A Digital Humanities Approach to Study Women's Perceptions of Safety in the City of Milan: The STEP uP Project

Lamia Abdelfattah[1(✉)], Lily Scarponi[1], Federico Messa[1], Carmen Guida[2], Floriana Zucaro[2], Gerardo Carpentieri[2], Florencia Andreola[3], Azzurra Muzzonigro[3], Jim Walker[4], and Andrea Gorrini[4]

[1] Fondazione Transform Transport ETS, Via Lovanio 8, 20121 Milan, Italy
l.abdelfattah@transformtransport.org
[2] TeMA Lab, Dipartimento di Ingegneria Civile, Edile e Ambientale, Università degli Studi di Napoli Federico II, Piazzale Tecchio, 80, 80125 Naples, Italy
[3] Sex and the City APS, Via della Moscova 58, 20121 Milan, Italy
[4] Walk21 Foundation, 24 Moorend Road, Cheltenham GL530HD, Gloucestershire, UK

Abstract. This paper outlines the data collection phase of the "STEP UP - Walkability for Women in Milan" project, funded by Fondazione Cariplo (Grant No. 2022-1643), which aimed to assess walkability for women in Milan using geospatial data analysis methods. The project, which ended in February 2024, involved three phases: use case definition, data collection, and data analysis through advanced GIS mapping techniques. This paper focuses on the data collection phase, employing a digital humanities approach to gather location-based open data, Wher app data, survey questionnaires, and focus group discussions. Location-based open data and Wher app georeferenced data about women's safety perceptions in the city of Milan were subsequently used to develop a statistical GIS analysis through a Geographically Weighted Regression model aimed at understanding the relationship between ratings of women's safety perceptions and the built environment. Over 70,000 ratings and 4000 comments submitted by female users through the Wher app provided insights into their perceived security while walking. Additionally, a survey questionnaire and focus groups were conducted to evaluate participants' needs and perceptions of safety. These data aimed to facilitate a detailed spatial analysis of nighttime safety perceptions for women and gender minorities in Milan, emphasizing a human-centered approach to gender-inclusive mobility research and addressing gaps in data collection for complex socio-spatial issues.

Keywords: Data Collection · Digital Humanities · GIS Analysis · Gender Inclusion · Walkability

1 Introduction

Increasing urbanization has led to a heightened focus on issues related to spatial inequality and urban safety, including the safety perceptions of women in cities. "STEPUP: Walkability for Women in Milan" addresses one of the critical challenges in tackling

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14823, pp. 49–65, 2024.
https://doi.org/10.1007/978-3-031-65329-2_4

this sensitive topic: the collection and selection of data. Women's safety on city streets is a complex and multifaceted issue that requires a deep understanding of the most significant indicators influencing such perceptions.

One of the main challenges addressed in this research is the difficulty in collecting specific data on this subject. Institutions often show a lack of interest in gathering detailed information on safety issues specific to women, creating a gap in overall understanding. This indifferent attitude results in a lack of dedicated tools and resources, limiting the ability to fully comprehend the dynamics that generate urban safety factors. Another significant challenge pertains to the absence of sector-specific data focusing exclusively on women's experiences. Generic data often does not allow for a detailed analysis of gender differences concerning specific aspects of the urban environment. This lack of sector-specific data limits the ability to formulate targeted policies and specific interventions that can effectively address safety issues affecting women.

In this context, the project aims to critically examine the challenges related to the collection and selection of data on urban safety for women, emphasizing the importance of addressing these gaps to develop more effective strategies for targeted user groups. Through in-depth analysis of multisource data, the goal is to explore women's perceptions of safety in the city, contributing to a more complete and accurate understanding of urban safety dynamics and walkability in relation to gender.

1.1 Inclusive and Safe Cities

The vision of inclusive and safe cities represents a fundamental paradigm in contemporary urban design, aiming to create spaces that meet the needs of the entire population, promoting accessibility, equity, and a sense of belonging. Inclusivity, in this context, refers to the design of urban environments that are welcoming to all citizens, regardless of their individual characteristics (Whitzman et al., 2013). The design of inclusive cities involves a significant commitment to removing physical and economic barriers. This means ensuring that urban spaces are accessible to people with physical disabilities but goes beyond, considering economic, cultural, and social accessibility. This includes creating accessible pedestrian environments and establishing public services that cater to everyone, irrespective of age, gender, or socio-economic background.

Walkability, namely referring to how friendly the urban environment is for walking, is gendered. Women generally experience the city differently than men, partly because they are more concerned with safety issues related to aggression and harassment (Koskela, 1999). Safety perception disparities tend to be especially exacerbated at night, when women's experiences tend to be significantly limited by fears of violence. This often leads to travel behavioral modifications, as evidence shows that women often suppress or modify their routes as a result (Loukaitou-Sideris, 2014). Transport planning tools and methods have often been slow in recognizing fundamental differences between the genders in their mobility patterns, largely overlooking women's needs in transport environments (Levy, 2013).

2 Methodology

2.1 A Digital Humanities Approach

Data collection in the framework of the STEP UP scientific research project follows a digital humanities approach, combining traditional and digital data collection and interpretative techniques to develop a transdisciplinary research at the intersection of computing, digital technologies and the disciplines of the humanities (Berry, 2011). This approach was deemed essential in the study of complex socio-spatial phenomena, such as the perception of safety of a specific demographic group, namely women.

