National College of Ireland

Project Submission Sheet

Student Name:	Shreeraj Santosh Sangle
Student ID:	23283254
Programme:	MSCAI1B **Year:** 2024-25
Module:	Data Analytics for Artificial Intelligence(H9DAI)
Lecturer:	Giovani Estrada
Submission Due Date:	11/12/2024
Project Title:	Annual Bicycle Counts for Key Dublin Locations(2023)
Word Count:	2744

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 references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Shreeraj Santosh Sangle

Date: 11/12/24

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

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

AI Acknowledgement Supplement

[Insert Module Name]

[Insert Title of your assignment]

Your Name/Student Number	Course	Date
Shreeraj Santosh Sangle	MSCAI1B	11/12/24

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
[Insert Description of use]	
[Insert Sample prompt]	[Insert Sample response]

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence:

[Place evidence here]

Additional Evidence:

[Place evidence here]

Annual Bicycle Counts for Key Dublin Locations(2023)

Shreeraj Santosh Sangle (23283254)
Data Analytics for Artificial Intelligence – H9DAI
MSCAI1B
School of Computing
National College of Ireland

1 Background Research

Cycling really is an important means of sustainable urban transport, which in turn yields to a number of benefits that include carbon footprint reduction, health improvements, and reduced traffic congestion. In addition to these, bicycle traffic volume monitoring exposes uses of infrastructure to urban planners and policymakers so that they may warrant better investment decisions for active transport. This article investigates patterns of cycling in Dublin using hourly data counts of cyclists in 2023 collected from automated counters at various locations. Data were acquired from Dublin City Council and from the National Transport Authority (NTA) using state of the art Eco-Multi Counters and Totem counters on Grove Road.

Monitoring location sites include Drumcondra, Charleville Mall, North Strand Road, Guild Street, Richmond Street, and Grove Road, to represent urban, residential, and arterial types of cycling. This information continuous collection creates temporal and spatial trends in cycling where high use times could become useful in effective cycling policy formation. With Dublin extending bike lanes and implementing programs in active transport, this would be one example of an applicable approach to maintain and enhance the cycling infrastructure of the city towards the promotion of sustainable urban mobility.

2 Data Analytics

The data analytics process undertaken for the dataset of hourly cyclist counts across Dublin. Dublin's all-hour cycling counts depict an elaborate process that is the systematic extraction of key transportation insights. Thorough cleaning and treatment of missing values are always preceded by temporal and categorical data standardization. Exploratory data analysis has revealed the most important features of cyclists' behaviors regarding peak hours, seasonal fluctuations, and environmental effects. Advanced statistical and machine learning models identify predictive relations between cycling volumes and external factors such as weather, timing, and urban characteristics. To ensure that the methodology is up to scratch, it encompasses rigorous statistical validations and cross-validation to create trust in the model. Complex data visualizations yield intuitive graphical representations which improve understanding by urban planners and policymakers. It is thus the eventual outcome that gives the effects for improvements of cycling infrastructure, traffic management, or even sustainable urban transport strategies to show the capabilities of transformation through data-driven decision-making in modern city planning.

2.1 Data Exploration and Preparation

Using the pandas.read_csv() method, the dataset is imported, and the exploratory prying of the dataset carries on through various paths like .head(), .tail(), .shape(), and .info(). These methods give information about the size of the dataset, column names, and types of data. It may also prove to be helpful to detect anomalies such as missing values or non-standard formats of data. Most critical and initial activity towards ensuring data quality is to identify missing values using .isnull().sum(). This provides the number of null values in each column, which then is represented as percentage with respect to the whole dataset in order to know how much data is missing.

Most analyses and model performance outcomes are widely impacted by the changes that come from missing data. Here, missing values are generally substituted by zeros using .fillna(0). It thereby ensures data consistency when calculations are performed leaving no opportunities for errors resulting from zeros in completion when aggregating into totals. Missing as such have simplified interpretation as not having cyclists thus making it too simple for some cases. Mean or median or even other more advanced means of estimating the missing values can be appreciated depending on the context of the dataset itself. The duplicate rows, if any, bias the output and inflate the measures. This has been achieved deriving aggregated metrics or trends.

