Deep Learning and Machine Learning to improve the operation of a shared bike system, a case study of DublinBikes

MSc Research Project
Artificial Intelligence

Raghav Raghu
Student ID: 23272341

School of Computing
National College of Ireland

Supervisor: Lavish Thomas

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Raghav Raghu
Student ID:	23272341
Programme:	MSCAI1B **Year:** 2024-25
Module:	……………………………………………………………………………………………………….
Supervisor:	Lavish Thomas
Submission Due Date:	11 August 2025
Project Title:	Deep Learning and Machine learning to improve the operation of a shared bike system, a case study of DublinBikes
Word Count:	5655 **Page Count**: 16

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Raghav Raghu

Date: 11 August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deep Learning and Machine Learning to improve the operation of a shared bike system, a case study of DublinBikes

Raghav Raghu
23272341

Abstract
Bike sharing is one of the most important modes of transport, DublinBikes is one of the most used services in Dublin for the same. There has been various research done on historical data and there have been multiple evidence that machine learning algorithms help identify a lot of flaws and issues in the system, and with that we can help improve the service for a smooth operation of a service so heavily dependent for the people of Dublin. If we also include deep learning algorithms to the already implemented machine learning results, we can find a deeper more accurate or even insightful results to help identify more problems if not already found. Finally simulating the bike usage patterns can help present the issues to the common man in a simpler way.
Keywords: DublinBikes, Machine Learning, Deep Learning, Simulation

1 Introduction

With increasing population in cities and the emerging environmental concerns, the world is turning to develop sustainability and efficient means of transport. A good example is Dublin Bikes- a well-known bike sharing program that gives residents and tourists in Dublin, Ireland, an easy and environment friendly mode of transport. Although the program has enjoyed massive success, it is vulnerable to the standard operational issues, such as poor availability of bikes, overcrowding of stations, and logistic inefficiencies in terms of redistribution of bikes.

To manage this type of system effectively, data-based information on when and where the demand is the greatest is needed. The foreknowledge of peak usage periods, areas that are either under- or over- served and the strategic planning of future station locations are essential in determining the reliability of the system and the satisfaction of the users.

The current project will help to research on these challenges by developing a predictive and spatial analysis framework. Based on the historical data of the Dublin Bikes stations between January 1 and May 2025, the project uses deep learning in predicting the number of bikes used per hour per station based on the Gated Recurrent Unit (GRU) neural network. Besides time-series forecasting, the project carries out in-depth studies of peak-hour use patterns and full-capacity incidents, providing information about the times of the day when a station is most likely to be overloaded or unavailable.

Also, the project includes geospatial clustering with DBSCAN and haversine distance measures to detect areas of high demand around current stations. On this analysis, it suggests new station locations that might be used to ease the burden on the system and coverage. The visualization of the existing station distribution, the most active ones, and the proposed new locations are created in an interactive map, which helps in planning and decision-making strategies.

The proposed project is research on the Dublin Bikes bike usage data to find out if historical data can help us to be future ready in terms of standard operation, efficient demand analysis, trying to suggest expansion of bike stations and thus improve user experience.

1.1 Research Questions

Research Question 1: How well can we forecast the bike usage of a busy bike station in Dublin based on historical data to be future ready?

Research Question 2: Can we suggest new locations for bike stations around busy stations?

2 Related Work

2.1 Spatio-Temporal Forecasting in Urban Mobility

Zhang et al. (2017) presented Deep Spatio-Temporal Residual Networks (ST-ResNet) to predict the crowd flows in cities and set the stage to cope with the spatial and temporal dependencies in demand forecasting. Their study showed that proper forecasting should include the observation of local station dynamics as well as that of the city in general.

In this project, although station-specific forecasting was conducted (not over a citywide grid), temporal seasonality like time-of-day and weekday effects were included as well. Though the spatial flows between the stations are not modelled in this project, the concept of capturing periodic usage behaviour over time is core and transferred directly to such body of work.

2.2 Recurrent Neural Networks and GRU for Time-Series Forecasting

Li et al. (2018) proposed a Diffusion Convolutional Recurrent Neural Networks (DCRNN) that utilizes graph structures with GRU in traffic forecasting. Their conclusions support GRUs as efficient models of sequential temporal data that do not suffer the problems of vanishing gradients that vanilla RNNs do.

The GRU-based forecasting used in this project is at the station level and hourly bike usage is considered as a time series. Cyclical demand patterns are captured by using time-related features (hour, day_of_week) and sliding window inputs in the model. Although the graph convolutions were not employed, the temporal modeling ability of GRU is a direct implementation of the findings of Li et al.

2.3 Deep Learning Frameworks for Urban Demand Prediction

Yao et al. (2018) revisited the spatio-temporal similarity in the urban mobility prediction by using deep learning combined with multi-view learning and attention. Their contribution underlined the need to consider multiple perspectives (e.g., temporal, spatial, contextual) in modeling shared mobility.

This project reflects that philosophy by treating different forms of patterns differently: temporal demand using GRU, spatial station clustering using DBSCAN, and event-based analysis using full-usage detection. The modular separation of concerns, which you do with demand forecasting, usage patterns, and spatial planning, is methodologically similar, even though you have a simpler architecture.

2.4 Short-Term Forecasting Using Spatio-Temporal Deep Learning

Ke et al. (2017) suggested a spatio-temporal deep neural network (CNNs and GRUs) to predict short-term demand of ride-hailing. The results presented by them emphasize the

applicability of deep learning in very dynamic mobility settings particularly when the historical usage trends are used to train the deep learning algorithms.
In the same way, this project trains on the months of historical bike use and extrapolates into the near future (May and June). The concept of training on recent history to forecast the demand of operation under similar cyclical patterns is based on the short-term forecasting method by Ke et al.

2.5 Machine Learning Models

As the name suggests the machine learns patterns from the data to understand the trends of a certain feature. Huthaifa et al. used Random Forest (RF) and Least-Squares Boosting (LSBoost) at each bike stations in their research and found slightly lower error than when there was a network of stations involved. In the San Francisco Bay Area study done by Ashqar et al. found out that using the above-mentioned Machine learning algorithms with Partial Least Squares Regression (PLSR) gave an upper hand in flexible and scalable prediction of bike usage. Xu et al. made a hybrid model of a Self-Organizing Map (SOM) with Regression Trees to achieve better accuracy than individual models.

2.6 Deep Learning Models

For working with sequential data for e.g. Time-series data Deep Learning (DL) models are used by researchers. Collini et al. used Bidirectional LSTM (Bi-LSTM) to yield a reliable prediction model to predict the usage of bikes. Mehdizadeh el al. combined Convolutional Neural Networks (CNN) with LSTM for when the patterns might shift drastically which also gave them good accuracy. CNN-LSTM models can extract spatial and temporal features as well to help with the better suggestion of new stations to be added on the map.

2.7 Statistical Models

One commonly used approach is the time-series analysis. One of the most used methods for that is the Auto-Regressive Integrated Moving Average (ARIMA) model. For instance, this model is used by Klattenbrunner et.al to forecast the number of available bikes for any station in the upcoming minutes or hours. Ahmen Jabel et al. used both Seasonal ARIMA and Exponential Smoothing (ES) models to examine bike-sharing usage over a five-year period. The findings were as expected, during the summers the bike-sharing was most used and in general during the weekdays the bike sharing system is most used. Their comparison of the models concluded that SARIMA provided a better prediction score than ES. Yoon et al. developed a spatiotemporal ARIMA model to forecast the number of bikes available for each bike stations in both short and long term.