The advent of digital technology has given rise to new research environments, offering new mediums for exploration using hybrid tools of data analysis that effectively mediate between new and conventional research mediums, and provides integrated and transdisciplinary methods for studying complex urban phenomena. Emerging digital technologies and data collection methods have significantly impacted the means of data production and interpretation, opening up numerous avenues for reading and understanding the urban experience through a user-centric approach (Zhang et al., 2024).

2.2 Objectives of the Research

The aim of this paper is to detail the collection and validation process for a series of relevant datasets useful to carry out the analysis in the framework of the goals of the STEP UP project. The paper is divided into four main sections discussing the collection methodology and outputs for four different data types and tools utilized in the research: Location-based Open Data (Sect. 3), Wher App Data (Sect. 4), Survey Questionnaires (Sect. 5) and Focus Groups (Sect. 6). It is important to note that these four branches of the data collection process are not mutually exclusive but are often interlinked and combined to refine one another.

Gender-disaggregated data sets collected through the four data groups are consequently analyzed through an extended Geographic Information Systems (GIS) analysis and advanced statistical techniques, as well as a series of quantitative and qualitative analyses for non-locational data, aimed to support and enrich the mapping results with further evidence from women users of the city. Techniques used to collect and analyze each data type are detailed in the following sections.

3 Location-Based Data

A series of relevant geolocated datasets were retrieved, sorted, validated, and filtered from open and proprietary data repositories and geoportals with the aim to populate a spatial regression model. The process outlined below shows the selection process of indicators leading to a final selection of eight indicators that were used to calibrate the Geographically Weighted Regression (GWR) Model to further analyze the relationship between said indicators and safety perception votes as observed through the Wher app. Data collected ranges from open to proprietary data from a wide range of sources.

3.1 Data Selection Methodology

In this study, proxy indicators were used to represent factors that are not directly observable among the list of safety factors identified as relevant in the literature in earlier phases of the STEP UP project. Within the framework of the STEP UP project, a total of 19 safety factors (SF_L2) were defined, classified into three types of safety factors (SF_L1): spatial features, city use and hotspots (Scarponi et al., 2023). A full list of these safety factors can be reviewed in Table 1 in the Annex. The set of proxy indicators identified in this research was cataloged and gathered by considering the following identified criteria: availability in the study area, alignment with the safety factors, reliability of the sources, sub-urban scale and spatial data. The complete dataset of proxy indicators was further reduced through a verifiable selection process as shown in Fig. 1.

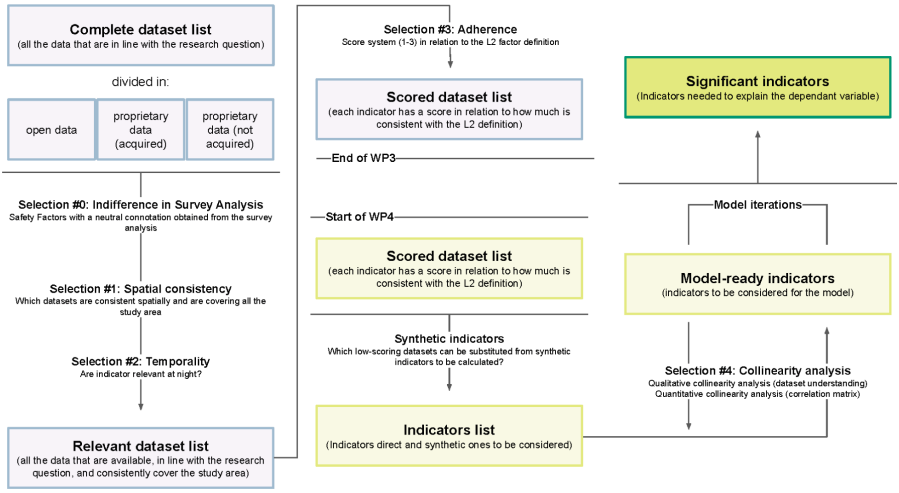

Fig. 1. Data gathering and selection methodology.

Identifying selection criteria can be a crucial step in reducing a set of indicators that describe urban characteristics more effectively and meaningfully. This process helps focus attention on more relevant indicators, thereby improving the quality and clarity of information. The selection criteria defined in the research were (i) Survey Analysis, focusing on indicators that emerged as significant in the survey results; (ii) Spatial Consistency, selecting indicators that are represented consistently in different parts of the city; (iii) Temporality, only selecting indicators that are relevant to nighttime (8pm–6am), thereby excluding activities that are only active during the day; (iv) Representativeness, which refers to the relevance of certain indicators as supported by the literature on the topic; (v) Adherence to safety factors definition, as defined in earlier phases of the research (see Table 1 in Annex). The final selection of 16 indicators to be imported in the GWR model is shown below:

1. Public Lighting
2. Space Syntax - 300m connectivity
3. Count of Food & Beverage destinations
4. Perpendicular Residential Building Curtains
5. Space Syntax - 300m choice
6. SOS Devices
7. Resident Population
8. Worker Population
9. Yearly Night Strava Users Flow
10. Overground PT Frequency
11. Railway Station Buildings
12. Discos and Night Clubs
13. Green Areas
14. Abandoned places
15. Industrial Buildings
16. Police Headquarters

4 Wher App Data

Data regarding the perceived level of security of women while walking was collected through 'Wher', a route planner application operated by Walk21 Foundation. Wher is an effective tool for collecting geo-localized data about women's perceived level of security while walking in urban areas (Carpentieri et al., 2023a, 2023b). It is a crowd-sourced mapping tool allowing women in a particular city to rank geo-referenced locations according to their perceived level of safety. As of the 1st of April 2019, in the city of Milan, the app had 5667 ratings, including 3850 textual ratings, collected from a total of 7844 users and 1296 reporters. Over 50% of the city extension (in kilometers) was rated through the application.

