Safe Routing for Pedestrians

Ross Bennfors

Student ID: 710035913

Abstract

This project investigates whether safer walking routes can be generated by integrating pedestrian infrastructure and crime data into route planning. Using OpenStreetMap data and crime reports from Police.uk, a custom pedestrian routing system was developed. Pedestrian infrastructure features, such as lighting, CCTV coverage, road types, and crossings, were extracted and weighted based on their perceived contribution to safety. Crime hotspots were spatially analysed using an Inverse Distance Weighted (IDW) approach. Both datasets were normalised and combined into a routing weight assigned to each path segment, enabling traditional shortest path algorithms to prioritise safety over distance.

The system was tested across 200 randomly generated origin-destination pairs in Exeter. Statistical analysis showed that safe routes consistently achieved lower risk scores compared to shortest routes, while typically increasing travel distance by less than 30%. Individual route analysis confirmed that the model favoured better-lit, lower-risk streets even when this required modest detours. Despite limitations such as incomplete OpenStreetMap data and limited crime data history, this project demonstrates the shortest route is rarely the safest. Future improvements, such as temporal crime modelling and mobile application development, could be made to improve the system's real-world usability. This work highlights the potential for data-driven routing to improve urban walkability and pedestrian safety.

Declaration

I acknowledge the following uses of GenAI tools in this assessment:

☒ I have used GenAI tools to:

- ☐ develop ideas.
- ☒ assist with research or gathering information.
- ☒ help me understand key theories and concepts.
- ☒ identify trends and themes as part of my data analysis.
- ☐ suggest a plan or structure for my assessment.
- ☒ give me feedback on a draft.
- ☒ generate images, figures or diagrams.
- ☐ proofread and correct grammar or spelling errors.
- ☐ generate citations or references.
- ☐ Other: [please specify]

☐ I have not used any GenAI tools in preparing this assessment.

I declare that I have referenced all use of GenAI outputs within my assessment in line with the University referencing guidelines.

I certify that all material in this dissertation which is not my own has been identified.

Signature: ______________________________

Contents

1 Introduction

1.1 Project Motivation

Urbanisation is happening everywhere and ensuring that these urban areas are safe and walkable is crucial. There are many ways planners and developers can enhance safety when developing. Urban planners and developers can implement safety features when designing cities, but the challenge for pedestrians is still there: how will a pedestrian know what areas are statistically the safest and know how to navigate them? Traditional navigation apps only provide the shortest or fastest route meaning they neglect important factors for pedestrian safety; such as lighting, high crime areas, or fast roads.

This presents a real threat for groups more vulnerable to street-level risks, including women, tourists, and late-night commuters, many of whom report changing their routes or travel times to avoid potential danger. A 2014 study by Beth Livingston found that 85% of surveyed women in the U.S. had altered their walking route to avoid harassment, and 67% adjusted the time they travelled for the same reason[1]. Similarly, Loukaitou-Sideris said in a paper that women frequently adjust their travel due to safety concerns by avoiding certain routes, changing their time of travel or taking alternate modes of transport[2]. These findings highlight the psychological burden and real safety concerns many pedestrians face when navigating urban environments.

Another study by a group of scientists at Johns Hopkins Bloomberg School of Public Health highlighted how the lack of infrastructure can contribute to pedestrian harm. They found that poor signage, lack of crossings, and inadequate pedestrian planning were key factors in pedestrian injuries across urban areas[3]. The study heavily advocates for better measures to be taken to ensure safety, such as more informative tools, better education, and engineering. This is supported with findings by the World Health Organization, which identified poor infrastructure as a leading cause of pedestrian injuries worldwide[4]. These findings support the development of a navigation system, that help users avoid unsafe walking conditions by leveraging crime and infrastructure data.

By leveraging open data sources such as OpenStreetMap and crime statistics from Police.uk, the tool developed in this project will be able to generate safer route alternatives that consider both pedestrian infrastructure and real-world risk factors.

1.2 Aims & Objectives

Research has shown that pedestrians often prioritise walkability factors such as infrastructure quality, amenities and perceived safety, over distance when choosing their route[5]. This highlights the need for a routing system that accounts for more than just distance.

The aim of this project is show the value a data-driven routing tool could give pedestrians, particularly women and tourists, to navigate cities areas more safely. The system combines two key datasets: pedestrian infrastructure from OpenStreetMap (OSM) and crime data from the Police.uk API. Together, these data sources support the development of a routing system that accounts for key safety-related features such as street lighting, CCTV coverage, road speed limits, pedestrian crossings, and proximity to crime hotspots.

Each segment of road or path will be scored based on the cumulative presence of these features, creating a pedestrian safety weighting system. Crimes will be spatially analysed using an Inverse Distance Weighted (IDW) approach to identify nearby crime risk levels. These scores will then be normalised and integrated into a custom routing algorithm designed to prioritise safety over distance.

Once implemented, the tool will generate route comparisons between the safest and shortest paths for randomly selected origin-destination pairs. A key goal for this project is to show that the shortest route between two walkable points in urban spaces, is usually not the safest. By analysing differences in distance and safety scores, the project will evaluate whether safer routes can be offered with minimal compromise to original journey length.

1.3 Background / Related Work

Routing with the goal to improve safety has been done before in a study by Galbrun, Pelechrinis, and Terzi where they explored safe urban navigation by incorporating crime data into routing algorithms. Their system leveraged OpenStreetMap data to model urban networks and applied Gaussian Kernel Density Estimation (KDE) to assign risk scores to road segments, using publicly available crime data, with their chosen cities being Chicago and Philadelphia[6]. They treated the task as a bi-objective optimisation problem, balancing safety and distance, and applied classic algorithms like Dijkstra's alongside recursive alternatives to generate a range of route options with different safety-efficiency trade-offs.

Although this study demonstrated that crime data could be used to highlight high-risk areas and build efficient navigation tools, it also had limitations. Its sole reliance on crime data meant it was affected more by data quality, particularly with offences like harassment or minor assaults that often go unrecorded. Additionally, the model treated all crimes with equal weight, which gives no distinction between the severity of each crime. Most notably, the system did not consider other spatial features that influence pedestrian safety; such as road lighting, pedestrian crossings, or street usage.

This project takes inspiration from Galbrun's methodology by using pedestrian infrastructure, from OpenStreetMap, alongside crime data . By doing so, it addresses key gaps in prior work, providing a more all round, safety focused routing algorithm for real-world use.

Another advanced study in this space was Sharon Levy's development of SafeRoute[7], which applied deep reinforced learning to learn safe pedestrian routes in rough urban areas. This system trained a model on city maps from OSM alongside crime data, allowing it to make route choices based off distance and risk. The system trained itself to avoid areas near crime by giving lower rewards for unsafe paths and higher rewards for shorter, safer ones. It then uses what it has learned to suggest routes. While this approach showed strong performance, it required substantial model training and tuning and results were not always easily interpretable.

On the other hand, this project aims to use a more lightweight solution using available crime and infrastructure data to assign roads weights. This allows for easier integration into conventional routing algorithms like Dijkstra's, allowing it to be more easily scaleable and adaptable to different datasets, removing the need of computationally expensive and complex neural models.

2 Methodology

2.1 Datasets and Preprocessing

This project integrates two main open-source datasets to model pedestrian safety across urban environments: OpenStreetMap (OSM) and crime records from the Police.uk API. Together, they provide a spatial and statistical foundation to evaluate safety across the road network.

