Enhancing the Dublin Bikes System: An Impact Analysis to Identify Pre, During, and Post-Pandemic Trend

MSc Research Project
Data Analytics

Pragya Jha
Student ID: x22211047

School of Computing
National College of Ireland

Supervisor: Prof Furqan Rustam

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Pragya Jha
Student ID:	x22211047
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Prof Furqan Rustam
Submission Due Date:	12/08/2024
Project Title:	Enhancing the Dublin Bikes System: An Impact Analysis to Identify Pre, During, and Post-Pandemic Trend
Word Count:	3465
Page Count:	27

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:	
Date:	16th September 2024

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):	

Enhancing the Dublin Bikes System: An Impact Analysis to Identify Pre, During, and Post-Pandemic Trend

Pragya Jha
x22211047

Abstract

The sudden attack of COVID-19 outlined the need for a resilient and adaptable public transportation system. Since the pandemic impeded the dynamics of urban mobility, causing overcrowding and thus lockdowns the systems traditionally operating were not able to put up with the challenge showing that more flexible transport alternatives are called for. There is extensive research on resilience within public transportation systems; however, the contribution of bike-sharing systems in maintaining urban mobility during such crises remains underexplored. This paper is a detailed analysis of the Dublin bike-sharing system given its potential to be an effective mode of public transport when there might arise a repeat pandemic or events of that nature. The critical analysis, extending up to data received in 2018 to 2024, proposed the use of variants of predictive models, namely Random Forest Regressor, Linear Regression, ARIMA, LSTM, SARIMA, and Prophet for bike availability and usage pattern prediction. To arrive at this methodology, exhaustive data cleaning and processing had been done to offer the expected and accurate level. After such data cleansing, each model was trained and tested to estimate their predictive performance. Of these, the Random Forest model proved the most accurate, with an R-squared value of 0.93. The findings are indicative of the fact that utilization patterns of the bikes changed in crucial ways during the pandemic, in the form of increased demand and peak shifting. The results indicated that the bike-sharing system in Dublin played a fundamental role in replacing demand during a time when the traditional transport systems were choked. The author has suggested several recommendations based on findings that would make the system resilient: dynamical reassignment of bikes by evolving demand, developing infrastructure that would flex or adapt to new circumstances quickly, ramping up sanitization protocols to keep the public safe and clearer messaging with customers about bike supply and ways to maintain health. The paper also identifies some possible future directions of research by which, through the integration of health data and collaboration with public health authorities, further improvement in system adaptability and resilience may be achieved. Such measures would make the bike-sharing system of Dublin more efficient but also substantially better prepared to serve as a reliable means of public transport in any future public health crisis.

1 Introduction

In 2024 ”Smart City” is a topic of great interest. Designing mechanisms for a more sustainable style of life includes three major parameters for consideration pollution, public health, and transportation Juan (2021) The Department of Transport, Ireland recorded a 23% decrease in the total vehicles driven, and particularly the rate of public transport passengers fell by 54% in 2020. This is a testament that the COVID-19 pandemic caused extraordinary disruptions in urban mobility, drastically affecting transport systems around the world. Lockdowns, social isolation, and public health concerns had a significant impact on commuting behaviors, forcing a rethinking of public and shared transportation options

1.1 Background

The drastic changes in urban mobility patterns during the COVID-19 pandemic showed a significant decrease in the usage of traditional transport usage. A staggering 54% drop shows that during a public health crisis, conventional commuting methods are not reliable and are negatively impacted. This says that there is a need for more adaptable and sustainable transportation alternatives. Amidst these disruptions, studies revealed that the most suited mode of transportation to fight against future pandemics (similar to COVID19), and pollution for Ireland can be DublinBike. But despite its potential, the system took a hard hit during COVID-19. Figures from NTA revealed a 46% drop in usage of the bike during the pandemic and also that 10,000 people canceled their subscriptions mid-pandemic National Transport Authority (2023) Wilbur et al. (2023) . This indicates that from the onset of the pandemic (early 2020), there was a sharp decline in actual bike usage, which is likely to be Answer To Examiner's Questions a result of initial lockdown measures and public apprehension. Post this, there is a noticeable rise which could be attributed to individuals seeking alternatives to public transport due to social distancing concerns, and also a little of taking advantage of periods of eased restriction. The data revealed by the Dublin City Council showed that there were around 2,001,810 trips by passengers in 2022. If we compare this to the COVID timeframe, which can be considered as 2019 (3,816,652) the trips are down by around 1.8 million. Cycling has a critical impact on the community as it presents economic and ecological benefits. For instance, according to the European Cyclists' Federation, saving emissions by cycling is greater than 16 million tons of CO2 per year in the EU (which represents 600 to 5.630 million € savings). According to WHO, Cycling is among the most efficient and sustainable means of transportation World Health Organization (2022).

1.2 Problem Definition

The author conducted this study with the aim of addressing several key questions that are critical for the future of bike-sharing systems in Dublin.

The Dublin bike-sharing system, like many others globally, experienced dramatic changes in usage patterns during the pandemic. The study revolved around questions like,

- How did the usage patterns of the Dublin bike-sharing system change during the COVID-19 pandemic?

- What are the key factors influencing bike availability and user demand during different phases of the pandemic?
- How can predictive modeling improve the management and optimization of the bike-sharing system?

The author in order to find answers to the above-listed question considered the primary objectives of this research to analyze historical and usage data to identify trends and patterns in bike-sharing behavior and then evaluate the effectiveness of various predictive models in forecasting bike availability and usage. This would help create a layout of actionable recommendations to optimize Dublin's bike-sharing system and ensure its reliability during future pandemics.

This author also aims to ensure that this work contributes to the growing body of literature on urban mobility and public transportation resilience. With the help of predictive modeling techniques, this study offers a detailed analysis of the Dublin bike-sharing system's response to the COVID-19 pandemic. The findings of this analysis provide valuable insights for urban planners, policymakers, and transportation authorities, highlighting the importance of flexibility, data-driven decision-making, and proactive planning in managing bike-sharing systems.

1.3 Motivation

This research is motivated by the critical need to evaluate the strength and adaptability of the DublinBike system in the context of the COVID-19 pandemic and beyond. The previous studies on cycling have highlighted the ecological and economic benefits. With this study, the author aims to assess how bike-sharing systems can sustain urban mobility during public health crises and contribute to post-pandemic recovery. By analyzing usage patterns, user demographics, and operational challenges before, during, and after the pandemic, this research seeks to provide actionable insights for enhancing the DublinBike system. The findings of this analysis are intended to guide future urban planning efforts, ensuring that bike-sharing systems like DublinBike can become integral, resilient components of urban transportation networks, capable of withstanding future disruptions. This study also aims to promote bike-sharing as a viable, socially distanced mode of transport, contributing to sustainable urban mobility in both normal and crisis conditions.

Research Question: How have usage patterns, user demographics, and the role of Dublin Bikes in urban transportation evolved before, during, and after the COVID-19 pandemic, and identifying ways for future enhancements to the bike-sharing system?.

1.4 Report Structure

This research paper is structured such that section 2 helps in understanding the Dublin bike system and why it is a good choice of public transport during public health crises and for smart city enhancement and also presents a study of similar analysis in different geographic region. Section 3 talks about the methodology of the study. Section 4 talks about how the study is implemented and the final section discusses the results in length, provides reccommendation, and proposes future work.

This study underscores the critical role of adaptive and data-driven approaches in managing urban bike-sharing systems, ensuring their effectiveness and reliability during unprecedented challenges like pandemics.

2 Related Work

This section talks about all the existing literature work that has helped the author cause a base for this analysis and also learn from existing shortcomings.

2.1 Understanding the Dublin Bike System

The DublinBike is a self-service bike-sharing system that provides rental bikes to everyone above the age of 13. The bike stations are distributed throughout the city center and every station has a minimum of fifteen stands. DublinBike runs in partnership with JCDecaux. The system currently has around 1,600 bikes made available at 115 stations around the city. According to major reviews from the users the bikes are quite nice, gear range isn't the best- but for the flat Dublin City, it is perfect. What makes it an optimal choice is the affordability that it offers. The bikes are free for the first 30 minutes and also have subscription plans for each individual with the highest being annually for 35 euros. It is also noticed that the service is easy to set up. The major drawback revealed in the initial analysis is that the number of stations is very limited. There are several studies conducted on the bike-sharing system in different places. These studies help the author form an outline structure of work to be put into the research.