4.1 Methodology

Two types of data from the Wher App were analysed: (*i*) point data to be integrated in a GIS model, and (*ii*) corpus of textual comments. For the point data analysis, a road graph made using OpenStreetMap was joined to the point Wher App data to visualize ratings distribution and identify areas with the highest number of ratings, to provide insights into the spatial distribution of user perceptions. In the comments analysis, textual data was anlaysed through textometric analysis using IRAMUTEQ software to uncover common themes, and a sentiment analysis using Kimola software to derive sentiment classifications and identify frequently occurring terms.

Data distribution Analysis
The Wher App dataset is structured as single ratings, with the information regarding the creator of the comment and the timeslot of the day. Each score is positioned by the user on the geometric barycentre of the arc forming the road graph from OSM (OpenStreetMap), so that the geometric consistency of the dataset is a set of points. The same rating can be attributed to more than one point on the map, giving the user the chance to select a larger area with the same security perception characteristics, by including similar street segments considered similar to the main one. This divides the ratings in two typologies, the ratings associated with a single point on the map and the ratings associated with more than one point on the map (indicating an area that can vary). Most street segments are characterized by more than one comment or rating and a representation of the rating distribution across the city (see Fig. 2b).

App Rating Procedure and User Profiles

Users are asked to review a selected stretch of road on the application's map through a user rating and optional comment. They are asked to: (i) select a time slot between day, evening and night; (ii) rate the lighting from "Not at all" to "A lot"; (iii) specify the level of crowding from "Deserted" to "Very crowded"; and (iv) select advisability, with respect to perceived safety (Avoid, be careful and go easy). The ratings are provided with geolocation and time reference (date and time of transcription) and the points are distributed across three time segments (day, evening, and night, 06:00–20:00, 20:00–24:00, and 24:00–06:00, respectively).

Basic biographical information is also collected through the application, allowing to have some segmented information about city users. Each data point includes the following information: User ID; Creation time; Birthday; City; Occupation; Transport; Mapped meters; Ratings; Vote on perception; Vote on light; Vote on crowd; Timeslots; Rating data; Longitude; Latitude. In particular, the analysis summaries on user data, occupation, age and distribution of ratings over time, shows some data bias towards younger age groups (81.7% of users are in the 15–30 age group), and a concentration of comments in the year 2019 (2822 comments in 2019 representing over 71% of total comments collected on the App between 2017 and 2023).

4.2 Wher Data Analysis

Wher Data Analysis for GWR Model

The dataset available in the city of Milan consists of data from 2018 to 2023 with a coverage of the road graph of over 50% of rated roads. Considering the availability of data and the direction of interest of the project, the Wher App dataset submitted to further analysis is the data relative to the timeslots of evening and night (8:00pm–6:00am) from the year 2019, where two thirds of the data are concentrated. Only ratings within the boundaries Municipality of Milan are considered.

The Wher Road Graph Map

The Wher App point dataset was joined with the OpenStreetMap graph, only selecting data relevant to pedestrian flow. Arcs of sidewalks were split to a maximum length of 100 m, to be able to represent the Wher Data accurately. The resulting graph contained univocal information of each rating composed of the ID of the user, the time and date of the rating and the univocal information of each arc of the graph, such that overlapping arcs are equal to the number of ratings that were produced for each arc.

Two different maps were obtained, the first based on the ID of each rating (ID of the user and time and date of said rating), the second based on the ID of each arc of the graph. Figure 2a shows the distribution of the ratings on the city of Milan, with the overlapping of different votes and ratings of different dimensions. Ratings go from a vote of 1 to 3 and don't necessarily refer to one point. Within this image it is possible to see the overlapping of votes: some areas are characterized by similar votes, resulting in a strong red or a strong green, whereas some areas are characterized by diverging results (see Fig. 2a). Figure 2b shows the total sum of ratings for each arc, ranging from a minimum of one rating to a maximum of over 160 ratings.

