

Exploring the Impact of Environmental Factors on Pedestrian Footfall in Dublin City

Alphons Zacharia James

Msc in Data Analytics
National College of Ireland
Dublin, Ireland
x23169702@student.ncirl.ie

Ankith Babu Joseph

Msc in Data Analytics
National College of Ireland
Dublin, Ireland
x23185813@student.ncirl.ie

Abhilash Janardhanan

Msc in Data Analytics
National College of Ireland
Dublin, Ireland
x23121424@student.ncirl.ie

Abstract—This paper presents an analysis of pedestrian footfall data in Dublin city while correlating it with weather conditions and air quality indices to understand and explore the relation between environmental factors and footfall. Using advanced analytics programming techniques and database management systems we process large datasets to extract actionable insights. Our method involves extracting and transforming data within a Docker container, this approach makes sure that our approach are scalable and can be consistently replicated. Our study uses visualizations to identify trends, changes and irregularities in footfall data across different times and locations. Our findings indicates significant correlations between pedestrian traffics and both weather patterns and pollution levels. This study can help with urban planning, public health, and environmental changes. This paper helps us to present a new way to use footfall data in Dublin city to understand human mobility and environmental effects. It shows how it can combine database technology and analytics programming for the society.

Index terms—Weather, Air Quality Index, Pedestrian Footfall, MongoDB, PostgreSQL, Docker, Dagster, ETL-Extract Transform Load

I. INTRODUCTION

In an era where urban spaces are highly monitored for efficiency and safety, understanding the dynamics of human mobility becomes very crucial. Pedestrian footfall data serves as a significant indicator of urban planning and economic activity thus offering insights into how people interact with environment under varying conditions. This research shows the intersection of environmental factors like weather conditions and air quality and pedestrian movement within urban places. Measure footfall is thus important for our project. Dublin city Council has placed Pyro Box Figure 1 at several places to collect footfall data.

Recent advancements in data collection and analytics have enabled more detailed observations of urban environments

therby allowing researchers to analyze patterns. The integration of big data technologies and analytical programming facilitates a deeper understanding of complex datasets, uncovering relationships that can inform public policy and urban planning.



Figure 1: Pyro Box, Henry St, Dublin (Source: Google Maps)

A. Research Objective

The objective of this study is to do a visualization of the impact of environmental variables on pedestrian footfall. By using extensive datasets of weather patterns, air quality indices, and footfall across multiple urban locations, this paper aims to:

- Identify trends in pedestrian traffic relative to environmental changes.
- Evaluate the influence of air quality and weather conditions on urban mobility.
- Provide actionable insights that can assist policymakers and urban planners in designing more sustainable infrastructure.
- Analyze the impact of air quality levels on pedestrian density in different urban locations, highlighting how pollution may affect pedestrian behaviors.
- Map spatial and temporal footfall variations, providing urban planners with detailed insights about pedestrian traffic patterns across different urban places.

The significance of this study lies in its potential to contribute to the sustainable development of urban areas and enhancing the well being of residents and promoting environmentally friendly urban planning practices.

II. RELATED WORK

In this paper we explore the intersection of urban footfall and environmental factors in Dublin city using datadriven approaches to analyse pedestrian traffic patterns. Our work is informed by similar studies in the field, each contributing different insights into how urban environments influence pedestrian behaviors and wellbeing.

One such study conducted by Anita Ratnasari Rakhmatulloh, Diah Intan Kusumo Dewi, and Dinar Mutiara Kusumo Nugraheni in Semarang, Indonesia, focused on how urban design influences pedestrian activity. They employed ArcGIS for spatial mapping and SPSS for data analysis, concluding that pedestrian frequency is higher in areas with diverse attractions and well-planned spaces. These findings align with our investigation into Dublin's urban layout, reinforcing the importance of strategic urban planning in enhancing pedestrian experiences and environmental quality.[1]

Similarly a study by Babatunde Olasunkanmi Folasayo and Abimbola A. Babatunde examined the impact of environmental pollution on pedestrians in Lagos, Nigeria. They utilized the Thermo Scientific MIE pDR-1500 instrument to measure air quality across 20 local government areas, finding that six areas exceeded acceptable pollution standards. The results, substantiated by a one-sample T-test with a t-value of 22.226, underscore the significant effect of air pollution on pedestrian health and underscore the need for urgent governmental interventions like restricted vehicle hours and enhanced urban greenery.[2]

III. METHODOLOGY

In our project we use a method called Knowledge Discovery in Databases (KDD) first described by Fayyad et al. in 1996 [3]. This approach outlines a clear set of steps for turning raw data into useful information. Our study follows this method closely by using a well organized process that starts with collecting data and ends with extracting valuable insights. The updated figure Figure 2 in our paper shows these steps thus ensuring our research is useful for improving and understanding environment's effect on pedestrian footfall.