OpenStreetMap
OpenStreetMap provides detailed geospatial data contributed by a large user community. For this project, OSM was used to extract pedestrian-relevant infrastructure including:

- Road types and footpaths (footway, pedestrian, residential)
- Speed limits
- Public lighting data
- CCTV coverage
- Nearby shelter or cover (e.g. shops)

The data was retrieved using the OSMnx Python library, which extracts a graph of features from OSM. These are separated into two GeoDataFrames:

- Nodes: representing intersections or points of infrastructure (e.g., CCTV or lighting)
- Edges: representing street/path segments between nodes

Police.uk Crime Data
Crime data was gathered via the Police.uk API, which allows access to crime reports at the street level. This data was filtered to include only crimes relevant to pedestrian safety, such as:

- Anti-social behaviour
- Robbery
- Violent crime
- Theft from the person
- Possession of weapons
- Public order offences

Each crime record includes latitude, longitude, and category information. These were transformed into spatial points for geospatial analysis.

Preprocessing
Before weight calculations and routing could be performed, both the crime and pedestrian datasets underwent preprocessing steps to ensure they were complete and accurate to the situation at hand. Crime data retrieved from the Police.uk API was first filtered to exclude irrelevant entries and narrowed to categories most relevant to pedestrian safety, such as violent crime, robbery, and anti-social behaviour. OpenStreetMap data did not always have all feature annotations (e.g. lighting or CCTV coverage) directly associated with path segments. To address this, spatial joins were performed using geometric buffers, assigning features like street lamps or surveillance points to the nearest pedestrian edge within a 50-meter threshold. Finally, all spatial datasets were projected to the same coordinate reference system (CRS) to maintain consistency in distance calculations (EPSG:4326). These preprocessing steps formed an important stage in the project to ensure that data was complete and consistent before proceeding with model calculations and construction.

2.2 Technical Methods

The pedestrian feature and crime weights used in this project were manually defined based on personal judgement and understanding of pedestrian safety. Although they were not derived from empirical studies, the values were chosen to reflect the relative importance of each factor. These values could be refined further through user testing or informed by future research.

2.2.1 Pedestrian Weight Calculations

Each edge in the pedestrian network was assigned a safety score calculated from a set of corresponding pedestrian infrastructure features collected in the preprocessing step. These features included attributes such as `highway=footway`, `lit=yes`, `highway=crossing`, `maxspeed`, and proximity to `shop` or `man_made=cctv`. Each feature was assigned a weight based on its assumed contribution to pedestrian safety and later summed to give a total pedestrian safety weighting. For instance, a lit footpath with CCTV coverage and low-speed traffic received a higher score than an unlit, high-speed road segment. The final pedestrian score for each edge was stored in a `pedestrian_weight` column in the edges GeoDataFrame and was decided by summing all relevant edge attributes.

These weights were partially influenced by existing evidence that features such as street lighting has a positive effect on crime and public safety. A review by Welsh and Farrington found that improved street lighting significantly reduces crime rates in public spaces[8], reinforcing the choice of prioritising well lit routes in routing decisions. Furthermore, Ewing and Handy showed that pedestrian favour areas that are pleasant and have good pedestrian infrastructure[9], supporting the positive weightings of features such as crossings and pedestrian only footpaths.

2.2.2 Crime Weight Calculations

After crime data from Police.uk was filtered to include only categories directly relevant to pedestrians (e.g., robbery, violent crime, anti-social behaviour), each crime category was assigned a weight to reflect its severity, with violent or weapon-related incidents scored higher than public order.

A custom Inverse Distance Weighted (IDW) scoring function was developed to calculate the influence of each crime on nearby areas. This function considers the weight of each crime (w_i), its distance (d_i) from the evaluated point, and a power parameter (p=2) to reduce the influence of distant events. The function was used to compute a score for every node in the graph by taking a weighted average of nearby crime points within a specified radius (100m), with closer crimes contributing more heavily. crime happened to be exactly on the node (d_i=0) then the formula would return the weight of the crime in order to avoid division by 0.

$$\text{IDW}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} \frac{w_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} & \text{if } d_i > 0 \text{ for all } i \\ w_i & \text{if } d_i = 0 \text{ for any } i \end{cases}$$

Where:

x is the location for which the IDW score is calculated
w_i is the weight of crime point i
d_i is the distance from x to crime point i
p is the power parameter
n is the number of crime points within the distance threshold

Each edge's crime score was then calculated as the average of the IDW scores of its terminal nodes, providing an indication of how risky that edge is based on nearby crime data.

*The pedestrian feature and crime weights used in this project were manually defined based on personal judgement and understanding of pedestrian safety. Although they were not derived from empirical studies, the values were chosen to reflect the relative importance of each factor. These values could be refined further through user testing or informed by future research.

2.2.3 Weight Normalisation and Merging

After computing separate crime and pedestrian safety scores for each edge, these two metrics needed to be combined into a single unified safety score to support routing decisions. However, since the two score types were on different scales (crime scores ranged from 0–3, and pedestrian weights from 0–6), normalisation was required to align them on a common [0, 1] scale. Additionally, the two metrics were directionally opposite: higher pedestrian scores indicated increased safety, while higher crime scores implied increased danger.

As a result, crime scores were first inverted and normalised, such that a high crime score resulted in a lower normalised value. Let c_i be the crime score of edge i, computed from IDW interpolation, with values in the range $[0, 3]$. This score is inverted and normalised:

$$c_i^{\text{norm}} = \frac{3 - c_i}{3}$$

Let p_i be the raw pedestrian infrastructure score of edge i, ranging from $[0, 6]$. We normalised this value by:

$$p_i^{\text{norm}} = \frac{p_i}{6}$$

These normalised scores were then combined into a single safety score, S_i using a weighted average, allowing the relative influence of crime vs. pedestrian features to be adjusted. For this project, equal weighting was used (w_c =0.5, w_p =0.5), although this could be tuned depending on user preference:

$$S_i = w_c \cdot c_i^{\text{norm}} + w_p \cdot p_i^{\text{norm}}$$

To be usable as a cost in shortest path algorithms (e.g. Dijkstra's), where lower values are preferable, this score was inverted to create the final routing weight for edge i is defined as:

$$w_i = 1 - S_i$$

This routing weight was added as a column to the final edges GeoDataFrame which enabled the algorithm to prefer safer routes (with lower weights), while still maintaining compatibility with traditional pathfinding logic.

2.3 Routing Methods

Once safety scores were assigned to edges, the final road network graph was reconstructed using the `graph_from_gdfs` function from the OSMnx library. This was done using the new, updated edges dataframe that has the new `routing_weight` attribute which will be used to find the safest paths.

To find a start and end point in our graph, the chosen coordinates were used to identify the closest corresponding nodes within the graph using OSMnx's `nearest_nodes` function.

With a graph complete, using NetworkX's `shortest_path` function two different routes could be considered:

- **Shortest Route:** Calculated based on edge length (`length` attribute in meters), representing the path with the minimal physical distance.
- **Safest Route:** Calculated using the inverse of the normalised safety score (`routing_weight`). Since higher safety scores indicate safer paths, inverting this value allows the algorithm to treat safer paths as lower-cost options in the graph, thereby preferring them during traversal.

This setup enabled the comparison between traditional shortest routes and their safer alternatives, all computed on the same graph structure. It also made it possible to evaluate route quality through metrics such as total length and safety score, and large scale testing and evaluation across randomly selected node pairs.

Additionally, the routing calculations using NetworkX's `shortest_path` function were consistently fast, typically completing in under a second. This makes the system well-suited for real-time use in future mobile or web-based applications.