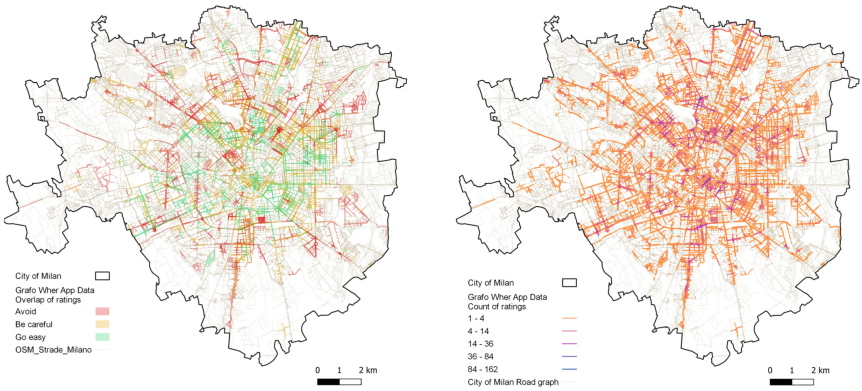

Fig. 2. (a) Wher App Data - Overlap of ratings; (b) Wher App Data - Count of ratings

Textual Comments Analysis

Textometric Analysis

Textometric analysis, also known as textual analysis or text mining, is a methodology used in the field of computational linguistics and natural language processing (NLP) to analyze and extract meaningful information from large bodies of text. The goal of textometric analysis is to uncover patterns, trends, and insights within textual data (Carpentieri et al. 2023a, 2023b). In this research, the open-source software tool IRAMUTEQ was used, which stands for Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires (Interface of R for Multidimensional Text and Questionnaire Analysis). IRAMUTEQ can be a powerful tool for qualitative text analysis, helping to uncover valuable insights from user ratings on Wher App.

The textual data from Wher ratings was preprocessed through data cleaning, tokenization, stopword removal, and stemming and lemmatization. Once imported into the software, the data was explored to give an overview of the word frequency and identify any immediate patterns, classifications, or themes.

IRAMUTEQ uses various statistical techniques for text analysis, including Principal Component Analysis (PCA), Multiple Correspondence Analysis (MCA), and Hierarchical Cluster Analysis (HCA). The findings were represented graphically using IRAMUTEQ's visualization tools. By clustering and grouping similar ratings or themes within the dataset, the tool helped identify commonalities and differences in women's experiences while walking. Figure 3a shows a factor analysis corresponding to the increasing frequency of forms and lemmas, where Factor 1 and Factor 2 represent the values and percentages of each class. Figure 3b visualizes the textometric analysis word graph based on word clustering.

Sentiment Analysis

Sentiment analysis describes the field of study concerned with analyzing the opinions, attitudes and emotions of individuals towards entities such as products, services, organizations, locations and events (Liu, 2012). Over the last two decades, the field has become increasingly active given the vast real-world applications to a plethora of disciplines. For the first time in history, researchers have access to huge volumes of freely accessible data

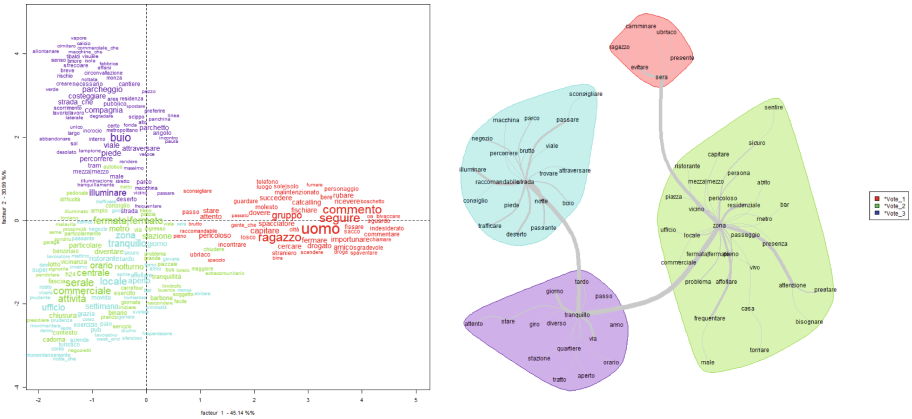

Fig. 3. (a) Textometric Analysis Classes Graph; (b) Textometric Analysis Word graph.

published by individuals online, can be analyzed through sentiment analysis (Roberts et al., 2018).

This analysis is performed using Kimola, a platform dedicated to social research and machine learning, primarily designed for obtaining consumer insights and enhancing marketing strategies. Automatic classifiers are generated, grounded in prevalent comment themes, with associated Sentiment Classifications. The analysis resulted in the following classifiers: (i) Crowded and busy Streets; (ii) Homeless people; (iii) Drug activity; (iv) Poor lighting; (v) Sex workers or the activity of sex work; (vi) Quiet residential areas; and (vii) Safety concerns at night. Figure 4 shows the sentiment polarity linked to each classifier as well as the geographic variation for the classifier "Crowded and busy Streets" by neighborhood NIL (Nuclei d'Identita Locale), the neighborhood unit in the city of Milan.

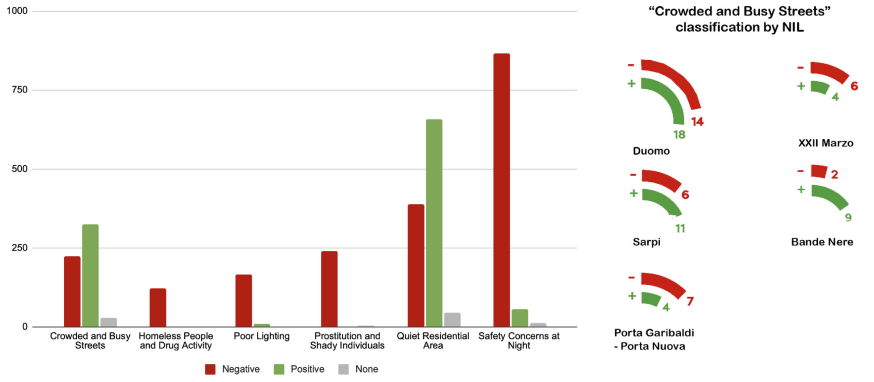

Fig. 4. Sentiment polarity total classification count and distribution by NIL for the classifier "Crowded and Busy Streets" (Kimola analysis)

5 Survey Questionnaires

A survey questionnaire was designed to intercept the participants' priorities in relation to perception of safety while walking, as well as to enrich the data collected through the Wher App filling in gaps about women's perceptions of safety, in relation to intersectional differences between respondents. The use of surveys as a data collection method for intersectional analysis has been highlighted in previous research (Fernandez et al., 2016; Hughes et al., 2022; Safer Parks Consortium, 2023).