A. Data Selection:

The initial stage of our project involves selecting appropriate datasets. This process requires consideration of various factors like data quality and relevance. We use historical pedestrian footfall data which is collected from SmartDublin in CSV format along with historical weather and air quality data, which we gather through APIs from OpenMeteo [4]–[7]. This two approach allows us to combine different types of data thereby giving us a clear picture of historical trends. This method make sure that we have a rich dataset that captures both the human

activity in urban spaces and the environmental conditions they experience.

B. Storage and Preprocessing/Transformation:

Once we collect the data we store it in MongoDB, which is a type of NoSQL database known for its ability to efficiently manage large amounts of unstructured or semi-structured data. We then use Python scripts to prepare the data for analysis. This preparation involves cleaning the data to remove any irrelevant information by dropping columns where the percentage of null values or zero values in more than eighty, for others null values are replaced by imputing 0 to the respective columns. These null values or zero values maybe due to the improper working of the sensors. This step is crucial because it makes sure that the data is accurate and organized thereby setting the stage for reliable analysis and insights.

C. Storage and Visualizations:

After preprocessing we transfer the cleaned and structured data into a PostgreSQL database which is an SQL database known for its strong capabilities in data warehousing and handling complex queries. This is a critical step towards deeper data analysis allowing us to perform analysis and create detailed visualizations. This transition is not just a technical step it's about moving closer to our goal of understanding and visualizing complex patterns within the data. By using PostgreSQL we can dig deeper into the data and uncover insights that can inform decisions. This enables us to turn raw data into meaningful visual stories that can illustrate trends, challenges, and opportunities in urban environments.

Parameters	Description (Unit)
id	Unique identifier for each data entry
date	Date and time of the recorded data.
temperature_2m(°C)	Temperature at 2 meters above ground level in Celsius.
relative_humidity_2m(%)	Relative humidity at 2 meters above ground level, expressed as a percentage
dew_point_2m(°C)	Dew point temperature at 2 meters above ground level in Celsius.
apparent_temperature(°C)	Perceived temperature, taking into account factors like humidity and wind, in Celsius
precipitation(mm)	Total precipitation in millimeters
rain(mm)	Amount of rainfall in millimeters
snowfall(cm)	Amount of snowfall in centimeter
cloud_cover(%)	Percentage of sky covered by clouds
wind_speed_10m(km/h)	Wind speed at 10 meters above ground level in kilometers per hour

sunshine_duration(Seconds):	Duration of sunshine in seconds
-----------------------------	---------------------------------

Parameters	Description (Unit)
pm10($\mu\text{g}/\text{m}^3$)	Particulate Matter (PM10) concentration in micrograms per cubic meter.
pm2_5($\mu\text{g}/\text{m}^3$)	Particulate Matter (PM2.5) concentration in micrograms per cubic meter
carbon_monoxide($\mu\text{g}/\text{m}^3$)	Carbon Monoxide (CO) concentration in micrograms per cubic meter
nitrogen_dioxide($\mu\text{g}/\text{m}^3$)	Nitrogen Dioxide (NO2) concentration in micrograms per cubic meter
sulphur_dioxide($\mu\text{g}/\text{m}^3$)	Sulphur Dioxide (SO2) concentration in micrograms per cubic meter
dust($\mu\text{g}/\text{m}^3$)	Dust concentration in micrograms per cubic meter.
european_aqi	European Air Quality Index (AQI) calculated based on various pollutant concentrations.
european_aqi_pm2_5	European AQI specifically calculated for PM2.5.
european_aqi_pm10	European AQI specifically calculated for PM10
european_aqi_nitrogen_dioxide	European AQI specifically calculated for nitrogen dioxide (NO2)
european_aqi_ozone	European AQI specifically calculated for ozone (O3).
european_aqi_sulphur_dioxide	European AQI specifically calculated for sulphur dioxide (SO2)