2.2 Data Cleaning and Feature Engineering

Feature engineering takes the analytical power of the dataset further by adding derivatives to input that reveal further insights. Such is the ‘Total_Cyclists’ column, which aggregated counts of cyclists from various monitoring locations into a single, all-encompassing view of urban cycling activity. This all-encompassing view goes beyond those that are sometimes skewed by being location-specific, allowing a better reconstruction of broader cycling trends and patterns for researchers and urban planners.

Transformation of raw data through cleaning, standardization, and feature engineering serves multiple critical purposes. The first is to reduce the computational complexity; the second is to minimize the analysis errors that may occur; and the third is to make visible latent insights, which will probably be obscured in the original datasets. It is through systematic restructuring and enrichment of the data that analysts can develop more carefully nuanced models of dynamics in urban cycling. These preparatory steps for varied outcome transformation are not mere technical processes-formularized methods of analysis that decipher pooled numerical records into semi-definite, publicly usable intelligence for urban transport planning and infrastructure development.

2.3 Handling Outliers

The procedure implemented within my data pre-processing will handle outliers in a more advanced manner by applying the Interquartile Range (IQR) method to numerical columns. This was done by looping through integer columns and floating-point columns, determining the Q1 (or first quartile) and Q3 (or third quartile), and then determining the Interquartile Range for each numerical value. This allowed me to define a statistically driven range for each numerical feature, thus giving a more refined consideration for extreme values. Implementation was carried out in identifying quartile points, then calculating IQR, and defining upper and lower boundaries by a standard formula using 1.5x IQR from quartile points. The standardization was achieved by applying this method to constrain all values within these defined boundaries without discarding extremes. It is a technique that completely preserves the vague content of the dataset while keeping it from distorting the model, thus maintaining the integrity of information by adjusting outliers slightly and geocaching subtle, nuanced information preserved in original data without necessarily breaking altogether.

2.4 Descriptive Statistics and Initial Insights

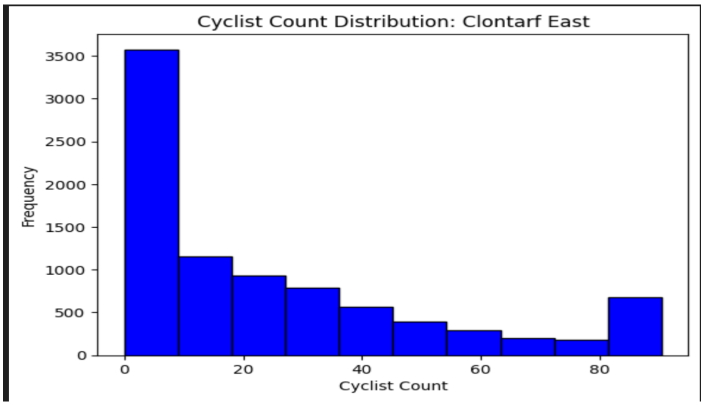
Such is also true for the world of data and descriptive statistical evaluation; the heartbeat of a dataset is a diagnostic tool, and thus, it should be able to employ describe before unfolding the complete caveat of data personality with all its different central tendencies, spread, and underlying rhythms. A clear indication of some true readings being a data detective, revealing good insights in a rapid analysis, shows that the mean and median would give some indication of normal cyclist counts, while the standard deviation would provide information on variation in the data. "Deliver also the minimum and maximum-displayed values as context for the extremes of urban cycling activity coverage." Overall, it probably was that focusing on numeric columns through select_dtypes(include="number") is the equivalent of walking over a street to bring the most relevant instruments in an orchestra.

Then, a collection of histograms and box plots are used to turn abstract numbers into graphics. It tells more than just numbers of cycling resources; as to cycling paths, it has revealed distributions, outliers, and various perceptions that would otherwise be lost in raw numbers. This is a systematic process to convert raw cyclist counts into insightful information for transforming raw data into usable understanding for the city transport department.

