTA help desk is to regularly update the static data to prevent discrepancies between real-time and static GTFS data. Identified this issue during the initial stages and labeled it as the static data refresh problem. Developed a solution at the same stage, which involves creating a dedicated table for mapping bus route IDs to bus route names and indicating active routes. The visual representation of this solution to the static data refresh problem is as follows:

Figure 2 Solution for static data refresh problem.

A diagram of a data processing process

Description automatically generated

This solution can be implemented within the NTA static loader module, responsible for reading static data dumps from the NTA and loading the data into the master database. Originally intended for one-time execution, the static loader module will now be triggered each time a refresh occurs in the static data due to the static data refresh problem. The NTA static loader module can be integrated into the extract engine, which operates at periodic intervals. Whenever an exception arises due to a mismatch between existing route IDs in the master database and route IDs from the latest dump, control is transferred to the NTA static loader module. Here, the module identifies the mismatch and proceeds to clear all static tables before updating them with the latest data from the NTA static dump. Upon completion of the process, the NTA static loader updates the route mapping table, maintaining records of all route IDs to date, and marking the current route IDs as active. The advantage of this solution is that the dashboard will have access to both historic and current delay information.

Another crucial dataset required for the delay dashboard is weather data. Numerous weather services offer forecasts or current weather conditions for any given location. Open-Meteo is one such weather service that provides up to 10,000 API requests per day for free of cost. Open-Meteo collaborates with national weather services to provide open data with high resolution, ranging from 1 to 11 kilometers. The concept is to gather weather data from Dublin city concurrently with the collection of GTFS-R data. The periodic data will encompass various types of weather information at different times of the day. To ensure data consistency, ensure that each delay entry is accompanied by corresponding weather data. This weather data can also be inserted into the same database used for storing the GTFS-R data.

* 1. **DATA FORMAT**

GTFS data adheres to a specific structure, which is also followed by the NTA dataset. However, when it comes to the data requirements for the delay dashboard, the data must be stored in a specific format that can be easily retrieved and processed for displaying on the dashboard. The goal is to maintain data integrity, ensuring that each delay entry is identified as a distinct occurrence. To achieve this, each entry must be assigned a unique identifier, serving as an index. Additionally, every delay entry must include a timestamp indicating when the delay occurred. To display delay information for routes, the route\_id, which serves as the unique identifier for a specific route, is required. Additionally, the direction of the route where the delay information is calculated must also be included. When it comes to the data format of weather information, it is straightforward to utilize the data supplied by Open-Meteo. This data can be paired with the entry ID, serving as the unique identifier for the corresponding delay information. Additionally, the location from which the weather is obtained, and the timestamp of the data entry must also be included. The weather data supplied by Open-Meteo includes temperature, relative humidity, apparent temperature, day or night, precipitation, rain, showers, snowfall, weather codes, cloud cover, pressure, surface pressure, wind speed, wind direction and wind gusts. The weather codes follow WMO (WORLD METEOROLOGICAL ORGANIZATION) Weather interpretation codes:

|  |  |
| --- | --- |
| **Code** | **Description** |
| **0** | Clear sky |
| **1, 2, 3** | Mainly clear, partly cloudy, and overcast |
| **45, 48** | Fog and depositing rime fog |
| **51, 53, 55** | Drizzle: Light, moderate, and dense intensity |
| **56, 57** | Freezing Drizzle: Light and dense intensity |
| **61, 63, 65** | Rain: Slight, moderate, and heavy intensity |
| **66, 67** | Freezing Rain: Light and heavy intensity |
| **71, 73, 75** | Snow fall: Slight, moderate, and heavy intensity |
| **77** | Snow grains |
| **80, 81, 82** | Rain showers: Slight, moderate, and violent |
| **85, 86** | Snow showers slight and heavy |
| **95 \*** | Thunderstorm: Slight or moderate |
| **96, 99 \*** | Thunderstorm with slight and heavy hail |

