Analyzing Qualitative Data with AI to Improve the Streets of Harlem

Kofi Boakye, Rajesh Banik, Yumna Al-Sati, Simanto Sen

Center for Smart Streetscapes (CS3)
My Streetscape Summer Research Institute 2026

INTRODUCTION

Traditional research methods of data collection from qualitative research have been time consuming and tedious whether through note taking, the use of voice memos, or manual analysis. As a result, such methods may be exhausting and cause errors in various aspects. This report provides insights on qualitative research methods of data collection. The project studied in this report resulted in the development of a mobile application designed to help researchers collect questionnaire responses related to senior-citizen pedestrian mobility, travel behavior, and perceived safety. Research for this project included long-form interviews with pedestrians combined with their geographic locations. An analysis of our data identifies patterns in pedestrian mobility, accessibility concerns, and safety-related behaviors.

Collecting data on pedestrian mobility will help identify problems in regard to streetscapes within the city. We found that most pedestrians are concerned about the dangers posed by motorcyclists and cyclists, the lack of quality sidewalks, and lack of maintenance for public transportation facilities. These findings suggest that there is a need for stricter regulations for various vehicles on the road, and for constant maintenance of sidewalks and public transportation facilities. Lastly, our project allowed us to see if Artificial Intelligence (AI) can serve as a better tool than traditional research methods to analyze data on pedestrian safety.

LITERATURE REVIEW

We conducted secondary research with the key terms “pedestrian safety” and “analyzing qualitative data with LLMs.” In searching for literature on pedestrian safety, there was little about Harlem specifically. However, there was a substantial amount of research on analyzing qualitative data using LLM’s. In the article, *Integrating AI language models in qualitative research: Replicating interview data analysis with ChatGPT* (2024) by Mohammad S. Jalali and Ali Akhavan, the authors wrote about using AI to transcribe interviews and organize data. This research also speaks to the ethical concerns that come with using these tools. Another article, *Generative Artificial Intelligence in Qualitative Analysis: A Critical Examination of Tools, Trust and Rigor* by Joaquim Jose, Carvalho Proença and Carmen Ramos Vera specifically focused on the ethics of using artificial intelligence.

We expanded our search with the key words “artificial intelligence,” "qualitative data," “interviews,” and “language learning models.” We found many more articles related to the technical dynamics of our project. An article titled *Automated Transcription of Interviews in Qualitative Research Using Artificial Intelligence: A Simple Guide* (2024) by Zubair Mojadedd and Jacob Rosenberg gave tutorials, opinions and actual research done by other organizations. In *The Design and Development of the AI Interviewer System* (2026) by Xinzhou Ye et.al, the authors’ research showed multiple ways AI can be used to improve the efficiency of interviews, while also being able to test the competence of job candidates.

The articles we reviewed regarding the use of AI for the analysis of qualitative data have similar conclusions. They touch on the ethical aspect of using AI for conducting research and they all agree that if used properly, LLMs can be useful and time-saving for researchers.

DATA AND METHODS

Our research project was composed of two portions. The first portion of the project was to collect data on the User Interface (UI) design of our mobile application. The project's second portion was using the mobile application to make a database of insights on pedestrian safety. The research methods we used to create the database were surveys, one-on-one interviews, focus groups, and walkshops. In the collection of data for the second portion of the project, which tested the capabilities of our mobile application, we used the app as well as manual methods of data collection. This portion of the project was designed to compare the use of the app to manual methods of recording interviews, composing transcripts, and coding responses. Our research included 19 respondents to a survey, 8 interviews, 3 focus groups, and 5 walkshops, which are essentially interviews taking place outside while discussing streetscape issues in real time.

Field research for the project's first phase, which focused on improving the UI of the app, consisted of a survey that we administered at 120th Street and Broadway. This first phase of the project collected data on what users of the app would find preferable in terms of the visual display as well as what information participants using the app would be comfortable sharing with researchers. For instance, the majority of survey respondents stated that they would want transcripts of their interviews after the interview concluded (See Figure B.2), which is something we originally did not think would be important to interviewees. The data collected in this first phase of the project expanded our insights to also identify key issues needed to make the app accessible to everyone.

Conducting research in Harlem for the project's first phase allowed us to collect data from respondents across a diverse population, being that the composition of the Morningside Heights neighborhoods is ~34% Hispanic, ~25% White, ~25% Black, and ~12% Asian (NYU Furman Center). Including perspectives from a broad range of people helped to give insight to the development of our app in considering the different ways users understand and interact with technology based on culture and background. For example, associations with different colors and symbols may vary (such as the red octagon that has become associated with "Stop" in the United States, but whose meaning is not universal) as well as how text is read on a screen (as in the case of the the English language, which is read from left to right, while other languages read text from right to left such as Hebrew and Arabic.) Becoming knowledgeable of these potential differences was crucial in order to create an app that is able to be used by everyone.