There has been a lot of research done in the past regarding the bike-sharing systems and the bike usage analysis. Many statistical methods were employed, machine learning methods were also employed to perform the analysis to understand bike usage behaviour. In our project and research analysis we have mostly used deep learning and geo-spatial analysis, but to cover all aspects of our analysis and for the sake of understanding our analysis, we must have a look at other approaches as well. For our analysis, we did not use ARIMA as it might be effective with time-series data, we are using a GRU based network model to evaluate and predict future bike usage because it is a more advanced method and has been giving better results in the past. The analysis being narrowed down to top 5 most used bike stations or basically hotspot bike stations of Dublin it doesn't cover a citywide analysis yet.

3 Research Methodology

The research methodology consists of seven phases, Data collection and pre-processing, feature engineering, High demand station identification, Peak hour and full usage analysis, GRU based time-series forecasting, trying to forecast for the month of June based on the given data, geospatial clustering & expansion suggestions.

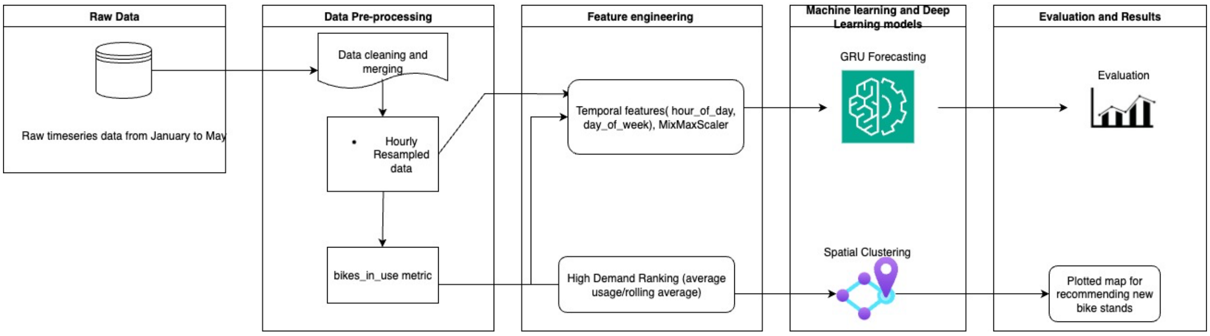

Fig.1. Research Methodology

Phase 1: Data collection and pre-processing. It involves combining 5 months of bike usage data from the months January to May 2025 and making one master dataset. Then the timestamp column is converted into 'datetime' format to be able to fetch the important information from the timestamps. Also, any column or row which was not necessary for analysis was removed from the data. Removed the columns 'system_id', 'is_installed', 'is_renting', 'is_returning', 'short_name', 'region_id' because they had no significance to the bike usage analysis. The data was collected from the 'Smart Dublin' website where the DublinBikes DCC[1] datasets are present from the month of January2025 to May 2025.

Phase 2: Feature engineering involves creating useful features which will be important for our analysis further and removing null values or values not useful. Time features were extracted from the timestamp like hour_of_day which captures the time of the day, day_of_week to distinguish weekdays from weekends, is_weekend which is a Boolean flag to tell if a day is a weekend or not. Then re-sampling of the data was done to hourly intervals for consistency and null values were removed if present. A key feature calculated over time was 'bikes_in_use' which is basically the difference between station capacity and available bikes. These features are widely used in deep temporal modelling (Zhang et al., 2017; Li et al., 2018). Features were also normalized using 'MinMaxScaler' for model convergence for the upcoming phases of the research.

Phase 3: To find the most crowded bike-sharing stations, the dataset was aggregated by 'station_id' and averaged over all timestamps to determine the average hourly usage of each station. The method offers a stable and interpretable measure of total demand since it indicates the frequency of the use of bikes at each location regardless of the choice of time windows. The 10 top stations were then ordered according to their mean values of bikes in use. This is in place of the previous methodology that was based on the peak-hour filtering and rolling averages. This approach provides a more general picture of strain on the infrastructure and long-term demand as it is based on station-level mean usage over the full five-month period. Also, there was a comparison of weekday and weekend use conducted.

[1] https://data.smartdublin.ie/dataset/dublinbikes-api

The average hourly demand curves were plotted separately on weekdays and weekends which enabled the planners to track changes in usage pattern with respect to hour of the day.
Phase 4: The maximum stress periods at a station level were determined by flagging all the events that the number of available bikes at use matches the maximum capacity of the station. These activities were mapped on a timeline in the area of operation visibility and planning. These types of anomalies and peak detection can be used to facilitate insight on latent system bottlenecks. Basically, we tried to find out that out of the top 5 used bike stations we found out in the previous phase, which bike station gets empty more often.
Phase 5: The PyTorch was used to create a Gated Recurrent Unit (GRU) network to capture the temporal relationship. The model was comprised of a single layer of GRU and an output layer that was linear. The training goal was to reduce the Mean Squared Error (MSE), where Adam was optimized down to 400 epochs. The model has been trained using January to April data and tested on May data. The comparison of forecasts and actual May usage was performed with the help of MAE, RMSE, and R 2. Such an evaluation method is aligned with previous deep learning models in demand forecasting (Li et al., 2018; Ke et al., 2017). Using Matplotlib the temporal trends were plotted to show the comparison between the prediction of the model and the actual trend based on the already present data.
Phase 6: To test how the model would perform in a real-life scenario, it was retrained on the entire January to May dataset and was applied to predict the hourly demand in June 1-30, 2025. Since we do not have the data for the month of June this is just a prediction to demonstrate the use of this research. The model used in the last 24 hours (May) as the seed input value and produced 720 hourly forecasts (30 days x 24 hours) using a recursive process. The next cycle of prediction used each predicted value as an input to the next iteration of the prediction-a recursive or rolling forecasting technique.
Phase 7: DBSCAN was used to find nearby high demand clusters with a radius of 600 meters in each top station. New station locations were proposed to be the cluster centres at least 300 meters away of the existing station. The proposed concept of spatial demand clustering is a vital precondition to dynamic infrastructure planning that confirms the ideas of Wu et al. (2019). Using 'Folium' a map was generated to show the suggested new bike stations based on our analysis.

4 Design Specification

4.1 Average hourly bike usage, weekday vs weekend:

The module Weekday vs Weekend Usage and Top Station Ranking is intended to give temporal and spatial information on the Dublin bike-sharing demand. It works with the cleaned aggregated dataset (df_all) that comprises station-level usage, timestamp and an is_weekend indicator. In the temporal analysis, the data is separated into weekday and weekend subsets by this flag and average hourly bike use in each subset is computed using a mean aggregation by hour. These are then plotted on a comparative Matplotlib line graph labelled as: Average Hourly Usage: Weekday vs Weekend to allow visual differentiation between the spikes of the weekday usage, which is commuter based, and the evenly distributed weekend usage. In the spatial ranking, the data is categorized by station_id and name and the average number of bikes utilized per station is calculated and arranged in descending order to determine the busiest places. The top 10 busiest stations are printed in ranked list on console which is a crucial input in subsequent modules of forecasting and spatial analysis. Incorporated in Python with Pandas to process the data and Matplotlib to visualize it, the component makes sure that the operational planning is enriched with the pattern identification of the demand and prioritization at the station level.