(\*) Thunderstorm forecast with hail is only available in Central Europe

* 1. **DELAY CALCULATION**

In GTFS-R, trip delays are linked to the TripUpdate API. The TripUpdate API response includes details about all active trips at the current moment, along with timestamps. This information encompasses the scheduled start time, schedule relationship (e.g., SCHEDULED or CANCELLED), direction ID indicating the trip's heading, and route ID specifying the operating route. Additionally, the response includes stop time updates, which detail arrival and departure delays at upcoming stops along the bus route. The objective is to utilize these delay values to compute the average delay of Dublin buses or the average delay on a specific route at any given moment. The average delay of Dublin buses is calculated using the following equation:

Let 𝐷𝑖​ represent the delay value for the 𝑖𝑡ℎ bus trip, where 𝑖 ranges from 1 to 𝑛. The average delay is determined by adding together the delay values from the 𝑛 bus trips and dividing by the total number of bus trips.

The delay for a specific route is calculated by first finding the average delay of each trip, and then determining the cumulative average of trips on that route. Here's the mathematical equation to calculate the average delay of a bus route:

Let 𝑇𝑖 represent the number of trips on the 𝑖𝑡ℎ bus route, where 𝑖 ranges from 1 to 𝑛 and 𝐷𝑖,𝑗 represent the average delay of the 𝑗𝑡ℎ trip on the 𝑖𝑡ℎ bus route, where 𝑗 ranges from 1 to 𝑇𝑖. The equation calculates the average delay for each bus route by taking the cumulative average of average delays of each trip on that route. The calculated delays are stored in the master database, along with unique entry IDs and timestamps, to facilitate easy retrieval for display on the dashboard.

* 1. **DASHBOARD REQUIREMENTS**

The main objective of the dashboard prioritizes functionality over design. The dashboard should effortlessly retrieve delay information from the database and display it seamlessly. It should be able to present the average delay of Dublin buses across all routes, as well as the current delay status from all routes. Additionally, the dashboard should provide historical delay data and be capable of plotting delays on graphs. Weather parameters should also be included for correlation analysis with delays. Users should have the ability to filter the graphs based on their preferences, selecting either all-time data or data from the current day. Furthermore, the dashboard should provide insights into delays through visualizations such as delays occurring at different times of the day, on different days of the week, and in various weather conditions. It should also display the distribution of weather conditions across Dublin city. These visualizations should be user-friendly, allowing users to filter the data based on parameters such as all routes, specific routes, all-time data, current day data, or selecting a particular date.

When it comes to development of the dashboard, simplicity and efficiency are paramount. The dashboard should operate seamlessly without any downtime and deliver data with all possible haste. To achieve these objectives, the methodology involves using Python Dash for dashboard development. Dash is the original low-code framework for rapidly building data apps in Python (‘Dash Python User Guide’, 2024). It's designed to make it easy to create complex, web-based data visualizations and interactive dashboards using Python, HTML, and CSS. Dash allows users to create interactive, web-based data visualizations and applications entirely in Python, without needing to know JavaScript or HTML. Dash can create interactive dashboards that update in real-time, respond to user input, and can be deployed as standalone web applications or integrated into larger web applications. It's particularly popular in data science and analytics circles for creating interactive data visualizations and dashboards that can be shared and accessed through web browsers.

1. **DESIGN**

The solution design prioritizes efficiency, reliability, and robustness. The main concept is to smoothly transfer data from the sources to the system, ensuring it is in the required format, and retrieve the data efficiently.

Figure 3 Design Diagram

A diagram of a computer system

Description automatically generated

The design diagram provides a clear visualization of the proposed solution. It outlines the seamless flow of data from the source systems to the destination, ensuring delivery in the desired format. Key components of the design include data sources, NTA Static loader, periodic extract engine, format converter, master database, and the dashboard engine.