The second phase of the project focused on creating a database on pedestrian safety. In order to make this database we conducted field research at two sites: the Cannan Senior Center and the A Phillip Randolph Senior Center. We started off at the Canaan Senior Center, located at the corner of 117th Street and 5th Avenue in East Harlem. Research at the Canaan Senior Center was done to test the effectiveness of the app with one-on-one interviews. The second field site for research in the project's second phase was at the A Phillip Randolph Senior Center at 147th Street and Malcolm X Boulevard in upper Manhattan. We used this site visit to test the app not for one-on-one interviews, but for focus groups. Some of the people we interviewed at both of the senior citizen centers have been living in their respective neighborhoods for multiple decades, which gave good insight on how NYC streetscapes have changed over time. The use of the app for these interviews and focus groups did not only supply us with beneficial data around pedestrian safety, but it was also an opportunity to find the weak points in the app and areas that

needed to be worked on such as the microphone being able to register sound at different levels of volume.

We tested the efficiency of our app at both of these senior citizen centers by comparing the processes of recording, transcribing, and coding by the app with that of manual, traditional research methods. Researchers were split into two groups. One group used the app to record interviews and focus groups while the second group used voice memos on an iPhone to record their interviews and focus groups as well as traditional note taking. The first group downloaded the app and uploaded the interview questions into the app. The app was then able to display these questions during the course of the interview and focus group. Once the first group interviewed all of the participants, they could then download the organized transcript from the app and code it easily. The first group was also able to download the audio recordings in order to listen to ensure that the written transcript was transcribed correctly.

The second group simply used the voice memos app on our iPhones and carried a notebook with us for notes. Once we recorded their interviews, the second group downloaded a transcript of the interview that was completed automatically on our iPhones. This second group then organized the transcription, and listened to the audio to ensure the transcript was correct. Finally, we coded the transcription manually.

Both groups then crossanalyzed our respective methods of one-on-one interviews, focus groups, and walkshops. Using the app for the one-on-one interviews worked exceptionally well. It easily transcribed interviews and walkshops from the first group making it easier for the team to go back and look at the data they collected. The app also analyzed the data giving reports on our findings through a code called BERTopic that the LLM uses. BERTopic is an algorithmic coding model where we put our data into code and it then gives us a HTML file of all our data, as well as a visual overview such as charts and graphs. This made data collection easy and simple.

The manual research methods that were used by the second group were easy to set up but did not collect the data properly at all. The second group recorded interviews as voice memos on an iPhone and supplemented this with notetaking by pen and paper. A transcript was given from the recordings but it was unorganized and parts of the conversation were missing making it hard to read and analyze. The second group also had to spend a considerable amount of time listening to the audio to make sure that the transcriptions were correct. The second group ultimately had to use a third party transcription service because the transcription was so inaccurate. Notetaking was also difficult to do during the interviews and focus groups while still paying attention to ensure no data was lost.

There was no way to put the data we had gathered from the iPhone voice memos into BERTopic, so the second group had to organize and code the transcript by hand. In order to ensure that there were no errors, the second group had to continuously check their work, making the process more time consuming. Analyzing the data was also extremely time consuming and difficult. Comparing the two methods used by the two groups of researchers showed how useful the app was for the collection of qualitative data. Even with this, the app had issues in recording one of the focus groups. The focus group session lasted almost an hour and the app did not have the capability to fully process and transcribe an audio file of this length of time, so the app

unfortunately deleted the file. As a result, the manual methods of phone-recorded voice memos and notetaking proved more successful for the conducting of longer focus groups, even though it may have been more time consuming and tedious.

FINDINGS

After conducting field research, we were able to analyze the data, finding many concerns that were frequently mentioned as well as some specific locations in Harlem that interviewees wanted addressed for the improvement of pedestrian safety. Firstly, the dangers posed by the recklessness of motorcyclists and cyclists was a frequently mentioned topic due to the magnitude of speed by many of them. This type of speeding puts many pedestrians in danger, and many interviewees mentioned the danger it poses for cyclists such as collisions with cars. As a result, some potential solutions for this topic include the increase of restrictions and regulations for cycling, and the implementation of licensing. Secondly, another issue mentioned by many interviewees was the lack of quality of sidewalks in New York City. The lack of maintenance and abundance of cracked, uneven, and unkept sidewalks proved to be a huge concern throughout the data collection. These conditions have proved to be a huge safety concern for many pedestrians as shown in our data. Solutions for this issue could be the implementation of constant maintenance and an increase in the frequency in cleanups to keep the sidewalks clean.