3 Results

3.1 Random Creation of Routes

To be able to conduct analysis of the improved safety of routes and to compare performance over a range of scenarios, 200 random origin–destination (OD) node pairs were generated within the study area (Exeter). By doing this, the aim was to simulate a range of realistic walking journeys, while covering different neighbourhoods, road types and distances.

Each node was selected from the graphs full set of nodes, and then checked and filtered to make sure the safe route of the nodes fell in between a distance threshold of 350 to 2000 meters. This constraint was implemented to exclude unreasonably short or long walking routes and instead focus on realistic pedestrian journeys.

With a selection of random OD node pairs identified there were two more key steps. First, for each node pair both a shortest route (optimised for distance) and a safest route (using the custom `routing_weight`) were computed. Routes without valid connections (due to graph disconnection or routing failure) were discarded. Then, each route was then passed through a custom metric extraction function, which calculated the total path length, cumulative safety weight, and a derived safety score (defined as safety risk per kilometre). These routes statistics were stored in a dictionary alongside other route data including route type and the origin and destination node ID's. A `route_id` field was assigned to link corresponding shortest and safest routes from the same node pair for easy identification.

The resulting dataset was compiled into a Pandas DataFrame, forming the basis for the subsequent statistical analysis.

3.2 Statistical Comparisons

From the dataset mentioned in Section 3.1, a number of different graphs were plotted to help visualise the comparison of short and safe routes.

First is the Route Classification Comparison chart (Figure 1), which categorises routes into four classes: Short and Risky, Long and Safe, Short and Safe, and Long and Risky. With 'safe' having a safety score below 10 and 'short' being a route under 1km. The number of shortest routes heavily outweighed safe routes in Short and Risky or Long and Risky, highlighting a clear safety compromise in many cases. In contrast, the safe routes overwhelmingly fell into the 'Long and Safe' category, showing the algorithm is able to identify safer paths but that it will generally result in longer routes. The low count of 'Short and Safe' routes, especial of short routes, also reinforces the idea that the shortest route is rarely the safest.

This trade-off shown in the Safety Gain vs. Length Penalty plot (Figure 2). The scatter plot shows that most routes cluster in the lower-to-mid range of length increases (under 30%), indicating that many routes achieve substantial safety improvements with realistic increases in distance. Only a small number of cases showed extreme distance forfeit, overall reinforcing the algorithm's efficiency in offering safer alternatives without significant travel cost.

Figure 3 shows safety scores across different distance groups. This plots shows a consistent trend: for nearly every distance range, safe routes yielded lower safety scores than their shortest route pair. This suggests that the safety focused routing algorithm scales well and performs consistently across short, medium, and long distances, not just with individual examples.

An outlier of this trend is the slight reversal in the final length bucket (1825–1930m), where the median safety scores for the safe routes increase slightly, having a larger spread than that of the short routes in the bucket. This may be explained by the fact that longer safe routes are inherently more likely to fall into this distance range due to the nature of their detours around high-risk areas. As a result, the sample of "safest" routes in this bin is larger, leading to larger variation in safety scores.

Overall the consistent gap in safety scores across bins supports the idea that weighted routing method is effective across different scales of pedestrian routes.

