

UNIVERSITY OF TARTU
Institute of Computer Science
Smart Mobility Data Science and Analytics

María Alejandra Rodríguez Sánchez

**Crowdsourced Perception of Neighbourhood Safety: A
Case Study Using Airbnb Reviews**

Master's Thesis (30 ECTS)

Supervisors:

Victor Henrique Cabral Pinheiro, PhD

Amnir Hadachi, PhD

Katarzyna Sowinska, MSc.

Tartu, 2025

Crowdsourced Perception of Neighbourhood Safety: A Case Study Using Airbnb Reviews

Abstract:

This thesis explored the potential of crowdsourced data to capture urban safety perception, focusing on Airbnb reviews as a source of subjective insight. The study aimed to assess whether these reviews contain meaningful information about neighbourhood surroundings and whether such perceptions vary spatially across the city. For this purpose, Stockholm was used as a case study, and the analysis was structured into three main stages: neighbourhood topic classification, sentiment analysis, and spatial/comparative evaluation. In the first stage, different unsupervised clustering techniques were tested to distinguish between neighbourhood-related and unrelated reviews. Due to limitations in the clustering results, a heuristic filtering approach was developed using TF-IDF vectorisation, SBERT embeddings, and cosine similarity to extract relevant reviews. The second stage applied three sentiment classification methods (VADER, TextBlob, and SBERT-based clustering) to determine whether neighbourhood-related reviews were positive, neutral, or negative. Most reviews expressed positive sentiment across all models, although classification performance varied depending on the approach. In the final stage, the results were spatially aggregated using a hexagonal grid placed over the map of Stockholm to explore spatial patterns, and then compared with police report data through a Spearman correlation analysis. The results showed weak and statistically insignificant correlations between perceived safety and reported crime incidents, suggesting that Airbnb reviews and police data may be capturing different aspects of urban experience. This shows the value of user-generated content for capturing subjective urban perception, and its limitations when compared to formal safety indicators.

Keywords:

Urban safety perception, Airbnb reviews, sentiment analysis, crowdsourced data, machine learning, spatial analysis, sbert

CERCS: P170 Computer science, numerical analysis, systems, control

Naabruskonna turvatunde rahvahankeandmete põhine käsitus: juhtumiuuring Airbnb arvustuste põhjal

Lühikokkuvõte:

Käesolev magistritöö uuris rahvahankeandmete potentsiaali linnakeskkonna turvatunde tajumise kaardistamisel, keskendudes Airbnb arvustustele kui subjektiivse info allikale. Töö eesmärgiks oli hinnata, kas need arvustused sisaldavad tähenduslikku teavet naabruskonna kohta ja kas turvatunnetus varieerub linna eri osades. Uuring viidi läbi Stockholmi näitel ning analüüs jagunes kolme etappi: naabruskonnateemaliste arvustuste tuvastamine, meeleoluanalüüs ning ruumiline ja võrdlev analüüs. Esimeses etapis testiti erinevaid järelevalveta klasterdusmeetodeid, et eristada naabruskonna-teemalisi ja muid arvustusi. Klasterdustulemuste piirangute tõttu töötati välja heuristiline filtreerimisviis, kasutades TF-IDF vektoriseerimist, SBERT-i ja koosinussarnasust. Teises etapis rakendati kolme meeleoluklassifikaatorit (VADER, TextBlob ja SBERT-i põhine klasterdus), et määrata arvustuste sentiment positiivse, neutraalse või negatiivseks. Kõigis mudelites ilmnis valdavalt positiivne hoiak, kuid mudelite täpsus varieerus. Kolmandas etapis koondati tulemused kuusnurkse ruudustiku abil, mis asetati Stockholmi kaardile, et uurida ruumilisi mustreid, ning seejärel võrreldi neid politseiraportite andmetega Spearmani korrelatsioonanalüüsi abil. Tulemused näitasid nõrka ja statistiliselt ebaolulist seost tajutud turvalisuse ja registreeritud kuritegude vahel, viidates sellele, et Airbnb arvustused ja ametlikud andmed kajastavad linna erinevaid kogemuslikke tahke. Töö rõhutab kasutajapõhise sisu väärtust subjektiivsete linna tajude kaardistamisel, kuid toob esile ka selle piirangud võrreldes ametlike turvalisuse näitajatega.

Võtmesõnad:

linnaturvalisuse tajumine, airbnb arvustused, meeleoluanalüüs, rahvahankeandmed, masinõpe, ruumianalüüs, sbert

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Contents

1. Introduction.....	5
2. Background.....	7
2.1. Overview of Study Area	7
2.2. Airbnb Reviews and Complementary Police Report Data	8
2.2.1. Airbnb Reviews – Crowdsourced Data.....	8
2.2.2. Airbnb Reviews – Data Collection Process	10
2.2.3. Police Reports– Complementary Safety Dataset	11
2.2.4. Police Reports – Data Collection.....	12
2.3. Machine Learning Background for Classification and Sentiment Analysis.....	13
Neighbourhood Topic Classification Methods	14
Sentiment Analysis Methods	18
2.3.1. Spatial and Comparative Analysis Methods.....	20
3. Methodology and Experiments.....	23
3.1. Data Preprocessing of Airbnb Reviews	23
3.1.1. Language Filtering	23
3.1.2. Text Preprocessing Pipeline.....	24
3.2. Work framework.....	27
3.2.1. Stage 1: Neighbourhood Topic Classification	27
3.2.2. Stage 2: Sentiment Analysis	30
3.2.3. Stage 3: Spatial and Comparative Analysis	30
4. Results and Discussions.....	34
4.1. Stage 1: Neighbourhood Topic Classification	34
4.1.1. Vectorisation	34
4.1.2. Truncated SVD Component.....	34
4.1.3. K-Means.....	36

4.1.4.	BIRCH	38
4.1.5.	Final Review Selection for Sentiment Analysis	39
4.2.	Stage 2: Sentiment Analysis	40
4.2.1.	VADER Sentiment Intensity Analyser	40
4.2.2.	Sentence-BERT (SBERT) Clustering.....	41
4.2.3.	TextBlob	42
4.2.4.	Model comparison	42
4.3.	Stage 3: Spatial and Comparative Analysis	44
4.3.1.	Hexagonal Grid Spatial Framework	44
4.3.2.	Correlation between Sentiment and Safety: Comparative Analysis	45
5.	Conclusions.....	47
	References.....	49
	Appendix.....	52
5.1.	Mapping of Original Crime Types to Grouped Categories	52
5.2.	Full Text of Sample Airbnb Reviews	54
5.3.	Enlarged Spatial Sentiment Maps.....	55
5.4.	Enlarged Police Report Category Maps.....	58
	License	59

1. Introduction

Understanding how people perceive safety in urban environments is crucial for promoting liveable, inclusive, and resilient cities (Qin et al., 2025). While traditional sources such as police reports provide factual accounts of crime and public safety incidents, they often overlook the subjective experiences that influence how safe individuals feel. This discrepancy (often referred to as “perception bias”) has been shown to produce mismatches between reported crime rates and individuals’ visual perception of safety in urban environments (Liu et al., 2021). In recent years, the rise of crowdsourced data has opened new possibilities for capturing public perception at a more granular and human-centred level (Huang et al., 2024). Crowdsourced data has thus emerged as an alternative for assessing human perception, as it eases the collection of subjective feedback and enhances inclusivity (Jacoby et al., 2024).

Among the many sources of user-generated content, Airbnb reviews represent a valuable dataset for exploration. While these reviews are primarily intended to describe short-term stays and the experience within the rented property, they often include remarks on the surrounding area, such as comments on safety, walkability, or the overall atmosphere of the neighbourhood (Culotta et al., 2023). This thesis, developed in collaboration with Your Way Home, investigates how Airbnb reviews can be leveraged to extract and classify perceptions of urban safety, and how these perceptions vary across space.

The primary analysis in this research is based exclusively on open data from Airbnb reviews. As a complementary step, police report data was used to explore whether patterns of perceived safety in the reviews align with the spatial distribution of officially reported crime incidents.

The main objective of this thesis is to evaluate the potential of using Airbnb reviews to extract and analyse perceptions of neighbourhood safety. The study applies natural language processing and unsupervised learning techniques to identify neighbourhood-related content, classify sentiment, and explore spatial patterns. The broader aim is to contribute to the understanding of urban safety perception through the use of publicly available, user-generated data.

To address the objective of this thesis, the following research questions were proposed:

- Can machine learning models accurately classify whether user reviews refer to neighbourhood environments?
- How do the overall sentiments vary across different neighbourhood?
- Are there spatial patterns in neighbourhood sentiment that correlate with objective safety indicators, such as the number of police reports?

The methodological approach was structured into three main analytical stages. **Stage 1** identified reviews referring to the neighbourhood environment. **Stage 2** applied sentiment analysis to those filtered reviews. These two stages form the core of the analysis. **Stage 3** was developed as a complementary step, introducing police report data for a comparative spatial analysis and exploring whether subjective perceptions align with objective safety indicators.

The structure of the thesis is as follows:

- **Chapter 2 – Background:** Introduces the theoretical bases and relevant literature for the study, including key concepts, description of the data sources, and an overview of sentiment analysis and clustering methods.
- **Chapter 3 – Methodology and Experiments:** Presents the three-stage analytical framework followed in this research: **Stage 1**, the classification of neighbourhood-related reviews using a clustering-based approach; **Stage 2**, sentiment analysis using VADER, TextBlob, and SBERT-based models; and **Stage 3**, spatial aggregation and a complementary comparative analysis with police report data.
- **Chapter 4 – Results and Discussions:** Summarises the outcomes of each stage, including classification performance, sentiment and spatial patterns, and the comparative findings.
- **Chapter 5 – Conclusions:** Concludes the thesis by highlighting key findings, discussing methodological limitations, and suggesting possible directions for future research.

Generative artificial intelligence, specifically ChatGPT, was used to improve the readability and clarity of certain paragraphs in this thesis. It was employed for language editing purposes, such as refining grammar and improving sentence flow. Additionally, ChatGPT was used as a support tool during the coding process, helping to troubleshoot Python scripts and improve code readability.

2. Background

This chapter outlines the theoretical background behind this thesis, beginning with an overview of the study area and the data sources used. Then, it introduces the analytical framework across the three stages: **Stage 1** classification of neighbourhood-related reviews, **Stage 2** sentiment analysis, and **Stage 3** spatial and comparative evaluation using police report data. Although Stage 3 extends the analysis scope, it builds upon insights generated from the previous stages.

2.1. Overview of Study Area

The present study focuses on Stockholm, Sweden. This area was chosen as the case study city, as Your Way Home, the startup with which this thesis was developed in collaboration, is currently operating in this city. This focus ensures that the findings could be directly relevant to the startup's ongoing and future projects, particularly those related to user perception and urban safety analysis in Stockholm. Furthermore, Stockholm's focus on increasing mobility for pedestrians and cyclists (City of Stockholm, 2012) make it an ideal setting for a study involving crowdsourced perception data and spatial safety indicators.

The study area includes the 13 municipal districts that form the administrative core of Stockholm, such as Södermalm, Norrmalm, Kungsholmen, and Östermalm, as well as adjacent areas like Solna and Norra Djurgården, as shown in Figure 1.



Figure 1. Study Area Overview

2.2. Airbnb Reviews and Complementary Police Report Data

This section introduces the two main datasets used in this research: Airbnb reviews, which served as the primary source of analysis, and police reports, used in a complementary comparative analysis. Together, they allowed for exploration of how user-generated perceptions (crowdsourced data) align or diverge from officially reported urban safety conditions in the study area.

2.2.1. Airbnb Reviews – Crowdsourced Data

The primary data used in this study was the Airbnb Reviews, a non-traditional, crowdsourced data source. It was explored to understand how user-generated content can support urban analysis, particularly in identifying perceptions of safety.

To define the nature of this dataset, the following definition of crowdsourcing was adopted:

“Crowdsourcing involves gathering work, information, or opinions from a large group of people who submit their data via the Internet, social media, and smartphone apps” (Hargrave, 2025).