5.1 Survey Methodology

The survey combines quantitative and qualitative questions with the aim to understand the relationship between user profiles, mobility patterns and their perceptions of safety in the city. It was disseminated in Italian and English, to target diverse populations in the city of Milan. The online survey was active for a duration of 4 months, between July 2023 and October 2023, with the aim to gather at least 500 responses from users of the city of Milan. It was disseminated primarily through social media platforms as well as by email to the researchers' extended networks, targeting active NGOs and community groups in the city of Milan as well as academic networks and Universities. The actual number of responses collected reached 1,808 responses (Italian version: 1762, English version: 46) by the end of the dissemination period.

5.2 Survey Structure

The questionnaire was designed around three main sections: (i) "Walkability in Milan profile" of each respondent collecting mobility data and general perception of safety in the city of Milan; (ii) view on perception of safety in relation to the predefined list of safety factors, elaborated on through a Likert scale, and (iii) extensive demographic data section including information on participants' age, sex, gender, sexual orientation, citizenship, physical and cultural ethnicity, and disability, among others. Some of these aspects were further investigated, and the response to each of these categories was always kept optional. Open-ended questions were also included to gather further insight into unstructured input from the participants.

5.3 Survey Results

Survey responses were cleaned to remove inconsistencies, and English responses were translated to Italian to unify the dataset for analysis. Some analyses cover the full 1,808 responses, while others focused only on responses from people who identified as women or gender minorities (1,667 responses).

Quantitative Analysis
Demographic Analysis
Detailed demographic data was gathered through the survey, and consequently regrouped into combined categories where sample sizes were too small, considering a benchmark

of a minimum of 15 people for each group to ensure representativeness (see Fig. 5). Demographic diversity of the respondent base is limited in several areas, especially with respect to nationality (96% Italian), ethnicity (95% white) and gender identity (92% women and gender minorities). On the other hand, the respondent sample base showed a fairly representative distribution in terms of age, sexual orientation, education level and economic status, as shown in Fig. 5.

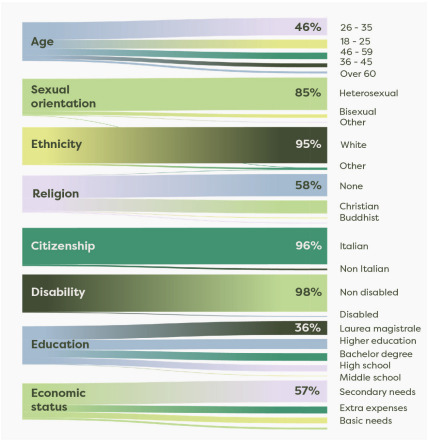

Fig. 5. Demographic distribution of survey participants (Graphics: Amina Salama).

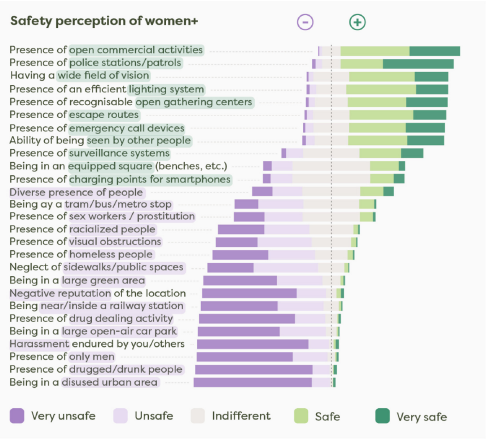

Fig. 6. Safety perception of women and gender minorities by factor (Graphics: Amina Salama)

Mobility Analysis

Within the women and gender minorities participant group, walking and metro were the most popular transport modes used. Around 90% of women and gender minorities reported walking in the city at least 30 min per day, most commonly during the evening

(7–11pm). These results are significant in showing the importance of walking in women and gender minorities' daily routines and thereby, the importance of enhancing their sense of safety around the clock.

Safety Perception Analysis
Results highlight significant difference between perception of safety of women and gender minorities compared to men during both night and day. This difference is especially accentuated during nighttime hours where 57.05% of women and gender minorities report feeling unsafe while walking compared to 8.27% of men. Men are also more likely to go out in the evening and nighttime, and to feel safe doing so. Significant findings regarding women and gender minorities' safety perceptions include:

- A significant portion of women and gender minorities experienced sexual harassment (63%) and violence (17%) while walking, leading to heightened feelings of unsafety when walking alone at night (8 in 10, compared to 6 in 10 for those who rarely or never experienced harassment or violence). This is a personal factor that has an a priori influence on perception of safety irrespective of the physical environment for walking.
- Older women are significantly more likely to have a higher perception of safety both by day and night than younger groups (25.7% of the Over 60 group reported high perceptions of safety, compared to 2.4% of those between 18 and 25 years old). Older age groups are also nearly twice as likely to feel safe or mostly feel safe in their own neighborhoods (Over 60: 54.2%; 18–25: 27.1%).