4.2 Full Usage and Peak hour analysis:

Full Usage Events and Peak Hour Analysis module was designed to confirm the top five busiest bike stations based on the evaluation of the level of their strain and the periods when they are most used. In this analysis, the most important stations in terms of average demand are considered which are IDs 89, 111, 47, 103, and 104, as determined in the list of top stations. The data is filtered in such a way that only the relevant records are taken per station and the average bike usage is obtained by the hour of the day. The peak hour is the one with the highest average usage which is recorded, and the corresponding mean demand value is recorded as the peak hour of the station. Bar charts are used to visualize these results in terms of hourly distribution of demand. In an attempt to further validate the intensity of demand in the stations, the module records full usage incidents which are defined as those instances when the number of bikes in circulation at a station is equal to the maximum capacity recorded by the station. These are depicted as time stamp markers giving a time perspective of capacity limits. The results of this module, peak hour detection, full usage frequency and visual confirmation, are crucial to the verification of the results of the demand analysis and the making of operational decisions related to the redistribution of bikes and the expansion of capacity.

4.3 GRU Forecasting Model:

The GRU forecasting model was applied to forecast the hourly bike usage on high-demand stations based on the historical data. The steps involved are first, data preparation where important features are extracted and normalized, which includes the number of bikes in use (y), the hour of the day and the day of the week. The time series is then divided into overlapping sequences with a 24-hour sliding window that predicts the bike demand of the following hour. These sequences are separated into training and testing sets, and the model GRU was trained on data of January-April and validated in May. The GRU model has one hidden layer consisting of 64 units and the Adam optimizer with learning rate of 0.005 for 400 epochs. Model accuracy is evaluated using evaluation metrics like MAE, RMSE, and R 2 score. The effectiveness of the model was illustrated using visual representation of plots between the real and predicted values and it was also extrapolated to provide the recursive forecast on the month of June using the previous output as the input of future.

4.4 Spatial Demand Clustering using DBSCAN:

To determine the areas of infrastructure deficiencies in the Dublin Bikes system, a spatial demand analysis module was applied. The design starts with a filter of the dataset, so that it contains only the commuter peak hours (7-9 AM and 4-6 PM) and an optional filter of weekday or weekend pattern, depending on a user-defined flag. Hourly bike usage at each station is then smoothed with a 12-point rolling average and the five most popular stations are chosen based on maximum mean rolling demand. On each of these high-demand stations, the demand points in a 600-meter radius are extracted and clustered with the DBSCAN algorithm with haversine distance, so it is possible to identify the natural demand hotspots. To achieve meaningful expansion, cluster centers that are at least 300 meters away with the original station are taken as valid recommendations of new station locations. The map of all existing stations, crowded areas, and suggested places of expansion are presented as an interactive Folium map, with clustering on markers to form groups of nearby suggestions. The outcome is stored as an HTML file which is easy to explore. The design facilitates data-informed infrastructure planning through a synthesis of time-dependent modeling of demands, spatial clustering, and geodesic distance filtering, which guarantees that proposed sites will be high-demand and non-redundant.

5 Implementation

The last stage of this project implementation has led to the full-fledged data-driven system of predicting and analysing Dublin bike-sharing demand. It resulted in the production of structured outputs such as predictive models, interactive maps, transformed datasets, and evaluation visualizations all of which aided the implementation and decision-making application of the system.

The important artifact of this step was a trained Gated Recurrent Unit (GRU) neural network model of time-series prediction that could predict the hourly use of bikes at each station. The model has been implemented on the five busiest stations which have been determined during demand analysis and calibrated with the historic data of May 2025. Along with the training of the model, recursive forecasting was also applied to estimate the bike demand in June 2025, purely on the basis of the data that was observed in the past. The outputs of the forecast were represented graphically in the line plots of predicted versus actual demand, which allowed quantitative analysis with the help of MAE, RMSE, and R 2 indicators.

On the geospatial front, an interactive station recommendation map was created, which was used to show areas of high demand and possible new station locations. The map involved cluster methods to examine localized demand around congested stations and geodesic distance constraints were implemented to make sure that suggested locations were not repetitive. The result was a fully interactive map (in HTML format) based on Folium, which indicated all existing stations, most demanded hubs, and suggested new station points, which were color-coded according to their importance.