Under this definition, Airbnb reviews provide valuable insights into how individuals experience their staying experience. In this research, a key focus was to explore whether these reviews also reflected perceptions of the surrounding neighbourhood or urban environment. Specifically, the analysis aimed to assess the extent to which Airbnb comments—written initially about a lodging experience—could be reliably used to capture and describe broader neighbourhood-level perceptions.

However, it is important to note some limitations of crowdsourced data. The total number of users and producers of crowdsourced data represents only a small fraction of the population (Niu & Silva, 2020). In the case of Airbnb, reviews are primarily written by short-term visitors, not long-term residents, which may potentially bias the data towards tourist-centric perceptions. These perspectives may not fully reflect the day-to-day experiences of local communities.

For example, in the dataset used in this study (explained in detail in Section 3.1), most listings had an average minimum stay of between 1 and 17 nights, as shown in Figure 2. However, the average maximum stay ranged predominantly between 338 and 375 nights, suggesting that while most guests stay for short durations, listings may allow longer stays Figure 3.

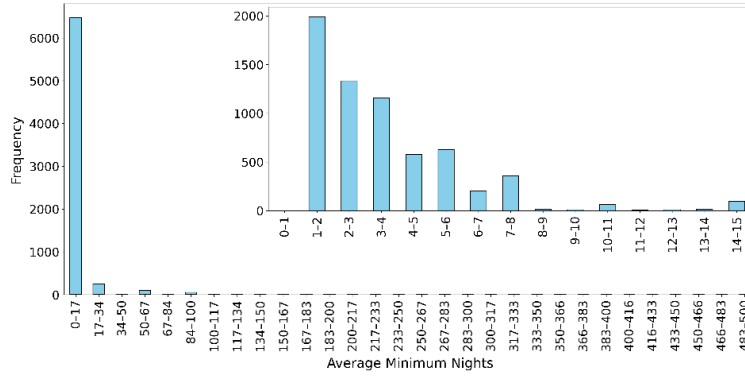


Figure 2. Distribution of Average Minimum Nights in the Full Dataset

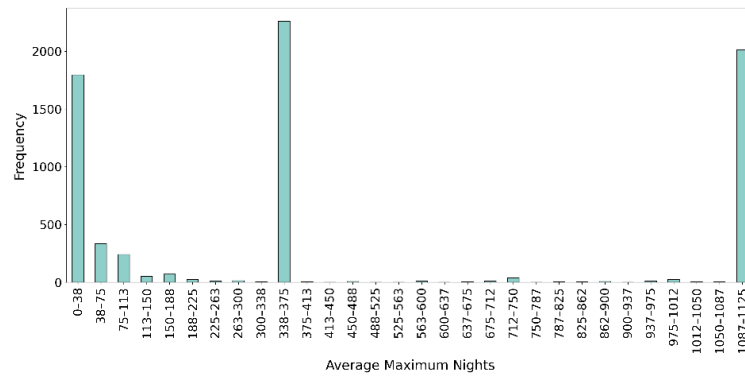


Figure 3. Distribution of Average Maximum Nights in the Full Dataset

To verify whether this pattern held in the final cleaned dataset, a similar analysis was conducted. The results showed consistency; most listings still had a short average minimum stay of fewer than 17 nights Figure 4. However, for maximum nights, a broader range was observed, with most of the listings having bookings between 1,087 and 1,157 nights Figure 5.

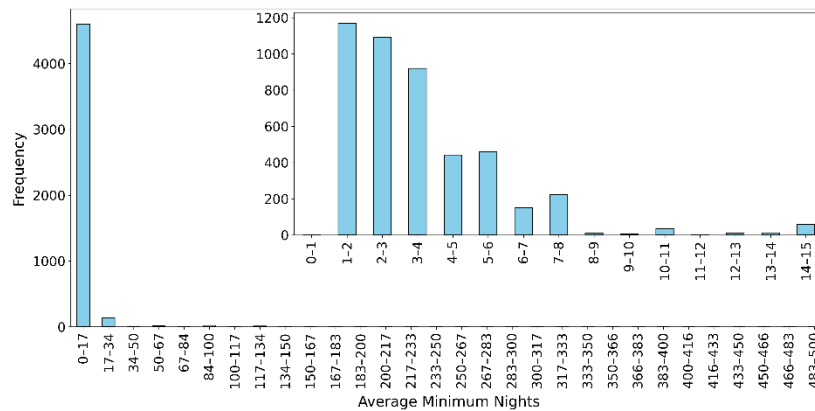


Figure 4. Distribution of Minimum Nights in the Cleaned Dataset

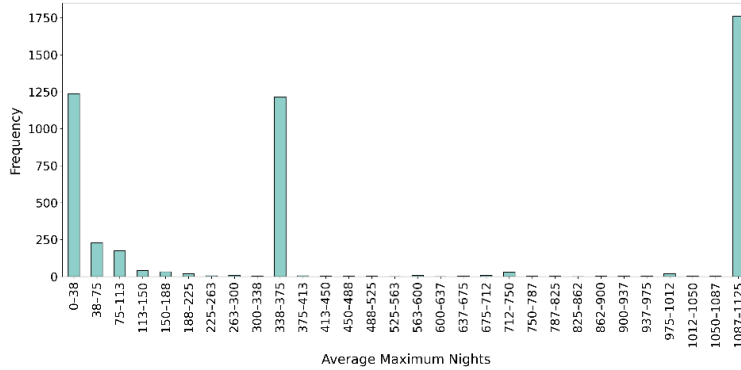


Figure 5. Distribution of Maximum Nights in the Cleaned Dataset

2.2.2. Airbnb Reviews – Data Collection Process

For this research, data was extracted from the publicly available Airbnb Open Data Portal (Inside Airbnb). The analysis focused exclusively on listings located in Stockholm.

The dataset consisted of four data snapshots corresponding to four different periods between September 2023 and June 2024. Each snapshot was stored in a separate folder and included four compressed .gz files: calendar.gz, listings.gz, neighbourhoods.gz and reviews.gz.

A brief description of each file is presented below:

- **Calendar:** Records the pricing, daily availability, and other details from the listing's calendar for each day of the next 365 days (Inside Airbnb).
- **Listings:** Includes detailed metadata for each property, including listing ID, room type, price, host details, and geographic coordinates. It is important to note that the geographic coordinates provided by Airbnb are anonymised for privacy purposes. The location shown in the data set can deviate up to 150 meters from the actual address, following Airbnb's internal privacy policies.
- **Neighbourhoods:** Provides the names and spatial boundaries of the neighbourhoods in GeoJSON format. This file is useful for mapping and spatial aggregation of listings and reviews.
- **Reviews:** Includes user-generated textual reviews for each property, along with the corresponding listing IDs and timestamps. This file served as the primary source for textual and sentiment analysis.

Although four datasets were available per snapshot, only the reviews and listing files were used in the analysis. These were linked using the `listing_id` field, which enabled the geolocation of reviews for the spatial analysis conducted in the final stage of the methodology.

The dataset contained 469,535 rows (each corresponding to a single review) and six features: `listing_id`, `id`, `date`, `reviewer_id`, `reviewer_name`, `comments`. For the purpose of this analysis, only two features were retained: `listing_id` and `comments`.

After removing duplicate entries across different snapshots, 145,610 unique reviews were retained from the listings file. Figure 6 provides an overview of how these reviews are distributed throughout the city.

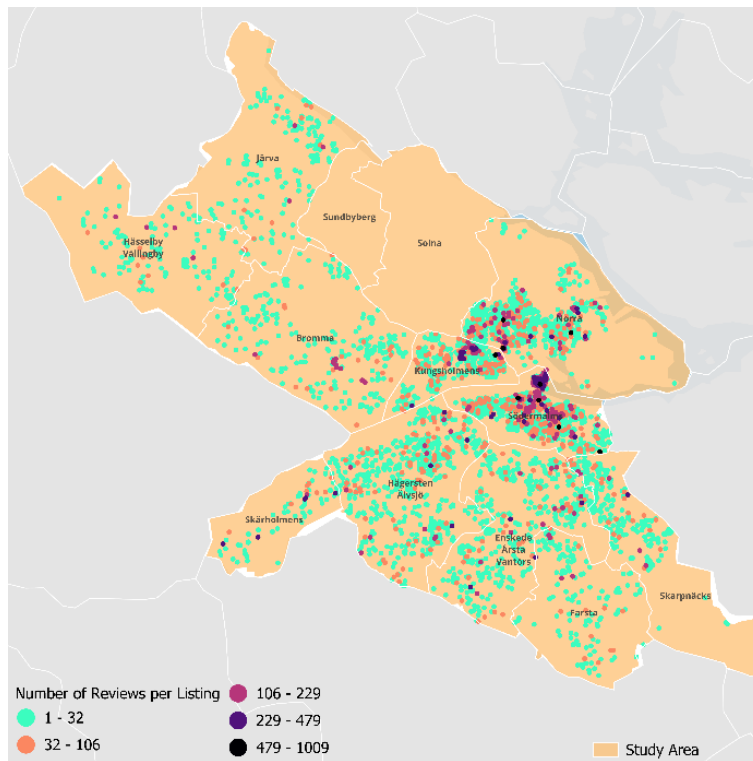


Figure 6. Spatial Distribution of Airbnb Reviews in Stockholm

The next steps followed during the data preprocessing stage are described in Section 3.1.

2.2.3. Police Reports– Complementary Safety Dataset

As mentioned earlier, this study primarily focused on analysing Airbnb reviews to capture a subjective view of neighbourhood safety perception. To complement this analysis, police reports were incorporated as an objective reference for actual incident data. Although they may not capture

every incident occurring in the city, they provide insights into events that have impacted the community and have been reported to or by the police. In this way, they are considered a reliable source for assessing patterns in urban safety over time and across locations.

2.2.4. Police Reports – Data Collection

Police reports were retrieved using a public API provided by <https://brottsplatskartan.se>. This platform publishes real-time police report data, including timestamps, geolocation, a headline, and a description of the incident. Prior to using the data, Your Way Home contacted the website administrators to inform them of the intended academic use and received confirmation for non-commercial usage.

The data retrieval was performed in Python using a simple API request structure. The study area was limited to Stockholms län, and the requests were paginated to ensure the complete collection of events. In total, 27,918 reports were retrieved, encompassing 94 distinct crime types. The spatial distribution of these reports across Stockholm is shown in Figure 7. It is important to note that any data falling outside the defined study area was excluded from the analysis.

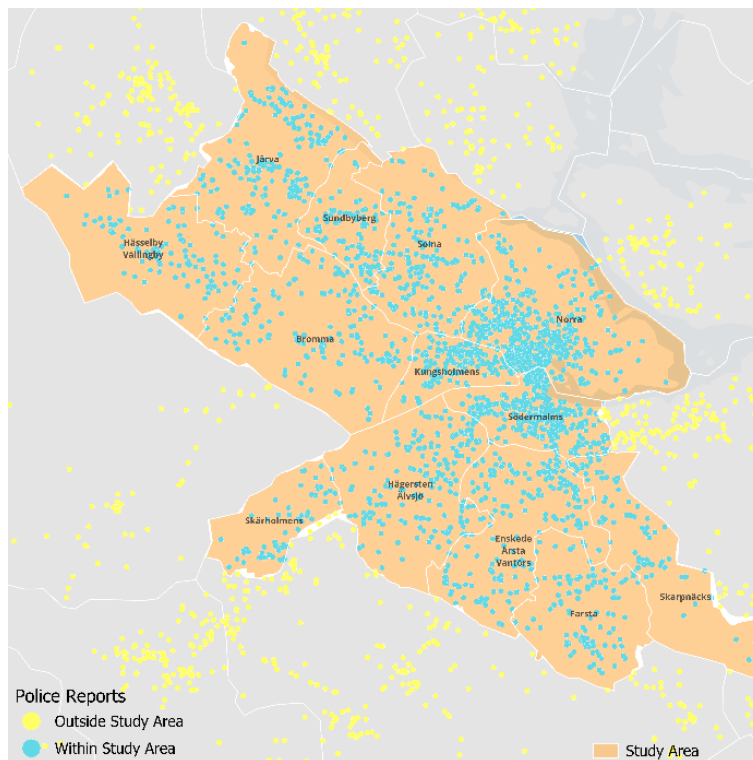


Figure 7. Spatial Distribution of Police Reports in Stockholm

The dataset includes reports dated between 2 February 2018 to 17 November 2024. The most frequent crime types are presented in Figure 8 and the temporal patterns are shown as daily and hourly distributions in Figure 9 and Figure 10, respectively.