Additionally, we also collected data on the opinions of public transportation services such as buses and subways provided by the New York City Metropolitan Transit Authority (MTA). There were about 19 recurring topics, or general suggestions, that stood out in our findings. Firstly, over 20 participants expressed that their preferred mode of public transportation was the bus compared to the subway for various reasons. Our data shows that many people are unhappy with the quality of accessibility for the subway compared to the bus with some interviewees having opted to live in specific areas in order to rely solely on buses for transportation. The lack of elevators and reliance on stairs at train stations has been a huge factor in causing seniors to prefer buses over the subway. For this reason, seniors would rather avoid neighborhoods that are reliant on subway transportation. Many respondents also favored using the bus over the subway because they preferred to be above ground in the case of emergencies that may require urgent exits.

Nonetheless, participants also criticized the bus service throughout the city. To our surprise, 21 participants mentioned that there needs to be stricter regulations in terms of priority seating as it is neglected by many younger people. Our original expectation was that challenges of public transportation for seniors would primarily be around accessibility issues, yet the interviews revealed that there were many social dynamic failures. Younger passengers often did not get up to give their seats to the elderly, and had to be told to do so. This finding highlights a place where societal structures as well as those that are physical play a large factor in senior pedestrian safety. As a result, participants suggested that bus drivers should refuse service whenever someone breaks these laws.

Interestingly, there were about 20 participants who mentioned the lack of punctuality and amount of buses moving in New York City. Wait times from 15 to 30 minutes were mentioned as well as calls for the increase in the number of buses as many get very congested with riders. Additionally, many bus stops do not have benches, leaving senior citizens having to stand

uncomfortably for an extended amount of time while they wait for a bus. For this reason, more buses may help with buses becoming more punctual as well as solve the problem of overcrowded buses.

Lastly, an issue mentioned for both the bus and subway was the lack of security present or difficulties in locating law enforcement at MTA facilities. Due to this, about 24 of our participants expressed the fact that they felt unsafe at times when using public transportation because they were unaware of where they would be able to find police at their respective bus or subway stations.

DISCUSSION

Many of the recurring themes and topics mentioned by research participants would require official regulations or laws. Motorcyclists and other cyclists had already proven to be an issue, as an interviewee expressed that the rules for where bikes are meant to be ridden is unclear. One interviewee stated, “some of them will tell you, oh, get out of my way. But the sidewalk is for pedestrians. Ride the bike in the street because now you're pushing us off the sidewalk where we're supposed to be safe.” As a result, there is a need for stricter regulations to keep bikes and motorcycles on the road, away from pedestrians. However, this issue can be further explored in the sense that many streets do not have bike lanes, causing some cyclists to feel unsafe integrating into the main road with cars. Therefore, a compromise to regulating bikes to ride in the street would need to be the creation of more bike lanes throughout the city. Additionally, bikes and motorcycles are also known for excessive speeding as expressed by one of our participants saying, “So, we're supposed to be able to walk on the sidewalk, but they're riding these bikes and it's not, they're not riding them slowly. They're speeding.” The excessive speeding of motorcycles and bikes have posed dangers for pedestrians as well as for cyclists as they also can be victims of collisions. Overall, stricter regulations to keep cyclists and pedestrians safe would prove to be helpful, in addition to more bike lanes allowing them to use the main road and not the sidewalks.

Another topic that resonated among many of the interviewees was the lack of quality and cleanliness in the streets and in public transportation facilities. As a result, one of the main calls to action from interviewees was constant cleanup and maintenance of streets and facilities. One interviewee said “[m]y recommendation to them would be more maintenance. Every time you have a problem, fix it right away.” The lack of maintenance throughout the city has been a result of extended periods of construction on the streets and at train stations. Although it is true that constant maintenance could improve the quality of facilities throughout the city, for the subways specifically, maintenance has proven to be an issue as 25% of subway delays are a result of maintenance at train stations (DiNapoli, 2025). However, such maintenance would probably result in facilities that last longer, even though it may cause delays in public transportation. Overall, the call to focus on quality is important because constant repairs and maintenance could subdue malfunctions and improve the quality of life in New York City.

A final issue we found in our data is a lack of accessibility. One interviewee who is a senior citizen and was in a car accident within the past year said “[i]t's hard for me to go up and down the steps.” Travel by train is nearly impossible for this interviewee considering how Harlem has

a subway inaccessibility rate of ~71.4%. There are 14 subway stations in Harlem, and out of the 14 are only 4 accessible stations (MTA). Although those who aren't affected may believe something like this does not matter, this issue remains significant to a lot of senior citizens and disabled people who have to work around these challenges of transportation.