2.2 Bike sharing systems Exploratory Analysis

Ngo and Martin (2023) examined the effects of the COVID-19 pandemic on public transportation systems across various U.S. cities. The study combines quantitative data analysis of transit ridership records with qualitative survey data from transit users to assess changes in accessibility and user behavior. The quantitative analysis employs regression models, and the qualitative component includes a thematic analysis. The results of the study show a significant decrease in ridership, attributed mainly to concerns about safety and changes in commuting patterns. The study provides insights into the immediate impacts 2 of health crises on public transport. The study helps the author guide the Dublin Bike system in adapting to changes in user behavior and commuting patterns. Bergantino et al. (2021) paper provides an empirical analysis of various influencing factors for potential Bike-Sharing Users. The region of analysis is Italy and the study employed an online survey to understand the factors affecting people's decision to use bike-sharing. The examination conducts factor analysis, ordered logit, and probit regressions to analyze the survey data collected nationally in 2020. This study identified several key determinants such as environmental concerns, health benefits, and changes in consumer habits due to COVID-19 restrictions. The results of the study showed an increased awareness of health benefits and reduced pollution significantly motivates the adoption of bike-sharing among potential users. The study also provides insights that could assist local regulators and bike-sharing operators in promoting this sustainable mode of transportation, emphasizing the changed consumer habits in response to the pandemic. This study acted as a fundamental outline to support the motivation of the author's choice to conduct the proposed analysis. The paper Chibwe et al. (2021) provides a framework for the analysis of what factors influence the usage of bike-sharing systems in London, focusing on various elements such as the number of docking stations, the availability of bikes, user preferences, and broader patterns of use influenced by the city's infrastructure and seasonal changes. The paper specifically explores how events like "COVID-19" pandemic have impacted these patterns, offering insights into the strength of bike-sharing systems during external shocks and public health crises. The paper uses a binomial model, which helps in understanding how different variables affect bike-sharing usage. From temporal patterns i.e. how the usage of bikes changes over time, demographic influence i.e. who uses the service and how their behavior varies, and the system's integration within the broader urban transport network. The method used also forecasts future trends and helps make better-informed decisions for system management and expansions. The author aims to use a similar analysis to provide a blueprint for enhanced Dublin bike-sharing services enabled through specific findings.

2.3 Comparative Analysis of Bike-Sharing Usage Patterns and Demand Prediction Models

Across Different Cities Hu et al. (2021) paper explores how the usage patterns of bike-sharing users in Chicago evolve spatially and temporally in response to the pandemic. The study analyses two years of bike-sharing trip data and the model employs generalized additive models. This model examines the relationships and nonlinear temporal interactions among the different influencing factors such as socio-demographics, land use, transportation features, and COVID-19 infection rates. The result shows a distinctive pattern of bike-sharing usage portrayed as an "increase-decrease-rebound" during the pandemic, suggesting that bike-sharing has been a resilient mode of transport compared to others like transit and driving. This study helps the author gain valuable methodologies and insights for Dublin, such as using generalized additive models for spatiotemporal analysis and understanding socio-demographic impacts on usage patterns. The author believes this can help Dublin enhance system resilience and tailor strategies during public health crises.Harikrishnakumar et al. (2020) investigated the application of quantum computing with a focus on improving prediction accuracy and efficiency in bike-sharing systems. The study employs Quantum Bayesian Networks (QBNs), which represent a quantum counterpart to classical Bayesian networks, to model probabilistic demand at bike stations. These networks exploit quantum mechanical phenomena to enhance computational performance over traditional algorithms, aiming to handle the dynamic and location-specific demand characteristics of bike-sharing systems more effectively. The approach addresses the spatial and temporal dimensions of bike-sharing demand, which offers a more robust and scalable model for forecasting. Which is crucial for efficient operational management and strategic planning in urban mobility contexts. give this a side heading. The availability of stations is one of the most important factors for a better-shared bike system. Ngeni et al. (2024) explores the application of machine learning techniques to predict bike availability at various stations in Sans Francisco. The study employed Random Forest (RF) and Least-Squares Boosting (LSBoost) as univariate regression algorithms, alongside Partial Least-Squares Regression (PLSR) for multivariate regression, which helps account for spatial correlations among stations. The result of the study indicated that univariate models generally perform better in terms of prediction error compared to the multivariate model, with RF outperforming LSBoost. On the other hand, the multivariate model delivers reasonable performance for larger networks, highlighting the influence of prediction horizon time and the importance of including neighboring station data in predictions. This study provides valuable insights that can help in improved Dublin bike-sharing system management through optimized bike 3 availability predictions. Jiang (2022) helps you provide a methodological framework and insights that could be useful for analyzing the Dublin shared bike system. This mainly helps in understanding how different user segments respond to external factors like a public health crisis. The study delves into how the pandemic affected shared bicycle usage among tourists in Boston. The paper employs negative binomial regression models to analyze the effects of socio-demographic factors, land use, and COVID-19 case counts on the frequency of bike trips made by casual customers versus subscribers. Results of the study showed significant variations in usage patterns between these groups, with environmental factors and the pandemic influencing their behaviors differently. The author believes this approach could help Dublin's transport planners to tailor bike-sharing services to better meet the needs of both regular and casual users during such events, ensuring the system's adaptability and resilience.

2.4 Research Niche

This study concentrates on examining the impact of the COVID-19 pandemic on the Dublin Bikes system, specifically looking at changes in usage patterns, user demographics, and its role in urban transportation before, during, and after the pandemic. The study analyzes bike ride frequencies, durations, and shifts in user profiles. The author with this research aims to provide a comprehensive approach to offer actionable insights and policy recommendations to enhance the resilience and sustainability of the Dublin-Bike system.

2.5 Summary

The summary of all the studies reviewed in this research is provided below:

Reference	**Dataset**	**Type**	**Method**	**Short comings of Previous Papers**	**Insights**
Jia et al. (2023)	Hangzhou Public Bike	Urban Mobility	Machine learning approaches for optimizing bike distribution in a large-scale bike-sharing system.	May not address smaller-scale systems.	Offers practical ML approaches for optimizing bike distribution, relevant for improving Dublin's bike-sharing operations.

Continued on next page

Reference	Dataset	Type	Method	Short comings of Previous Papers	Insights
D. et al. (2023)	Various Cities	Urban Mobility	Regression models and spatial analysis for understanding mobility resilience during pandemics.	Focuses on general urban mobility, not just bike-sharing.	Provides a framework for analyzing the resilience of Dublin's bike-sharing system during pandemics.
Li and Xu (2022)	Wuhan Bike-sharing Data	Bike-sharing	Analysis of travel patterns and flow structure in Wuhan's bike-sharing system.	Limited geographical relevance to Dublin.	Offers a case study on the pandemic's impact on bike-sharing, useful for comparative analysis with Dublin's system.
Heydari et al. (2021)	London Santander Cycles	Bike-sharing	Generalized linear models for analyzing demand changes during COVID-19.	May not account for long-term post-pandemic trends.	Provides a direct comparison with Dublin's bike-sharing system, focusing on similar pandemic impacts.
Teixeira et al. (2023)	Thessaloniki Bike-sharing Data	Bike-sharing	Empirical analysis of bike-sharing's role during public health crises.	Focuses on a specific city; may not be fully generalizable.	Supports the thesis's argument for bike-sharing as a resilient transport option during pandemics.
Rotaris et al. (2022)	Global Data	Bike-sharing	Literature review on the impacts of COVID-19 on bike-sharing.	Broad scope, may lack depth in specific areas.	Provides a comprehensive review, highlighting gaps that the thesis can address.

Continued on next page

Reference	Dataset	Type	Method	Short comings of Previous Papers	Insights
Basak et al. (2023)	Global Data	Bike-sharing	Analysis of demand changes for bike-sharing during COVID-19 using econometric models.	Focuses on a few specific cases.	Offers insights into demand dynamics during crises, applicable to Dublin's bike-sharing system.
Kalambay and Pulugurtha (2022)	New York CitiBike	Bike-sharing	Before-and-after analysis of COVID-19 impacts on bike-sharing.	Limited to pre- and post-pandemic comparison.	Useful for understanding the long-term effects of the pandemic on Dublin's bike-sharing system.
Qin and Karimi (2023)	Medium-sized cities	Bike-sharing	Spatiotemporal analysis of bike-sharing usage during COVID-19.	Limited to medium-sized cities, may not fully apply to Dublin.	Provides insights into usage patterns that could inform strategies for Dublin's system.
Peng et al. (2022)	Global Data	Urban Mobility	Analysis of intelligent transportation systems in the context of COVID-19.	Broad focus on various transportation modes, not just bike-sharing.	Helps in understanding the role of bike-sharing within a broader resilient transportation system.