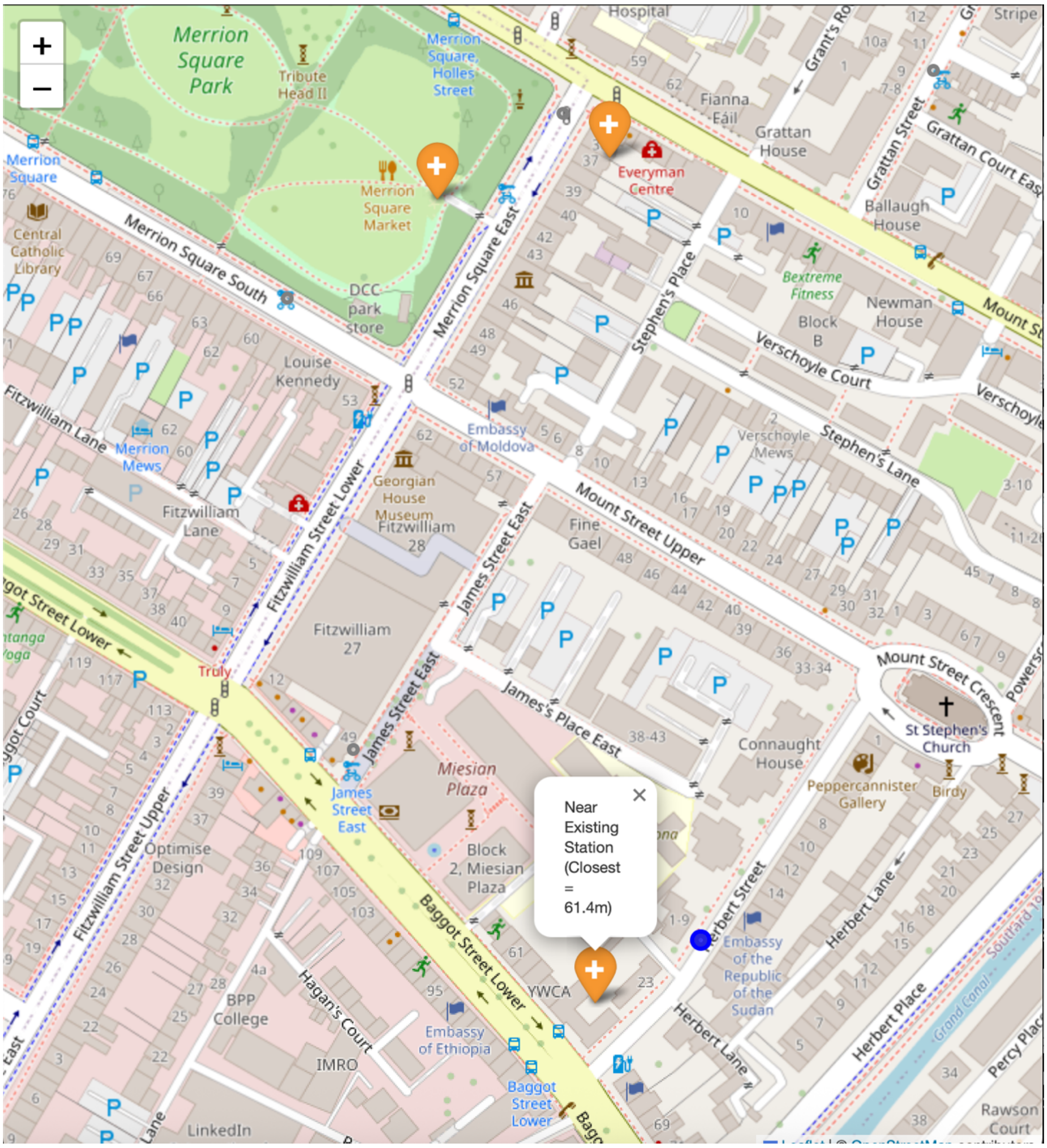

Fig. 2. Station recommendation map

The processing steps that were performed on all datasets prior to model training included time-based feature engineering, scaling, rolling average computations, and hourly resampling. Additional insights provided by exploratory analysis were on weekday vs weekend usage and station-level full-capacity events that were represented in the form of customized plots and charts.

5.1 System Architecture

The architecture of the system of the Dublin Bike-Sharing Demand Forecasting and Spatial Analysis project is implemented in the form of a multi-stage pipeline that reflects the flow of the activities outlined in the implementation. It starts with the Data Acquisition Layer, which gathers and consolidates historical bike-sharing data of the period January-May 2025 into a single structure. This makes sure that all attributes of interest- station IDs, timestamps, counts of available bikes, and geospatial coordinates are formatted and consistent to be used in downstream processing. This phase is the direct input to the transformation work as shown in the implementation whereby these raw data sets were cleaned up ready to be used in temporal forecasting and spatial analysis.

The Data Processing and Analysis Layer converts these raw datasets to the forms ready to be used as a model, which is in accordance with feature engineering, demand smoothing, and filtering procedures applied in the project. Features that are based on time like hour of the day and day of the week are extracted and rolling averages are calculated to capture smoothed demand patterns. In this layer, the GRU-based time-series forecaster module is used to forecast the hourly bicycle use of the most popular stations, which were determined in the previous analysis steps. In tandem with this, the spatial demand analysis module uses DBSCAN clustering and geodesic distance filtering to determine new potential station locations without duplicate infrastructure. The implementation of this stage also includes the logic of identifying full-capacity usage events and comparing the weekday and weekend demand, which are also characterized as intermediate outputs.

Visualization and Output layer is the equivalent of the final outputs outlined in the implementation whereby results are converted to more understandable visual formats. The outputs can be forecasted and presented as line plots of predicted versus actual demand, and evaluation measures, like MAE, RMSE, and R 2. The spatial findings are presented in the form of interactive Folium maps, with existing stations and busiest hubs clearly distinguished, as well as suggested points of expansion. There are other charts indicating variations in demand during weekdays and weekends, the ranking of the top stations, and the cases of maximum occupancy of stations. Implementation of the system in Python, with the help of libraries, such as Pandas, NumPy, Scikit-learn, PyTorch, Matplotlib, Folium, and Geopy, makes sure that the architecture is reproducible, scalable, and can be expanded in the future (e.g., long-term demand forecasting or integration with live data feeds).

6 Evaluation

The purpose of this section is to provide a comprehensive analysis of the results and main the analysis of this project includes the quality of the GRU model predictions, as well as the efficiency of spatial analytics in defining the areas of high demand and proposing the locations of new stations. This section presents some main findings, performance measures, validation methods, and practical implications.

6.1 Average hourly bike usage results

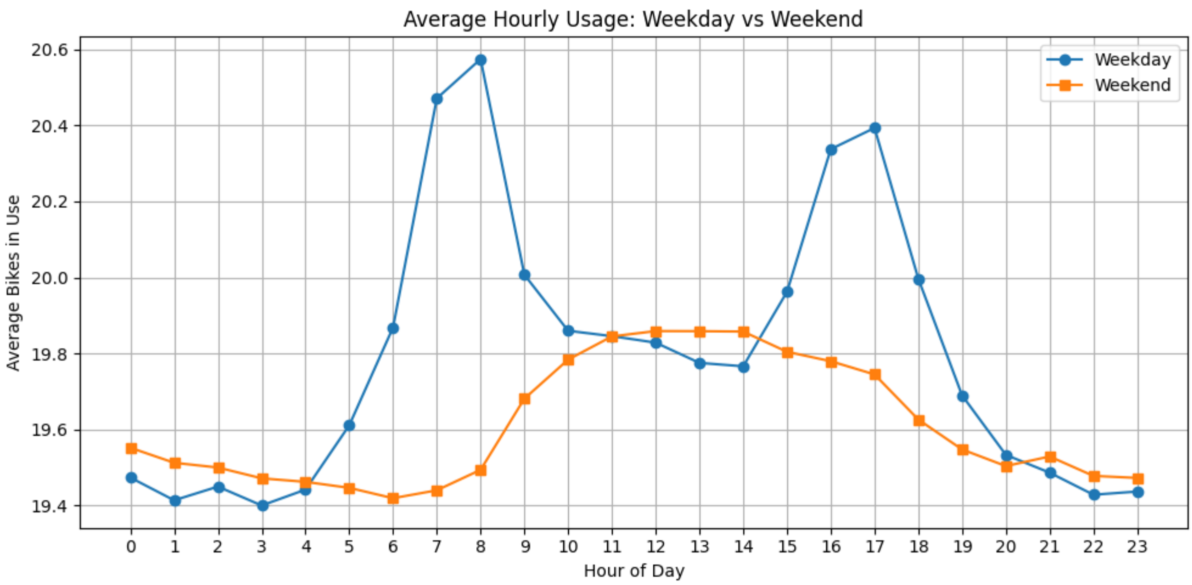

Fig. 11. Average hourly usage: weekday vs weekend

Assessments were performed to understand bike usage behaviour by analysing weekday versus weekend usage. It was found based on the analysis that the highest level of bike utilization was much more evident during weekday travelling hours especially at 8:00 AM and 5:00 PM which is in line with the common working hours travel. Conversely, weekend demand was better spread out during the day which fits the leisure-focused travel behaviour and the lack of time pressure in comparison to weekdays.

```
Top 10 Busiest Stations (Avg Bikes in Use):
 station_id                         name  bikes_in_use
         89       FITZWILLIAM SQUARE EAST     36.783915
        111          MOUNTJOY SQUARE EAST     36.136629
         47                HERBERT STREET     35.157621
        103    GRANGEGORMAN LOWER (SOUTH)     34.293994
        104  GRANGEGORMAN LOWER (CENTRAL)     33.036122
         78                MATER HOSPITAL     32.883808
         53                  NEWMAN HOUSE     32.534337
        113          MERRION SQUARE SOUTH     31.649051
        117             HANOVER QUAY EAST     30.613361
        102                   WESTERN WAY     30.608792
```

Fig. 12. Top 10 busiest stations

We also found the top 10 most used bike stations in Dublin on the basis of the average number of bikes in use.