Num	Locations
1	Aston Quay/Fitzgeralds
2	Baggot st lower/Wilton tce inbound
3	Baggot st upper/Mespil rd/Bank
4	Capel st/Mary street
5	College Green/Bank Of Ireland
6	College st/Westmoreland st
7	D'olier st/Burgh Quay
8	Dame Street/Londis
9	Grafton st/Monsoon
10	Grafton Street / Nassau Street / Suffolk Street
11	Grafton Street/CompuB
12	Grand Canal st upp/Clanwilliam place

13	Grand Canal st upp/Clanwilliam place/Google
14	Mary st/Jervis st
15	Phibsborough Rd/Enniskerry Road
16	North Wall Quay/Samuel Beckett bridge West
17	O'Connell st/Princes st North
18	North Wall Quay/Samuel Beckett bridge East
19	Richmond st south/Portabello Harbour inbound
20	Richmond st south/Portabello Harbour outbound

D. Knowledge:

The final goal of our process is to extract knowledge that can be used in decision making. We get this by converting complex data sets into clear, easy to understand visualizations. These graphs not only make it easier to spot patterns and insights thereby making informed decisions. These visualizations are explained in the next section “Data Visualization”

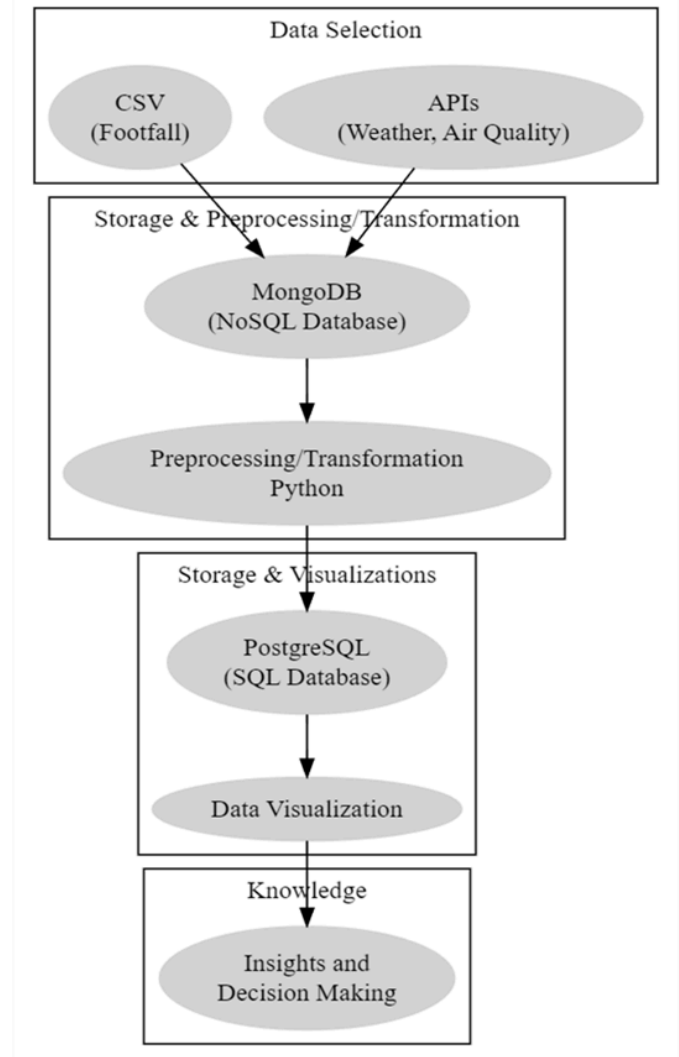


Figure 2: KDD Lifecycle

The way we use the Knowledge Discovery in Databases (KDD) lifecycle is iterative which means we continuously refine our methods of analysis. Each step of this process is carefully detailed in Figure 2. This approach is about more than just processing data infact every step in our process builds upon the previous one.