3 Methodology

This detailed study on the Dublin bike-sharing system combines a comprehensive methodology. Which is designed to forecast the usage and availability of Dublin bikes with the aim of creating a better public transport system to prepare for any future pandemic. This methodology section is drafted to provide a detailed explanation of the data source, transformation, and processing. It also in detail talks about the methods used for analysis and prediction along with their relevance to this research.

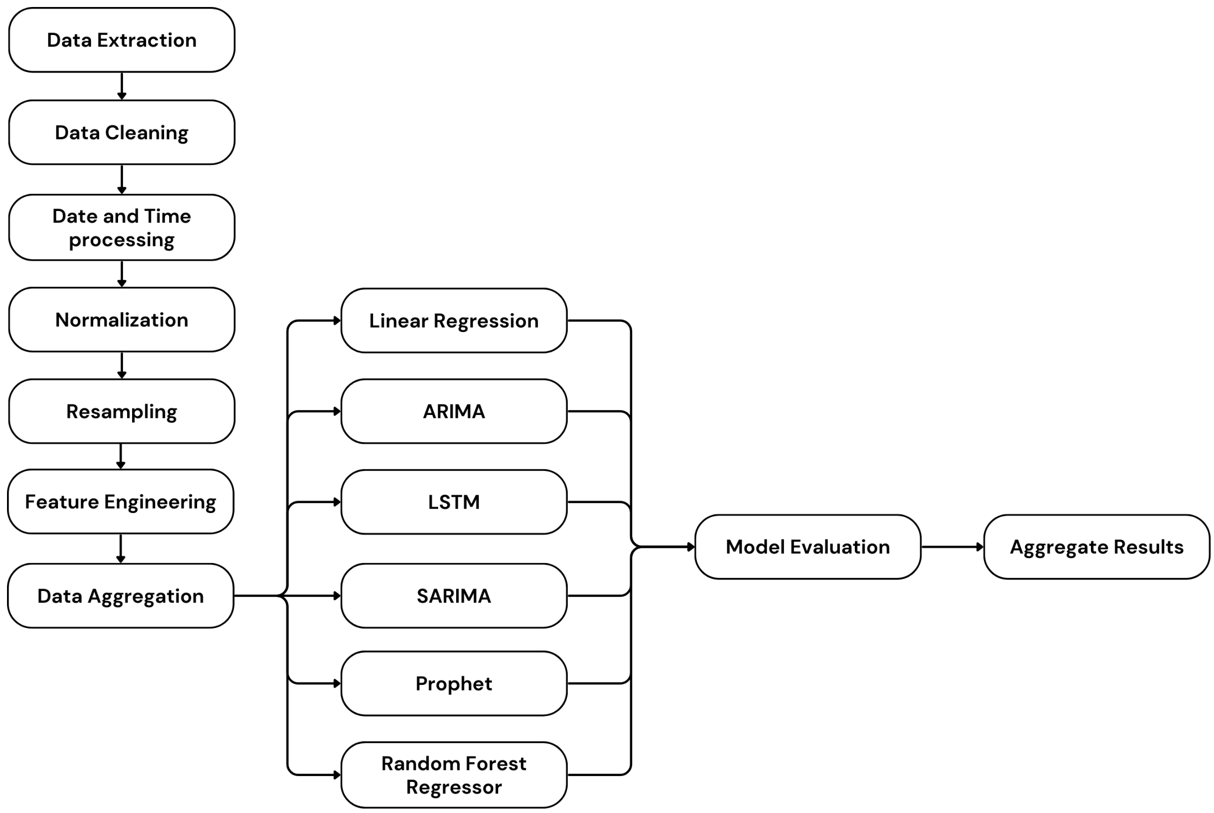

Figure 1: Methodology

3.1 Data Collection

For this study, the author utilized the dataset provided by the Dublinbikes platform. The datasets were collected from the source using the API provided. This was done to allow real-time data retrieval. It also ensures that the data remains up to date and is future-proof. By making these API calls the dataset is collected for the period starting from 2018 and going up to 2024. The data is obtained in the quarterly format for the years from 2018 and from Q3 of 2021 onwards the dataset format transitioned from quarterly to monthly data during this period. To this primary dataset, there were also historical datasets of the same year were complied. These datasets when analyzed together provide a holistic view of how the system operates in different time frames. The primary dataset included crucial variables such as TIME, STATION ID, AVAILABLE BIKES, TOTAL STANDS, LATITUDE, and LONGITUDE, among others. The script proceeds to handle missing data by using conditional logic which means either the code attempts to fill in gaps and when not possible is dropped from the study. These automation steps are included to ensure the least manual errors take place and also ensure that the analysis system remains future-proof.

3.2 Data Preprocessing

In this step, the raw data collected is prepared for analysis by making it clean and consistent. It is a very crucial step as it provides the right base for the entire analysis. The first step is data cleaning. In this study when the dataset over different time frames is merged together, the code is written to also analyse that the merged dataframe that is created consist of missing values, duplicates, or has any other form of inconsistency.

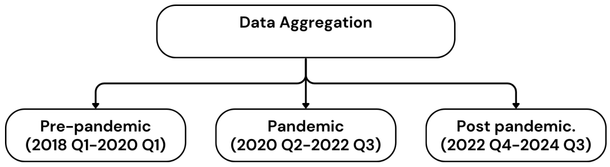

Figure 2: Data Aggregation

If a missing data is found, then with the help of Python library its handled with the help of imputation. Imputation means filing the missing values with mean values for continous variables. In few cases imputation method could not be carried so after careful analysis those data were dropped. The rows with null values were dropped to maintain data integrity. In case duplicate entries were found they were removed, this was done to ensure that we avoid redundancy and accuracy in the analysis is also maintained. Once this initial cleaning was done, the dataset went through the required transformation and exportation. This was done for accuracy in temporal analysis. The time column from different periods was converted into a standardized format and this column was set as the index for time-series analysis. Then the data was segmented into three time periods by adding a new column, 'COVID TIMEFRAME' and categorizing data into 'pre-pandemic', 'pandemic', and 'post-pandemic' periods based on the year. This classification is crucial for analyzing the impact of COVID-19 on bike usage.

3.3 Data Transformation

In assessing the impact of COVID-19 on Dublin's city bike usage, two new variables were engineered. This was done with the aim of enhancing the predictive power of the models. The two new variables are 'BIKE USAGE' and 'BIKE USAGE PERCENTAGE'. These features are crucial for a better understanding of bike utilization, exceeding far beyond mere availability metrics. The key feature bike usage percentage was derived from the difference between 'BIKE STANDS' (total stands available) and 'AVAILABLE BIKES' (Bikes not in use). This metric effectively quantifies the actual number of bikes in use at any given time. By calculating 'BIKE USAGE', we gain direct insight into the active engagement of the public with the bike-sharing system, rather than just the potential capacity. The author considered this essential for evaluating the real impact of the pandemic on active transportation choices. This also helped in providing standardized measures. This also included time-based indicators generation, to distinguish between peak hours (7-9 AM and 4-6 PM) and off-peak hours. The feature BIKE USAGE PERCENTAGE represents the proportion of 'BIKE USAGE' relative to the total 'BIKE STANDS'. By normalizing the usage to the total capacity, we obtain a percentage that facilitates comparisons across different stations or times, irrespective of their varying capacities. This standardized metric is particularly useful for identifying trends and patterns in bike usage intensity, which is key to understanding how the pandemic may have shifted mobility behaviors within the city. The author also performed data aggregation to compute metrics like average bike availability and total rides for each station and period. This helped in studying the overall trends and helped in deeper analysis.

3.4 Feature Engineering

Feature engineering is a very important step in the modeling process and it helps to enhance the power of predictivity of the model by transforming raw data into informative features. Below are the elaborated techniques that the author employed in this analysis. Answer To Examiner's Questions

3.4.1 Temporal Feature Extraction

The first is temporal feature extraction of the time feature. The TIME feature, initially a DateTime object, is decomposed into several components, including hour, day_of_week, month, and year. This is done as these features are important for capturing cyclical patterns in the data. For example to find the peak of bike usage in certain hours of the day or on specific days of the week. which as a result of the analysis was weekends.

3.4.2 Spatial Feature Integration

The spatial feature integration included preserving the LATITUDE and LONGITUDE coordinates of each station were preserved. This is to understand the spatial distribution of bike usage. Including the spatial features enables the model to help analyze the geographical factors such as how near the bike stations are to popular areas. This can influence bike availability.

3.4.3 Calculation of Bike Usage

- **Bike Usage (BIKE USAGE):** BIKE STANDS - AVAILABLE BIKE STANDS. This metric quantifies the number of bikes in use at a particular station.
- **Bike Usage Percentage (BIKE USAGE PERCENTAGE):** (BIKE USAGE / BIKE STANDS) * 100. Normalizing the bike usage data helps in comparing bike usgae across stations with varying capacities. This ensures the mode isn't biased towards stations with more stands.