6.2 Full usage and peak hour analysis of top 5 used bike stations

For demonstration purposes, showing the full usage and peak hour analysis of station 111 (one of the top 5 most used bike stations in Dublin)

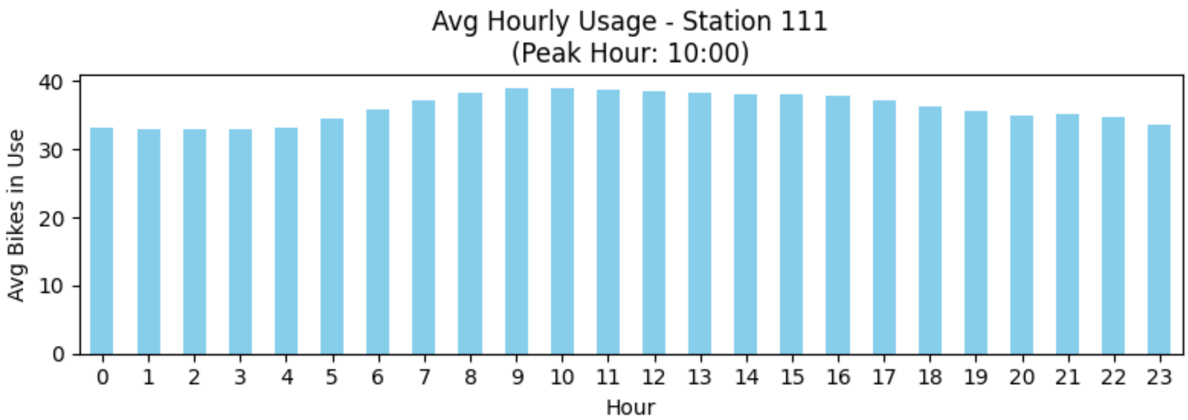

Fig. 13. Average hourly usage of station 111

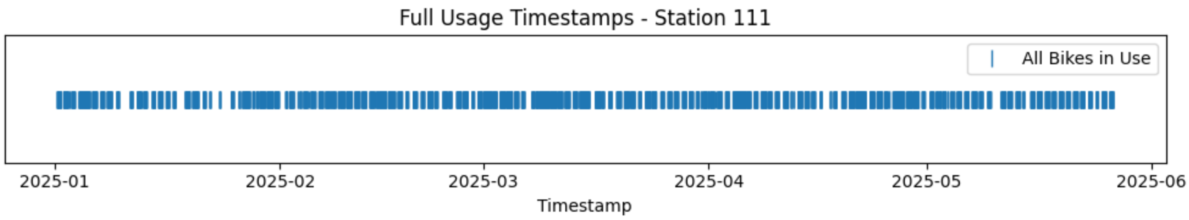

Fig. 14. Full usage timestamps of station 111

Figure 13 shows us the average hourly bike usage for station 111 and figure 14 shows us the exact timestamps when the bike station is completely empty.
The top 5 most used stations have many instances where the bike stations are empty. Such saturation of capacity events indicates possible operational difficulties and points to the possibilities of introducing specific redistribution solutions or expanding capacity in priority areas. This behavioral observation does not only confirm the correctness of the station ranking based on the demand analysis but also offers important non-model-based findings that can support the practical decision-making process aimed at increasing the efficiency of the system and the satisfaction of its users.
To determine which station has a full usage event, we counted the number of such instances and found out the following:

```
Times When All Bikes Were in Use:

Station 89 - 6840 occurrences:

Station 111 - 5330 occurrences:

Station 47 - 5923 occurrences:

Station 103 - 5417 occurrences:

Station 104 - 6052 occurrences:
```

Fig. 15. Full usage event count output

Figure 14 shows that station 89 has the most occurrences when the bike station is empty. For reference all the findings and graphs shown above are for all the 5 months of data i.e. from the month of January to May.

6.3 GRU Forecasting accuracy

The forecasting performance was measured by training the GRU model on the data of the period between January and April 2025 and testing the model on the data of May 2025. The model was implemented on the five busiest stations (IDs: 111, 104, 89, 47, 103) which were determined by the past average usage.
Each model's output was compared to the actual hourly demand for the month of May using standard regression metrics:

- Mean Absolute Error (MAE): Gave a common average size of errors in predictions.
- Root Mean Squared Error (RMSE): Weighted greater errors and assisted in determining reliability.
- R 2 score: Indicated the degree of the model fit to explain the variability in the real data.