* 1. **DATA SOURCES**

The data sources consist of NTA services and weather services. NTA services encompass both GTFS and GTFS-R data. GTFS represents static data, constituting a one-time dump containing text files storing trip, route, and stop details. These files are read by the NTA static loader and deposited into the database. To tackle the static data refresh issue, the NTA static loader continuously monitors static files for any alterations. Upon detecting changes, it updates the static data in the master database. The GTFS-R data is extracted by the Periodic Extract Engine at regular intervals and forwarded to the format converter for additional processing. GTFS-R contains delay information for all active trips at the current moment. This data, typically in JSON format, can be pulled using Python scripts at regular intervals. Just as with GTFS-R data, weather data can also be retrieved using Python scripts at the same intervals as GTFS-R data extraction. This weather data encompasses various parameters such as temperature, relative humidity, apparent temperature, day or night conditions, precipitation, rain, showers, snowfall, weather codes, cloud cover, pressure, surface pressure, wind speed, wind direction, and wind gusts.

* 1. **NTA Static loader**

The NTA static loader module is pivotal for incorporating GTFS static data dumps into the master database. It retrieves the static files from the NTA portal, processing each file sequentially, and then inserts the data into the master database. These files encompass essential information such as agency details, calendar schedules, route shapes, and stop times. To address the static data refresh problem, which arises from changes made to the static data by NTA, a solution is implemented. This involves maintaining a separate mapping between route names and route IDs. Upon detecting changes in the static data, the NTA static loader module updates the master database, accordingly, ensuring synchronization between the data sources and the database.

* 1. **Periodic Extract Engine**

The Periodic Extract Engine plays a crucial role in retrieving delay information and weather data at regular intervals. The Python code for the Periodic Extract Engine is designed to request Trip Updates services from NTA for GTFS-R data. Upon receiving the response, it processes the data with the help of the delay calculator module, ensuring efficient processing. The delay information for each route is then calculated and forwarded for further processing, ultimately being stored in the master database. Simultaneously, weather data is fetched from Open-Meteo services. The extracted weather parameters are processed and directly sent to the database. Ensure that both datasets are synchronized with the same timestamp to maintain data consistency throughout the process. The Extract engine's configuration allows it to execute at specified intervals, typically set to 30 minutes for optimal performance considering the data load and available computational resources. However, this interval parameter can be adjusted for more precise data or if higher computational power becomes available.

* 1. **Format Converter**

The format converter module plays a crucial role in ensuring compatibility between source and destination data types. While some data types from the GTFS-R format may differ from those in the master database, the format converter ensures that they are transformed into the appropriate format for insertion. Upon examination, it was noted that only timestamp fields require conversion, while other data types remain unchanged. Timestamps from both GTFS-R and weather services are in Unix time format. The Unix timestamp is a way to track time as a running total of seconds. This count starts at the Unix Epoch on January 1st, 1970, at UTC. Therefore, the Unix time stamp is merely the number of seconds between a particular date and the Unix Epoch (‘What is the unix time stamp?’, 2014). Therefore, the format converter handles the conversion of these timestamps to ensure consistency in the data. To achieve this, the code can be implemented to convert the Unix timestamp to a local timestamp compatible with the SQL database. This conversion involves localizing the timestamp based on the time zone of Dublin, ensuring that the entry time reflects the local time in Dublin without conflicting with the time zone of the database. Additionally, this approach resolves issues related to different times due to daylight saving changes.

* 1. **Database**

As per the design, the database should be a relational database. Relational databases are preferred in this scenario because they offer structured storage of data, with each piece of information associated with specific tables. This ensures data integrity throughout the data. This structure allows each entry to be treated as individual objects, facilitating efficient retrieval of data using SQL queries when required. The GTFS schedule data can be stored as it is without making many changes. The format for each static table is as follows:

* Agency

A screenshot of a computer

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **agency\_id** | agency\_name | agency\_url | agency\_timezone |
| Integer | Varchar | Varchar | Varchar |

* Calendar

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **service\_id** | monday | tuesday | wednesday | thursday | friday | saturday | sunday | start\_date | end\_date |
| Integer | Integar |  |  |  |  |  |  |  |  |

* 1. **Dashboard Engine**

tables

TECHNOLOGIES/ IMPLEMENTATION

-DATABASE

-PYTHON

-TOOLS USED

\_LIBRARIES

-Cronjobs

-Deployment

RESULTS

Project work tracking - Trello

CHALLENGES FACED – static data refresh problem and time zone bug

ETHICAL CONSIDERATION

CONCLUSION

-limitations future study