E. Technologies:

In our database and analytical programming project a variety of technologies are used to simplify data processing and analysis. We organised our coding efforts to maximize efficiency and maintainability using python's adaptability and strength. In python we incooperated several libraries like Pandas, dagster, Pymongo, sqlalchemy, bokeh, Panel etc for scripting the analysis. The central orchestration tool for our workflow is Dagster which seamlessly facilitates the Extract, Transform, Load (ETL) process. The dagster guarantees the orchestrated flow of data throughout the various stages. MongoDB a robust NoSQL database is used to store the raw data that are extracted from API calls and csv file read. We used MongoDB over other because of its document-oriented approach which simplifies data representation thereby reducing development complexity. PostgreSQL which is reliable relational database known for its superior querying capabilities and structured data management is to store the data after wrangling and merging processes which is used for further analysis. PostgreSQL is our first preference as it ensures robust ACID compliance, data integrity and reliability even in complex transactional scenarios. We integrated mongo-express and PG admin within the environment for visually interacting with databases. We wrapped our whole setup in Docker containers to ensure portability and uniformity across several computer settings. By merging these tools, we were able to establish a robust database management and analytical programming ecosystem capable of providing effective data driven insights.

IV. DATA VISUALIZATION

After loading the data successfully into PostgreSQL we have a created a dashboard using panel library in python for visualizing and bokeh for plotting graphs. We have also incooperated a jupyter notebook for our analysis and vizualisation. This visualization strategy simplifies our analysis and therby making it more interactive by allowing users to engage directly with the data through dynamic visual tools. This interactive method of data visualization clearly shows how urban environments affect people's movement.

A. Dashboard Features:

The introduction page shows the environmental variables and final counter locations. Dataset page which shows the com-

bined sample dataset of weather, air quality and footfall. Relationship between variable shows a scatter plot with all the environmental variables and counter locations. The distribution of variables pages shows a line graph with all the environmental variables and counter locations. The project report page has our project report displayed.

B. Pedestrian Traffic Distribution Using Bar Chart:

From the footfall dataset analysis we created a bar chart to visualise the average footfall at various locations across Dublin city thus revealing key pedestrian hotspots. This bar chart titled "Average Footfall by Location" Figure 3 clearly shows which areas experience the highest pedestrian traffic with locations like "Aston Quay/Fitzgeralds" and "Baggot st lower/Wilton tce inbound" leading in footfall. These insights can be crucial for urban planners and government authorities as they can design strategies to enhance urban mobility and optimize city spaces for better pedestrian flow and can also is a key metric to optimise your business performance in that locations. The visualization help to identify and address areas of high pedestrian activity thus helping in decision making processes related to urban development and infrastructure planning.

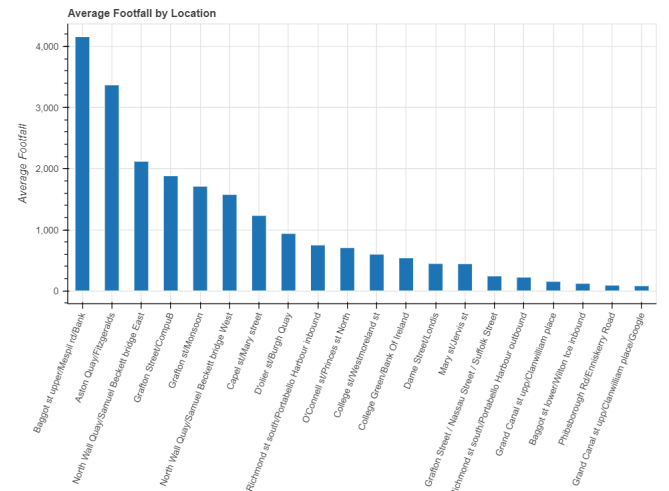


Figure 3: Average Footfall by Location

C. Environmental Impact On Footfall Using Scatter Plot:

The scatter plot "Temperature vs Pedestrian Traffic at College Green/Bank Of Ireland" Figure 4 shows a relatively dense clustering of points in the midrange of temperatures therby suggesting a potential correlation where footfall increases with moderate temperatures. The distribution is wide indicating fluctuations in footfall which could be due to other factors like time of day or specific events.