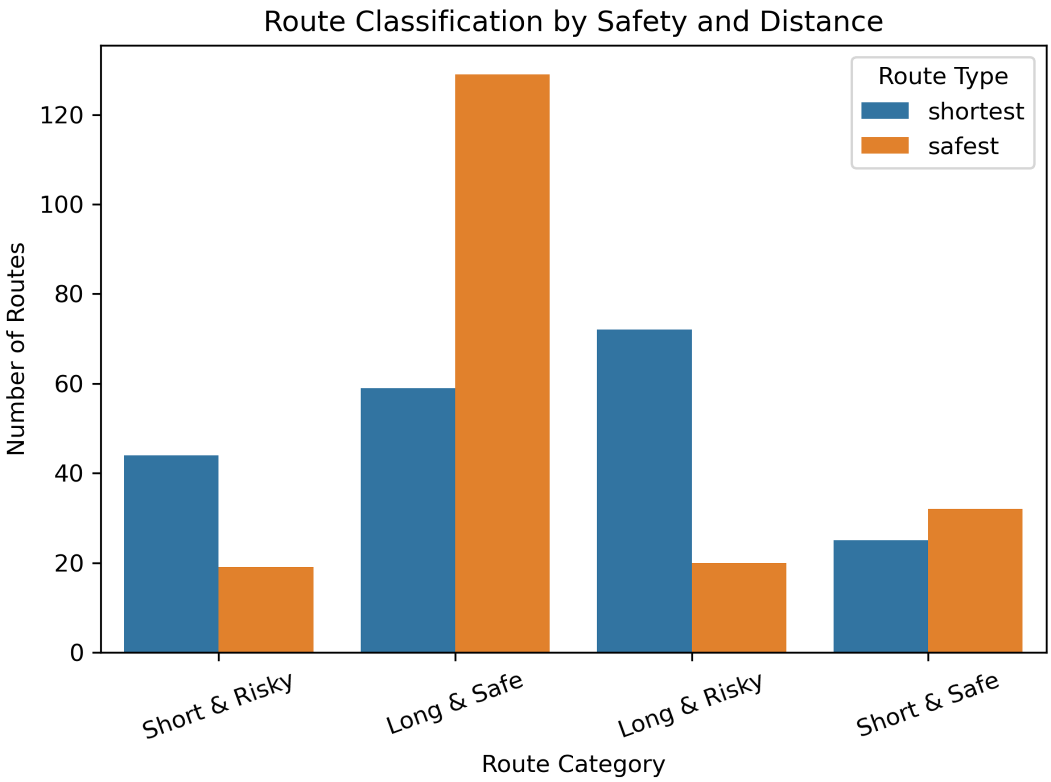

Figure 1: Route Classification Comparison

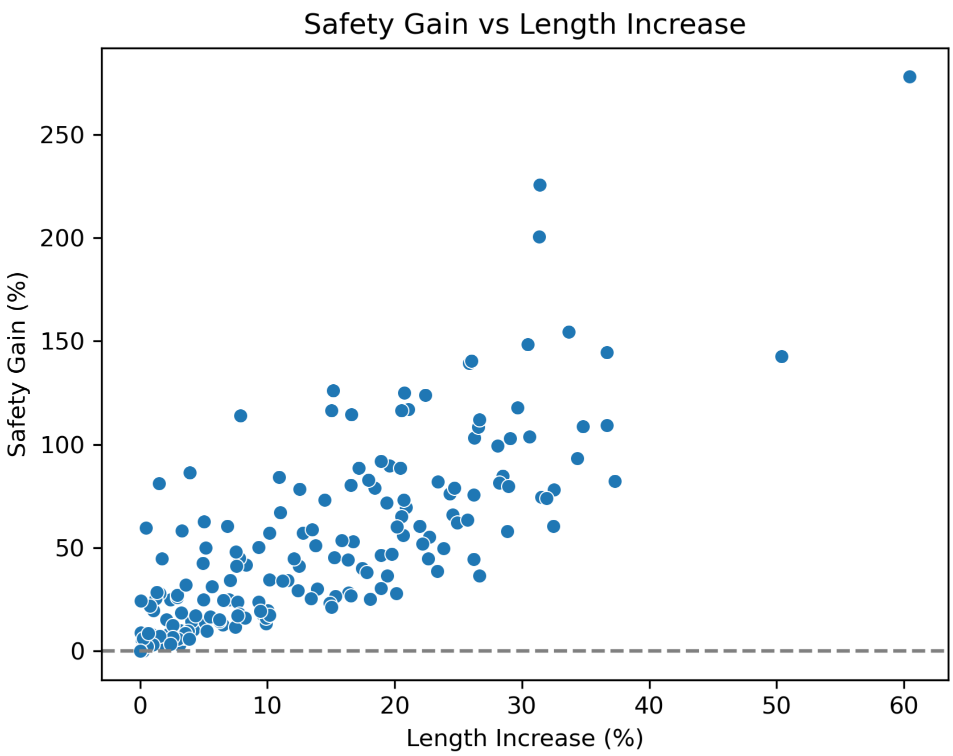

Figure 2: Safety Gain vs. Length Penalty

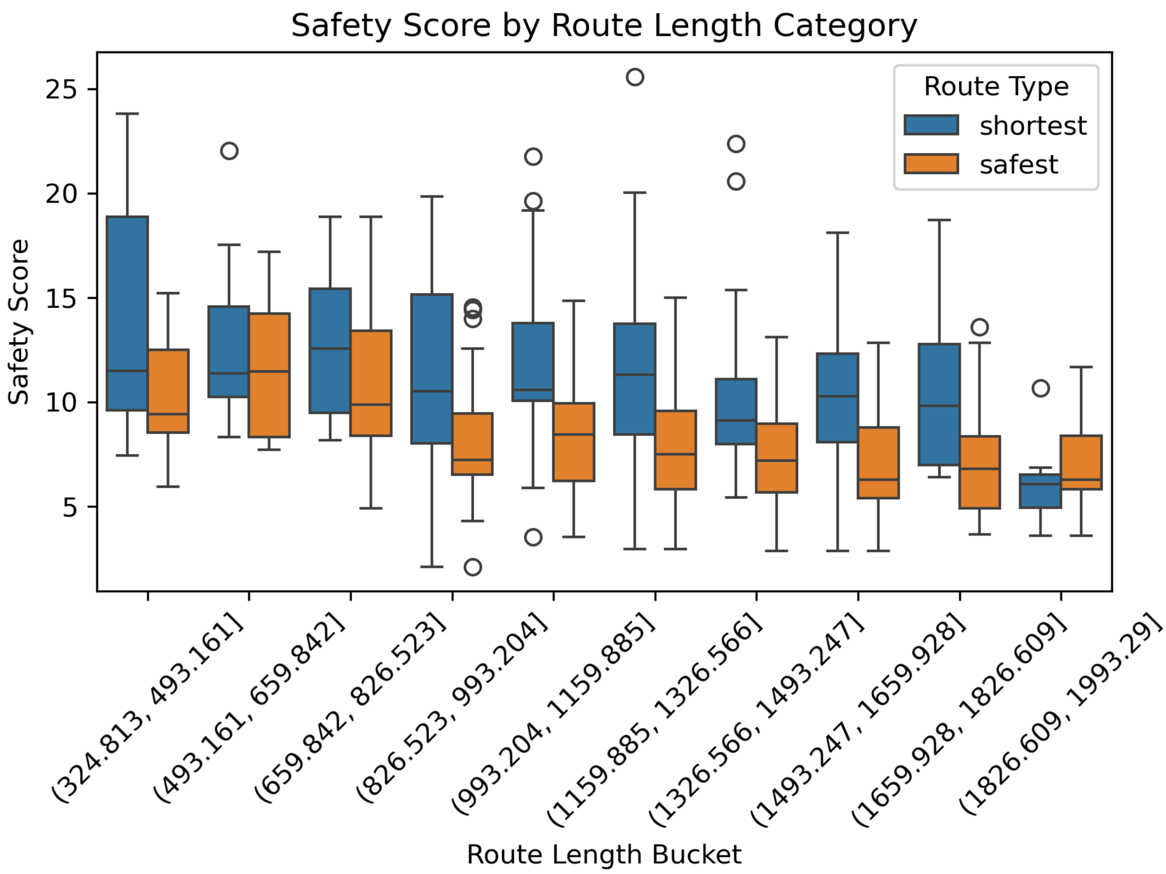

Figure 3: Safety Score Across Length Buckets

The Safety Score Comparison by Route Type boxplot (Figure 4) also reinforces this idea. The safe routes consistently demonstrated lower median safety scores with reduced variability, while the short routes not only showed higher average risk but also a wider spread in safety scores. This indicates that shortest routes are not just riskier on average but that they are also less predictable in terms of safety.

The Percentage Difference in Length histogram (Figure 5) helps demonstrate can mostly deliver safer routes without excessive detours. In nearly 75% of cases, the safest route required less than a 30%, and in many cases even less than a 15%, increase in total length compared to the shortest option. A few results displayed a major increase in distance, likely due to having to navigate around large, high risk crime hotspots, however the majority of safety-optimised routes remained efficient and walkable. This supports one of the core aims of the project; that, even without enforcing a strict cap on maximum route length, the algorithm can consistently identify safer alternatives that are still practical for real-world navigation.

The Total Length vs. Safety Score (Figure 6) further examines the relationship between distance and safety score. Here, short routes consistently trend towards higher safety scores across most lengths, also having a much larger number of routes above a score of 15 of which there is only a handful of safe routes. Meanwhile, safe routes maintain lower scores regardless of total distance. The general spread of safe routes across the plot confirms that the safety improvements are not simply a result of longer paths but rather that they are a result of calculated routing decisions based of infrastructure and crime data. Although there is seen to be a slightly higher concentration of safe routes in the longer route length areas this is expected given that safer paths involve intentional detours to avoid high-risk areas.