CONCLUSION

We learned through this project's research some of the core issues that affect senior pedestrians in Harlem communities. Our data displayed common themes around how to improve senior pedestrians' experience with public transportation in Harlem. Insights from our research can help to generate potential solutions to problems of pedestrian safety. For example, regular city cleanups, construction of elevators, and new laws to further protect pedestrians and other commuters can improve safety within the city. Our research project also found that our app for qualitative research was efficient in the collection of data for one-on-one interviews, but not well for focus groups. The app was extremely useful in our research as it applies to transcription and analysis. Also, by comparing the use of the app with the manual methods of traditional voice memo recordings and note taking, we were able to conclude that although traditional methods of conducting qualitative research can allow researchers to better synthesize and understand their data, the use of LLMs and the BERTOPIC model pave the way for easier analysis of data while also providing visuals that effectively convey research findings. Use of the app for this project allowed researchers to collect and analyze qualitative data more quickly and efficiently than traditional methods. As AI becomes more common in the conducting of qualitative research, our project shows how AI can be used to produce research that aims to improve NYC streetscapes.

Appendix

1. Survey Questions
    a. Do you trust AI helping with analyzing data in apps?
    b. In the app would you want the transcript of the interview?
    c. What difficulties would stop you from using an app?
    d. How much does the color scheme of an app tie into your interest in using it?
    e. Would you use this app with its current interface?
    f. How important do you think street safety is to the well-being of the city?
    g. What personal information are you comfortable giving to an interviewer?
    h. Demographic questions:
        i. Gender
        ii. First Language
        iii. Place of residence?
        iv. Duration of residence?
        v. Disability Status
        vi. Age
    i. We asked out survey questions about the application and the user interface. The data we collected was for feedback and advice so that we could improve and change our app accordingly.

2. Interview Questions
    a. One on one:
        i. What borough and neighborhood do you live in?
        ii. Do you usually walk in your neighborhood and why/why not?
        iii. Is there anything that makes it hard for you to walk in your neighborhood and how so? (e.g., cracked sidewalks, garbage, construction, e-bikes or delivery bikes, flooded curbs, etc.)
        iv. Which mode of transportation in NYC do you prefer and why?
        v. Tell me a story about an issue you had using public transportation?
        vi. Do you feel streets in Harlem are safe for pedestrians to walk in terms of accidents?
        vii. Has there been anything that has happened to you that made you feel unsafe as a pedestrian (e.g., cars, buses, e-bikes, motorcycles, scooters, etc.)?
        viii. How do you think public transportation can be improved?
    b. Walkshop:

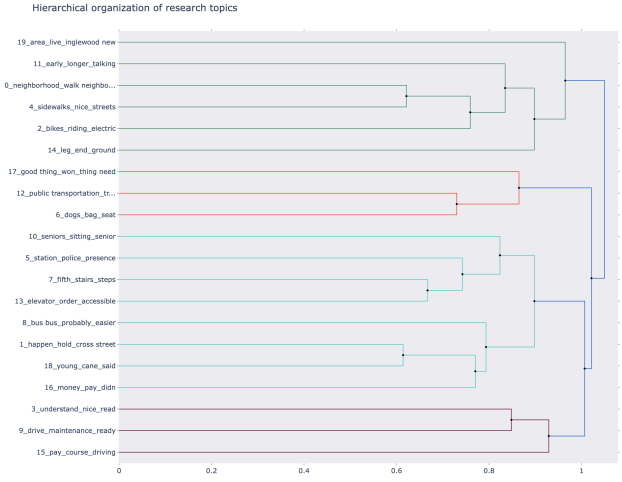
i. How do you think the condition of the sidewalks, crosswalks, and intersections affects your walking experience?
ii. Which travel mode do you usually use to navigate/usually use to navigate the city?
iii. What is something you don't like about public areas such as bus stops, train stations, sidewalks, crosswalks, roads, intersections?
iv. What is one thing you think could improve public transportation, including buses and trains?
v. Where do you think we should have more seating areas and shades (trees) throughout the city?
vi. What additional public services, ramps, pedestrian plazas, crossing buttons, do you think could help pedestrian comfort or safety?
vii. What public services should be more accessible?
viii. Should e-bikes, e-scooters have more regulations?
ix. Do you think intersections make it difficult to walk?
x. Do you think crosswalk time should be longer?
xi. Do we need more bus-exclusive lanes (where only buses can run)?
xii. What do you think about Congestion Pricing implemented on Jan 5, 2028?
xiii. Are you OK to see Waymo Robotaxi moving around the city?

c. Focus Group:
i. Do you usually walk in your neighborhood and why/why not?
ii. Is there anything that makes it hard for you to walk in your neighborhood and how so? (e.g., cracked sidewalks, garbage, construction, e-bikes or delivery bikes, flooded curbs, etc.)
iii. Which mode of transportation in NYC do you prefer and why?
iv. Tell me a story about an issue you had using public transportation?
v. Do you feel streets in Harlem are safe for pedestrians to walk in terms of accidents?
vi. Have any vehicles made you feel unsafe as a pedestrian (e.g., cars, buses, e-bikes, motorcycles, scooters, etc.)?
vii. How do you think public transportation can be improved?