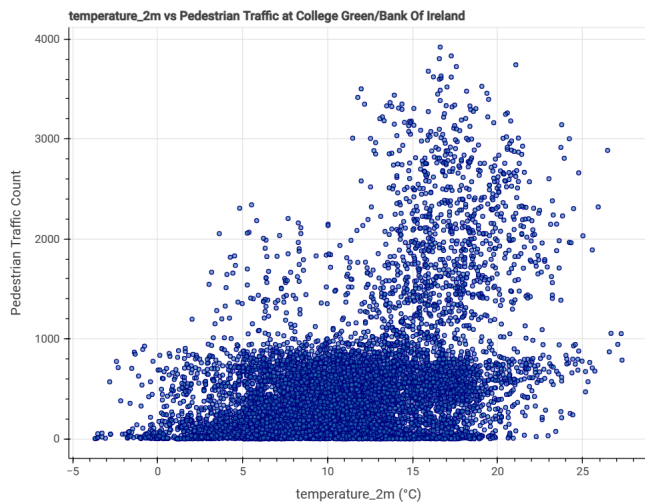


Figure 4: temperature_2m vs Pedestrian Traffic at College Green/Bank Of Ireland

The scatter plot “Temperature vs Pedestrian Traffic at Baggot St Upper/Mespil Rd/Bank” Figure 5 presents a very dense data points in the mid to higher temperature ranges which could suggest higher pedestrian activity in warmer conditions. This plot can serve as an excellent example to discuss the influence of pleasant weather on human mobility patterns.

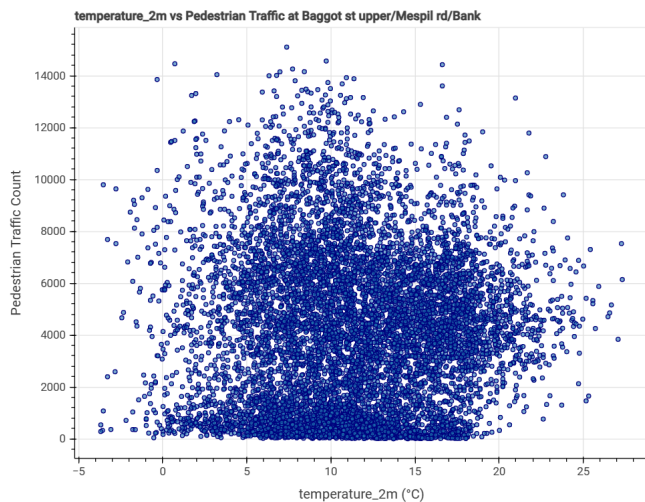


Figure 5: temperature_2m vs Pedestrian Traffic at Baggot st upper/Mespil rd/Bank

These scatter plots are excellent representations of how temperature variations influence pedestrian movements in urban areas. The College Green/Bank of Ireland plot for example could indicate that more people are likely to walk in moderate temperatures which aligns with comfortable walking conditions. On the other hand, Baggot St Upper/Mespil Rd/Bank plot shows higher footfall during warmer temperatures thus possibly indicating a preference for outdoor activities or walking during warmer days.

In addition to the influence of weather on pedestrian traffic as shown in the scatter plot for Baggot Street Upper and Mespil Road near the bank, the area’s unique blend of commercial activities and historical significance might contribute to its footfall patterns. This location is known for its historic Georgian architecture and can attract both locals and visitors especially in favorable weather conditions.

The scatter plot for “European AQI PM2.5 vs Pedestrian Traffic at O’Connell St/Princes St North” Figure 6 shows a visible pattern that there is a dense clustering of data points at lower AQI levels and then it begins to spread out as AQI values increase. This suggests that more people like to walk in the area when the air quality is good and there are fewer people on road as pollution worsens. This pattern shows the how air quality impact peoples decision to participate in outdoor activities.