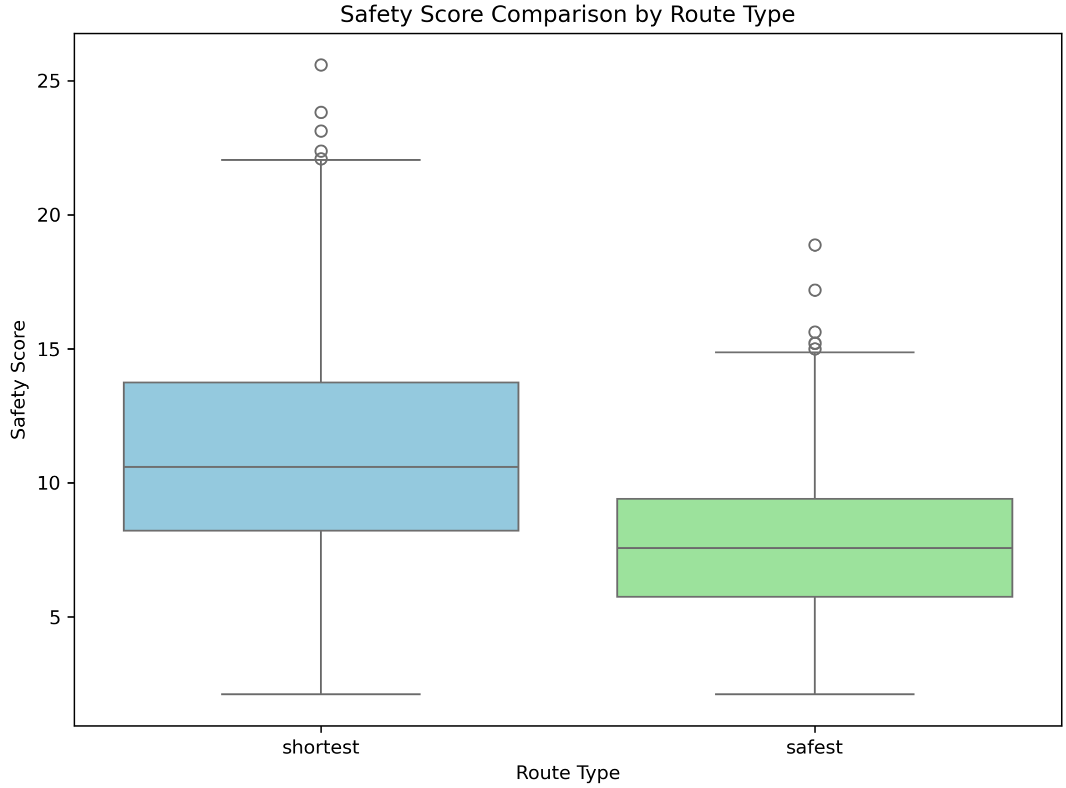

Figure 4: Safety Score Comparison by Route Type

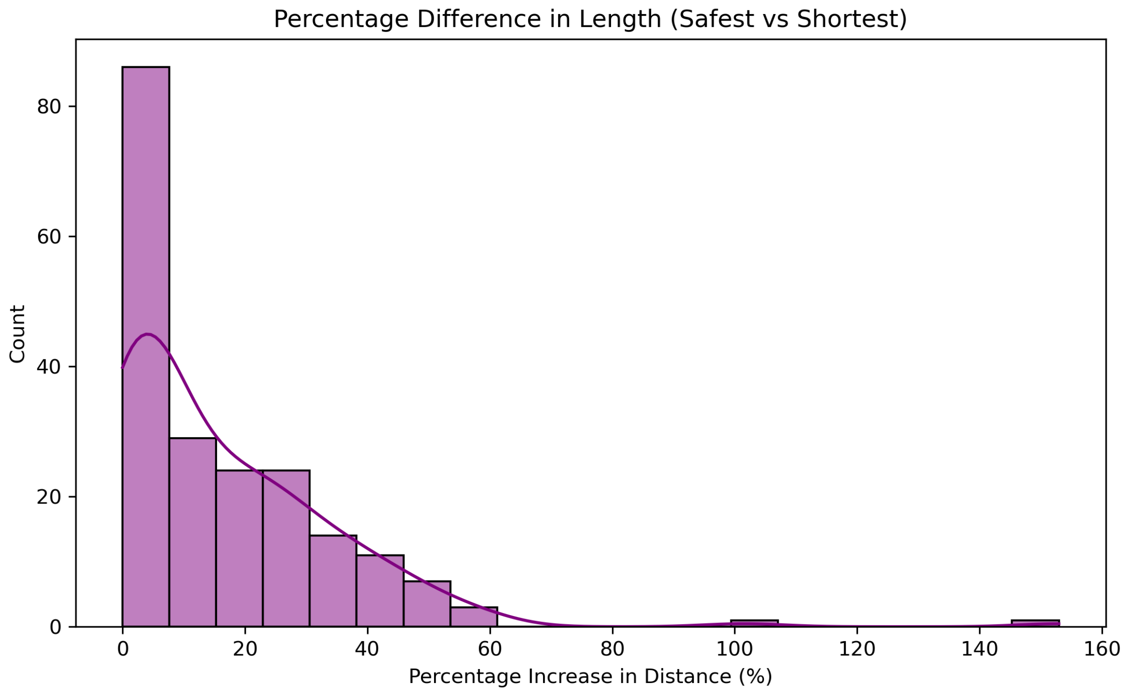

Figure 5: Percentage Difference in Length

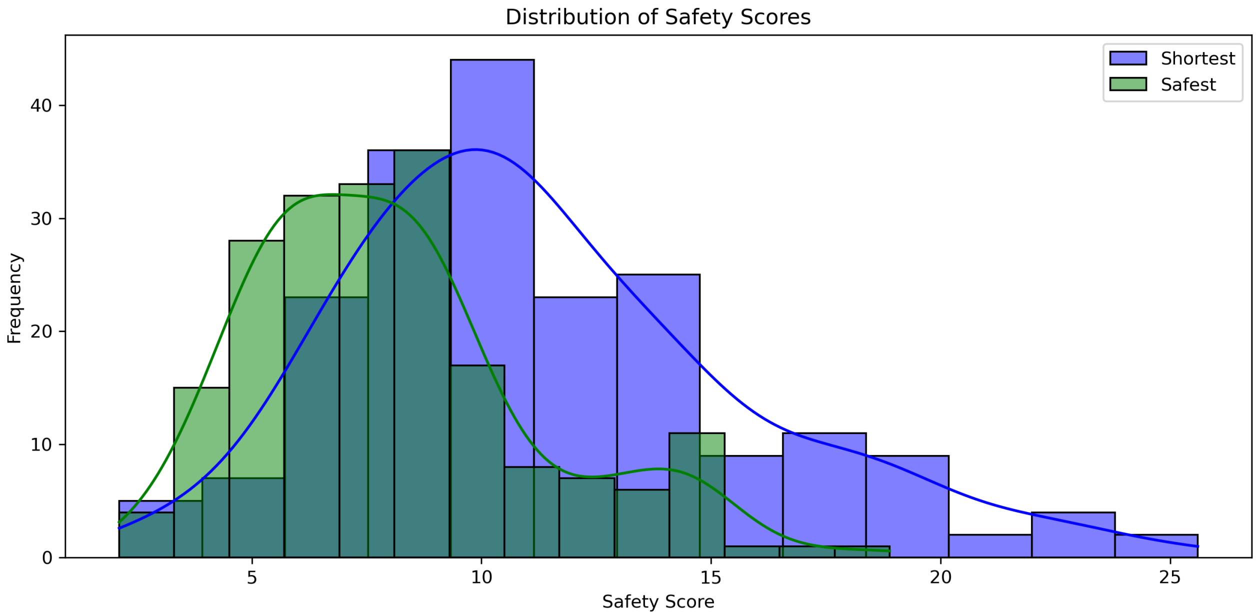

Figure 7: Distribution of Safety Scores

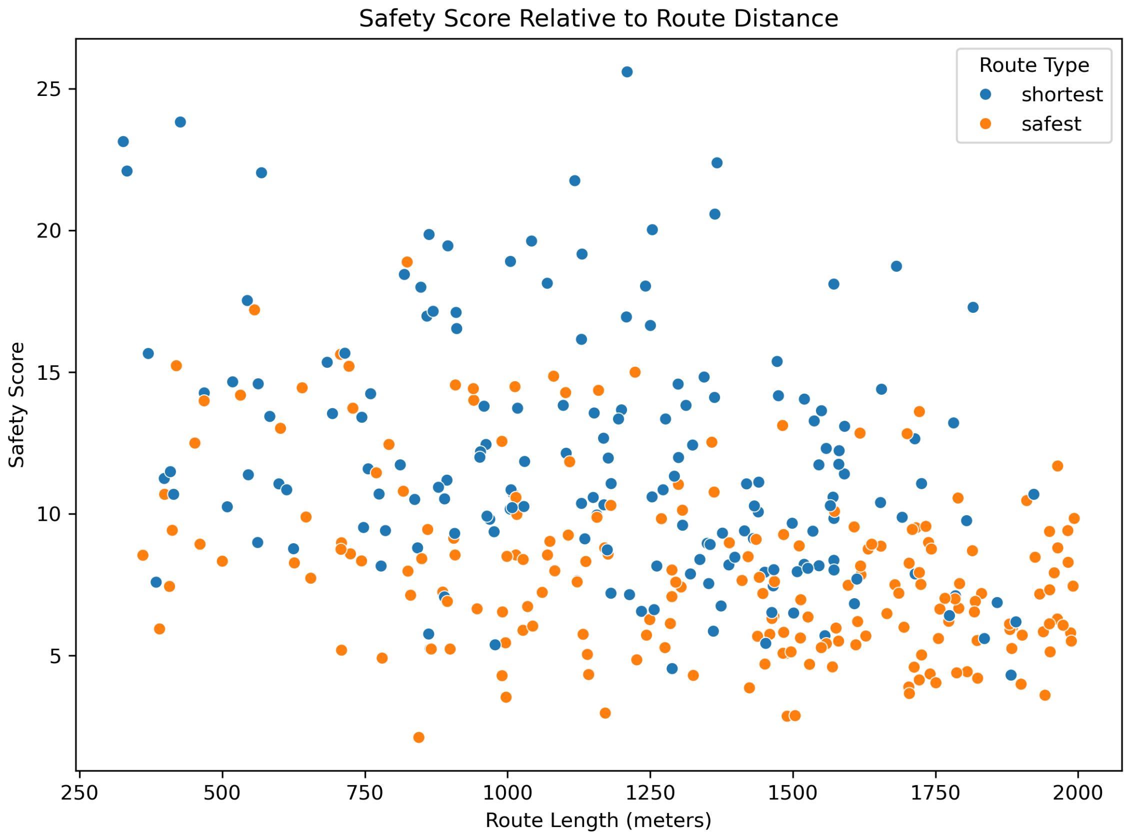

Figure 6: Total Length vs. Safety Score

Finally, the Distribution of Safety Scores (Figure 7) illustrates the overall effectiveness of the approach. The safe routes (green) showed a tight distribution of lower safety scores, showing that the algorithm successfully and consistently identified high risk areas. In contrast, short routes (blue) had a broader and less concentrated distribution, with many routes scoring a considerably higher in risk. This again supports one of the projects key objectives to demonstrate that the shortest route is not necessarily the safest. The clear separation between the two distributions highlights the effectiveness of the scoring and weighting methods. This consistency is important is the system is to be used in real-world application.

These visualisations back up the project hypothesis that he shortest route is often not the safest. More importantly, the data shows that safer routes can typically be offered with a reasonable and realistic trade-off in travel distance. This demonstrates the potential value a real-world routing system, that incorporates pedestrian infrastructure and crime awareness, can give to improve safety and confidence for pedestrians.

3.3 Routes Analysis

Figure 8 displays two generated routes between a selected origin-destination pair from the random dataset used earlier: the safest route (green) and the shortest route (blue). These routes are analysed alongside two supplementary maps: a colour coded crime risk map, which visualises risk using IDW crime scores (ranging from blue = safe to dark red = high risk), and a feature map, which highlights pedestrian infrastructure (e.g. lit road segments, CCTV coverage, and shop/cover locations). Together, these visualisations help explain why the routing algorithm picked its chosen detour.