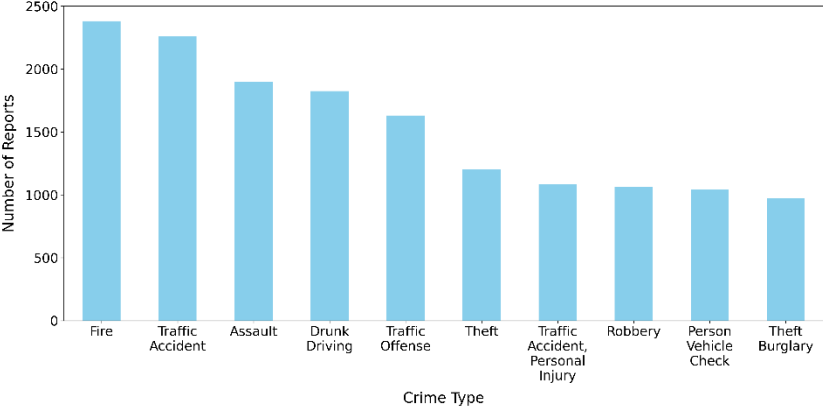


Figure 8. Most Frequently Reported Crime Categories

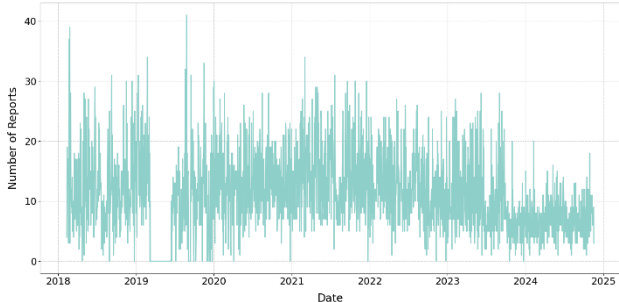


Figure 9. Day Distribution of Crime Reports

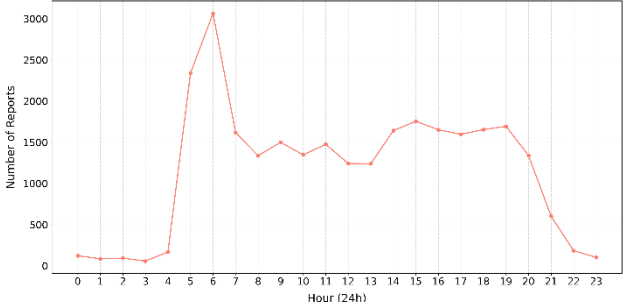


Figure 10. Time Distribution of Crime Reports

2.3. Machine Learning Background for Classification and Sentiment Analysis

The following subsections presents the theoretical and algorithmic foundations of the analytical pipeline used in this thesis. This section focuses on the background section knowledge and justification for the machine learning techniques applied, which served as the ground floor for the methodological implementation detailed in Chapter 3.

Specifically, this section introduces the key principles behind the classification of neighbourhood-related content, sentiment analysis, and spatial aggregation. Each subsection outlines the concepts and models used to extract meaning from Airbnb reviews, structure unlabelled data, and explore spatial safety perception

Neighbourhood Topic Classification Methods

The first stage focused on identifying whether each Airbnb review referred to the surrounding neighbourhood or solely to the accommodation itself. This distinction was essential for filtering the dataset into neighbourhood-related and non-neighbourhood-related content.

As the reviews lacked predefined labels and were only cleaned for language and formatting (as described in Section 3.1), an unsupervised learning approach was used to discover underlying patterns without manual annotations. Thus, various clustering methodologies were explored, starting with the vectorisation and dimensionality reduction of the reviews, followed by the application of the K-Means and BIRCH algorithms. For each algorithm, different parameters were tested to understand their impact, and various strategies were explored to optimise clustering performance.

Vectorisation

Since raw text data cannot be directly processed by most machine learning algorithms, which require numerical input, vectorisation is a key step in natural language processing (NLP) (Abubakar et al., 2022). Vectorisation transforms each review into a numerical feature space, enabling the measurement and modelling of semantic similarity between texts.

For this purpose, two methodologies were analysed: Term Frequency-Inverse Document Frequency Vectoriser (TF-IDF Vectoriser) and Count Vectoriser.

- **Count Vectorizer:** This approach converts text into fixed-length vectors by counting the number of times each word appears in a comment (Shandeep, 2019).
- **TF - IDF Vectorizer:** This approach also produces fixed-length vectors but incorporates two steps. The first (Term Frequency) measures how frequently a term occurs in a document, while the second (Inverse Document Frequency) evaluates the importance of a term by penalising words that are overly common across the entire dataset (Shandeep, 2019).

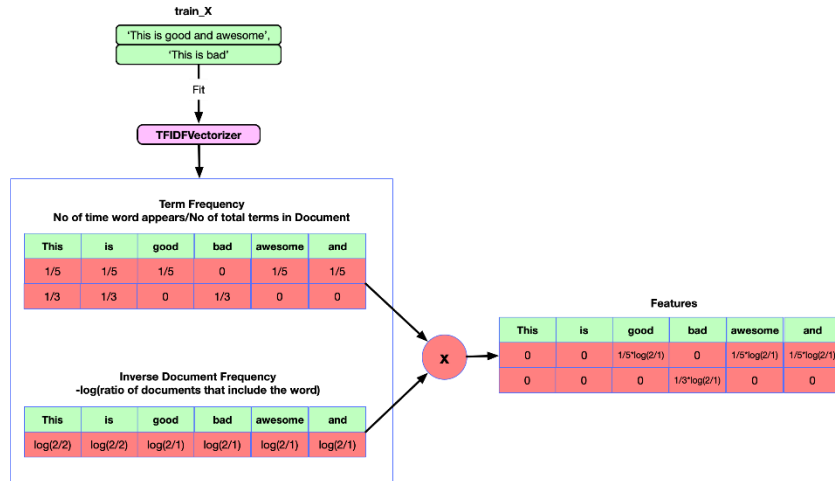


Figure 11. TF-IDF Vectorizer methodology (Shandeep, 2019)

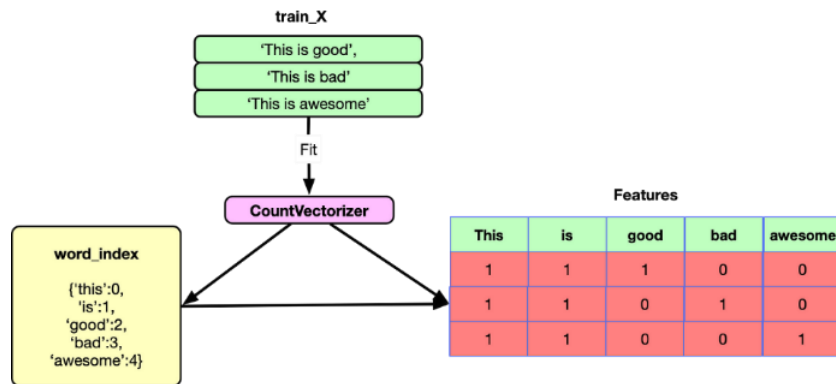


Figure 12. Count Vectorizer methodology (Shandeep, 2019)

As shown in Figure 11, TF - IDF Vectorizer normalises these frequencies by penalising overly common terms by measuring their significance, while, as shown in Figure 12, Count Vectorizer produces a raw word frequency.

The TF-IDF score is computed as the product of Term Frequency (TF) and Inverse Document Frequency (IDF), as shown below:

$$TF - IDF = TF \times IDF$$

$$TF - IDF = TF \times \log\left(\frac{N}{DF}\right)$$

Where, **TF** is the term frequency of a word in a document, **N** is the total number of documents in the corpus, **DF** is the document frequency of a word in the corpus (i.e., the number of documents that contain the word).

While Count Vectorizer is simple and effective for capturing word frequency, it treats all terms equally and often overemphasises common, less informative words (e.g., “place”, “stay”). TF-IDF, on the other hand, balances frequency with distinctiveness by penalising overly common terms across the dataset. This makes it better suited for tasks like clustering and topic detection, where the representation of distinctive terms plays a key role in grouping semantically meaningful content (Sampath & Vignesh, 2024). Based on this, TF-IDF was selected for the analysis.

Table 1 illustrates the result of applying the TF-IDF vectorisation process to three sample Airbnb reviews. As can be observed, each row corresponds to a review, and each column corresponds to one of the extracted terms.

Table 1. TF-IDF Weighted Feature Matrix for Three Sample Reviews¹

Review	Features																		
	alright	apartment	appartement	appreciate	bakery	bed	bedroom	block	bright	clean	comfortable	grocery	includes	incredibly	kitchen	large	look	lot	lovely
1	0,228	0,000	0,228	0,000	0,000	0,000	0,000	0,000	0,000	0,174	0,000	0,000	0,228	0,000	0,000	0,000	0,000	0,000	0,000
2	0,000	0,349	0,000	0,000	0,116	0,116	0,116	0,000	0,116	0,088	0,232	0,088	0,000	0,116	0,088	0,000	0,116	0,088	0,116
3	0,000	0,000	0,000	0,231	0,000	0,000	0,000	0,231	0,000	0,000	0,000	0,175	0,000	0,000	0,175	0,231	0,000	0,351	0,000

Truncated SVD Components

Following the vectorisation process, the resulting TF-IDF matrix tends to be very large due to the natural characteristics of textual data. Long reviews, combined with varied vocabulary, lead to matrices with a high number of features. To reduce this complexity while retaining as much meaningful variance as possible, Truncated Singular Value Decomposition (Truncated SVD) was applied.

As defined by (Khadka, Truncated SVD, 2020):

¹ The full text of the reviews used in this example is provided in Appendix 6.2

“Truncated SVD is a popular technique in machine learning for reducing the dimensions of high-dimensional data while retaining most of the original information. This technique is particularly useful in scenarios where the data has a large number of features, making it difficult to perform efficient computations or visualize the data.”

A key parameter in this process is the number of components, which determines how many dimensions (features) are retained in the reduced matrix. Fewer components mean greater compression, but potentially at the cost of information loss; more components retain more variance but reduce dimensionality less effectively.

Table 2 illustrates the result of applying Truncated SVD to the TF-IDF matrix from the earlier example (from Table 1), using a reduced dimensionality of three components.

Table 2. Example Output of Truncated SVD with 3 Components

Review	Component 1	Component 2	Component 3
1	0,2822	0,9546	0,0958
2	0,7109	-0,2597	0,6536
3	0,7298	-0,1162	-0,6738

Each row corresponds to a review, and each column represents one of the semantic components extracted from the original high-dimensional feature space, and each component is a combination of words that tend to appear together in reviews. The numerical values indicate how strongly each review aligns with the corresponding component, helping to capture key patterns in the data.

K-Means

With the final vectorized and truncated matrix, the K-Means algorithm was applied to partition the reviews into clusters. The K-Means algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares (Pedregosa et al., 2011).

In order to check the clustering performance, it was used the Silhouette Score, as explained by (Lovmar et al., 2005):

In a Silhouettes calculation, the distance from each data point in a cluster to all other data points within the same cluster and to all data points in the closest cluster are determined. Thus Silhouettes provides a measure of how well a data point was classified when it was

assigned to a cluster by according to both the tightness of the clusters and the separation between them.

The coefficient's range is between -1 and 1, where higher values indicate a better clustering.

BIRCH

The second clustering algorithm explored was Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH). According to (GeeksforGeeks, 2020):

“BIRCH is a clustering algorithm that can cluster large datasets by first generating a small and compact summary of the large dataset that retains as much information as possible. This smaller summary is then clustered instead of clustering the larger dataset.”

BIRCH is particularly suitable for very large datasets, as it incrementally and dynamically builds a Clustering Feature (CF) tree to summarise the data. A key parameter in BIRCH is the threshold, which controls the maximum radius of subclusters: lower threshold values produce more compact and homogeneous clusters, whereas higher threshold values allow broader, more loosely defined grouping. This flexibility makes BIRCH well-suited for adapting clustering granularity to the specific needs of the analysis (Zhang et al., 1996).