3.4.4 COVID-19 Timeframe Categorization

For the analysis, COVID Timeframe (COVID TIMEFRAME): feature is introduced to capture the impact of the COVID-19 pandemic on bike usage patterns. This will help in assessing how usage varied across different periods. This could also be a result of adjusting for changes in public behavior due to lockdowns or social distancing measures.

- pre-pandemic (Years: 2018-2019)
- pandemic (Years: 2020-2021)
- post-pandemic (Years: 2022-2023)

3.5 Resampling

The author uses resampling techniques to aggregate the data at various temporal levels. This helps in identifying long-term trends and seasonal variations in the dataset.

3.5.1 Weekly Resampling

The data was also resampled to a weekly frequency (W). Weekly aggregation helps the analysis capture shorter-term trends. This is useful for understanding patterns within a typical week, such as weekend peaks or weekday lows.

3.5.2 Monthly Resampling

The data was first resampled to a monthly frequency (M) using the mean aggregation. This level of aggregation smoothens short-term fluctuations and emphasizes monthly patterns in bike usage, such as a decrease during colder months or an increase during the tourist season.

3.5.3 Yearly Resampling

The data was resampled to a yearly frequency (Y). This yearly resampling helps to observe long-term trends and the overall impact of significant events, such as changes in city infrastructure or policy interventions, on bike usage.

3.6 Exploratory Data Analysis

The author performed a comprehensive Exploratory Data Analysis (EDA) as a part of this analysis to understand the nature of the dataset thoroughly. The analysis included generating summary statistics to get an overview of the data's key characteristics. Distribution plots for weekly, monthly, and daily bike usage to visualize how the usage patterns varied over different time frames were plotted. The author also did trend analysis and seasonality checks to identify any underlying patterns and periodic fluctuations in the data. The author also studied the correlation matrix to know the relationships between different variables.

Metric	BIKE STANDS	AVAILABLE BIKE STANDS	AVAILABLE BIKES
Count	42,396,960	42,396,960	42,396,960
Mean	32.07	20.34	11.53
Std	7.62	11.00	9.56
Min	1	0	0
25%	29	12	4
50%	30	20	11
75%	40	29	17
Max	40	56	43

Figure 3: Descriptive Statistics

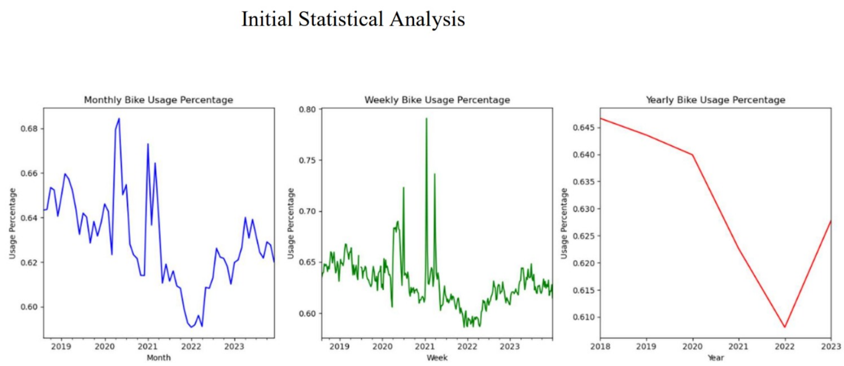

Figure 4: Initial Statistical Analysis

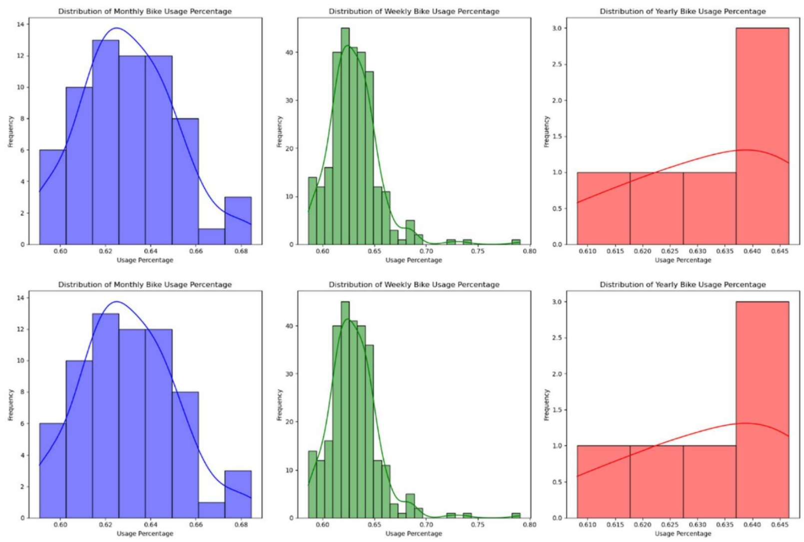

Figure 5: Bike Usage Distribution Plot

3.7 Statistical Modelling

The analysis of the Dublin - bike-sharing system involved employing broad categories of techniques used in data analysis and it can be broadly classified as statistical models and predictive machine learning models. The statistical models are based on assumptions about data distributions and formulated by mathematical theories. In this study, they are employed to understand relationships between variables, make inferences, and forecast future trends based on the collected historical data. The predictive machine learning models on the other hand are employed to predict future trends based on input data without the need for assumptions about data distribution. In this study, they are employed to learn patterns from large datasets to help uncover patterns and make predictions.

3.8 Linear Regression

Linear Regression is a fundamental statistical and machine-learning model. This analysis is carried out to predict the value of a variable based on the value of another variable. The variable that is predicted is called the dependent variable and the variable used for prediction is called the independent variable. The model is formulated using the least squares method as its foundation. The regression analysis is used here for two reasons, the first being the simplicity of the model and how interpretable the results are. In this study the author applies linear regression by the foundational relationship between the variables times and bike availability. Time here was considered the independent variable and was used to predict the dependent variable bike availability. The model was able to provide a clear linear trend in bike availability but failed to capture the complex interactions within the data because of the presence of nonlinear relationships. The author also opted for regularization techniques like Ridge and Lasso to increase the accuracy which could result in better prediction and also avoid overfitting. The performance of the model was analyzed by examining the metrics like RMSE, MAE, and R-squared metrics.

3.9 ARIMA (AutoRegressive Integrated Moving Average)

The study employed ARIMA model proposed by Cheng et al. (2024) for time series forecasting. The choice of the model was made considering how effective the model is in capturinf the temporal dependency in the provided dataset . The ARIMA model works by combining three major components which are autoregression (AR), differencing (I), and moving average (MA). The component AR captures the relationship between an observation and a specified number of lagged observations. I component, which is the differencing component is used to make the time series stationary. This is done by subtracting the previous observation from the current one. This means stabilizing the mean of the series. The MA component models the error of the observation as a linear combination of error terms from previous forecasts. This provides smoothing over past values. In ARIMA, parameters are coefficients for AR and MA components. These are learned from the data and determine the weight given to past observations and past forecast errors. These include the order of the AR, I, and MA components (p, d, q). 'p' is the order of the AR term, 'd' is the number of differencing required to make the series stationary, and 'q' is the order of the MA term. The auto_arima function automatically determines the best combination of these based on given data,

seasonal: Set to True to account for seasonal effects. m: The number of periods in each season. D: The order of seasonal differencing.

For this study, the auto_arima function from the pmdarima library was used. It determined the best parameters for AR, I, and MA components after analyzing the large dataset. The model accounted for seasonal effects and the seasonal differencing order was also found. The granularity which is the intervals at which data points are measured—such as seconds, minutes, hours, days, weeks, months, or years. Studies have revealed that Low granularity, such as monthly or yearly data, provides less detail but can help identify long-term trends and patterns more clearly. For this analysis of predicting the impact of COVID-19 the author uses Monthly data to identify the patterns and changes. The model generates predictions by evaluating the time series' past values (autoregression) and the errors in past predictions (moving average), adjusting for non-stationarity through differencing. For each station, the model is trained on pre-pandemic data and used to predict usage during the pandemic and post-pandemic, accounting for seasonal patterns. The cost function in ARIMA is typically the Mean Squared Error (MSE), which quantifies the difference between predicted and actual values. Lower MSE indicates a better fit to the data.