First looking at the crime map (Figure 9): we can see at the starting point (bottom left) the short route travels up through a high-risk road segment early on, identified by its red IDW shading. It then continues along busier main roads consistently marked with moderate to high crime scores. In contrast, the safe route avoids this area completely. Many segments on the safer path are marked in blue or green, indicating minimal or no crime risk. This demonstrates the algorithm's ability to identify lower-risk paths, even when they are not the most direct.

There is also a noticeable diversion from the main road toward the end of the route where the safe path diverts down a side road. While the crime risk here is moderate, the algorithm appears to have prioritised it due to the presence of street lighting and residential classification, both of which contribute to a better pedestrian weighting. This kind of trade-off shows that the algorithm doesn't favour avoiding all crime but fairly accounts for pedestrian features effectively.

On the feature map (Figure 10), this routing decision is reinforced. The safe route passes through more consistently lit segments, with visible yellow points (streetlamps) and green lines (lit roads), while with the short route we can only see one condensed area of lighting a a couple streetlamps. It also passed by a CCTV-covered area, which further boosts its score. Although not visualised on the feature map, another contributing factor is the algorithm's preference for residential roads and lower speed limits. These attributes were assigned more favourable weights during the pedestrian safety scoring phase, as they generally more pleasant and safer to walk.

Another important factor to account for when analysing these routes, is that the safe route is only 26% longer than the shortest option, increasing from 1,363m to 1,716m. Despite no explicit constraint on maximum route length, the algorithm has returned a realistic, walkable alternative.

Overall this route case study demonstrates the value of the system. It shows that safety-aware routing can integrate multiple data sources to produce safer routes that remain practical for everyday navigation. It also reinforces the project's finding that the shortest path is often not the safest.