6.3.1 Station 111

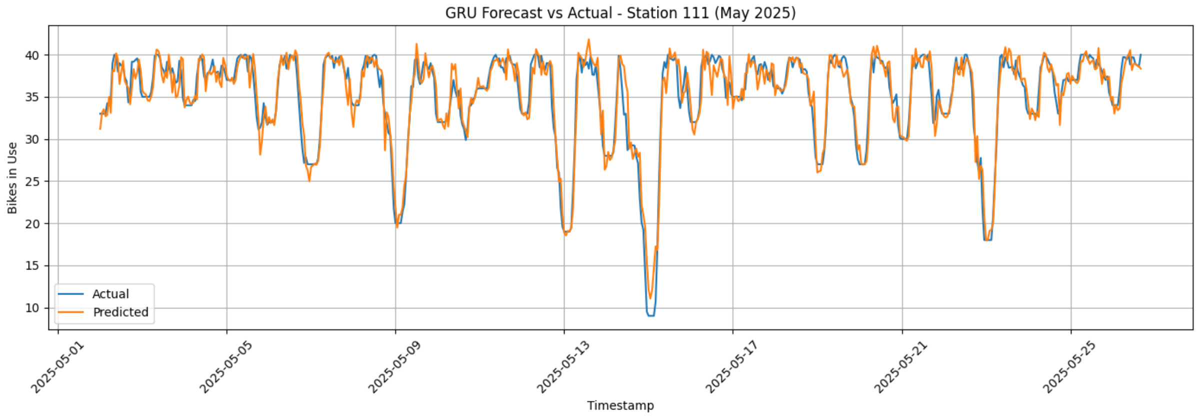

Fig. 3. Forecast vs actual comparison for station 111

MAE: 1.17, RMSE: 1.62, R2: 0.9170

6.3.2 Station 104

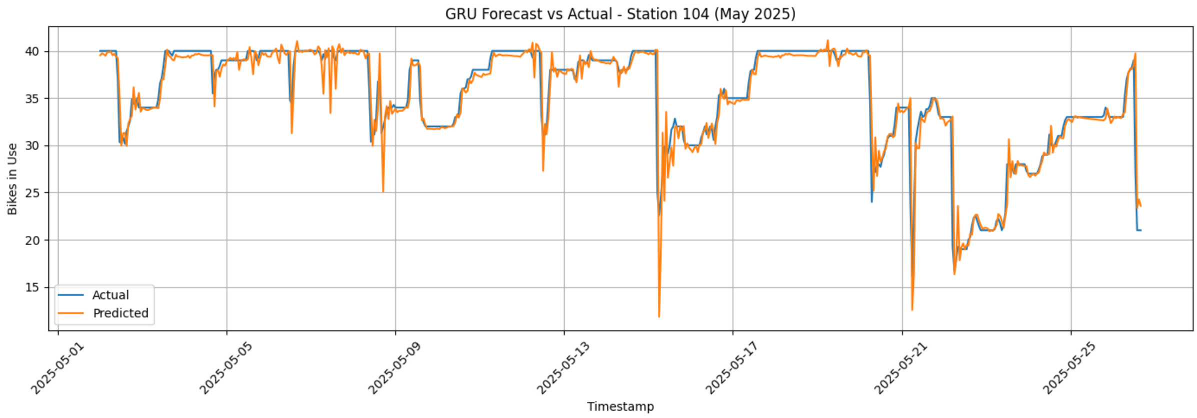

Fig. 4. Forecast vs actual comparison for station 104

MAE: 0.81, RMSE: 1.72, R2: 0.9003

6.3.3 Station 89

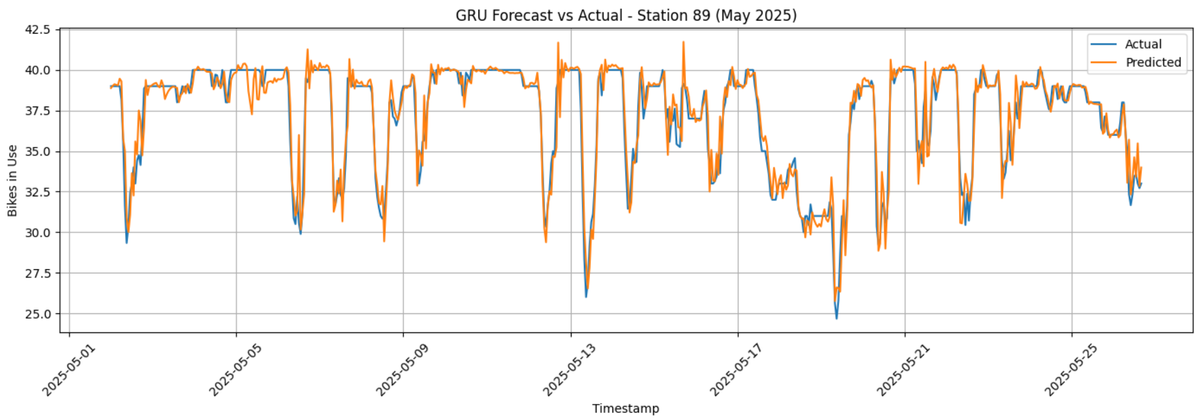

Fig. 5. Forecast vs actual comparison for station 89

MAE: 0.65, RMSE: 1.00, R2: 0.9055

6.3.4 Station 47

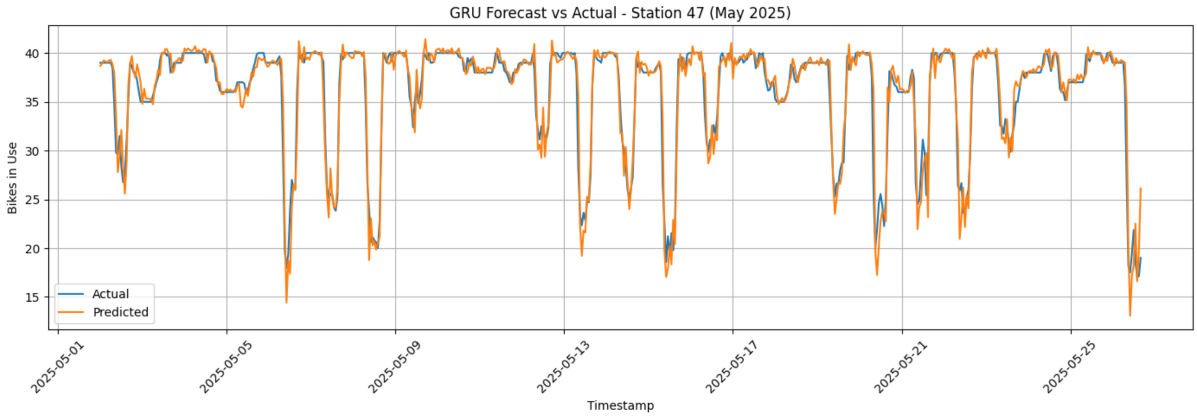

Fig. 6. Forecast vs actual comparison for station 47

MAE: 0.94, RMSE: 1.55, R2: 0.9205

6.3.5 Station 103

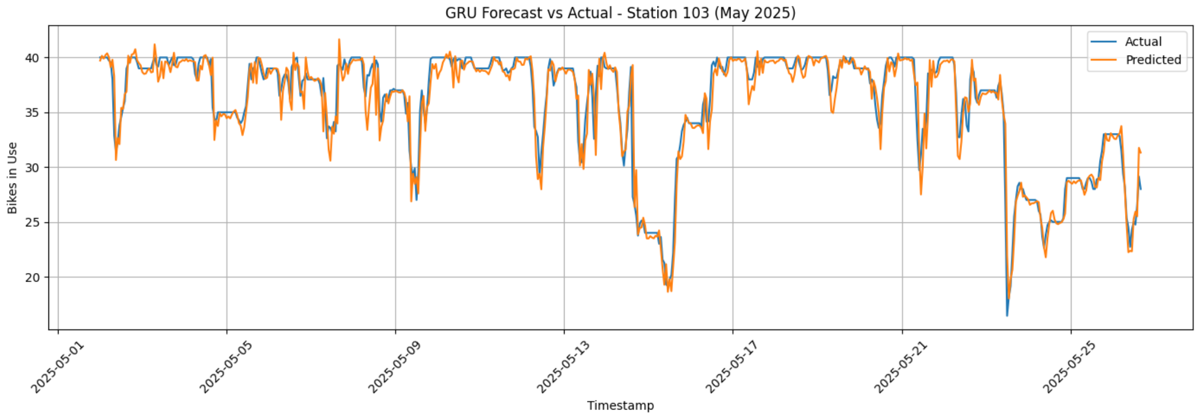

Fig. 7. Forecast vs actual comparison for station 103

MAE: 0.89, RMSE: 1.43, R2: 0.9214

The GRU model returned high R 2 scores (greater than 0.9) and low RMSE that validates that the model was able to capture seasonality and station-specific usage behavior. The plotted visuals of the predicted vs. actual usage confirmed that the model was tracking the demand curve, especially in the peak hours.

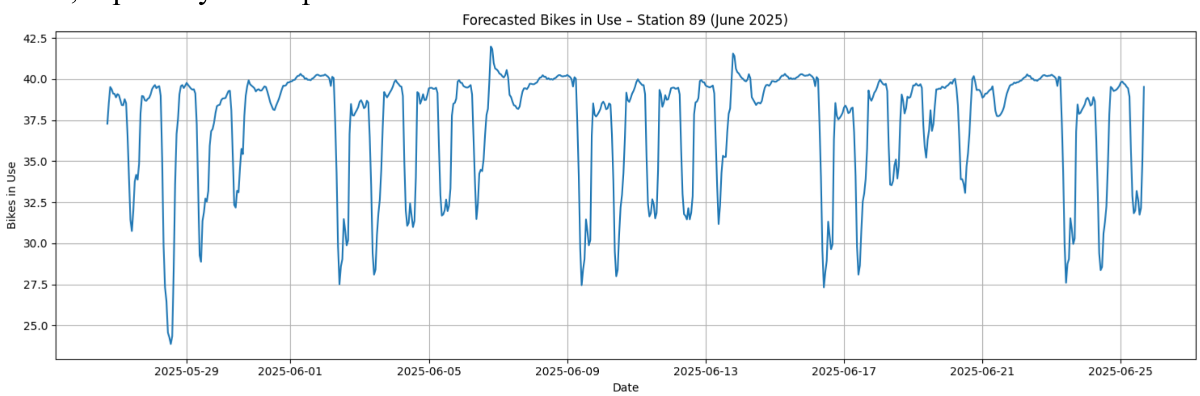

Fig. 8. June forecast for station 89

Moreover, the model was extrapolated to predict June 2025 demand, with a recursive input strategy. Although these forecasts lack ground truth in June, they can serve as a useful estimation in forward planning and the fact that it matches the May data closely implies good generalization capability.