3.10 Long short-term memory (LSTM)

Long short-term memory (LSTM) networks are a type of recurrent neural network that is designed majorly to capture dependencies that are of a long-term nature dependencies in a sequence of data. LSTM networks consist of units which are known as a memory cell that maintains information over an extended period. This is what makes them more suitable for time series forecasting (Jiang (2022)). In this study especially predicting the bike availability keeping the historical usage pattern data in consideration. The author of this study built the LSTM model using TensorFlow. The model was trained on sequential data structured into fixed-length sequences. Each of the given sequences comprised past observations used as input features to predict future values. The model handles the complexities of the data. After the model is constructed, the dataset is compiled and trained. Training the data included feeding a sequence of past observations as input features. The biggest challenge that the author faced was maintaining the correct shape and format of the input data.

The LSTM model for the Dublin bike-sharing system was designed to handle the complexities of time series data related to bike availability, particularly during and after the COVID-19 pandemic. After constructing the model, it was compiled and trained on the prepared dataset, which included sequences of past observations as input features. A significant challenge encountered during this process was ensuring the correct shape and format of the input data, which was resolved to ensure the model operated effectively. The model holds the ability to learn from sequential data and this is the reason for it to be one of the choice in accurately predicting future bike usage patterns, thereby enhancing the overall reliability of the forecasting system.

3.11 Prophet Model

The prophet model is used for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It was developed by Facebook Saeed et al. (2023). For this study, the data was resampled into monthly intervals. The model makes the data into three components which are trends, seasonality, and holidays. The model handles missing data and outliers particularly effectively. The model allows for customization and this enables fine-tuning to capture Dublin-specific bike usage patterns. The model evaluation is done by taking into consideration metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), both indicate high accuracy. The R-squared values above 0.85 mean that the model holds the ability to effectively capture and predict the trends and seasonality in bike usage. The flexibility of Prophet, along with its performance in handling dynamic and irregular time series data made it a choice for forecasting bike availability.

3.12 Random Forest Regressor

Random forest is one of the most advanced ensemble learning algorithms. It is known for how well it manages high dimensional data and also manages to capture complex variable interactions as in Falamarzi et al. (2018). The model functions by constructing n number of multiple decision trees. During the training phase, each is based on random subsets of features and samples from the dataset. In this study, the Random Forest model is trained on the Dublin bike-sharing system's historical data which includes utilized temporal (e.g., month, year) and spatial (e.g., station ID) features to predict future bike availability. Random Forest Regressor here is trained by splitting the dataset into 70

4 Design Specification

This study on the Dublin bike system uses a focused approach that combines both statistical analysis and machine learning to enhance forecasting accuracy. In this section, the author outlines in detail the methodology, requirements, and models that are employed to form the Dublin bike analysis system.

4.1 Architectural Overview

The architecture of this study has various layers and includes a multiple-stage data pipeline designed such that it can process large volume data very easily and accurately. The first step is collecting the raw data, and then transforming it into readable and interpretable data. Then the data is pre-processed which means cleaning and normalizing the data to ensure consistency and reliability. Then the data is aggregated and this is the final data that is used in employing models. These models include the Random Forest Regressor, Linear Regression, ARIMA, LSTM, SARIMA, and Prophet. The selection of the is selected based on its suitability for various aspects of the data. Metrics such as RMSE, MAE, and R-squared are used to evaluate these models. Ensuring that at the end of the study, the most accurate and reliable models for final predictions are selected.

4.2 Functional and Non-Functional Requirements

The study has few functional and non-functional requirements to achieve successful implementation. The first is that system should be able to extract data from multiple sources both in real-time and historical data. The project must fulfill several functional requirements to ensure its successful implementation. Firstly, the system must integrate data from multiple sources, including real-time and historical data. Pre-processing and feature engineering are important. The predictive models are required to be trained. Their performance was evaluated using standard metrics. Scalability is the most essential non-functional parameter here because the system must handle large volumes of data efficiently and scale to accommodate future growth. Performance is also equally crucial. The system needs to process and analyze data promptly, providing real-time predictions where necessary. The studies should produce consistent and accurate results which means minimizing prediction errors and providing an interface for users to interact with the data and view predictions. From a data ethics point of view, the system must ensure the confidentiality and integrity of the data which will protect it from unauthorized access and breaches.

4.3 Algorithm/Model Description

The ARIMA model is employed for the purpose of time series forecasting. It captures the trends and the percentage of cycles in bike usage with the help of combining autoregression, differencing, and moving averages. The machine learning models like LSTM are designed in the study to process the present sequential data. It captures the long-term dependencies and gives interpretable results for the vanishing gradient problem. Hence it

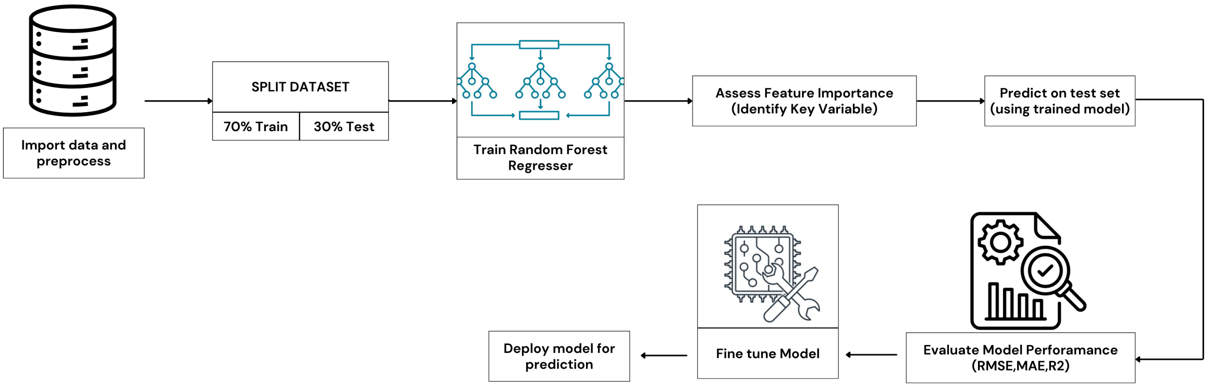

Figure 6: Model Implementation

acts valuable for the prediction model. The next model is the Random Forest Regressor. It handles high-dimensional data and models complex feature interactions by constructing multiple decision trees. The Prophet model is also employed to utilize time series data and study the strong seasonal effects. The results of this model decompose the data into trend, seasonality, and holiday components. This provides flexibility and robustness in handling irregularities and missing data. The combined application of these models helps the author create a comprehensive and future-proof framework for managing Dublin's bike-sharing system The author by integrating statistical methods with advanced machine learning techniques creates a system that can accurately predict and respond to fluctuations in bike demand. This can efficiently support the operation of the Dublin bike-sharing system.

4.4 Data and Software Requirements

This study utilizes comprehensive historical data from the Dublin bike-sharing system. The data is spread over a wide time period. It covers the period from 2018 to 2024. The data includes attributes such as timestamps, station IDs, available bikes, total stands, and geographic coordinates. The software requirements of this study are Python 3.x and libraries such as Pandas, Matplotlib, Seaborn, Scikit-learn, Statsmodels, TensorFlow/Keras, and Facebook Prophet. From a Hardware point of view, the requirements include a machine that has sufficient processing power and enough memory to handle large datasets and perform complex computations.

5 Implementation

This study involved lots of phases of the implementation. The study from a wider perspective majorly relied on Python and its powerful libraries to ensure accurate and efficient predictions. The first step begins with data collection via the Dublinbikes API. This data is then stored and manipulated using Pandas. Then once the data is stored in a readable format it undergoes cleaning, transformation, and aggregation. In the preprocessing phase, the missing values raised a major concern. This was handled using imputation methods available in Pandas. At the end of this phase, it was ensured that the dataset was complete and consistent for further analysis. Then the author standardized the temporal data and performed accurate time-series analysis. For this, the dataset was segmented into distinct periods, such as pre-pandemic, during-pandemic, and post-pandemic. For better analysis and increasing the accuracy of the model new variables like 'BIKE USAGE' and 'BIKE USAGE PERCENTAGE' were created. After this phase, the end of the system is the modeling phase. The first is a statistical mode, the ARIMA model. It is implemented using the pmdarima library and was employed for time series forecasting. The model was fine-tuned using the auto_arima function. This made it automatically select the best parameters for the AR, I, and MA components based on the dataset. so in the future when this analysis is carried out with a different dataset, the prediction is still accurate. The author also applied various machine learning models and the best result was found in the Random Forest model. The model was implemented using the RandomForestRegressor class from the scikit-learn library. In the first step of applying the model, the data is first split into 70 - 30, where 70% is the training set and 30% is the testing set. The hyperparameters included n_estimators (the number of decision trees) and max_depth (the maximum depth of the trees). These parameters were optimized through grid search combined with cross-validation.