Safety Factors Analysis
Questions about safety factors were structured according to the Likert rating scale, providing information on the relevance of each factor as well as the connotation (whether positive or negative). Inherent limitations of the Likert scale include the interpretation of the numbers as continuous variables (Jamieson, 2004) (Cohen et al., 2002) and ambiguity in the meaning of the neutral value (Nadler et al., 2015). Figure 6 shows aggregated results for nighttime safety perception by safety factor for all respondents identifying as women or gender minorities. To enable a comparison among different groups, the quantitative analysis was performed by analyzing the average values of the ratings, as visualized in Fig. 7 below.

- Safety factors are generally perceived more neutrally for men compared to women and gender minorities, with a variation of up to ± 1 scale points.

Intersectional differences among women and gender minorities include the presence of police stations/patrols which vary by sexual orientation, "presence of men", "railway station" and "tram/bus/metro station", which vary by age, the possibility to be seen by other people which various by ability/disability.

Qualitative Content Analysis
A qualitative analysis was carried out on questions with open-ended answers, providing insight into (i) factors contributing to a sense of safety in specific areas, (ii) factors contributing to a sense of unsafety in specific areas, and (iii) strategies adopted by



Fig. 7. Intersectional analysis of nighttime safety perception (Graphics: Amina Salama)

respondents to feel safer when walking alone at night. Only responses from women and gender minorities were considered here. The methodology for analysis is divided into two parts: (a) frequency counts relying on automated classification techniques; and (b) a qualitative interpretation of the content through a manual review of the responses.

The frequency count was carried out using the GPT-3.5 Open AI language generation model, recognized as a strong interpretative tool to classify thousands of responses through deep learning techniques. The model was trained on a representative sample selection of 50 responses, and consequently run on the entire list of responses, automatically classifying and labeling the unstructured text data. The GPT model could identify additional factors, expanding on the list of factors from the initial manual selection.

In sum, the results from open-answer questions in the survey largely complimented and reinforced the safety factors identified at the start of the research. Additional themes emerged from these user and respondent inputs that had not been included in the safety factor list derived from the literature. These include specific land uses, such as residential areas; pedestrian areas and high-end retail (all viewed as positive), city use features, such as neighborhood affluence or appearance of affluence (viewed as positive), and road traffic conditions, associated with a negative perception of safety.

Respondents also reported various strategies adopted to feel safer while walking at night. These can be split into three main categories: a) manipulating the urban environment, b) manipulating one's person or changing behavior, and c) social strategies, i.e. reaching out to others for help. It is also noteworthy that several of the strategies mentioned involve the use of some functionality of a personal mobile device (Apps, phone call, location sharing, etc.). This highlights the importance of handheld devices in enhancing women and gender minorities' perceptions of safety while walking in the city, emphasizing the potential role of digital communication and navigation services.

6 Focus Groups

6.1 Methodology

Area Selection Process
Areas selected for focus groups were based on inputs from the Wher App data and the survey responses. Whereas Wher App data gives us more accurate georeferenced data, the input from the survey is based on a wider diversity of respondents. Thus, it is important to consider input from both sources when selecting areas for focus group discussions. Given the lack of congruence between rated areas and NIL divisions, the areas defined for the purpose of this study are based on a wider neighborhood approach.

Three neighborhoods were selected in total, out of which two are among the most negatively perceived (Centrale and Loreto-Padova), and one is among the areas most positively perceived by users (Sarpi). This is done to allow for a cross-examination between both groups. The area selection process was based on the following criteria: (i) data coverage (95th percentile), (ii) score or safety rating, (iii) overlap of Wher and survey results, (iv) population diversity. Areas rated outside the city of Milan and data positioned in non-walkable environments (e.g. highways) were excluded.

Participant Selection Criteria
Participants were selected to encompass a wide diversity of women with respect to age, ethnicity and nationality, with the precondition of residing in or frequenting the neighborhoods under study on a regular basis. Non-heterosexual males were also considered to potentially offer insights aligned with the LGBTQI+ community. Furthermore, participants were required to have previously expressed their perceptions of safety through the survey discussed in Sect. 5. This ensured inclusion of individuals who had already articulated their viewpoints on the subject and have a good grasp of the framework of the project.

Structure of the Focus Group Discussions (FGDs)
The focus group sessions were designed as semi-structured discussions using participatory mapping tools. The sessions lasted about two hours each and involved 4–6 participants in each of three focus areas. The moderators facilitated the discussions to delve into factors that evoke feelings of discomfort or feelings of safety while walking in these neighborhoods as well as brainstorming solutions or potential intervention strategies where needed. Recurring patterns and noteworthy observations are highlighted in the results.