6.4 Spatial Analysis & Station Suggestion

The spatial analysis component was able to integrate the peak hour demand filtering, rolling average and geospatial clustering via DBSCAN to propose new station locations. High-demand stations were used as centre points and the surrounding demand clusters within the radius of 600 meters were identified as top 5 high-demand stations.

Clusters that were at least 300 meters distant to existing stations were tagged as an expansion area potential. These were presented on an interactive Folium map where users can easily differentiate between:

- Stations already in place (grey)
- Bustling centers (blue)
- Recommended expansion locations (green or orange according to distance)

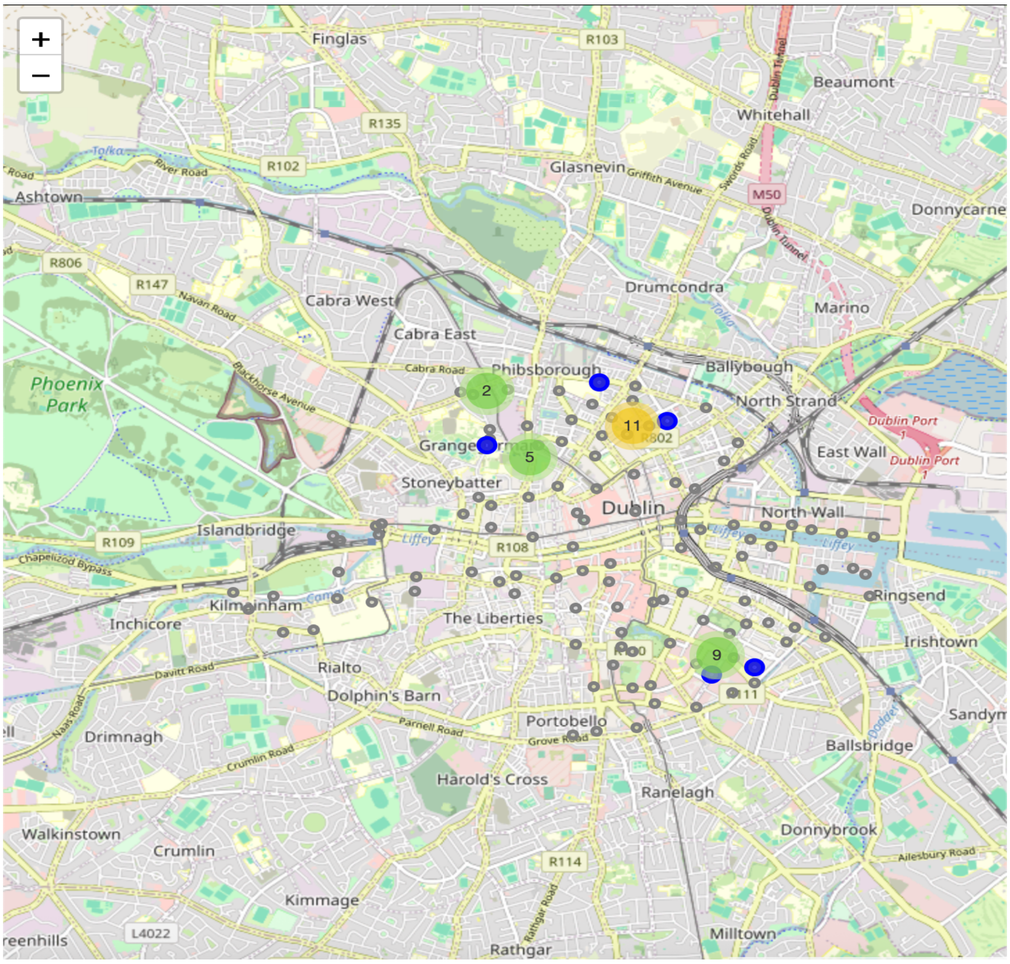

Fig. 9. Overview of bike stations in Dublin

The main findings of spatial analysis were that at least around the top 5 busiest stations there are enough bike stations present since there were no green markers found on the resultant map. The analysis either recommends already present bike stations or approximately close 100 meters to the busy or already present bike stations. This suggests that infrastructure-wise Dublin Bikes is well prepared.