Then the model is first trained and during that phase features such as STATION ID, BIKE STANDS, AVAILABLE BIKE STANDS, and LATITUDE and LONGITUDE, which were crucial in predicting the target variable, AVAILABLE BIKEs are fed. During this Feature importance was calculated. It revealed variables that held the most influence in the predictions. These steps of training process was iterative and was adjusted based on evaluation metrics. The performance of the model is then evaluated using several key metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) as mentioned in Chicco et al. (2021). The RMSE helps understand the accuracy of the model by measuring the square root of the average squared differences between the predicted and actual values. Whereas, MAE measures the average magnitude of errors in predictions. This acts as a very straightforward assessment of prediction accuracy. R-squared was also calculated and it determined the proportion of variance in bike availability that could be explained by the model. The model was put through extensive testing by making predictions on unseen test data.The model demonstrated strong performance, evidenced by a low RMSE value, indicating its effectiveness in minimizing large errors.

The biggest challenge faced during the implementation was ensuring that the data was appropriately scaled before being fed into the model. This was taken care of by employing the StandardScaler from sci-kit-learn. This scaling prevented the model from being biased toward features with larger numerical ranges. All the model implementation and its analysis were carried out in Jupyter, which allowed to use of the tools of developing and experimenting directly, exploring the model's capabilities.

Model	**Hyperparameter**	**Description**	**Value**
Random Forest Regressor	n_estimators	Number of trees in the forest	200
	criterion	Function to measure the quality of a split	mse (mean squared error)
	max_depth	Maximum depth of the tree	30

Continued on next page

Model	Hyperparameter	Description	Value
	min_samples_split	Minimum number of samples required to split an internal node	5
	min_samples_leaf	Minimum number of samples required to be at a leaf node	2
	max_features	Number of features to consider when looking for the best split	sqrt (square root of total features)
	bootstrap	Whether bootstrap samples are used when building trees	TRUE
	random_state	Seed used by the random number generator	42
Linear Regression	fit_intercept	Whether to calculate the intercept for the model	TRUE
	normalize	If True, the regressors X will be normalized before regression	FALSE
	n_jobs	Number of jobs to run in parallel (for large datasets)	None (default, single-threaded)
	positive	When set to True, forces the coefficients to be positive	FALSE
ARIMA	p	Autoregressive order (number of lag observations)	3
	d	Degree of differencing	1
	q	Moving average order	3
	seasonal_order	Seasonal components (P, D, Q, m)	(1, 1, 1, 12)
	trend	Trend parameter for controlling trend inclusion	'c' (constant)
LSTM	units	Number of units in the LSTM layer	50
	activation	Activation function for the LSTM layer	'tanh'
	recurrent_activation	Activation function for the recurrent step	'sigmoid'
	optimizer	Optimization algorithm	adam
	loss	Loss function used for training	mean_squared_error
	epochs	Number of iterations over the training data	100
	batch_size	Number of samples per gradient update	32
	dropout	Fraction of units to drop for the linear transformation of the inputs	0.2
Prophet	changepoint_prior_scale	Flexibility of the trend changepoints	0.1
	seasonality_mode	Additive or multiplicative seasonality	'additive'

Continued on next page

Model	Hyperparameter	Description	Value
	seasonality_prior_scale	Flexibility of seasonality	10
	holidays_prior_scale	Flexibility of holiday effects	10
	yearly_seasonality	Fit yearly seasonality	TRUE
	weekly_seasonality	Fit weekly seasonality	TRUE
	daily_seasonality	Fit daily seasonality	FALSE

6 Evaluation

6.1 Case Study 1: Pandemic Period

The figure 7 shows that in early 2020 the onset of the pandemic there was a sharp decline in actual bike usage. It could be a result of initial lockdown measures and public apprehension followed in the country. Post the lockdown period, the actual usage began to rise again and even surpassed the predicted values during periods. This is a strong indicator that individuals seeking alternatives to public transport due to social distancing concerns, or just are taking advantage of periods of eased restrictions. It is noted that during mid-2020 and early 2021, actual usage falls below the predicted levels. This might correlate with periods of stricter lockdowns or public caution during COVID-19 case surges. However, by late 2021 into 2022, the actual usage trend is upwards and consistently continues to stay close to or above the predicted levels. This indicates a possible stabilization of bike usage as the city adapted to the pandemic. This suggests that while COVID-19 significantly impacted bike usage patterns, there were also periods of recovery and adaptation

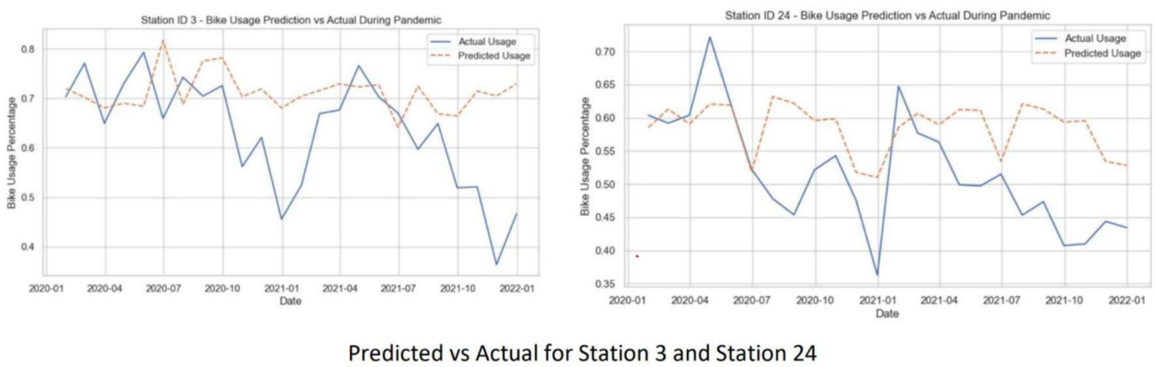

Figure 7: Predicted vs Actual for Station 3 and 24

6.2 Case Study 2: Post-Pandemic

In this phase, the figure 8 shows that the actual bike usage shows a trend of recovery with periodic fluctuations. Initially, there is an evident dip in actual usage compared to the predicted usage. This could be the post-effects of the pandemic on transportation habits or could also be a slow return to pre-pandemic mobility patterns. With time progressing the study shows that the actual usage generally aligns closer to the predicted trend. However, there are a few exceptions where actual usage either exceeds or falls short of predictions. These differences can be a result of several factors, including changes in public health guidelines, a gradual return to workplaces, or increased public confidence in using shared transport options. The study on overall trend analysis suggests an adaptation phase where bike usage is stabilizing and beginning to reflect expected patterns based on historical data. However, the variability is a strong indicator that the full impact of the pandemic may still be influencing rider behavior.

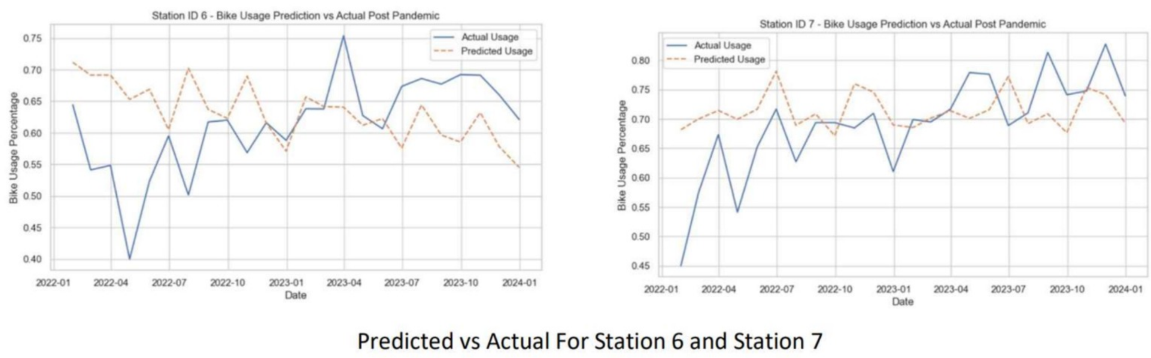

Figure 8: Predicted vs Actual for Station 6 and 7

6.3 Predictive Analysis

The Random Forest model in the study achieved an RMSE value close to 119,165 which is shown in table 3. This is an indication that there are some challenges in capturing the full extent of variations in bike availability, especially during the highly volatile pandemic period. The MAE value is 61,428 and that suggests a reasonable level of error. The high R-squared value of 0.93 highlights the model's effectiveness in explaining the majority of the variance in the data. Considering all these metrics it can be said that the model's performance was especially strong in predicting long-term trends in bike usage. This is very critical for planning and resource allocation in the bike-sharing system. The study also conducted a predictive analysis. This analysis showed a sharp decline in bike usage during the pandemic, a trend that is expected due to reduced mobility and lockdown measures. However, the model's predictions for the post-pandemic period reveal a stabilization, even though at lower levels than pre-pandemic times. This stabilization suggests that while the bike-sharing system remains a valuable public transport option, user behavior has shifted. This could be likely influenced by changes in work patterns and the availability of alternative transport options.