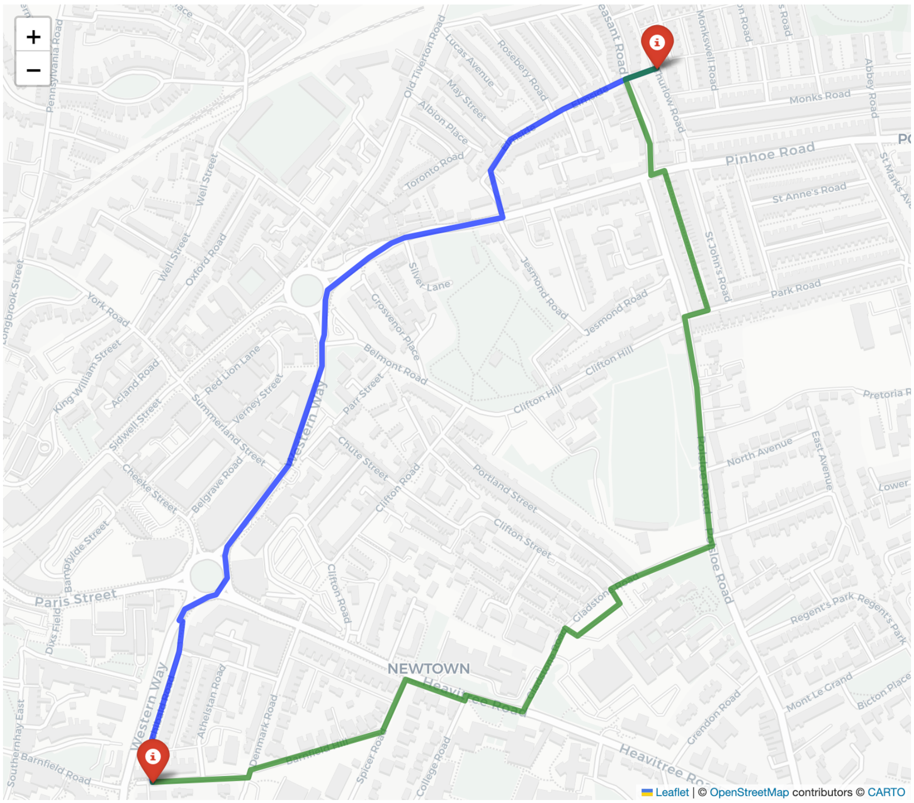

Figure 8: Routes

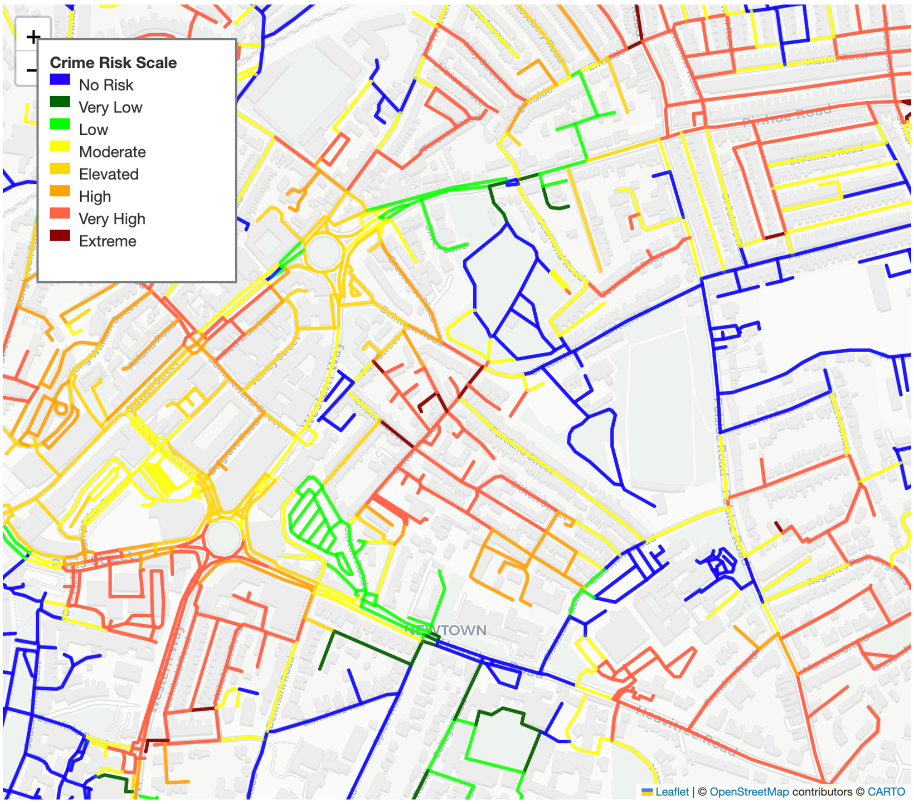

Figure 9: Colour Coded Crime Map

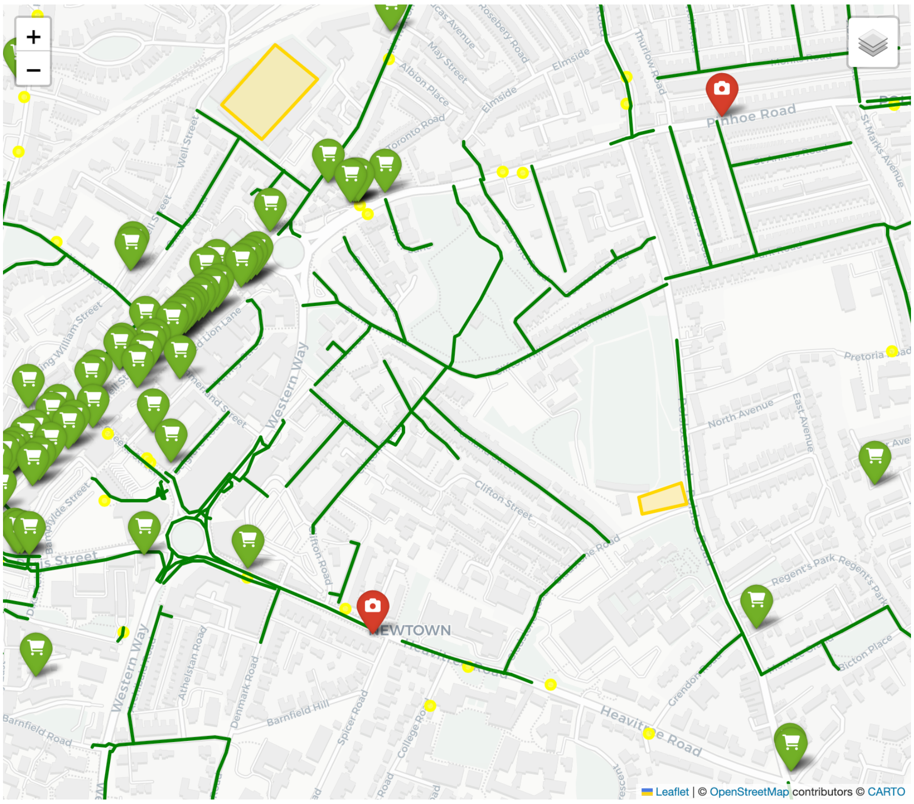

Figure 10: Feature Map

3.4 Discussion

The statistical analysis and case study comparison clearly demonstrates that the shortest route is not always the safest. Analysing data of shortest and safest routes from 200 randomly generated origin-destination pairs, safe routes consistently achieved better safety scores which validates the effectiveness of of this custom routing approach. These improvements in safety were often achieved with a manageable increase in walking distance, with nearly 75% of safe routes being within a 30% distance increase of the shortest route.

Additionally, the scatter plot (Figure 2) further showed this trade off with most safe routes clustered around the area of that graph that displays substantial risk reduction a with reasonable length increases. This balance between safety and walkability reinforces one of the project's key goals of being able to give safer alternatives without compromising on journey length.

The results were consistent across route length bins (Figure 3), route classifications (Figure 1), and different, random locations, suggesting the algorithm is generalisable and not dependent on specific examples. The boxplot analysis of safety scores (Figure 4) confirm that not only are shortest routes are more unsafe but are also have less predictable risk levels. This inconsistency implies how shortest routes tend to take direct routes that frequently pass through high-risk areas such as main roads or crime hot spots.

However, a number of limitations need to be considered. First is the completeness of OpenStreetMap (OSM) data, with many features such as CCTV and lit areas missing as well as missing or incorrectly labelled pedestrian highways. This gap in data likely affects the effectiveness of the weighting system and potentially weakened the influence of features on routing decisions. Additionally, while crime data from Police.uk gave valuable insight to common crime areas, it was limited in its depth with only recorded crimes since the beginning of 2022 available. The crime data also lacked timestamps which would enable time sensitive risk modelling such as increased nighttime danger. Importantly, all crime and feature weights were manually assigned based of personal assumptions, without empirical evidence. This adds a level of subjectivity and may affect scaleability to other cities. A potential improvement of this would be to conduct public surveys in order to obtain more objective weights. Similarly, if used in a mobile application, users could set weights themselves to tailor routes to personal preferences and needs.

Although the system performed well in producing routes with reasonable increases in distance, it is important to note that there was no explicit implementation to restrict how much longer a safe route could be compared to its shortest route pair. The fact most safe routes fell into a walkable range is promising, however, introducing a threshold to limit distance increase would enhance the usability of this in a real world application. Such constraints could be introduced by having a percentage cap of a safe route's distance relative to its shortest route counterpart.

It is also important to note that, while the randomly generated origin-destination pairs provide an informative and unbiased sample for evaluation, it does not necessarily reflect realistic routes a pedestrian may take. In reality, pedestrians usually start and finish their routes at a meaningful location, either their homes, jobs or public transport stops, rather than arbitrary points. This means that routes in this study may not reflect patterns typically seen in urban travel. Future studies could improve this approach by using points of interest to create more realistic OD pairs.

Despite these limitations, the algorithm showed strong performance in producing safer alternative routes quickly and that also generally remained practical in length. These finding showcase the validity of this approach and display potential for integration into a real time navigation tool for pedestrians. Furthermore, this work demonstrates how combining open source data can lead to useful improvements to pedestrian safety.

4 Future Improvements

Despite the promising outcomes of this project, several areas present opportunities for further development and refinement:

4.1 Increasing Data Reliability

A core limitation of this project was the incompleteness and inconsistency of OpenStreetMap (OSM) data. Important features such as lighting, CCTV coverage, and footway classifications were often sparsely tagged, especially in residential areas. This would have had an effect on the accuracy of route calculations.

To address this, future work could start by incorporating alternative geospatial datasets, such as council records of lighting inventories or CCTV maps. Crowdsourcing initiatives or partnerships with local authorities could also be a great way to improve completeness of OSM in high-footfall areas. Research by Goodchild and Glennon highlights this benefit, showing that crowdsourced information can provide valuable data when official sources are incomplete or slow to update[10].