Sentiment Analysis Methods

The second stage aimed to classify the neighbourhood-related reviews into positive, neutral, and negative sentiments. These categories were selected because most text classification models are commonly designed to distinguish between two or three classes (Lagutina, 2019). Moreover, many sentiment analysis studies specifically focus on identifying the polarity of reviews, typically classifying them as either positive or negative (Pang & Lee, 2008).

To achieve this, three sentiment analysis methods suitable for Airbnb review data were explored: VADER, SBERT, and TextBlob. The following subsections explain the theoretical basis of each method used in this stage.

VADER Sentiment Intensity Analyzer

The first methodology explored was the Valance Aware Dictionary sEntiment Reasoner (VADER), as explained by (Youvan, 2024):

“this lexicon is constructed from a combination of lexical features and sentiment-laden heuristics, which are fine-tuned based on human evaluations. It incorporates a comprehensive set of rules that account for grammatical and syntactical nuances, such as capitalization, punctuation, and degree modifiers (e.g., "very," "extremely"), which affect the intensity of sentiment. VADER's simplicity, interpretability, and effectiveness make it an attractive option for sentiment analysis in various applications.”

This method computes the polarity score (i.e. positive/negative), by four different (Ladyous, 2021):

- **pos:** The probability of the sentiment to be **positive**
- **neu:** The probability of the sentiment to be **neutral**
- **neg:** The probability of the sentiment to be **negative**
- **compound:** The normalized compound score which calculates the sum of all lexicon ratings and takes values from -1 to 1

By using the compound score, it can be easily determined the overall sentiment of a sentence, while the individual positive, negative, and neutral scores give a breakdown of the sentiment expressed (GeeksforGeeks, 2024). This score used a threshold as shown in Table 3

Table 3. Polarity Score Label VADER

Score	Label
≥ 0.05	Positive
> -0.05 and < 0.05	Negative
≤ -0.05	Neutral

These thresholds follow the standard VADER implementation and have been shown to effectively separate sentiment polarity in short and informal texts such as reviews (Bonta et al., 2019).

Sentence-BERT (SBERT) Clustering

The next methodology evaluated was the Sentence-Bidirectional Encoder Representations from Transformers (SBERT) model, which was used to generate sentence-level embeddings from the Airbnb reviews. SBERT encodes entire sentences or text fragments into fixed-size vectors, preserving this contextual meaning (Marqo, n.d.). Unlike traditional word embeddings, SBERT captures the semantic structure of entire sentences, making it particularly suitable for comparing, clustering, or classifying short texts (Reimers & Gurevych, 2019).

For this study, a pre-trained variant was used to generate embeddings optimised for clustering and sentiment analysis tasks. These embeddings provided a semantic representation of the reviews, facilitating the detection of variations in sentiment and content

After obtaining the sentence embeddings, K-Means clustering was applied to group the reviews into three clusters. Each cluster was then evaluated using the VADER Sentiment Intensity Analyser, as previously described, to assign a dominant sentiment label (positive, neutral, or negative). This combined approach of using SBERT for embedding generation and clustering for grouping has been adopted in previous studies across various domains, including applications on social media and subreddit datasets (Liu et al., 2021).

TextBlob

The third methodology evaluated was TextBlob, a lexicon-based and rule-based sentiment analysis tool. TextBlob calculates sentiment by using a predefined dictionary of word polarities and simple grammatical rules to assess the overall sentiment of a text. Specifically, it provides two key outputs: a polarity score (measuring how positive or negative a sentence is) and a subjectivity score (measuring how subjective or objective a sentence is) (Loria, n.d.).

The computed polarity scores were classified into sentiment labels based on the thresholds shown in Table 4.

Table 4. Polarity Score Label TextBlob

Score	Label
> 0.1	Positive
≥ -0.1 and ≤ 0.1	Negative
< -0.1	Neutral

These thresholds follow common practice in sentiment analysis using lexicon-based methods. They are applied in previous studies using social media data, such as (Qi & Shabrina, 2023), where similar classification rules were used to label tweet sentiments based on TextBlob polarity scores.

2.3.1. Spatial and Comparative Analysis Methods

The third stage was divided into two parts: spatial analysis and comparative analysis. The spatial analysis focused on examining the distribution of sentiment across the study area, while the comparative analysis aimed to contrast the Airbnb review sentiments with the police reports.

This stage was developed as a further step following the clustering and spatial filtering of neighbourhood-related reviews. The following subsections introduce the spatial framework, which is based on a hexagonal grid system, used to aggregate and visualise sentiment geographically. Additionally, the comparative analysis evaluates the relationship between the sentiment scores and the spatial distribution of reported crime incidents.

Hexagonal Grid Spatial Framework

To spatially aggregate both Airbnb sentiment data and police reports, the study area was divided into hexagonal grid cells. Hexagons were chosen over square grids due to their geometric properties, which provide more consistent spatial coverage with less distortion and fewer edge effects. Unlike squares, hexagons have equal distances to neighbouring cells in all directions and minimise variance in cell area, making them especially suitable for analysing connectivity and movement patterns (Sinha, 2020). These properties are especially relevant in the context of human-scale urban phenomena, such as pedestrian mobility and neighbourhood perception. This approach also aligns with established practices in spatial analysis, as hexagonal grids have been widely adopted in location intelligence frameworks such as Uber’s H3 system (Brodsky, 2018).

Figure 13, illustrates an example of how the study area can be divided using a hexagonal grid structure, which was used as the basis for spatial aggregation in the analysis.



Figure 13. Example of Hexagonal Grid Division in the Study Area

Correlation between Sentiment and Safety: Comparative Analysis

A Spearman correlation analysis was performed to explore the relationship between safety perception and police reports. In this sense, for each hexagon, the average sentiment score derived from Airbnb reviews and the total number of police reports were calculated, and based on these values, the association between them was evaluated.

Spearman's correlation coefficient (ρ , also signified by r_s) measures the strength and direction of association between two ranked variables (Laerd Statistics, n.d.). This correlation was selected over Pearson's correlation, as the data used violated several of the assumptions required for applying Pearson's method, such as (Scribbr, n.d.):

- Both variables are on an interval or ratio level of measurement
- Data from both variables follow normal distributions
- Your data have no outliers
- Your data is from a random or representative sample
- You expect a linear relationship between the two variables

Specifically, the crime count data is discrete, the sentiment scores do not strictly follow a normal distribution, and the spatial sampling of reviews is non-random and biased towards specific areas. Furthermore, the relationship between perceived safety and crime incidence is unlikely to be strictly linear.

3. Methodology and Experiments

This chapter outlines the methodological framework followed in this research. It describes the key steps taken throughout the process, including the data preprocessing, the analytical methods applied, and the comparative analysis conducted in the final stage.

The methodology was structured into three sequential stages, each building on the output of the previous one. First, it was essential to determine whether each review referred to the perceptions of the neighbourhoods or focused on other aspects such as the accommodation or the host. This distinction was crucial, as a negative comment about the host or the accommodation could be misinterpreted as a negative perception of the surrounding area, leading to inaccurate assumptions regarding the surrounding safety perception.

Subsequent stages focused on applying sentiment analysis to the neighbourhood-related review dataset and exploring spatial patterns using a grid-based aggregation approach. Finally, the sentiment results were compared with police report data to examine potential correlations between subjective perceptions and reported incidents.

3.1. Data Preprocessing of Airbnb Reviews

Based on the initial dataset described in Section 2.2, this section outlines the pipeline used to clean, filter, and prepare the Airbnb review data for analysis. The process involved language filtering to ensure consistency, followed by a comprehensive text preprocessing pipeline to transform raw comments into a structured and analysable format.

3.1.1. Language Filtering

As part of the cleaning pipeline, the language of each review was detected using the langdetect (Shuyo, n.d.) algorithm from Google. A total of 44 different languages were identified, with English being the most common, comprising approximately 71.4% of all reviews. To ensure consistency and interpretability in the analysis, only English-language reviews were retained. This resulted in a filtered dataset of 103,942 reviews from 4,821 properties.

Figure 14 shows the difference in the number of reviews per listing before and after filtering for English-language content. This was calculated by subtracting the number of English reviews from

the total number of reviews for each listing. Figure 15 highlights the Airbnb listings that had no English reviews and were excluded from the dataset after the language filtering step.

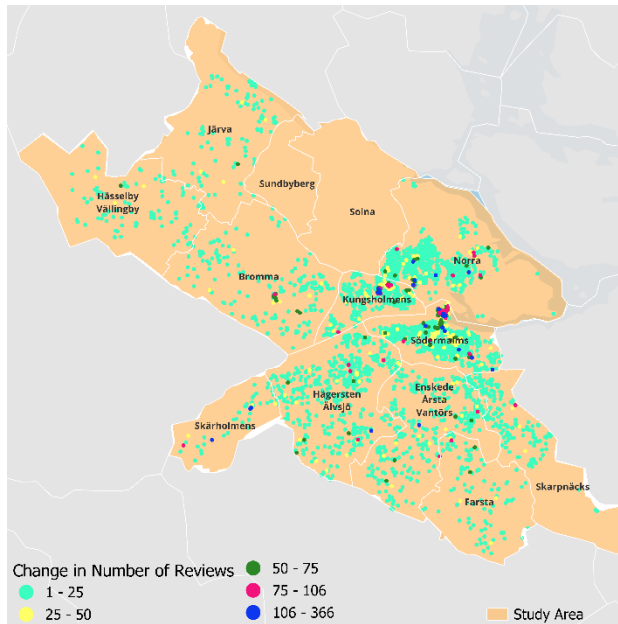


Figure 14. Change in Review Count per Listing After English Language Filtering

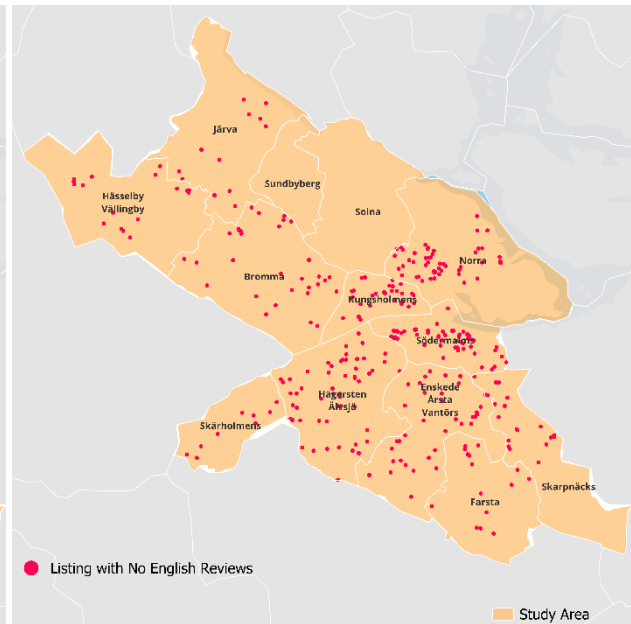


Figure 15. Listings Excluded Due to Absence of English Reviews

3.1.2. Text Preprocessing Pipeline

To prepare the reviews for analysis, a series of preprocessing steps were applied. These were designed to reduce noise, improve consistency, and convert the raw text into a more structured and analysable format. The following steps were performed on each comment:

1. Converted all text to lowercase.
2. Removed URLs, HTML tags, mentions (e.g. @user), and digits using regular expressions.
3. Removed special characters and punctuation, retaining only alphabetical characters and spaces.
4. Normalised repeated characters (e.g., “soooo” became “soo”).
5. Tokenised the text into words.
6. Removed common stopwords using NLTK's English stopwords list.
7. Removed tokens shorter than 3 characters.
8. Applied lemmatisation to reduce words to their base form (e.g., “running” → “run”).

This preprocessing standardised the review content and improved the quality of features used in vectorisation, clustering, and sentiment classification.