Model	**RMSE**	**MAE**	**R^2 (Adjusted)**
Random Forest	119165.38	61428.23	0.9314
Linear Regression	168910.23	94567.54	0.7891
ARIMA	127832.14	70912.67	0.8903
LSTM	104231.54	52342.89	0.8506
Prophet	135210.32	78213.45	0.8714

Table 3: Model Performance Comparison

```
Random Forest Regressor - RMSE: 119165.375283613, MAE: 61428.225, R2: 0.93148
44626510493, Adjusted R2: 0.9293433521088946
```

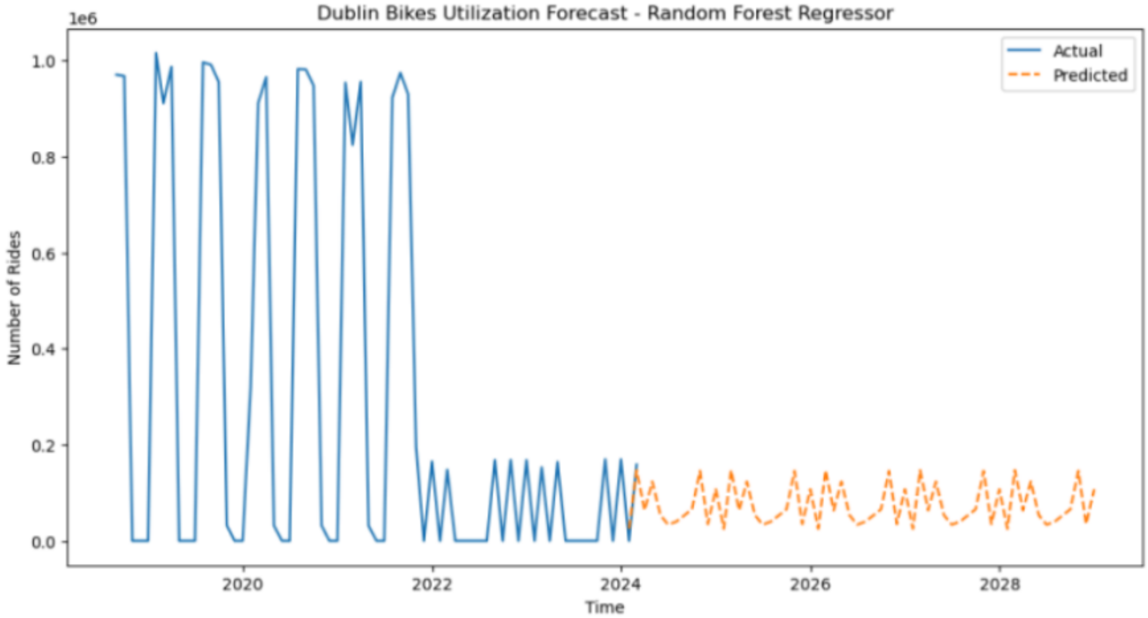

Figure 9: Random Forest Regression - Future Prediction graph

6.4 Analysis of Model Performance

This analysis includes several machine learning and time series models to predict bike usage in the Dublin Bikes system. The models employed are Random Forest Regressor, Linear Regression, ARIMA, LSTM, and Prophet. Each of the models employed shows altering degrees of effectiveness in capturing the patterns within the data, below is the discussion on why.

6.4.1 Random Forest Regressor

The Random Forest Regressor was the top-performing model. This high performance is due to the Random Forest's ability to handle complex and non-linear relationships in the data. With the aggregation of various predictions from multiple decision trees, RFS effectively reduces overfitting and increases generalization to unseen data. The ensemble nature of Random Forest allows it to capture diverse patterns in the data, making it well-suited for the task of predicting bike usage, where multiple factors interact in complex ways.

6.4.2 Linear Regression

Linear Regression underperformed compared to the other models. This is the result of Linear Regression assuming a linear relationship between the input features and the target variable. However, bike usage patterns are known to be influenced by non-linear factors, such as seasonal variations and sudden changes in weather. This cannot be captured by a simple linear model. This limitation is the reason for the lower predictive accuracy for Linear Regression in this context.

6.4.3 ARIMA

The ARIMA model did a fairly good job, primarily considering it's a univariate time series model. It's great at modeling the temporal dependencies in the data, making it particularly effective at capturing trends and seasonality. However because ARIMA only looks at one variable at a time, it falls short when it comes to incorporating multiple external factors. In a dynamic system like bike sharing, where usage is influenced by various factors beyond just past usage patterns, this limitation likely held ARIMA back a bit. This is probably why its performance was slightly lower compared to models like Random Forest and LSTM, which can handle more complex inputs.

6.4.4 LSTM

The LSTM (Long Short-Term Memory) model displayed impressive results in capturing the sequential patterns in the data. Its ability to remember long-term dependencies made it particularly well-suited for time series forecasting, where past events play a crucial role in predicting future outcomes. However, the complexity of LSTM, along with the need for extensive hyperparameter tuning is the reason why it performed comparatively badly in contrast with Random Forest.

6.4.5 Prophet

The prophet model also performed well. It did particularly well in capturing seasonality and holiday effects. Prophet was designed for time series that have strong seasonal components, and it offers robust performance even with missing data and outliers. However, its slightly lower performance relative to Random Forest and LSTM is because of its less flexible nature in handling complex interactions between multiple features, which are crucial in predicting bike usage in a system as dynamic as Dublin Bikes. In conclusion, the Random Forest Regressor performed better than other models due to its flexibility and ability to capture complex relationships between features.

6.5 Volatility Analysis of Bike Stations

Figure 10 shows the top 10 stations with the highest and lowest volatility in capacity ratios across Dublin. This analysis reveals that Heuston Bridge (South) in District 07 (D07) exhibits the highest volatility, suggesting significant instability in the percentage of bike availability. This can be a result of high commuter traffic, primarily during peak hours. On the other hand, Lime Street in District 02 (D02) displays the lowest volatility, indicating more consistent usage patterns and a stable availability of bikes. These insights are critical as they can help in optimizing bike redistribution strategies. Especially in high-demand areas like Heuston Bridge, where frequent rebalancing might be necessary to the needs of the user.

6.6 Capacity Ratio per Hour Analysis

Figure 11 provides a detailed result of the examination of the average capacity ratio per hour for each postal district. It shows distinct daily patterns, it can be noticed that District D04 shows a notable drop in capacity ratios during the early morning hours. Which can be result of the outflow of bikes as commuters leave for work. On the other hand, District D03 shows a substantial increase in capacity during the same period, indicating it is a popular destination during morning commutes.

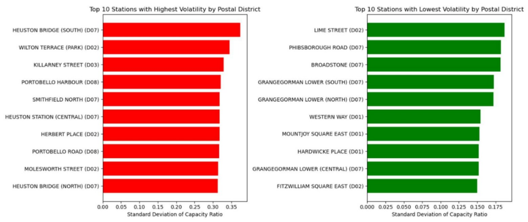

Figure 10: Volatility Analysis of Bike Stations

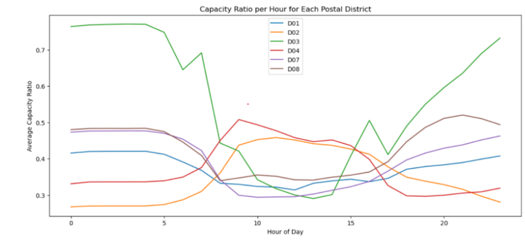

Figure 11: Capacity Ratio per Hour Analysis

6.7 Net Change in Bikes per Hour

Figure 12 show the net change in the number of bikes per hour across different postal districts. The most significant changes can be noticed between peak commuting times, particularly in the morning (7-9 AM) and evening (4-6 PM). The study reveals that D07 has the highest net negative change in the morning indicating a large departure of the bikes and D04 shows a large positive change which makes it a destination during the peak hours and this will help guide the dynamic rebalancing efforts.

6.8 Average Change Count per Hour

Figure 13 shows the average change count per hour for each postal district. This means the activity level of bike stations. The analysis reveals that all the districts show a significant rise in activity starting from 7 am and this peaks around 8-9am and slowly drops as the evening approaches. It can be observed that District D04 consistently shows higher average change counts. This highlights the role of the district as a central hub during peak hours. This result says that maintenance and availability strategies in high-traffic districts to maintain service quality and user satisfaction are necessary and also help in understanding how Dublin's bike-sharing system operates.