2.5 Visualization and Insights

Histogram:

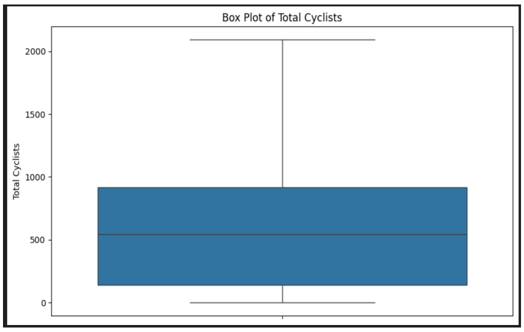
The histogram of Clontarf East translates into a blue tone visual narrative which is divided into ten bins, and overtly reflects the rhythmic pulse of moving cyclists. This vertical bar, clearly calibrated, transforms abstract numerical data into intuitive landscapes of urban mobility through opposing axes. Each one rises and then falls, implying bicycle traffic - very silence ranges from peaks to troughs, hence evincing intricate patterns of transit usually hidden in the raw rows of so-called statistics. Because of the graduated blue tone, it gives an impression of depth and dimension, thereby leading the audience as it were into the very meanderings of difference in cyclist frequency. In a manner like that of a topographical map of human movement, the histogram demarcates the flow of daily transportation while making contours invisible along the Transurban string of cycling. Cold numbers have been transformed into a very visual language talking about the unseen rhythms of metropolitan streets.



Box Plot:

The entire box plot of total cycling becomes a storyteller in modern statistics: a really good summary of the data structure but at the same time quite concise. A box, whiskers, dangling at ends, and possibly outliers bring the cycling activities in Dublin alive. The box within indicates the interquartile range which defines the middle fifty percent of data points; there is a line one sees inside this box that marks the median which indicates central tendency. Whiskers are merely lines of different lengths extended and caught most of the data but excluding the extremes; see how they are. The outliers display as single dots thereby telling different stories, each of which depicts either very high or low days in terms of cyclists' counts.

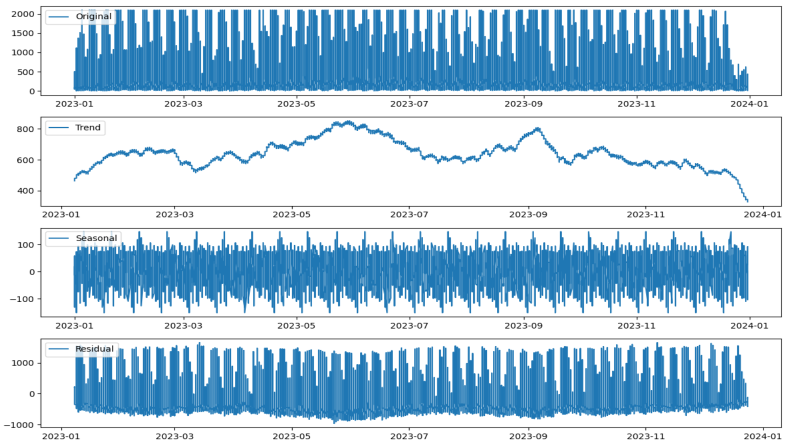
It would not only hold wonderful numbers in the visual gathering but too much more about the pattern and variability associated with cycling activity-those would synthesize trends and behaviours. The box plot is supposed to represent entirely Dublin's cyclists' rhythm- the most required-one for understanding and analysing this urban phenomenon- for describing those rhythms during various times of the day, due to weather impacts, events, infrastructure, etc.



Season Decomposition:

Seasonal decomposition is among the main tool for analysis because it decomposes the different components of time series data as such-the original data, trend, seasonal patterns, and residuals which articulate all of different influences present among them. Here is the method provided for realizing highly complex data like a musicologist going through a symphony. Now that we've got our Time column converted to datetime and indexed, we must, in the addition model sum of all those components representing your data, decompose by the additivity model ("365") for instance, complete to represent the yearly seasonality in daily data and examine this-pattering properly.

The trend component explains the long-term path of cycling activities, and seasonal patterns denote cyclical changes therefore influenced by recurrent factors-including the weather or commuting behaviours. Unexpected changes, thus, accrue as remains of any former variabilities due to sudden fluctuations. This method also clarifies the underlying rhythms of the mobility system, hence allowing urban planners and researchers to predict patterns for formulating sustainable policies on mobility in agglomerations, such as Dublin.