Examples of cleaned comments

Table 5 shows an example of raw versus cleaned review, demonstrating the effect of the preprocessing steps:

Table 5. Example of Raw vs Cleaned Reviews

Raw comment	Cleaned comment
Great little apartment in the perfect spot in town. There is a nice grocery store (ICA) around the corner, you are in walking distance of everything (Gamla Stan, Oster, Skansen, etc.) and a few public bike racks just down the street that are cheap and easy to rent. Avoid the stress and cost of a hotel and restaurants. This is a great way to see Stockholm.	great little apartment perfect spot town nice grocery store ica around corner walking distance everything gamla stan oster skansen etc public bike rack street cheap easy rent avoid stress cost hotel restaurant great way see stockholm

Although the preprocessing pipeline was designed to clean the text and make it suitable for analysis, it involved trade-offs. For example, tokens shorter than three characters were removed to reduce noise from stopwords, typos, or fragments that often lack meaningful content. While this might occasionally exclude short but valid words (e.g., "ok" or "go"), similar decisions are common in NLP workflows where simplifying input improves vectorisation and clustering results (Gasparetto et al., 2022; Kowsari et al., 2019). The aim was to retain as much relevant information as possible while ensuring the data remained consistent and manageable for downstream tasks.

As mentioned, the cleaning process introduced minor structural issues, particularly in reviews that contained excessive special characters or HTML-like formatting. An example of this case is shown in Table 6.

Table 6. Impact of Preprocessing on a Review with HTML Tags and Irregular Structure

Raw Comment	Cleaned Comment
Where do I begin???? I have been using Air B@b for 9 years. I have always stayed in the most fabulous apartments..from Athens, Las Vegas, Singapore, Australia and many others. But this accomodation is definitely the most fabulous I have chosen. The photos on the website do not do it justice. It is the	begini using air yearsi always stayed fabulous apartmentsfrom athens la vega singapore australia many othersbut accomodation definitely fabulous chosenthe photo website justiceit beautifully styled spacious apartment could wish many apartment include pianonils superhostthe fridge well stockedbeautiful wine suppliedcandleshe

most beautifully styled and spacious apartment you could wish for. (how many apartments include a piano) Nils is more than a Superhost...the fridge was so well stocked...beautiful wine supplied...candles... [...]	generousthe apartment peaceful hip neighbourhoodsituated metroi purchased weekly pas enabled zip around stockholm ease also able use pas tramsa block away huge supermarket maxi everything could wantupstairs supermarket liquor shop fabulous selection wine[...]
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Furthermore, the text length was also analysed between the raw comments and the processed text to assess the impact of the cleaning process. Figure 16 presents the distribution of comment lengths before and after preprocessing, highlighting the structural differences introduced by text standardization.

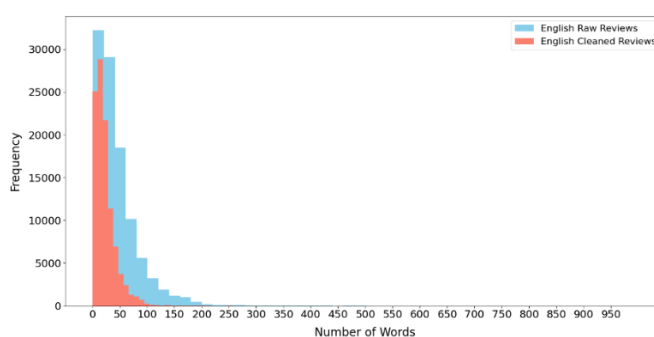


Figure 16. Comparison of Comment Length Distribution

In addition, a word frequency analysis was performed to get an initial overview of the most commonly used terms in the cleaned dataset. This step served two purposes: to verify that the cleaning process retained semantically meaningful content, and to better understand the general vocabulary used by Airbnb users. Figure 17 presents a visualisation of the 20 most frequent terms in the cleaned dataset.

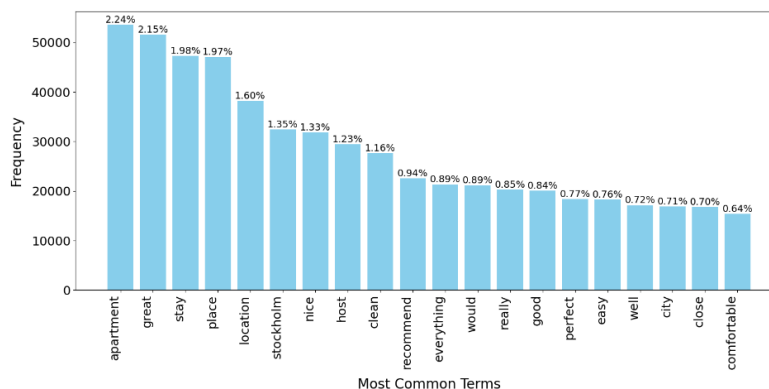


Figure 17. Top 20 Most Common Terms in Cleaned Dataset

3.2. Work framework

After the initial dataset was cleaned, the following steps were done following a methodology structured into three main stages: Neighbourhood Topic Classification, Sentiment Analysis, and Spatial and Comparative Analysis.

The diagram presented in Figure 18 provides a simplified overview of the full data processing pipeline.

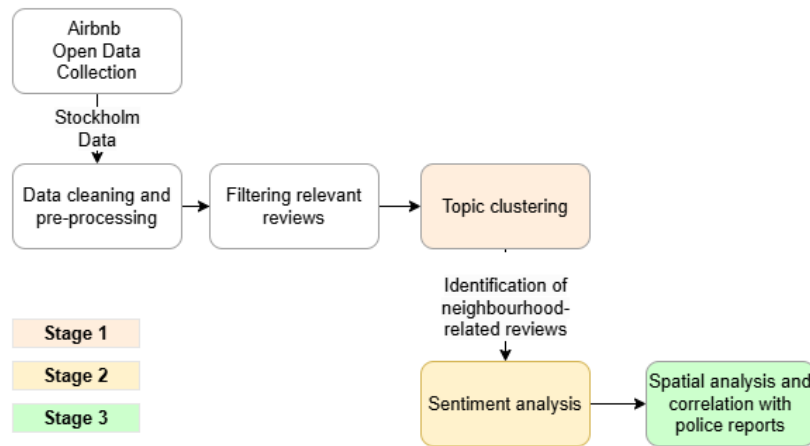


Figure 18. Overview of the Data Processing Pipeline

This structured workflow enabled a step-by-step approach in which each process built upon the outputs of the previous one. This ensured a coherent analysis flow, from filtering relevant reviews to extracting perception-based insights and ultimately comparing them with objective safety indicators.

3.2.1. Stage 1: Neighbourhood Topic Classification

The first stage involved identifying reviews that referred to the neighbourhood environment. As the dataset was unlabelled, a clustering-based approach was used to group the reviews based on their semantic similarity. The following sections describe the vectorisation process, and the clustering algorithms applied to separate neighbourhood-related from non-neighbourhood-related content.

Vectorisation

For the clustering tasks, the reviews were vectorised using the **TF-IDF** approach. This method was selected to highlight more informative and distinctive terms across the dataset, enabling more accurate groupings based on semantic similarity.

To further refine the feature space and reduce noise, specific parameters were set in the TF-IDF vectorizer:

- **max_df = 0.5:** Terms that appear in more than 50% of the reviews were ignored, as they are likely too common to be informative.
- **min_df = 5:** Terms appearing in fewer than five reviews were excluded to reduce sparsity and remove overly specific or uninformative words.
- **stop_words = 'english':** Common English stopwords (e.g., “the,” “and,” “is”) were removed to focus the analysis on semantically meaningful content.

Truncated SVD Component

To improve the quality of the clustering input, the dimensionality was reduced using Truncated Singular Value Decomposition (Truncated SVD). This technique reduced the number of features in the TF-IDF matrix while preserving the underlying semantic structure of the data.

One of the main hyperparameters to tune when using truncated SVD is the number of components (i.e., dimensions) to retain. This hyperparameter controls the amount of variance retained in the dataset (Khadka, 2020). To determine the optimal number of components, a sensitivity analysis was conducted to examine how different values influenced the cumulative explained variance.

In addition to the sensitivity analysis, a cumulative explained variance approach was implemented, guided by the elbow curve. This method involved calculating the proportion of variance explained by each component and incrementally summing these values until reaching a predefined threshold, typically between 70% and 95%, as suggested in previous studies (Falini, 2022). The number of components required to reach this threshold was selected as the optimal dimensionality. This approach ensured that the reduced feature retained most of the original information while significantly lowering computational complexity, making it suitable for large-scale clustering tasks.

K-Means

K-Means clustering was applied to the vectorised Airbnb reviews using two and three clusters. The choice of two clusters was driven by the hypothesis that reviews could be broadly divided into those discussing the neighbourhood and those focusing on other elements, such as the apartment itself or the host. To explore the possibility of more nuanced groupings, a third cluster was tested to evaluate whether an intermediate or mixed category would emerge and how the model would behave when this additional granularity was introduced.

BIRCH

BIRCH was applied to the review embeddings to explore its suitability for separating neighbourhood-related comments from others. A sensitivity analysis was performed on the threshold parameter to assess its impact on the granularity of clustering. This allowed the identification of a value that produced the most meaningful separation between neighbourhood-related and non-neighbourhood-related reviews, balancing specificity with cluster cohesion.

Final Review Selection for Sentiment Analysis

Although unsupervised clustering methods such as K-Means and BIRCH were initially tested to separate neighbourhood-related reviews, the results did not yield meaningful groupings. As discussed in Chapter 4, the low Silhouette Scores and overlapping keywords indicated limited cluster differentiation.

Due to these limitations, a heuristic filtering approach was applied to select the neighbourhood-related reviews for the sentiment analysis. First, the reviews were vectorised using a TF-IDF vectorizer to extract the full vocabulary. Then, a small list of seed words was defined based on terms likely to reference the surrounding environment.

These seed words were embedded using a pre-trained SBERT model, and the most semantically similar terms in the TF-IDF vocabulary were identified using cosine similarity. The final list of terms was used to filter the reviews through a regular expression pattern, and the reviews that matched at least one of these terms were flagged as neighbourhood-related. This process ensured a more targeted input for the sentiment analysis while avoiding the limitations found with clustering-based filtering.

3.2.2. Stage 2: Sentiment Analysis

The second stage focused on categorising the reviews' sentiments previously identified as neighbourhood-related. For this purpose, three sentiment analysis techniques were implemented to capture a range of linguistic and contextual insight: VADER, TextBlob and SBERT embeddings combined with K-Means clustering. Each model aimed to classify reviews into three polarities: *positive*, *neutral*, and *negative*.

VADER Sentiment Intensity Analyzer

VADER classifier was applied to the neighbourhood-related reviews using the Sentiment Intensity Analyzer from the NLTK library. In this context, the polarity scores were examined across the four available metrics: positive, negative, neutral, and compound. The final sentiment classification was determined based on the compound score, as explained in Section 2.3 where thresholds were applied such that a score above 0.05 indicated positive sentiment, a score below -0.05 indicated negative sentiment, and scores in between were considered neutral.

Sentence-BERT (SBERT) Clustering

SBERT – K-Means was applied to the neighbourhood-related reviews using the SentenceTransformer('all-MiniLM-L6-v2') model from the SBERT library. After encoding all reviews with this model, the resulting embeddings were clustered using K-Means with three clusters, each representing one of the sentiment categories: positive, negative, and neutral.

TextBlob

Finally, the third sentiment analysis approach applied was based on the TextBlob library. For each review, a polarity score was computed and then categorised into sentiment labels using predefined threshold values, as described in Section 2.3. Specifically, scores above 0.1 were classified as positive, scores below -0.1 as negative, and scores between these thresholds as neutral.

3.2.3. Stage 3: Spatial and Comparative Analysis

The final stage focused on representing the spatial distribution of sentiments identified in the neighbourhood-related reviews and comparing these with police report data. The following sections describe the methodology used to obtain, process, and integrate the police data with the

sentiment analysis results, enabling a comparative exploration of perceived versus reported urban safety.

Police Report Data

To be able to compare safety perceptions from Airbnb reviews and objective safety indicators from police reports, the crime data was pre-processed through a categorisation and weighting process. This transformation allowed for the aggregation of similar incident types and the adjustment of their impact based on perceived severity. The following sections describe the methodology applied to group, classify, and weight the police report data used in the final stage of the analysis.