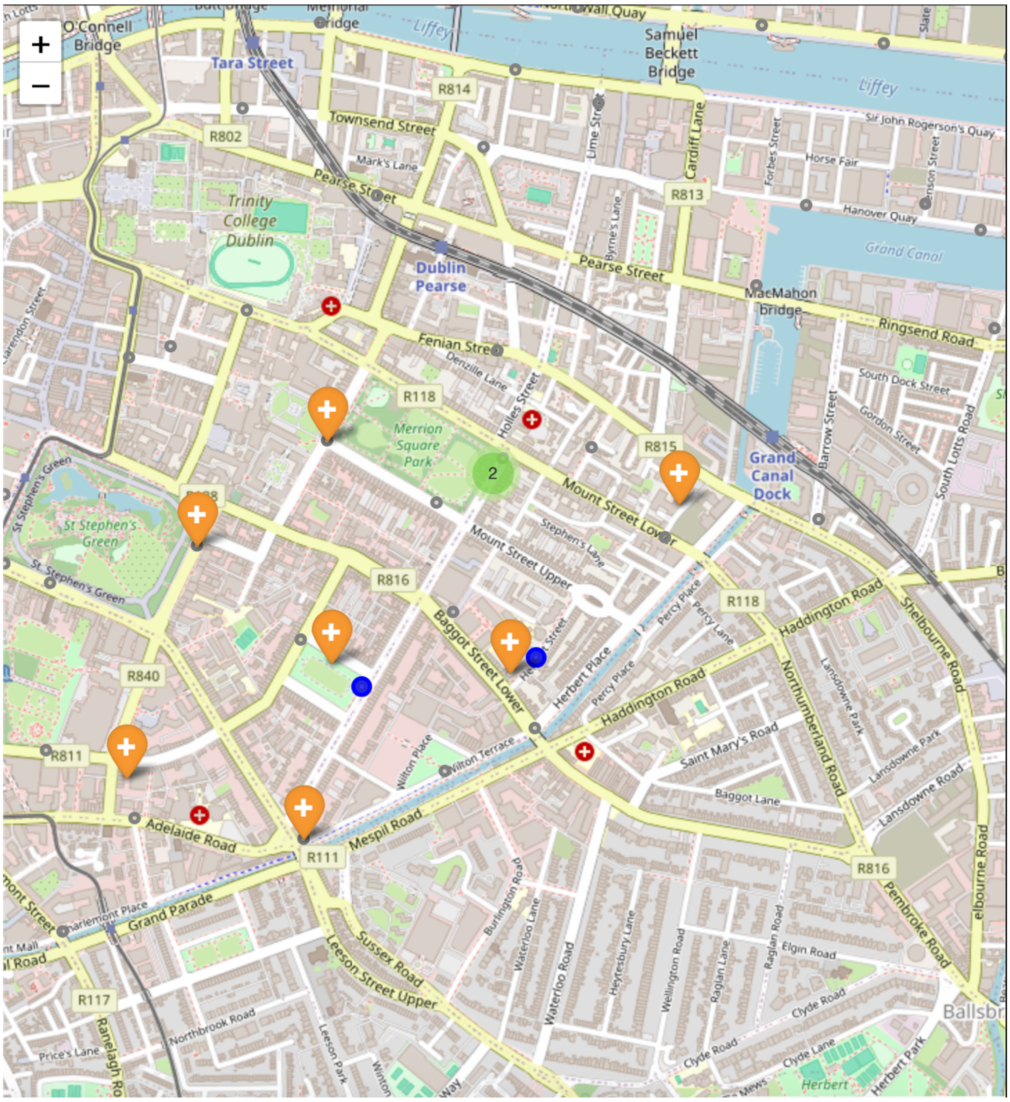

Fig. 10. A closer look at new bike station recommendation around busy stations

If we have a look closely at Fig. 10, we can see some of the orange suggestion pointers are directly on the already present bike station i.e. the grey markers.

6.5 Discussion

To evaluate the project, several dimensions were discussed in order to determine the efficiency of the predictive models and the soundness of the exploratory analyses. GRU based forecasting was subjected to the five busiest stations, and performance assessment was made in terms of MAE, RMSE and R 2 score. The graphs and the scores show that if the model is trained on historical data, in our case on the bike usage data of four months from January to April, it can closely predict the bike usage for the next month, and we come to that conclusion with a comparison of the prediction with the actual bike usage data i.e. for the month of May. We did station specific analysis on the top 5 most used bike stations and the performance metrics calculated show us that the model is a good fit for predicting the future bike usage for the specific station and if we perform such analysis on all the bike stations in Dublin, even with a few margins for error we can be prepared to better equip the stations with either extra bikes or better serviced bikes and thus provide a better service and more importantly be aware and prepared of the possible shortcomings.

The complementary analysis included weekday vs. weekend demand patterns and showed a significant peak of weekday usage during commuting hours and a more even spread of weekend usage, as expected of leisure travel. Full-capacity event detection at station-level further confirmed the high-demand station ranking showing repeated saturation points, indicating bottlenecks in the operation and areas of capacity improvement. And with the full capacity usage counts we can confirm that the average bike usage calculated hourly can tell us the overall usage patterns easily and it can be used as a dashboard for an overview of how the bike service is working.

The mapping of spatial demand revealed areas of high activity and proposed new possible places of the station, so the geography coverage is taken into account, as well as the time forecast. With this part of our research and analysis, it was concluded that infrastructure wise Dublin Bikes is well equipped, there are good number of stations spread evenly across the city so if a particular busy bike stand is empty the user doesn't have to walk for long to another station. This analysis though can be extrapolated, and we can do a spatial cluster analysis on each station to find out how busy or not the stations near the top busy stations are and how well are they maintained.

7 Conclusion and Future Work

The research questions were answered effectively in this project by combining the deep learning forecasting and spatial demand analysis. In the case of Research Question 1, the GRU-based time-series forecasting model was able to show a great predictive ability in the most popular bike stations in Dublin with an R 2 of 0.9170 in Station 111 and other leading stations also showing an impressive R 2. These findings support the conclusion that past usage can be used to predict demand in the future when supplemented by temporal features, like hour of day and day of week. The closeness of the model to represent the actual usage patterns during May implies that the model is a sound tool in operational planning and allocation of resources in subsequent months.

In the case of Research Question 2, spatial analysis was conducted through the rolling average demand, geodesic distance filtering, and DBSCAN clustering in order to determine whether more stations are needed close to high-demand locations. It was found in the analysis that the current stations are already distributed and located in such a way that they are close to satisfy the current patterns of demand. This means that no new station locations were advised to be established at this stage implying that the current station network of Dublin is already efficient in terms of space coverage of the busiest places.

Comprehensively, the project demonstrates that although forecasting can play the role of enhancing the operational readiness significantly, the spatial expansion decisions have to be made on the basis of data-driven assessment in order to avoid the spending of money on the unnecessary infrastructure. The general strategy is repetitive in nature, which is to monitor demand and then evaluate network coverage.

Future work may further lengthen the forecasting horizon to forecast demand further into the future (e.g., several months ahead) and test model performance during seasonal and non-typical demand situations (e.g., during a special event or weather anomaly). The addition of other contextual information such as real time weather, schedules of public transport, and city event schedules could also enhance the accuracy of prediction. In the spatial aspect, it was found that there may be no urgency in having new stations as of now, but continuous monitoring can establish to identify new hotspots that may arise due to urban growth or population change. Other spatial optimisation methods, e.g. graph neural networks or multi-objective location-allocation models, might be considered as well in order to facilitate strategic long-term network planning. Lastly, the forecasting model may be incorporated into a real-time decision support system and make it possible to dynamically redistribute bikes to maintain optimal levels of service during high demand periods. In the future this analysis can be expanded to do a citywide analysis, the geospatial analysis conducted can be expanded in that sense and the research that has already been done by others it can be expanded in the same way for Dublin.

References

A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, “Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system,” *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, 2010.

A. Jaber, B. Csonka, and J. Juhász, “Long Term Time Series Prediction of Bike Sharing Trips: A Cast Study of Budapest City,” in *2022 Smart City Symposium Prague (SCSP)*, Prague, Czech Republic, 2022.

J. W. Yoon, F. Pinelli, and F. Calabrese, “Cityride: A Predictive Bike Sharing Journey Advisor,” in *2012 IEEE 13th International Conference on Mobile Data Management*, Bengaluru, India, 2012.

H. I. Ashqar, M. Elhenawy, M. H. Almannaa, A. Ghanem, and H. A. Rakha, "Modeling bike availability in a bike-sharing system using machine learning," in 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 2017.

T. Xu, G. Han, X. Qi, J. Du, C. Lin, and L. Shu, "A Hybrid Machine Learning Model for Demand Prediction of Edge-Computing-Based Bike-Sharing System Using Internet of Things," IEEE Internet of Things Journal , vol. 7, no. 8, pp. 7345 - 7356, 2020.

E. Collini, P. Nesi, and G. Pantaleo, "Deep Learning for Short-Term Prediction of Available Bikes on Bike-Sharing Stations," IEEE Access , vol. 9, pp. 124337 - 124347, 2021.

Ke, J., Zheng, H., Yang, H. and Chen, X. (2017) ‘Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach’, *Transportation Research Part C: Emerging Technologies*, 85, pp.591–608.

Li, Y., Yu, R., Shahabi, C. and Liu, Y. (2018) ‘Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting’, *International Conference on Learning Representations (ICLR)*.

Zhang, J., Zheng, Y. and Qi, D. (2017) ‘Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

Yao, H., Tang, X., Wei, H., Zheng, G. and Li, Z. (2018) ‘Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Forecasting’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Wu, Z., Pan, S., Long, G., Jiang, J. and Zhang, C. (2019) ‘Graph WaveNet for Deep Spatial-Temporal Graph Modeling’, *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*.