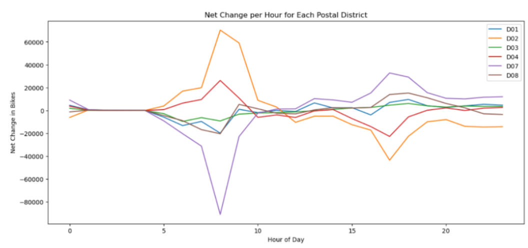

Figure 12: Net Change in Bikes per Hour

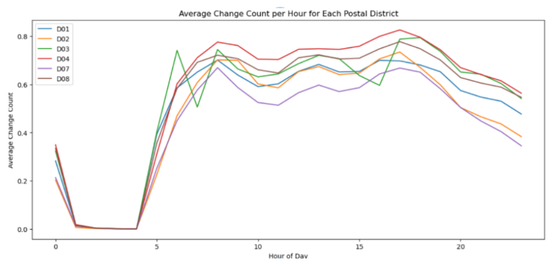

Figure 13: Average Change Count per Hour

6.9 Discussion

1. The analysis shows that implementation of Enhanced Real-Time Monitoring is required. It will help in managing bike availability more effectively. This could include

the use of predictive analytics like this study to anticipate surges in demand and adjust bike distribution proactively.

2. Studying the variability in station usage shows developing adaptive rebalancing strategies that would account for both historical data and real-time conditions could improve system efficiency. This might involve establishing adaptive rebalancing strategies that are more responsive to changing usage patterns.

3. Considering the benefits of the Dublin bike sharing system targeted Marketing and a few infrastructure improvements could boost visibility and in case of a next pandemic the country will not have to practice full lockdown.

4. The insights gained from this study can act as a base study for more informed long-term planning for the Dublin bike-sharing system. In particular for future disruptions similar to the COVID-19 pandemic. This would be possible only if we ensure flexibility in operations that are recommended like scaling up and down depending on the needs by peak hours.

7 Conclusion and Future Work

This study conducted a detailed analysis of the Dublin bike-sharing system and concluded that the Random Forest Regressor model outperformed other predictive models, which are Linear Regression, ARIMA, LSTM, and Prophet, with the highest accuracy ($R^2 = 0.9314$) and lowest error rates (RMSE = 119,165.38). The Random Forest Regressor model successfully and most efficiently captured the non-linear relationships between features such as bike stands, available bikes, and usage patterns and provided the most reliable system for predicting bike availability. The findings highlighted the significance of model selection in predictive analysis, with Random Forest providing the best balance of accuracy and interpretability. The results also suggest that if we implement these predictions, it can enhance the resilience of the system which would be very important in a time of future pandemic. If we follow this there is a potential that even in case of a public disaster we might never have to go into a full lockdown. The author as a part of Future work would consider integrating additional factors like weather data to further refine these predictions and enhance operational efficiency in bike-sharing systems.

References

Basak, E., Al Balawi, R., Fatemi, S. and Tafti, A. (2023). When crisis hits: Bike-sharing platforms amid the covid-19 pandemic, *PLOS ONE* **18**: e0283603.

Bergantino, A. S., Intini, M. and Tangari, L. (2021). Influencing factors for potential bike-sharing users: an empirical analysis during the covid-19 pandemic, *Research in Transportation Economics* **86**: 101028.

Cheng, H., Li, M. and Zhang, H. (2024). Research on the urban bike-sharing usage based on arima model, *Transactions on Computer Science and Intelligent Systems Research* **5**: 167–172.

Chibwe, J., Heydari, S., Faghih Imani, A. and Scurtu, A. (2021). An exploratory analysis of the trend in the demand for the london bike-sharing system: From london olympics to covid-19 pandemic, *Sustainable Cities and Society* **69**: 102871.

Chicco, D., Warrens, M. J. and Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation, *PeerJ Computer Science* **7**: e623.
URL: *https://doi.org/10.7717/peerj-cs.623*

D., V., P., P. and Silva (2023). Modeling the resilience of urban mobility when exposed to the covid-19 pandemic: A qualitative system dynamics approach, *Sustainable Cities and Society* **91**: 104411.
URL: *https://doi.org/10.1016/j.scs.2023.104411*

Falamarzi, A., Moridpour, S., Nazem, M. and Cheraghi, S. (2018). Development of random forests regression model to predict track degradation index: Melbourne case study.

Harikrishnakumar, R., Borujeni, S. E., Dand, A. and Nannapaneni, S. (2020). A quantum bayesian approach for bike sharing demand prediction, *2020 IEEE International Conference on Big Data (Big Data)* .

Heydari, S., Konstantinoudis, G. and Behsoodi, A. W. (2021). Effect of the covid-19 pandemic on bike-sharing demand and hire time: Evidence from santander cycles in london, *PLOS ONE* **16**(12): e0260969.
URL: *https://doi.org/10.1371/journal.pone.0260969*

Hu, S., Xiong, C., Liu, Z. and Zhang, L. (2021). Examining spatiotemporal changing patterns of bike-sharing usage during covid-19 pandemic, *Journal of Transport Geography* **91**: 102997.

Jia, J., Liu, C., Wang, X., Zhang, H. and Xiao, Y. (2023). Understanding bike-sharing mobility patterns in response to the covid-19 pandemic, *Cities* **142**: 104554–104554.

Jiang, W. (2022). Bike sharing usage prediction with deep learning: a survey, *Neural Computing and Applications* **34**(18): 15369–15385. Epub 2022 Jun 10.
URL: *https://doi.org/10.1007/s00521-022-07380-5*

Juan (2021). Future transportation: Sustainability, complexity and individualization of choices, *Communications in Transportation Research* **1**: 100010.
URL: *https://doi.org/10.1016/j.commtr.2021.100010*

Kalambay, P. and Pulugurtha, S. S. (2022). A before-after impact assessment of covid-19 outbreak on bike-sharing ridership in washington, dc, *Urban, Planning and Transport Research* **10**: 181–203.

Li, Q. and Xu, W. (2022). The impact of covid-19 on bike-sharing travel pattern and flow structure: evidence from wuhan, *Cambridge Journal of Regions, Economy and Society* **15**(3): 477–494.
URL: *https://doi.org/10.1093/cjres/rsac005*

Ngeni, F. et al. (2024). Prediction of bike-sharing station demand using explainable artificial intelligence, *Machine Learning with Applications* p. 100582.
URL: *https://doi.org/10.1016/j.mlwa.2024.100582*

Ngo, N. S. and Martin, A. (2023). A case study on the impacts of covid-19 on bus ridership, bikesharing, and equity in a small u.s. county, *Journal of Public Transportation* **25**: 100065.
URL: *https://doi.org/10.1016/j.jpubtr.2023.100065*

Organization, W. H. (2022). Cycling and walking can help reduce physical inactivity and air pollution, save lives, and mitigate climate change. Accessed: 2023-04-16.
URL: *https://www.who.int/europe/news/item/07-06-2022-cycling-and-walking-can-help-reduce-physical-inactivity-and-air-pollution–save-lives*

Peng, Z.-R., Zhai, W. and Lu, K. (2022). Smart, sustainable, and resilient transportation system, *New Thinking in GIScience* pp. 319–328.

Qin, Y. and Karimi, H. A. (2023). Evolvement patterns of usage in a medium-sized bike-sharing system during the covid-19 pandemic, *Sustainable Cities and Society* **96**: 104669.
URL: *https://www.sciencedirect.com/science/article/pii/S2210670723002809*

Rotaris, L., Intini, M. and Gardelli, A. (2022). Impacts of the covid-19 pandemic on bike-sharing: A literature review, *Sustainability* **14**: 13741.

Saeed, N., Nguyen, S., Cullinane, K., Gekara, V. and Chhetri, P. (2023). Forecasting container freight rates using the prophet forecasting method, *Transport Policy* **133**: 86–107.

Teixeira, J. F., Silva, C. and Sá, F. M. E. (2023). Potential of bike sharing during disruptive public health crises: A review of covid-19 impacts, *Transportation Research Record: Journal of the Transportation Research Board* .
URL: *https://doi.org/10.1177/03611981231160537*

Wilbur, M., Ayman, A., Sivagnanam, A., Ouyang, A., Poon, V., Kabir, R., Vadali, A., Pugliese, P., Freudberg, D., Laszka, A. and Dubey, A. (2023). Impact of covid-19 on public transit accessibility and ridership, *Transportation Research Record* **2677**(4): 531–546.
URL: *https://doi.org/10.1177/03611981231160531*