Crime categorization

The original dataset included 94 crime types, many with overlapping characteristics. For clearer analysis and interpretation, these were consolidated into nine broader categories, as shown in Table 7.

Table 7. Grouped Crime Categories and Total Reports²

Main Category	Total Reports
Traffic Incidents and Road Safety	8,202
Violent Crime and Personal Threats	5,900
Theft and Property Crime	3,216
Suspicious Activity and Potential Hazards	3,185
Fire And Rescue	2,647
Other/Miscellaneous	2,590
Public Disorder and Disturbance	1,820
Missing/Wanted Persons	193
Animal-Related Incidents	165

To further explore the spatial distribution of these reports, a cross-tabulation was performed to identify the most frequently reported crimes across different districts. It was observed that traffic-related incidents and violent crimes were consistently the most common categories throughout the study area. This distribution is illustrated in Figure 19.

² A full list of the 94 crime types and their grouping into the nine main categories is provided in Appendix 6.1

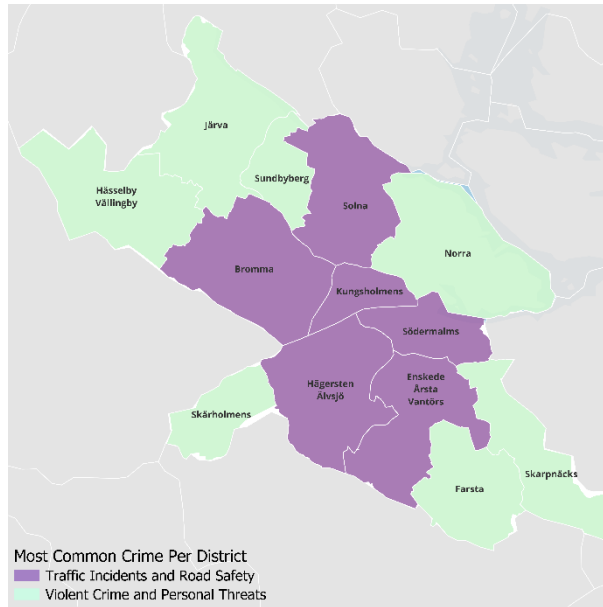


Figure 19. Most Common Crime Categories per District

Crime weighting

A weighting system was introduced since different types of crimes may influence public perception differently. Each crime category was assigned a score reflecting its perceived severity or potential impact on feelings of safety. These scores, listed in Table 8, were determined based on subjective judgment and aligned with general perceptions of crime impact. For future work, the weights could be refined through public surveys or expert consultations to more accurately reflect how residents perceive the severity of different incident types. These values were also used during the Spearman’s correlation analysis.

Table 8. Assigned Severity Scores per Crime Category

Main Category	Score
Traffic Incidents and Road Safety	5
Violent Crime and Personal Threats	10
Theft and Property Crime	8
Suspicious Activity and Potential Hazards	7
Fire And Rescue	2
Other/Miscellaneous	2
Public Disorder and Disturbance	6
Missing/Wanted Persons	4
Animal-Related Incidents	1

Hexagonal Grid Spatial Framework

To enable spatial comparison between sentiment and police report data, the study area was divided into hexagonal grid cells, each with a radius of 500 metres. This resolution was chosen to approximate neighbourhood-level spatial units while maintaining relevance to pedestrian-scale mobility and safe route planning.

The hexagons served as spatial containers to aggregate both Airbnb sentiment data and police report scores, allowing for a unified framework to compare subjective perception and objective safety indicators. Figure 20 shows a zoomed view of the study area, illustrating how hexagons encapsulate sequences of street segments, making them appropriate for this type of spatial analysis.

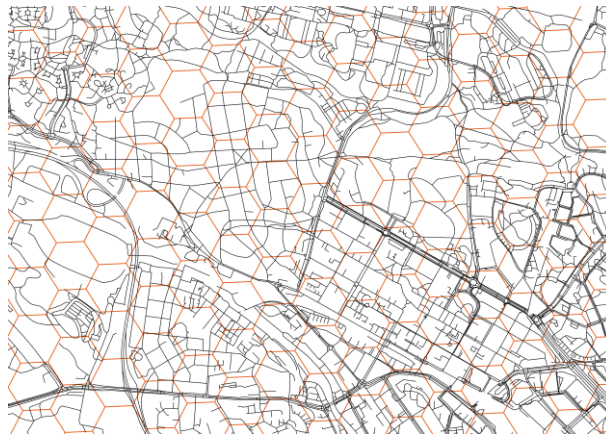


Figure 20. Hexagonal Grid Overview for Study Area

Correlation between Sentiment and Safety: Comparative Analysis

A comparative analysis was conducted using the final hexagonal grid, which included both sentiment classifications and crime severity scores. For each hexagon, the dominant sentiment category and the most frequently reported crime type were identified and mapped.

The correlation analysis used a unique hexagon ID to match corresponding sentiment and crime data across the grid. This analysis assessed the correlation by comparing the dominant sentiment and the dominant police report category in each hexagon using the Spearman correlation.

4. Results and Discussions

This chapter explains the main outcomes of the three-stage analytical framework. The results are structured according to each stage: the classification of neighbourhood-related reviews, the sentiment analysis of filtered comments, and the spatial and comparative evaluation with police report data.

4.1. Stage 1: Neighbourhood Topic Classification

This section presents the results of the neighbourhood topic classification. The results include the vectorised matrix, its dimensionality reduction through Truncated SVD, cluster distributions, top keywords per cluster, Silhouette Scores, and visual representations used to assess the quality of the grouping.

4.1.1. Vectorisation

After the vectorisation process, the dataset was transformed into a matrix of size $103,942 \times 8,715$, where each row corresponds to a review and each column represents a unique term retained after filtering. This means that each review was encoded as a numerical vector with 8,715 features, capturing the relative importance of each term across the dataset.

As expected with TF-IDF, the resulting matrix was highly sparse, with only 0.21% of the values being non-zero. This means that over 99.79% of the entries are zeros, showing that most reviews used only a small portion of the total vocabulary. Such sparsity is common in natural language data due to the nature of text data and reinforces the need for dimensionality reduction techniques (Shah, 2024), such as Truncated SVD, to enhance computational efficiency and clustering performance.

4.1.2. Truncated SVD Component

Following the vectorisation of the reviews, a sensitivity analysis was conducted to support the selection of the optimal number of components for dimensionality reduction. This was complemented by a cumulative explained variance approach, aiming to retain meaningful information while improving computational efficiency.

Based on the cumulative explained variance, the optimal number of components was determined to be 4,050, which accounted for approximately 95% of the total variance, as shown in Figure 21.

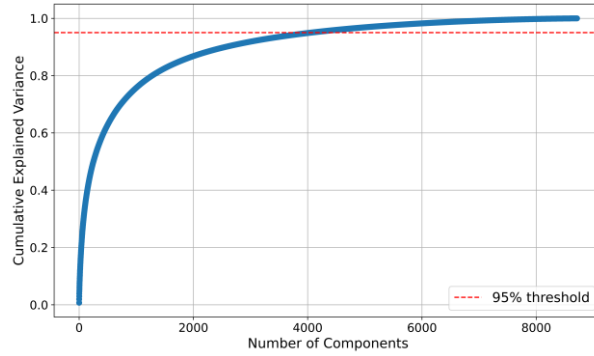


Figure 21. Cumulative Explained Variance “elbow curve” vs Number of Components

In parallel, a sensitivity analysis was carried out by testing six different values for the number of components. This analysis aimed to observe how the number of components affected clustering performance and variance retention. The results of this analysis are presented in Figure 22.

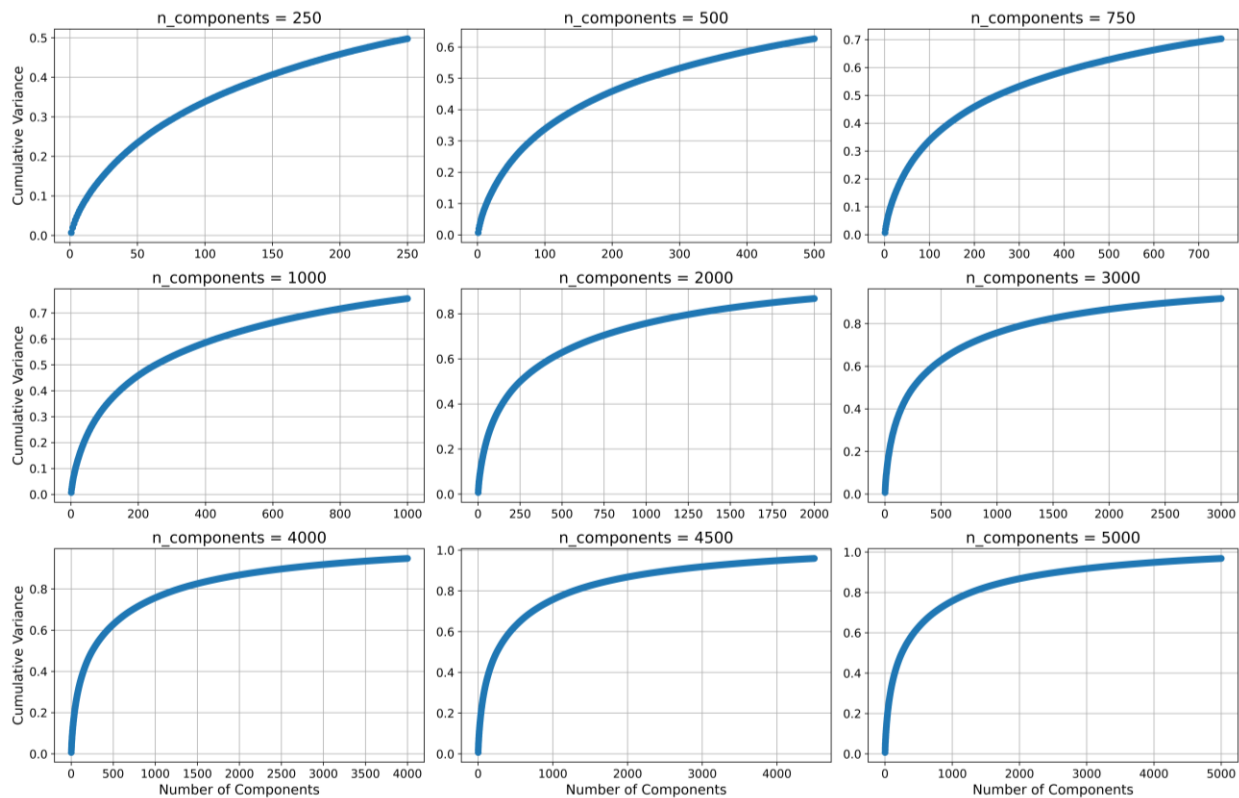


Figure 22. Sensitivity analysis of the number of Truncated SVD components and its impact on cumulative explained variance

As shown in the analysis, 4,050 components accounted for 95% of the cumulative variance. However, a lower threshold of approximately 75% was chosen to improve computational

efficiency. Based on this threshold, 1,000 components were selected, as they captured more than 75% of the cumulative variance while significantly reducing processing time.

4.1.3. K-Means

K-Means clustering was applied to the reduced TF-IDF matrix with two different values of k (2 and 3). Table 9 summarises the number of samples in each cluster, the top keywords that characterise each cluster, and the corresponding Silhouette Scores. These values were used to evaluate cluster quality and interpret the extent to which clusters captured semantically meaningful groupings within the review data.

Table 9. Summary of K-Means Clustering Results (k = 2 and k = 3)

K	Clusters Sizes	Top Words per Cluster	Silhouette Scores
2	C0 = 16,445 C1 = 87,497	C0 = 'great', 'location', 'place', 'stay', 'apartment', 'host', 'nice', 'clean', 'recommend', 'stockholm' C1 = 'apartment', 'place', 'stay', 'nice', 'location', 'stockholm', 'clean', 'great', 'host', 'good'	0.0351
3	C0 = 8,166 C1 = 15,545 C2 = 80,231	C0 = 'lovely', 'apartment', 'stay', 'great', 'place', 'location', 'host', 'stockholm', 'recommend', 'clean' C1 = 'nice', 'place', 'apartment', 'stay', 'location', 'great', 'really', 'clean', 'host', 'good' C2 = 'great', 'apartment', 'place', 'stay', 'location', 'host', 'stockholm', 'clean', 'good', 'perfect'	0.0137

To visually illustrate the clustering results, Figure 23 and Figure 24 presents a two-dimensional projection of the review data using Principal Component Analysis (PCA), coloured by cluster assignment for k = 2 and k = 3, respectively.