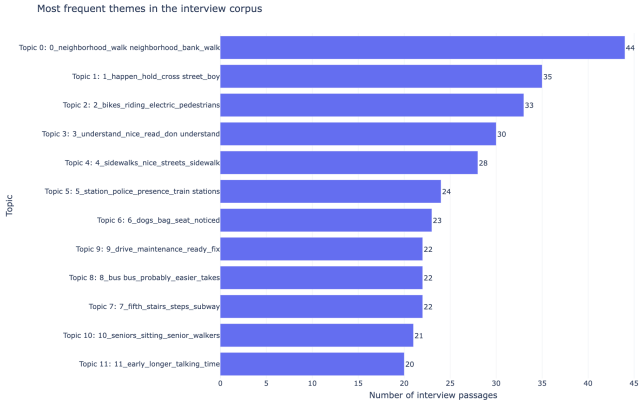
3. Observation Trackers
a. Google sheets for surveys and interviews where all interviews are transcribed and coded
b. Graph data for recurring themes in all the interviews and surveys

4. Charts and Graphs
    a. Data from interviews, focus groups and walkshop:

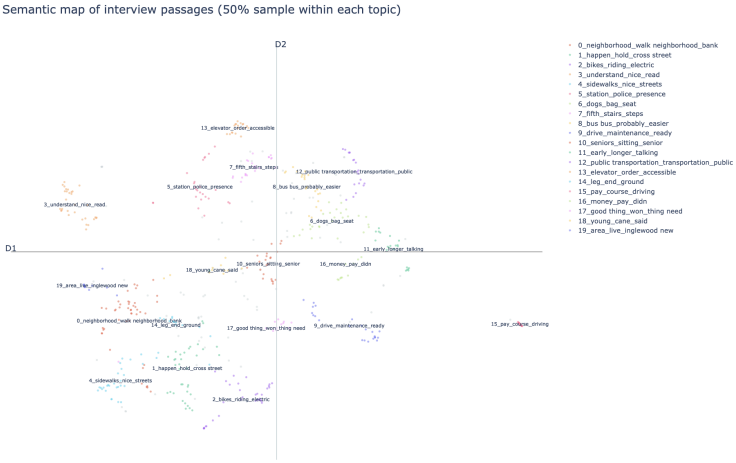
i.



ii.

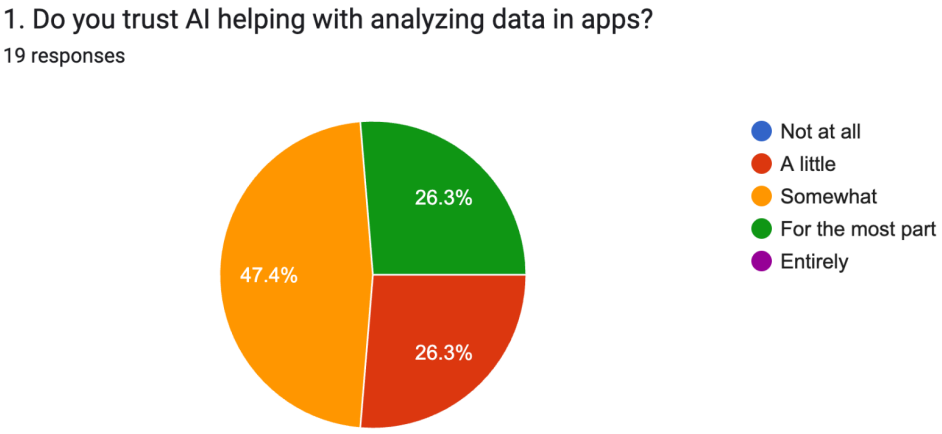


iii.

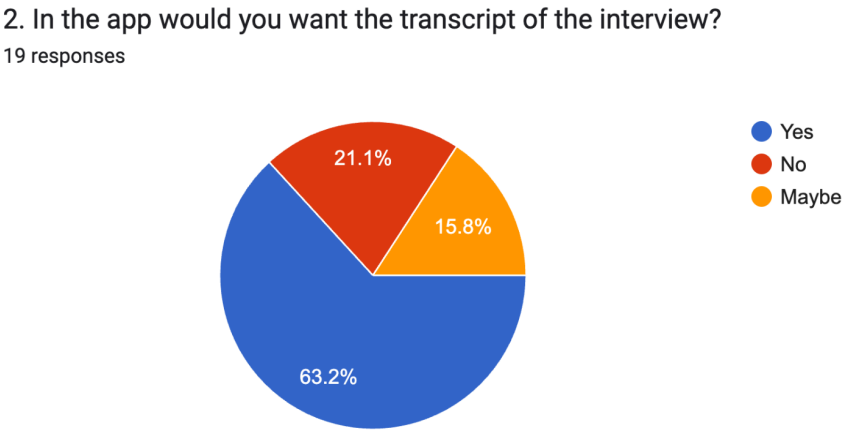


b. Data from survey:

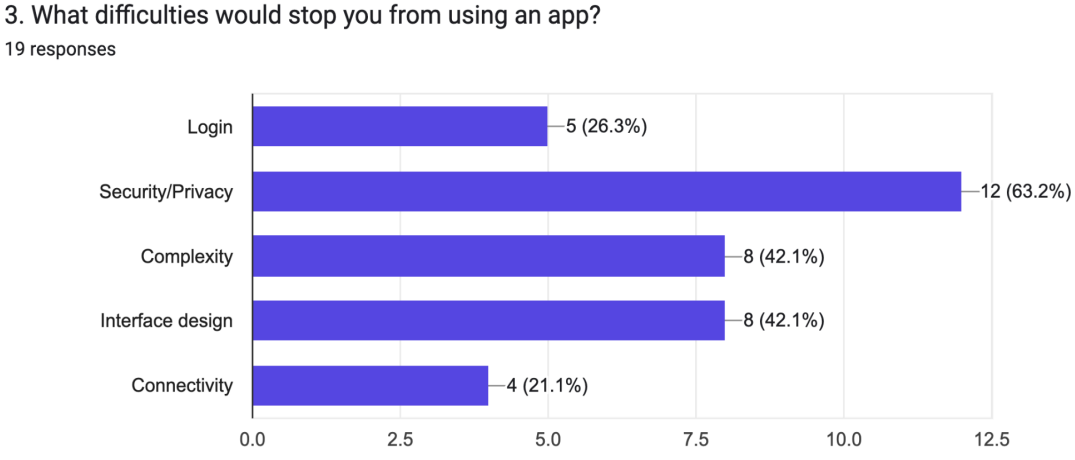
i.



ii.



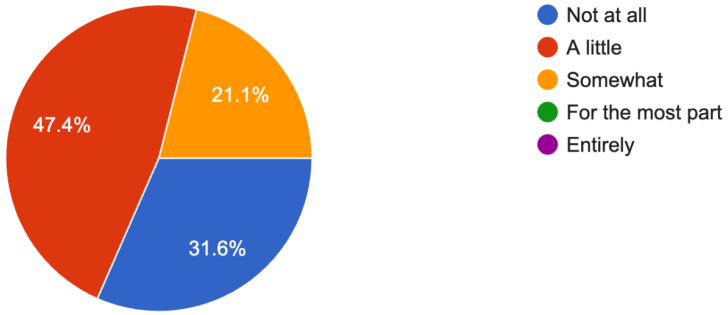
iii.



iv.

4. How much does the color scheme of an app tie into your interest in using it?

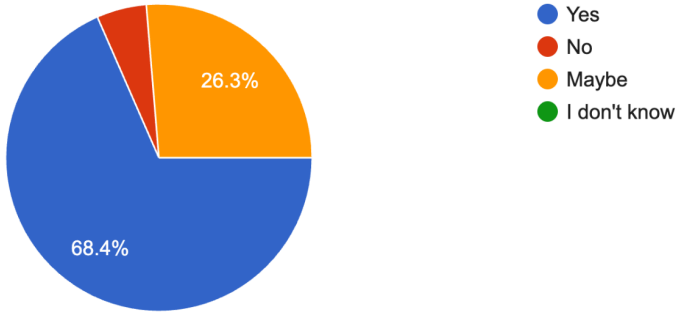
19 responses



v.