6.2 Focus Group Results

The three focus groups highlighted various positive and negative factors. Negative elements focused on problematizing other person groups, highlighting concerns about male only presence, homelessness, inebriated individuals, and (male) immigrant groups. These perceptions are often exacerbated by biased media portrayals and/or direct personal experiences. Car-centric environments were also highlighted hostile environments that impinge on personal safety. On the other hand, positive factors include vibrant, and

well-maintained spaces, well-lit areas at night, presence of security, pedestrian areas and community activities that foster community engagement. Solutions proposed by participants include revitalizing specific areas, improving lighting, and promoting active citizenship. Design solutions to enhance visibility, and infrastructure improvements such as dedicated cycle paths and parking facilities were also highlighted as measures to enhance safety perception.

7 Conclusions

The approach adopted in this research contributes significantly to enhancing the understanding of safety factors in urban environments, particularly in the context of space perception, city use, and hotspots. The process of defining and gathering potential proxy indicators for location-based data, through a multiple criteria selection, has resulted in a refined dataset of 16 indicators for statistical spatial analysis. The process of defining proxy indicators to represent safety factors posed challenges and limitations. The scarcity of and restricted access to certain proprietary data, and the limitations of spatial localization for others, presented limitations. Despite these challenges, a substantial, relevant list of datasets emerged, providing a foundation for a multivariate spatial analysis for consequent phases of the project.

The integration of technology through the Wher App brings a valuable dimension to the assessment of safety perception. The inclusion of various parameters, along with the option for users to provide free-form comments, enriches the data with geolocalized user-generated insights. Applied textometric and sentiment analysis techniques offer a deeper layer of understanding, unveiling common patterns and relationships within user comments. Sentiment analysis allowed for a nuanced exploration of spatialized sentiment variation of specific themes. The findings suggest that certain factors elicit a higher negative sentiment, providing valuable insights for urban planning and safety enhancement measures. This multi-faceted approach offers a robust framework for identifying common indicators across diverse urban zones.

The online user survey reveals some key findings about women and gender minorities' experiences while walking in the city, especially at night. The findings both confirm assumptions based on previously reviewed literature on the topic as well as highlight new themes, opening new avenues for exploration on the theme of women's safety in the night city. Beyond differences in perceptions of safety between genders, the survey also revealed differences by age, sexual orientation, and disability, emphasizing the importance of intersectionality for gender-responsive urban planning. Qualitative results shed light on additional safety factors that were not considered in the study. These range from urban characteristics to personal, subjective factors, such as proximity to one's home. Despite these findings, the survey's limitations in participant diversity restrict the broad applicability of its results. A more targeted dissemination approach could help to reduce or eliminate these biases.

The focus groups enabled a qualitative investigation providing deeper insight through dialogue and discussion. The restriction of focus group participants to individuals who (i) had completed the survey, and (ii) had a strong connection to the focus areas enhanced and focused the outputs of the FGDs. What emerges is an expertise of the participants

regarding their perceived safety based on their everyday life experience while walking in specific areas of the city. The discussion around potential solutions was successful in directing these findings and local knowledge into actionable outputs. One of the striking results of the FGDs is the alignment that emerged between sustainable mobility goals and perceptions of safety, highlighted through discourses around car-centric environments as hostile elements reducing women and gender minorities' urban sense of safety. Alternatively, pedestrianized areas and enhanced cycling infrastructure were noted as positive contributors to sense of safety. This result is also in line with results emerging from open-ended questions in the survey, which highlighted the role of transport environments in enhancing or inhibiting women's sense of safety.

7.1 Final Discussion

Overall, the data collection approach demonstrated in this paper follows a digital humanities approach that mixes between conventional and digital modes of data production and analysis, providing a solid foundation for the development of a detailed spatial analysis of the perception of safety of women and gender minorities in Milan. This mixed data collection methodology allowed the researchers to integrate between quantitative and qualitative insights from large sets of open and proprietary data, as well as compensate for limitations of data in terms of depth and bias. This also allowed for cross-analysis of data, whereby insights from one data type were leveraged to inform and refine the methodology of data collection from another data source. Finally, the data collected through the site-specific focus group discussions, as well as other qualitative data collection tools allowed for deeper qualitative insights, further used in consequent phases of the project to inform policy guidelines for the city of Milan and recommendations for gender-sensitive urban planning principles.

Acknowledgements. This research was part of a one-year funded scientific research project, titled 'STEP UP - Walkability for Women in Milan", funded by Fondazione Cariplo (Grant No. 2022-1643) under the call for proposals "INEQUALITIES RESEARCH - Generating new knowledge to reduce inequalities". The research was coordinated by Transform Transport in collaboration with TeMA Lab - Università degli Studi di Napoli Federico II, Sex and the City, and Walk21 Foundation. The analyzed data were treated according to the GDPR-General Data Protection Regulation (EU, 2016/679).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Appendix – Scored Dataset List