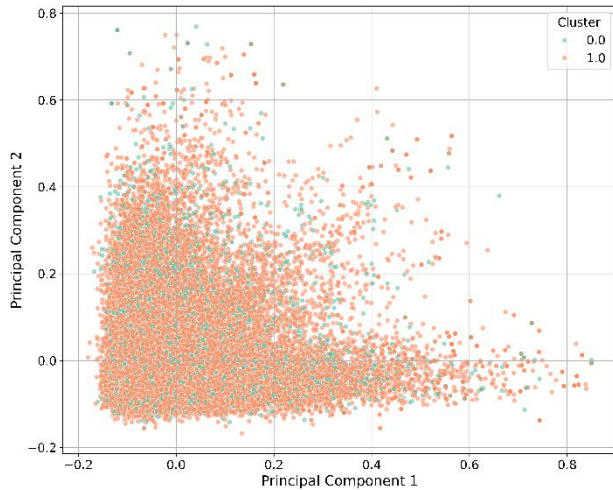


Figure 23. PCA Projection of Clustering Results for $k = 2$

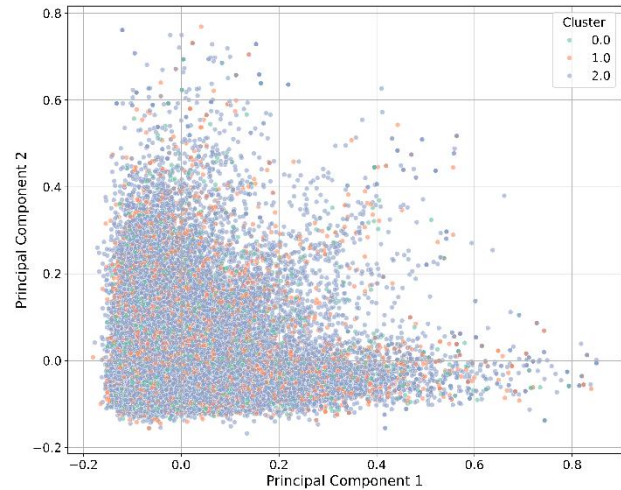


Figure 24. PCA Projection of Clustering Results for $k = 3$

The Silhouette Score for $k = 2$ (0.0351) was higher than for $k = 3$ (0.0137), indicating that the clusters were more distinct and internally coherent when only two groups were used. The lower score for $k = 3$ suggests that introducing a third cluster did not improve, and may have reduced, the quality of separation between groups. This is further supported by the overlap in top keywords across clusters, which indicates limited semantic differentiation.

Additionally, as observed in the plots, the clusters are highly intermixed in both cases, with no clearly separable boundaries. The visual dispersion and overlap between groups (particularly for $k = 3$) highlight the difficulty of extracting well-defined clusters from the review data using this method.

As an additional validation step, the Silhouette Score was calculated for different values of k to evaluate the clustering performance across varying numbers of clusters. As shown in Figure 25, increasing the number of clusters did not lead to an improvement in the Silhouette Score. In fact, the score remained low across all tested values, suggesting that the data does not naturally separate into well-defined groups using the K-Means approach.

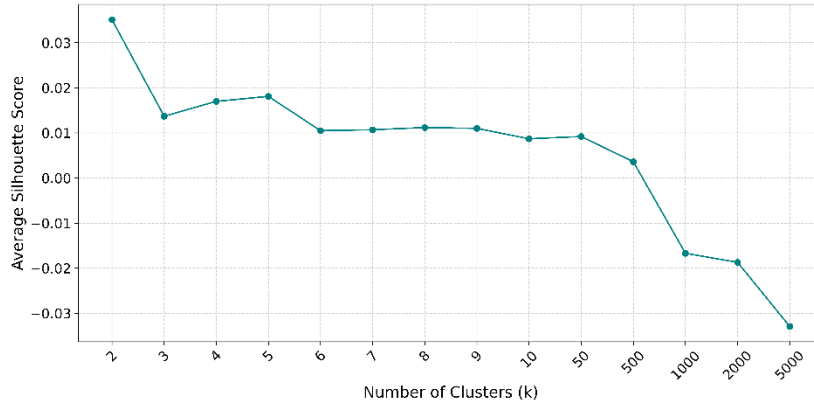
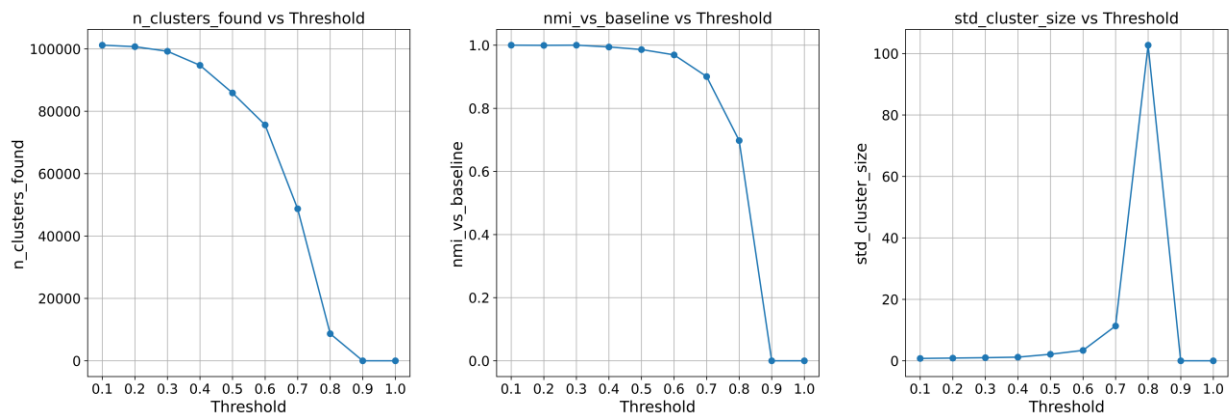


Figure 25. Average Silhouette Score vs. Number of Clusters

4.1.4. BIRCH

As a first step, a sensitivity analysis was conducted to determine the optimal value for the threshold parameter. This analysis evaluated how different threshold values (from 0.1 to 1) affected various clustering performance indicators, including the number of clusters formed, the Silhouette Score, the Normalised Mutual Information (NMI) score, and statistics on cluster size such as the standard deviation, minimum, and maximum. The goal was to identify the threshold value that granted a balanced number of clusters with cohesive internal structure and consistent sizes. Figure 26 presents the results of this sensitivity analysis.



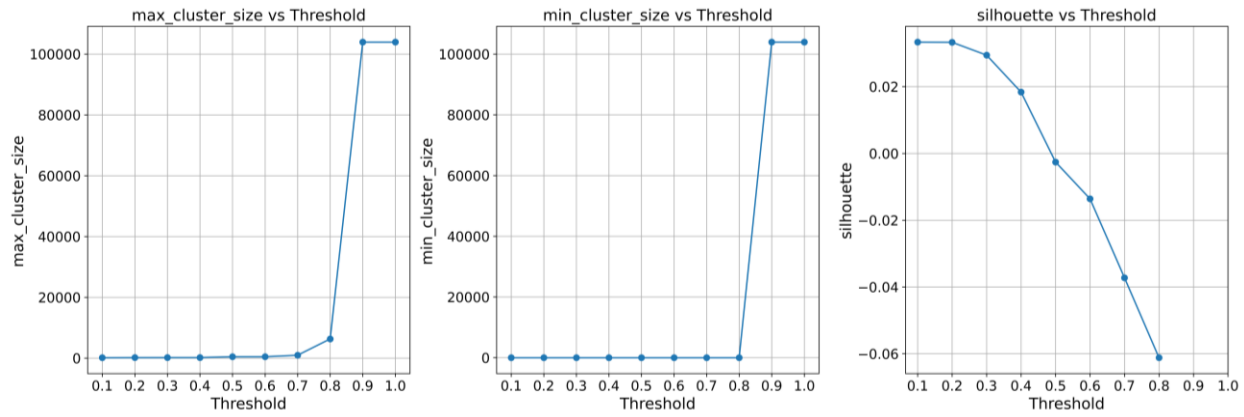


Figure 26. Impact of Threshold Parameter on Clustering Metrics

As observed, applying a higher threshold limited the number of clusters identified, resulting in broader and more loosely defined groupings. This led to higher standard deviations and larger cluster sizes. As shown in the Silhouette Score vs. threshold graph, the scores ranged between 0.03 and -0.06, reinforcing the earlier findings from the K-Means results: neither increasing the number of clusters nor allowing looser groupings led to meaningful separation of the review data. These results suggest that the dataset does not exhibit a strong inherent clustering structure using unsupervised methods.

4.1.5. Final Review Selection for Sentiment Analysis

Using the heuristic filtering method described in Section 3.2.1, a total of 83,344 reviews were identified as neighbourhood-related. This subset represents approximately 80% of the full dataset of English-language reviews. Each matched review contained at least one term from the expanded keyword list, which was generated through semantic similarity with a set of manually defined seed words.

The final seed terms used were: "neighbourhood", "area", "street", "location", "outside", "walk", "quiet", "dangerous", and "park".

This filtered dataset was then used as the input for the sentiment analysis stage, ensuring that only reviews relevant to the surrounding neighbourhood environment were included in the classification.

4.2.Stage 2: Sentiment Analysis

This section presents the results of the sentiment analysis applied to neighbourhood-related reviews. The aim of this stage was to determine whether each review expressed a positive, neutral, or negative sentiment. Three different sentiment analysis methods were evaluated: VADER Sentiment Intensity Analyzer, TextBlob, and SBERT with clustering. The following results compare the output of these models and assess how sentiment is distributed across the dataset.

4.2.1. VADER Sentiment Intensity Analyser

VADER was applied to the neighbourhood-related reviews to assess the sentiment expressed in each comment. Specifically, the compound sentiment score was used to classify each review as positive, neutral, or negative. In addition to the dominant sentiment label, probability scores for each category were also calculated to better understand the underlying composition of each review.

In total, 82,173 reviews were classified as having a positive dominant sentiment, followed by 664 negative and 507 neutral reviews. These results confirm the overall trend of the predominantly positive sentiment in Airbnb neighbourhood-related reviews.

Figure 27 shows the average proportion of positive, neutral, and negative sentiment components within each dominant sentiment label.

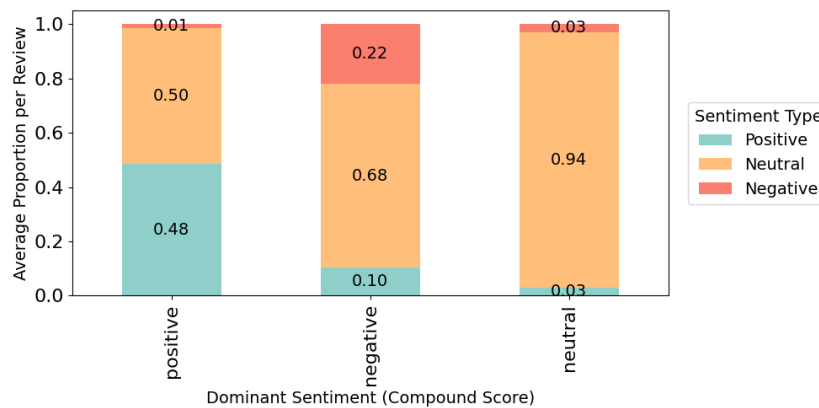


Figure 27. Sentiment Composition per VADER Classification

As observed, most reviews contained a significant neutral component, even when classified as either positive or negative, highlighting the nuance in how users describe their experiences.

4.2.2. Sentence-BERT (SBERT) Clustering

The second sentiment classification model applied to the neighbourhood-related reviews was the SBERT clustering approach. After encoding the reviews using the pre-trained SBERT model, K-Means was used to group the reviews into three clusters, intended to resemble the sentiment categories: positive, negative, and neutral.

he second sentiment classification approach used SBERT to generate sentence embeddings from the reviews. After encoding, K-Means clustering was applied with $k = 3$, aiming to group reviews according to sentiment: positive, neutral, or negative.