5. Would you use this app with its current interface?

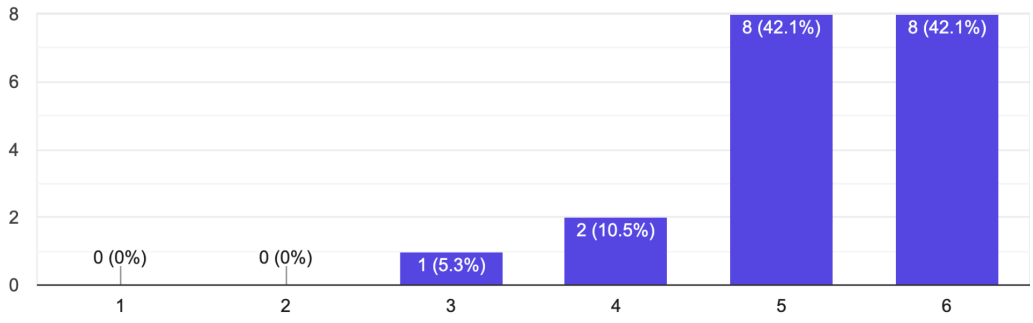
19 responses



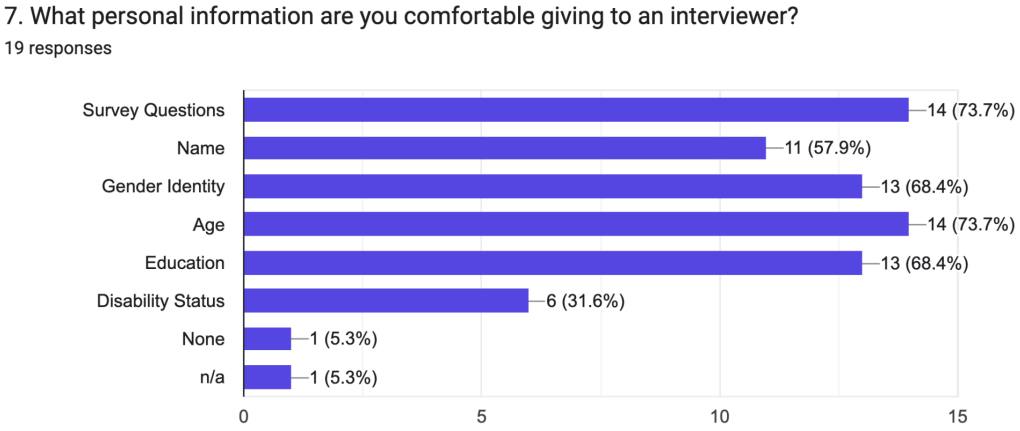
vi.

6. How important do you think street safety is to the well-being of the city?

19 responses

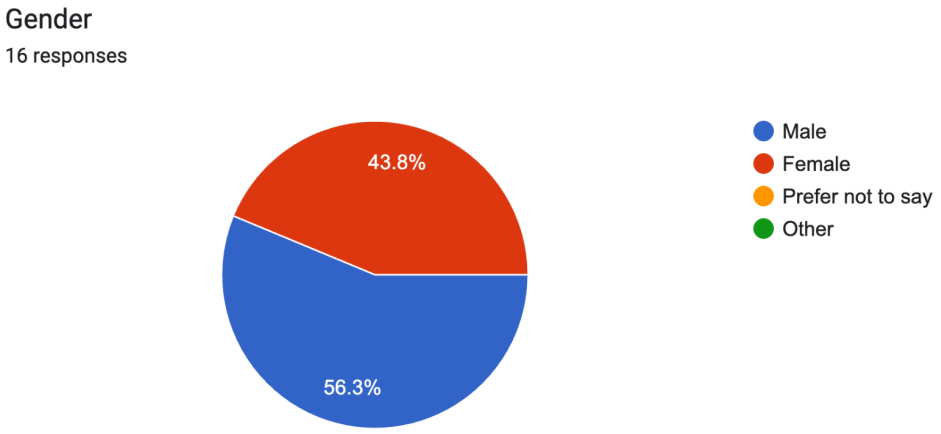


vii.

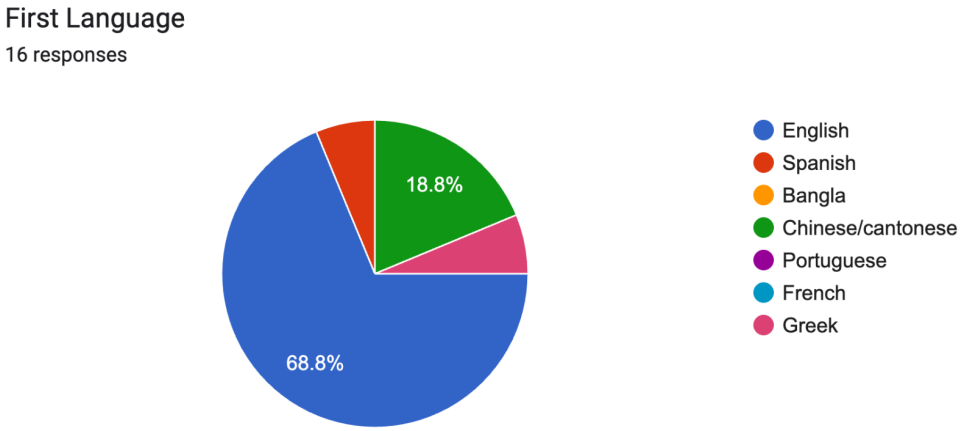


c. Demographic data:

i.

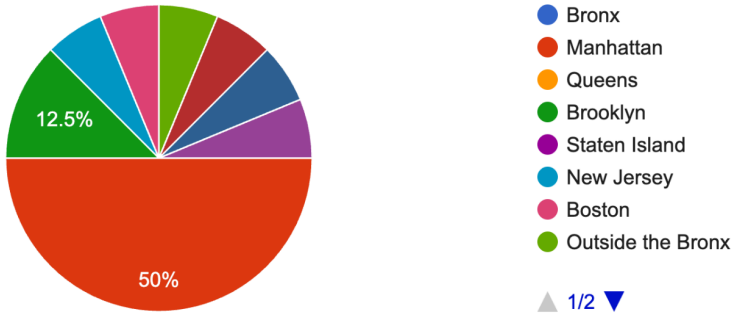


ii.