Table 1. Scored Dataset List

Safety Factor				Selection #0	Selection #1	Selection #2		Selection #3	
Level 1	Level 2	Indicator	Dataset Name	Survey Analysis Indifference	Spatial Consistency	Temporality	Score	Representativeness	Adherence to SF definitions
Spatial Features	Lighting	Count of Public Lighting	Public Lighting	No	x	x	2	x	3
	Openness	Ability to see and move in all directions	Roads Surface	No	x	x	2	No	1
	Visibility "eyes on the street"	Count of F&B	Food and Beverage Services	No	x	x	2	x	3
			Integrated Food and Beverage Services	No	x	x	2	x	2
	Obstructions to visibility	Count of residential buildings	Residential Buildings	No	x	x	2	No	2
		Count of partition elements	Partition Elements	No	x	x	2	No	1
		Count of isolated trees	Trees	No	x	x	2	No	1
	Architectural Barriers	Length of tunnels	Tunnels	No	x	x	2	No	1
		Length of bridges	Bridges	No	x	x	2	No	1
		Average Distance from Crossings	Highways	No	x	x	2	No	2
	Security devices	Count of Security Devices	SOS Devices	No	x	x	2	No	3
City Use	Presence of people	Count of Resident Population	Resident Population	No	x	x	2	x	2
		Count of Working Population (on selected Ateco codes)	Worker Population	No	x	x	2	x	2
		Count of F&B	Food and Beverage Services	No	x	x	2	x	3
			Integrated Food and Beverage Services	No	x	x	2	x	2
		Strava Metro Flows	Yearly Night Strava Users Flow	No	x	x	2	No	2
	Transport hub (bus/tram stop)	Count of Bus and tram stations	Surface Public Transport Stops	No	x	x	2	No	2
	Railway stations	Count of Railway Stations with external buildings	Railway Station Buildings	No	x	x	2	No	3
Hotspots	Businesses	Count of F&B	Food and Beverage Services	No	x	x	2	x	3
			Integrated Food and Beverage Services	No	x	x	2	x	2
	Spaces of gathering	Count of discos and night clubs	Discos and Night Clubs	No	x	x	2	No	3
	Parks	Count of Green Areas	Green Areas	No	x	x	2	x	3
	Open spaces	Count of abandoned and not activated places	Abandoned Buildings	No	x	x	2	No	2
			Areas of degradation	No	x	x	2	No	3
			Parking Lots With Area > 250sqm	No	x	x	2	x	2
		Count of industrial buildings	Industrial Buildings	No	x	x	2	No	2
	Law enforcement units	Count of Enforcement presence	Police Headquarters	No	x	x	2	No	2

References

Berry, D.M.: The computational turn: thinking about the digital humanities. Cult. Mach. **12** (2011)

Carpentieri, G., Guida, C., Gorrini, A., Messa, F., Abdelfattah, L., Büttner, B.: Digital data to support urban planning processes to develop women safety cities: an application to the city of Naples. TeMA – J. Land Use Mob. Environ. **16**(3), 595–608 (2023). https://doi.org/10.6093/1970-9870/10272

Carpentieri, G., Guida, C., Sgambati, S.: Textometric analysis on the ongoing academic spatial planning debate. TeMA-J. Land Use Mob. Environ. **16**, 197–223 (2023). https://doi.org/10.6093/1970-9870/9782

Cohen, L., Manion, L., Morrison, K.: Research Methods in Education. Routledge, London (2002). https://doi.org/10.4324/9780203224342

Fernandez, T., et al.: More comprehensive and inclusive approaches to demographic data collection (2016). https://doi.org/10.18260/p.25751

Hughes, J.L., et al.: Guidance for researchers when using inclusive demographic questions for surveys: improved and updated questions. Psi Chi J. Psychol. Res. **27**(4), 232–255 (2022). https://doi.org/10.24839/2325-7342.JN27.4.232

Jamieson, S.: Likert scales: how to (ab) use them? Med. Educ. **38**(12), 1217–1218 (2004). https://doi.org/10.1111/j.1365-2929.2004.02012.x

Koskela, H.: 'Gendered Exclusions': women's fear of violence and changing relations to space. Geografiska Annaler: Ser. B, Hum. Geogr. **81**(2), 111–124 (1999). https://doi.org/10.1111/j.0435-3684.1999.00052.x

Levy, C.: Travel choice reframed: "Deep distribution" and gender in urban transport. Environ. Urban. **25**(1), 47–63 (2013). https://doi.org/10.1177/0956247813477810

Loukaitou-Sideris, A.: Fear and safety in transit environments from the women's perspective. Secur. J. **27**, 242–256 (2014)

Nadler, J.T., Weston, R., Voyles, E.C.: Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. J. Gen. Psychol. **142**(2), 71–89 (2015). https://doi.org/10.1080/00221309.2014.994590

Roberts, H., Resch, B., Sadler, J., Chapman, L., Petutschnig, A., Zimmer, S.: Investigating the emotional responses of individuals to urban green space using twitter data: a critical comparison of three different methods of sentiment analysis. Urban Plann. (2018). https://doi.org/10.17645/up.v3i1.1231

Safer Parks Consortium: Safer Parks: Improving access for women and girls (2023). https://www.westyorks-ca.gov.uk/media/10332/safer-parks-final-050503-lr.pdf

Scarponi, L., et al.: Thematic review on women's perception of safety while walking in public space: the STEP UP project. Sustainability **15**(21), 15636 (2023). https://doi.org/10.3390/su152115636

Whitzman, C., Legacy, C., Andrew, C., Klodawsky, F., Shaw, M., Viswanath, K.: Building inclusive cities. Women's Safety and the Right to the City (2013)

Zhang, Z., Song, W., Liu, P.: Making and interpreting: digital humanities as embodied action. Human. Soc. Sci. Commun. **11**(1) (2024). https://doi.org/10.1057/s41599-023-02548-3