In addition, emerging techniques in computer vision and Google Street View (GSV)/satellite imagery analysis offer promising solutions for automated detection of urban features like crossings, footpaths, or lights, as mentioned in several urban computing studies[11][12]. For example, Li et al. demonstrated that combining GSV imagery and automated analysis detailed street level features, such as green spaces, displaying the potential for using GSV to capture urban features[13]. Improving data coverage and quality would directly enhance the accuracy and would allow for more reliable safety modelling.

4.2 Mobile app development

Building on the successful development of the underlying routing algorithm, a logical progression of this work would be to deploy a mobile or web-based application. A tool like this would enable real-world application of the methodology, translating academic findings into practical functionality. This aligns with the main objective of this project of improving pedestrian safety through accessible technology.

A proposed application could support:

- Real-time safety-focused routing, allowing used to chose routes based not just on distance but on chosen, dynamic safety indicators.
- Customisable safety preferences, enabling users to set their own personal weightings for features such as lighting or different crimes.
- Crowdsourced feedback system, such as user-reported incidents or unsafe areas, to refine the routing algorithm over time and improve trust in the system.

Previous work has explored personalised route planning. Nadi and Delavar developed a method that used Ordered Weighted Averaging (OWA) to dynamically weigh different route attributes based on preferences chosen by the user, offering alternative routes that match the users priorities[14]. While their work focused on features like travel time and scenic quality, this idea of a multi-criteria approach could be well adapted to this project. Allowing users to select their own routing criteria, whether they prefer lit areas or low road speeds, would enhance the system in a user-based application.

As well as its utility to individual users, the tool could support broader application such as aiding in safety planning of urban areas, helping local authorities to visualise risk concentrations and assess infrastructure gaps. Ultimately, such a platform would help bring together open data and community-driven navigation.

4.3 Temporal & Contextual Awareness

While this project models static spatial safety well, it currently lacks acknowledgement of temporal changes, an important aspect of real-world pedestrian risk. Crime risk isn't static and fluctuates over time, for instance areas that are safe during the day may be significantly more dangerous at night. Likewise, certain neighbourhoods experience increased criminal activity on weekends or during specific seasons. Introducing temporal awareness into this safety model would allow for a more accurate and responsive routing system.

Research supports the value of incorporating temporal patterns in crime distribution. Cory Haberman showed that while crime stays concentrated in small areas, the exact streets where crimes happen can change depending on the time of day or day of the week[15]. This means that a road that may be low-risk during the day may become a hotspot during the nighttime, highlighting the value having temporal awareness would have in a safety routing system.

An obvious improvement of this work would be to have adjusting route recommendations based on the time of day, day of the week, or even seasonally for areas with high seasonal changes. For instance, lighting features could be given increased weight when its dark, while crime scoring could factor in recent incidents to better reflect current risk. Incorporating real-time data could allow for dynamic updates and live response to changes in environment.

However, a limitation of the current system is the lack of substantial historical crime data. The Police.uk API only provides access to data dating back to early 2022, which means being able to model long-term trends and identify temporal changes in crime patterns is difficult. Access to a more in depth crime database would increase the accuracy of the IDW crime score and also allow for predictive modelling that anticipates risks based on historical trends.

5 Conclusion

5.1 Reflection on Previous Goals

This project set out to investigate whether the shortest route between two points in an urban environment is also the safest. The main goal was to develop a routing system that could factor in pedestrian infrastructure, such as lighting and CCTV coverage, and crime data. In doing so, the projected aim to identify safer routes that were still practical for everyday use.

Through the use of open data from OpenStreetMap and Police.uk, a custom routing weight was developed that merged infrastructure and crime data into a single metric. The results from randomised route generation consistently demonstrated that the shortest route was often not the safest.

While not all routes resulted in dramatic safety improvements, the overall trend showed clear benefits from a safety-aware routing approach. The routing system proved capable of accounting for a range of different pedestrian friendly features and the use of IDW based crime scoring added great insight to local crime affected areas. Additionally, the final routing method proved efficient in returning routes quickly making it suitable to real-world use.

5.2 Summary

Using 200 randomly generated origin-destination pairs, both shortest and safest routes were calculated and analysed. The results showed that safe routes consistently had lower risk scores than their shortest route pair. Moreover, most of these safer routes remained within 30% of the shortest route's length, with this realistic difference showing their ability for real-world navigation. Statistical analysis using safety score distributions, trade-off plots, and route classification breakdowns, further confirmed that the algorithm performed reliably across various route lengths and locations. Individual inspection of a selected route further illustrated how the algorithm favoured routes with better light coverage, residential roads and low roads speeds while also avoiding know crime hotspots.

Despite limitations such as the lack of completeness of OSM data and the limited availability of crime data from Police.uk, the results show promising potential. It sets a strong foundation that with future development in data coverage, user integration and other small improvements such as length caps, it could be the base of a mobile tool to support safer urban navigation.

References

[1] B. Livingston, M. Grillo, and R. Paluch, "Cornell international survey on street harassment," 2014.

[2] A. Loukaitou-Sideris, "Fear and safety in transit environments from the women's perspective," *Security journal*, vol. 27, pp. 242–256, 2014.

[3] K. M. Pollack, A. C. Gielen, M. N. Mohd Ismail, M. Mitzner, M. Wu, and J. M. Links, "Investigating and improving pedestrian safety in an urban environment," *Injury epidemiology*, vol. 1, pp. 1–9, 2014.

[4] M. M. Peden, *World report on road traffic injury prevention.* World Health Organization, 2004.

[5] Z. Guo and B. P. Loo, "Pedestrian environment and route choice: evidence from new york city and hong kong," *Journal of transport geography*, vol. 28, pp. 124–136, 2013.

[6] E. Galbrun, K. Pelechrinis, and E. Terzi, "Safe navigation in urban environments," in *3rd International Workshop on Urban Computing at KDD 2014, UrbComp'14*, 2014.

[7] S. Levy, W. Xiong, E. Belding, and W. Y. Wang, "Saferoute: Learning to navigate streets safely in an urban environment," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 6, pp. 1–17, 2020.

[8] B. C. Welsh and D. P. Farrington, "Effects of improved street lighting on crime," *Campbell systematic reviews*, vol. 4, no. 1, pp. 1–51, 2008.

[9] R. Ewing and S. Handy, "Measuring the unmeasurable: Urban design qualities related to walkability," *Journal of Urban design*, vol. 14, no. 1, pp. 65–84, 2009.

[10] M. F. Goodchild and J. A. Glennon, "Crowdsourcing geographic information for disaster response: a research frontier," *International Journal of Digital Earth*, vol. 3, no. 3, pp. 231–241, 2010.

[11] F. Biljecki and K. Ito, "Street view imagery in urban analytics and gis: A review," *Landscape and Urban Planning*, vol. 215, p. 104217, 2021.

[12] G.-Y. Wang, "Integrating street views, satellite imageries and remote sensing data into economics and the social sciences," *Social Science Computer Review*, vol. 42, no. 1, pp. 326–351, 2024.

[13] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using google street view and a modified green view index," *Urban Forestry & Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015.

[14] S. Nadi and M. R. Delavar, "Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 3, pp. 322–335, 2011.

[15] C. P. Haberman, E. T. Sorg, and J. H. Ratcliffe, "Assessing the validity of the law of crime concentration across different temporal scales," *Journal of Quantitative Criminology*, vol. 33, pp. 547–567, 2017.