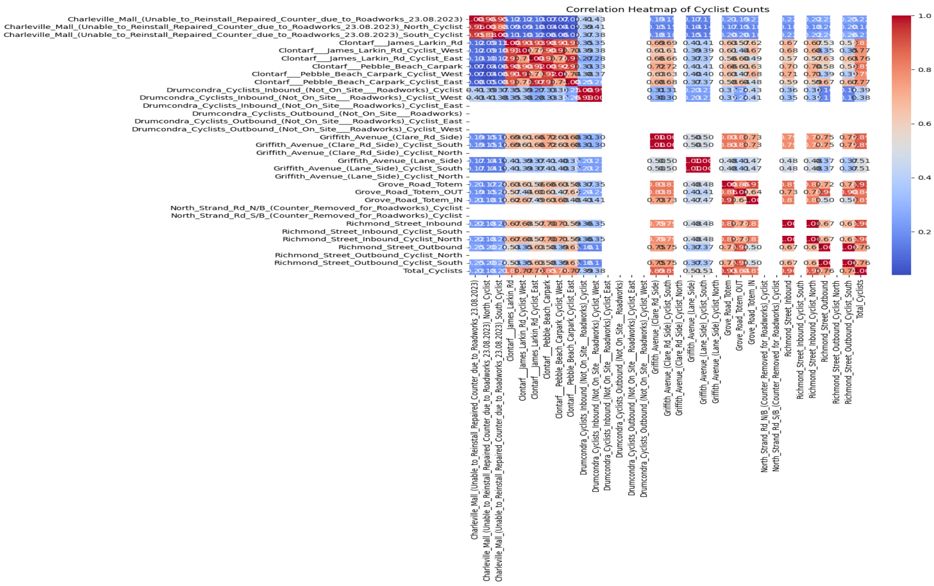


Heat Map:

A "coolwarm" colors scheme is used to unravel the relationship between all numeric variables of this dataset as it serves as the data detective in giving an electric heat map of correlation among cyclist counts. Create numeric columns first using `select_dtypes`, and then create the correlation matrix with `.corr()` to understand the strength and direction of relationships between the variables.

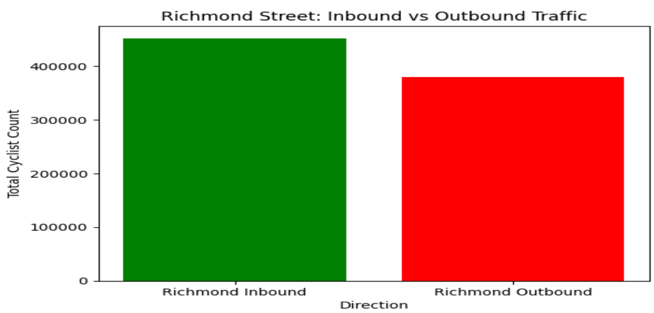
The relationships that will be shown in the heatmap built using `sns.heatmap` will be strong positive correlations in red, negative relationships in blue, and no relationships by white between the included variables in the map itself. Though the annotated values are informative, the "coolwarm" palette gives good contrast for easy inference.

In this intriguing pattern, complex statistical relationships translate into immediately intuitive formats that quickly visualize patterns and interdependencies of the data. It is evident as much as from the heat map that it does a splendid job in unlocking the hitherto hidden insight obstructions making the path to understanding the effects of cycling on influence very clear.



Bar Graph:

The careful bar graph collections of Richmond Street and the analysis by weekday create very wonderful pictures of urban cycling dynamics against the background of each other. The green and red bars depict movements by cyclists-inbound and outbound-on Richmond Street and starkly contrast directions. Summing and then plotting those travel counts on bicycles over the corridor transforms raw data into an intuitive picture of how the average cyclists navigate, which is immensely helpful for planning urban areas.