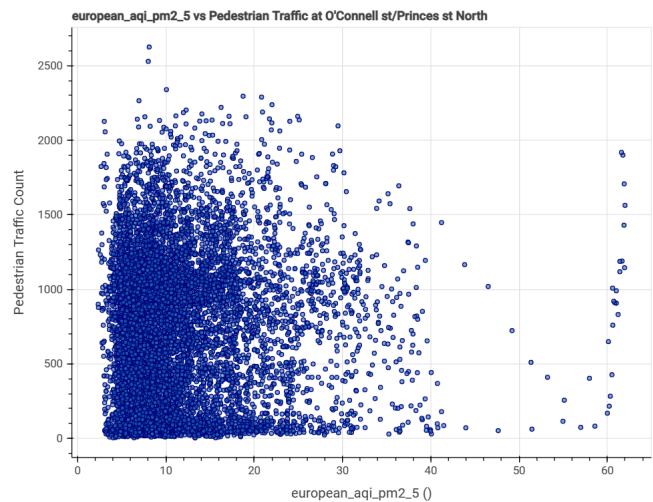


Figure 6: European AQI PM2.5 vs Pedestrian Traffic at O’Connell St/Princes St North

Similarly the scatter plot for “European AQI PM2.5 vs Pedestrian Traffic at College Green/Bank of Ireland” Figure 7 shows initially high density of data points at moderate levels of air pollution which then decreases as the AQI increases beyond a certain point that cant be bearable. This indicates that pedestrian traffic remains stable up to a certain level of air pollution but as soon the air quality decreases the pedestrian traffic also decreases. These trends are again very important for city planning and public health improvements.

D. Environmental Impact On Footfall Using Line Graph:

The graph showing the distribution of carbon monoxide and pedestrian traffic at Aston Quay/Fitzgeralds Figure 8 shows a interesting correlation, as carbon monoxide levels increase throughout the year so does the pedestrian count mainly peaking around September. This might suggests that if more pedestrians increase then vehicle activity might also increase in that area. This might be because both are interrealated and as vehicle increases carbon monoxide emissions might also increase.

Although this is just a hypothesis based on the observed data and generic understanding, it reflects the complex interaction between urban activity and environmental quality. Despite the decline in both carbon monoxide and pedestrian numbers towards the year's end—likely due to seasonal weather changes and reduced outdoor activities the area still remains a busy hub and that's possibly sustained by its business or residential appeal. This interesting relationship can be used for further investigation to fully understand these dynamics.

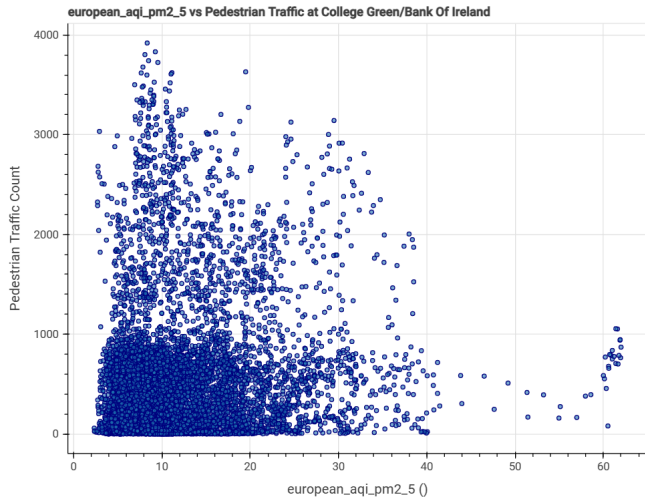


Figure 7: European AQI PM2.5 vs Pedestrian Traffic at College Green/Bank of Ireland

Similarly second line graph shows how sunshine duration and pedestrian traffic correlate at College Green/Bank of Ireland. This graph shows a clear pattern as the sunshine increases and reaches its highest in July more people are out on the road which suggests that sunny days encourage outdoor activities in that area. As the year progresses and the amount of sunshine decreases and so does the number of pedestrians. This highlights how good weather can positively affects the amount of pedestrian traffic outside.