Place of residence?

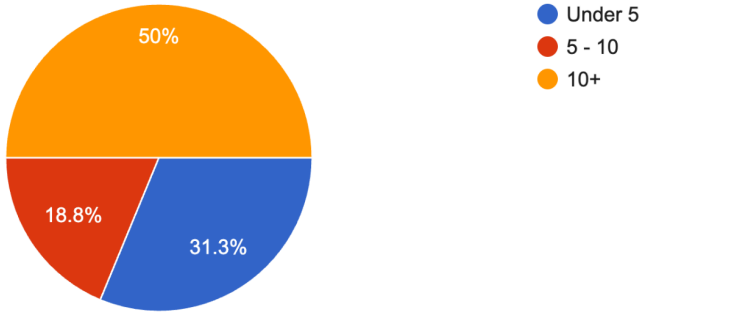
16 responses



iii.

Duration of residence?

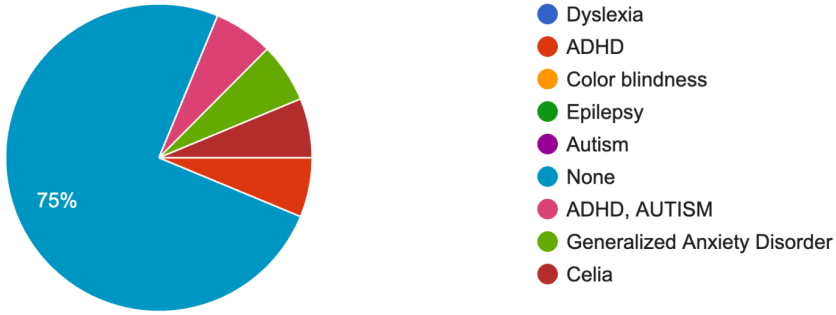
16 responses



iv.

Disability Status

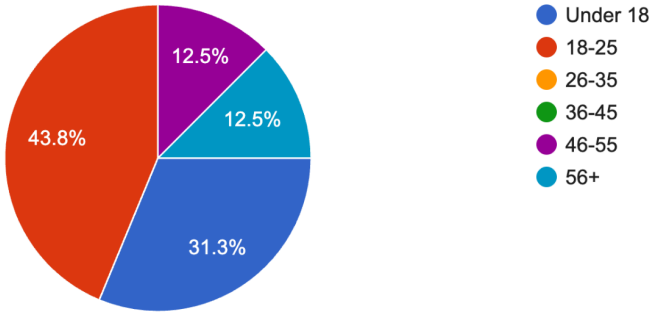
16 responses



v.

Age

16 responses



vi.

Bibliography

"DiNapoli Report Identifies Trends in Causes of Subway Delays." *Office of the New York State Comptroller*, 10 Sept. 2025, ny.gov. Accessed 10 Aug. 2026.

"East Harlem." *NYU Furman Center*, 13 Feb. 2026, furmancenter.org. Accessed 12 Aug. 2026.

Hd, Walji, and Martin-Ucar Ae. *Psychological Factors Affecting Recovery After Rupture of the Anterior Cruciate Ligament in Sports: A Review of Definitions, Impact, Risks and Treatments.* Accessed 13 Aug. 2026.

Jalali, Mohammad S., and Ali Akhavan. "Integrating AI Language Models in Qualitative Research: Replicating Interview Data Analysis with ChatGPT." *System Dynamics Review*, Wiley-Blackwell, May 2024.

Mojadeddi, Z. M., & Rosenberg, J. (2024). "Automated Transcription of Interviews in Qualitative Research Using Artificial Intelligence: A Simple Guide." *Journal of Surgery Research and Practice*, 2024 5(2), 1-6. Accessed 13 Aug. 2026.

"Morningside Heights/Hamilton." *NYU Furman Center*, 13 Feb. 2026, furmancenter.org. Accessed 11 Aug. 2026.

"MTA Accessible Stations." *MTA*, 2025, mta.info. Accessed 12 Aug. 2026.

"The New York City Department for the Aging." *Canaan Senior Service Center*, 2025, canaanssc.com. Accessed 12 Aug. 2026.

Proença, Joaquim Jose Carvalho, and Carmen Ramos Vera. "Generative Artificial Intelligence in Qualitative Analysis: A Critical Examination of Tools, Trust and Rigor." *Prometheus*, vol. 41, no. 1, Pluto Journals, 2026, doi.org. Accessed 13 Aug. 2026.

Ye, Xinzhou, et al. "The Design and Development of the AI Interviewer System." *Communications in Computer and Information Science*, Springer Nature Singapore, 2026, pp. 181–93, doi.org. Accessed 13 Aug. 2026.