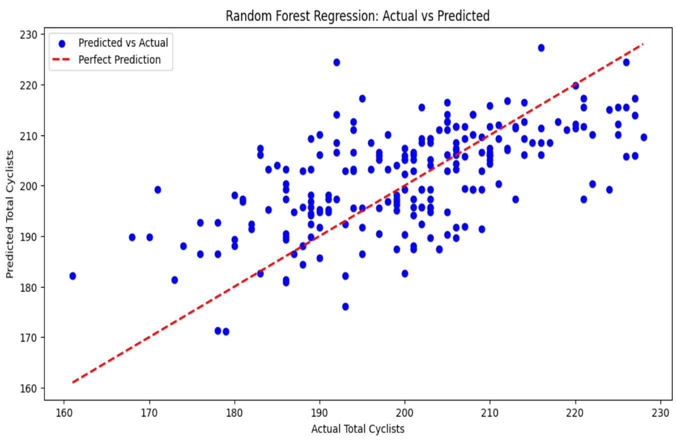


2.6 Advanced Data Analysis and Modeling

This code snippet is for analysis and prediction of time series using an ARIMA model built upon data of cyclist counts. It begins with loading the dataset and checking for duplicate indices, if any, to be removed. The following will be the Augmented Dickey-Fuller (ADF) test on the series termed as 'Total_Cyclists' to check for the necessity of differencing as an outcome inferencing whether the result suggests non-stationarity. In such cases, the input would be de-differenced and undergo stationarity check yet again. The next block generates figures of ACF and PACF to determine model parameters to be fed into the ARIMA model. Accordingly, this collection will apply the defined fitting of ARIMA upon data and output the summary of the model employed. Continue the forecasts for the next twelve periods, then along with actual values, and visualize them. The last is model assessment using RMSE after splitting data into training and testing, fitting the latter with the model, and making it available for prediction against test data.

3 Machine Learning Algorithms

This particular application configures a Random Forest Regressor to measure exactly how many cyclists make use of algorithms reconstructed from the dataset. It first imports all the necessary libraries and reads the dataset along with the target variable and features. The data is then split into a training set, comprising 80 percent of the data, and a testing set. A Random Forest model, 100 estimators strong, was built up and fit to the model. Models from this dataset were predicted on the test set for evaluation within the metrics of Mean Squared Error (MSE) and R^2 Score. Finally, a scatter plot represents the actual and model-predicted values to show the precision of the model.



4 Evaluation and Discussion

In a nutshell, the “Annual Bicycle Counts for Key Dublin Locations” Project represents the best sophisticated exploration of urban mobility through data analytics. It has methodological rigor in transforming raw counts of cyclists into meaningful insights. It combines these techniques with exploratory data analysis, feature engineering, innovative outlier management and predictive modelling to depict an advanced understanding of cycling patterns in Dublin. The very strength of this research lies in the multi-faceted approach it adopts: generating sophisticated tools such as pandas for data preprocessing, IQR method for handling outliers, and machine learning models including Random Forest Regressor and ARIMA time series analysis.

It enabled a deep dive into temporal trends and seasonality that show patterns in mobility dynamics while bridging technical frameworks for data analysis with practical urban planning needs. While valuable evidences have been provided by the research for continued infrastructure development and transportation policies, localization of results and also the need for data validation from time to time could be potential downsides. However, there are evidences which model the transformative potential of high-end analytics into understanding the increasingly complex urban transportation ecosystems and provide a robust framework for future-directed, data-driven mobility research applicable to strategy formation in future metropolitan development.

Multi-dimensions meeting in this study articulate the technical data frameworks with real-time needs in urban planning, thus providing evidence-based validation for infrastructure and transport policy development. The research methodology applied is holistic to study urban transportation ecosystems but recognizes the limitations of this study, such as results being localized and needing periodic validation of data. Thus, the work itself can be a manifestation of transforming high-end analytics into an enlightenment tool for producing data-based knowledge toward the metropolitan development strategy of the future.

5 Conclusion

In conclusion ,“Annual Bicycle Counts for Key Dublin Locations” project is exemplifying the transformative capacity of data for understanding and improving transportation systems within the city. The project translates a collection of raw cyclists' count data into the story behind the patterns that define Dublin's cycling ecosystem. Its bedazzling innovative dash is not only going to broaden understanding about the urban dynamics itself but also provide essential insights that policymaker and planners would need to design and develop sustainable strategies for transport in the city[4]. Finally, the project is an inspiring example of how data-driven decision-making completely reshapes urban transportation, creating a more connected, more efficient city that is also a much more vibrant living space for all its residents.

References

1. *Fishman, E. (2016). Cycling as transport. Transport Reviews, 36(1), 1-18.*
2. *Heinen, E., Maat, K., & Van Wee, B. (2010). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. Transportation*
3. *Research Part D: Transport and Environment, 15(2), 102-109.Pucher, J., & Buehler, R. (2017). Cycling towards a more sustainable transport future. Transport Reviews, 37(6), 689-694.*
4. ***Pucher, J. & Buehler, R.** (2017). Cycling towards a more sustainable transport future. Transport Reviews, 37(6), 689-694.*