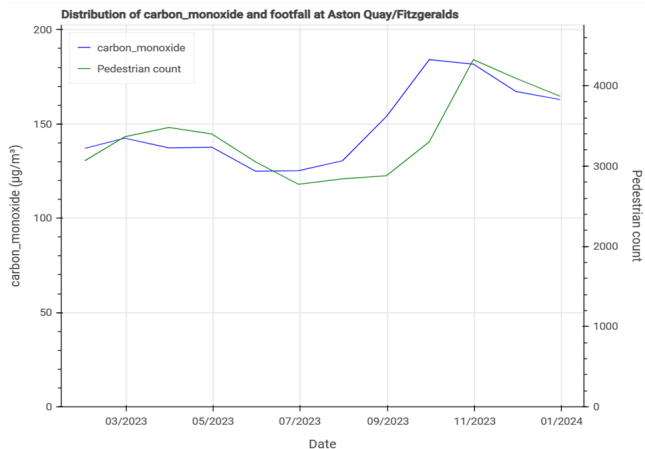


Figure 8: Distribution of carbon_monoxide and footfall at Aston Quay/Fitzgeralds

Similarly second line graph shows how sunshine duration and pedestrian traffic correlate at College Green/Bank of Ireland. This graph shows a clear pattern as the sunshine increases and reaches its highest in July more people are out on the road which suggests that sunny days encourage outdoor activities in that area. As the year progresses and the amount of sunshine decreases and so does the number of pedestrians. This highlights how good weather can positively affects the amount of pedestrian traffic outside.

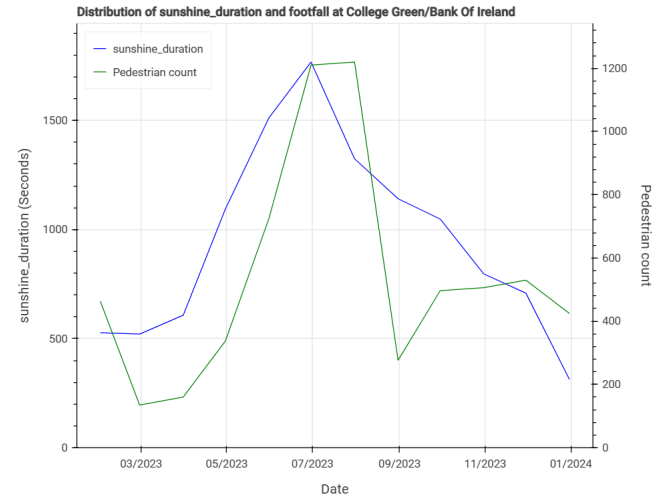


Figure 9: Distribution of sunshine_duration and College Green/Bank Of Ireland

V. CONCLUSION AND FUTURE WORK

Our analysis and visualization processes shows that the significant ways in which environmental factors influence pedestrian footfall, though it is also evident that other variables contribute to these dynamics. Like certain locations within our dataset had missing or incomplete data, which slightly made it difficult for our visual representations. There is always a scope for better analysis and visualization with more accurate data. Our work is just a starting point and more detailed research can be done given a larger set of accurate and detailed data. Such improvements could greatly benefit urban planning and public health, promoting a deeper understanding of how environmental conditions impact pedestrian behavior.

REFERENCES

- [1] A. R. Rakhmatulloh, D. I. K. Dewi, and D. M. K. Nugraheni, "The Built Environment and its Impact on Transit based Transportation Users Walking Activity in Semarang, Indonesia," *pertanika journal of science and technology*, 2021, doi: 10.47836/pjst.29.2.05.
- [2] A. Onifade, A. O. Onifade, B. Folasayo, B. Folasayo, A. Babatunde, and A. Babatunde, "ENVIRONMENTAL POLLUTION: EFFECT ON PEDESTRIANS WHILE WALKING IN URBAN STREET," *International Journal of Environmental Science*, 2022, doi: 10.47604/ijes.1436.

- [3] U. M. Fayyad, U. M. Fayyad, G. Piatetsky-Shapiro, G. Piatetsky-Shapiro, P. Smyth, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *Ai Magazine*, 1996, doi: 10.1609/aimag.v17i3.1230.
- [4] P. Zippenfenig, "Open-Meteo.com Weather API." [Online]. Available: <https://doi.org/10.5281/ZENODO.7970649>
- [5] H. Hersbach *et al.*, "ERA5 hourly data on single levels from 1940 to present." ECMWF, 2023. doi: 10.24381/cds.adbb2d47.
- [6] J. Muñoz Sabater, "ERA5-Land hourly data from 2001 to present." ECMWF, 2019. doi: 10.24381/CDS.E2161BAC.
- [7] S. Schimanke *et al.*, "CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present." ECMWF, 2021. doi: 10.24381/CDS.622A565A.