The resulting cluster distribution was:

- **Cluster 0:** 23,295 reviews;
- **Cluster 1:** 25,568 reviews;
- **Cluster 2:** 34,481 reviews.

Despite the low Silhouette Score (0.058), the PCA projection in Figure 28 suggests some visible separation among the clusters. This discrepancy arises because PCA is a dimensionality reduction technique that captures variance, not necessarily cluster quality, while the Silhouette Score evaluates cohesion and separation in the full high-dimensional space. Thus, clusters may appear distinct in 2D but still overlap when analysed in the full embedding space.

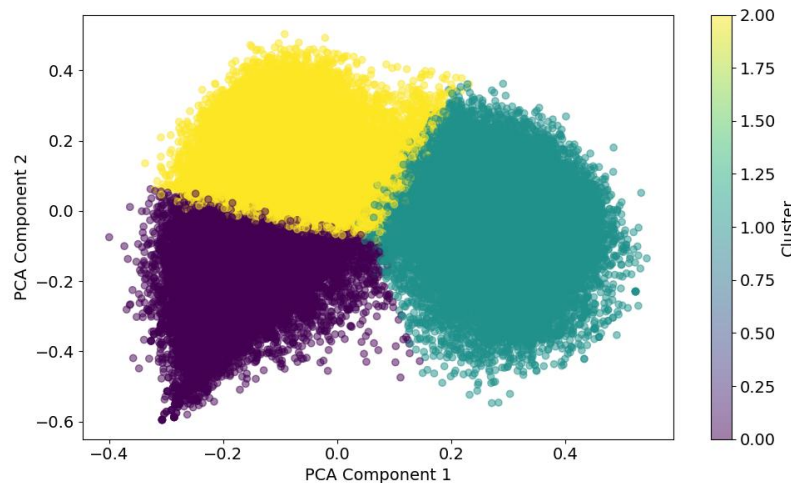


Figure 28. SBERT Embeddings Clustered with K-Means

To further interpret the clusters, the average sentiment scores (positive, neutral, negative) and the compound score were calculated for each using VADER. These are presented in Table 10.

Table 10. Average Sentiment Scores and Compound Values per Cluster

Cluster	Positive	Neutral	Negative	Compound	Likely Sentiment
1	0,44	0,55	0,01	0,90	Positive (highest compound + neutral balance)
2	0,44	0,54	0,02	0,85	Neutral-Positive
0	0,58	0,41	0,01	0,82	Positive (more emotional/stronger)

Cluster 1 had the highest compound score (0.90), indicating a strong overall positive sentiment, despite its reviews having a higher proportion of neutral content. Cluster 2 also showed a positive compound score (0.85), with a more balanced distribution between positive and neutral components. Cluster 0 had the lowest compound score (0.82), yet it contained the highest share of positive sentiment. This suggests that although Cluster 0 reviews had stronger explicitly positive wording, their overall tone may have been perceived as slightly less consistently positive than the other clusters.

4.2.3. TextBlob

The third model tested was TextBlob, which showed similar results to VADER. Using predefined polarity thresholds, the distribution was:

- Positive: 80,298 reviews
- Neutral: 2,769 reviews
- Negative: 277 reviews.

This skew toward positive sentiment matches the patterns observed with the other models.

4.2.4. Model comparison

To better understand how each sentiment model interprets the reviews, Table 11 presents a side-by-side comparison of six sample reviews and their corresponding sentiment labels.

Table 11. Sentiment Classification Comparison Across Models (VADER, TextBlob, SBERT)

N	Review	VADER	TextBlob	SBERT
1	Hedvig's flat is located right in the heart of Ostermalm, close to public transport and within a short walk from Stockholm's tourist landmarks. The flat itself is quiet, in nice condition, and cozy for a family of 3. Lots of nice restaurants and walking routes nearby. Hedvig was very responsive and helpful with recommendations for things to do. I'd stay here again.	positive	positive	positive-neutral
2	Johan's apartment is really nice. It has everything you can need to spend some days in Stockholm. Johan helped us in everything he could and was really nice with us, and amazing host. Not on city centre but just 15min - 20min away from it by bus, and bus stop just 1 min walking. Definitely 100% recomendable	positive	positive	positive-neutral
3	Cozy apartment in an apartment block- feels very safe and is convenient for places in Sodermalm, but also transport to get to the rest of Stockholm (the no. 2 bus is just round the corner from the apartment). It's quite small so I would suggest it is a place for a few nights in Stockholm, and suggest other options for longer stays.	neutral	neutral	positive-neutral
4	Super location and very spacious, however extremely noisy with people walking by the window the whole night, the restaurant underneath banging a metal door every 30 seconds during the evening to bring out the bin, and cleaning and garbage cars working in early morning hours were you think you get a heart attack in your bed.	neutral	positive	neutral-positive
5	I had a nice stay. The apartment is close to the metro so it doesn't take long to get in and out of the city (it took about 20 mins from T-centralen). It was quiet and the washing/drying facilities in the apartment helped a lot. The kitchen is missing a few basics like a kettle/coffee machine and didn't have any pans (only pots that weren't in the best shape). But there was a toaster which helped.	negative	positive	neutral-positive
6	The location of this apartment absolutely cannot be beat! Right in the middle of Old Town with restaurants and shops right outside the window! It is not a quiet location but not at all disturbingly loud. About what you would expect with people enjoying the city outside. There were all the necessary amenities but our stay was so short we didn't get to enjoy all of it.	negative	positive	neutral-positive

This comparison highlights that while VADER and TextBlob often agreed on clear cases, they differed when sentiment was mixed. SBERT tended to assign more nuanced, blended labels such as “neutral-positive,” reflecting its sensitivity to both pros and cons within a review. This suggests that while VADER and TextBlob are simpler and keyword-based, SBERT captures more contextual meaning.

4.3. Stage 3: Spatial and Comparative Analysis

This section presents the results of the spatial distribution of sentiment derived from neighbourhood-related reviews. The aim of this stage was to visualise the dominant sentiment within each hexagonal grid cell and to explore potential correlations between the subjective Airbnb review sentiments and police report data.

4.3.1. Hexagonal Grid Spatial Framework

Figure 29 shows the spatial distribution of the dominant sentiment across the study area, using the hexagonal grid framework introduced in Chapter 4. Each grid cell reflects the most frequent sentiment detected among neighbourhood-related reviews within its boundaries. Sentiment classifications are shown separately for the three models applied: (a) VADER, (b) SBERT, and (c) TextBlob.

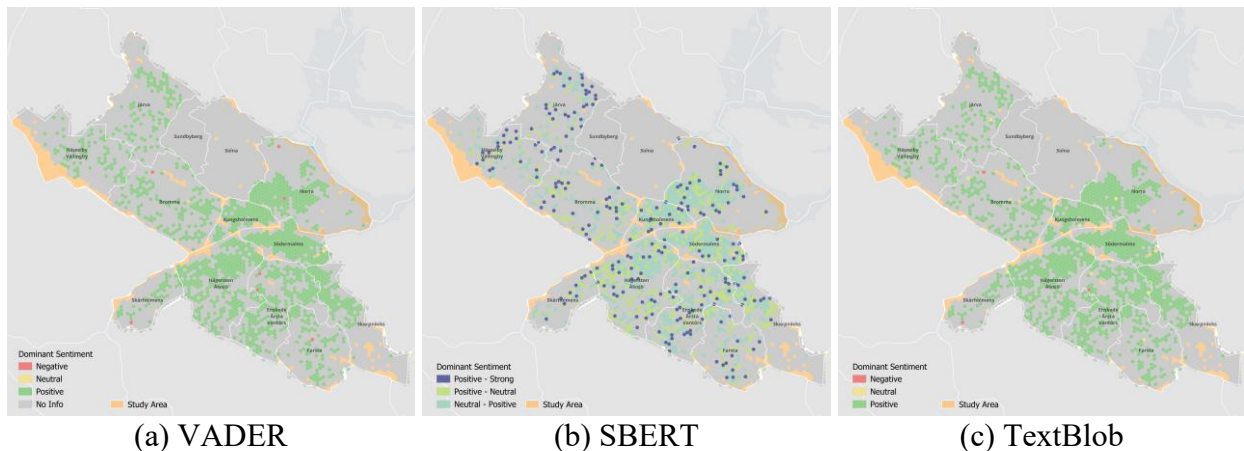


Figure 29. Spatial Distribution of Neighbourhood Sentiment Across the Study Area³

As observed, the overall sentiment distribution across the study area was consistent across all models, with most hexagons showing a dominant positive or positive-leaning sentiment. Both the

³ Enlarged versions of these maps are available in Appendix 6.3

VADER and TextBlob analyses displayed similar spatial patterns, where positive and neutral sentiments were evenly spread. However, slight differences emerged in some hexagonal areas that were classified as positive in the VADER analysis occasionally appeared as neutral in the TextBlob results. This suggests that while the models generally agree on the broader sentiment landscape, they may differ in how they interpret more moderately expressed reviews.

4.3.2. Correlation between Sentiment and Safety: Comparative Analysis

For the Spearman’s correlation analysis, it was used the dominant police report category within each hexagonal grid cell, as shown in Figure 30, and the most common sentiment classification per sentiment and per model as presented previously in Figure 29.

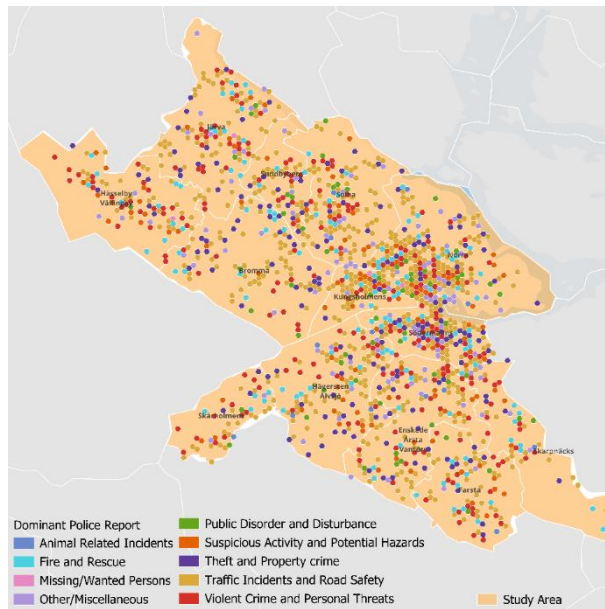


Figure 30. Most Common Police Report Category per Hexagon⁴

To compute the Spearman’s correlation between the sentiment categories and the dominant crime types per hexagon, it was necessary to convert the qualitative sentiment labels into numerical values. Table 12 presents the transformation values used for each sentiment category across the different models. For the police reports, the crime severity values were assigned based on the classification detailed in Section 3.2.3 (Table 8).

⁴ Enlarged versions of this map is available in Appendix 6.4

Table 12. Sentiment Category Encoding for Spearman’s Correlation Analysis

	Sentiment Category	Spearman’s Value
VADER and TextBlob Models	Negative	0
	Neutral	1
	Positive	2
SBERT Model	Positive – Neutral	1
	Neutral – Positive	1
	Positive – Strong	2

Table 13 presents the results of the Spearman correlation analysis, including both the correlation coefficients and corresponding p-values for each sentiment analysis method. These values reflect the degree of association between the dominant sentiment identified in Airbnb reviews and the most frequently reported crime type in each hexagon.

Table 13. Spearman Correlation Results Between Sentiment and Police Reports

Sentiment Analysis Method	Spearman Correlation	p-value
VADER	0.0753	0.0608
SBERT	0.0089	0.8243
TextBlob	0.0322	0.4233

The Spearman correlation analysis revealed weak and statistically insignificant associations between the dominant sentiment per hexagon and the dominant crime type from police reports. Among the models, VADER showed the highest correlation coefficient ($\rho = 0.0753$), but with a p-value of 0.0608, it remained just above the conventional significance threshold (0.05). The SBERT and TextBlob models showed even lower correlation values, suggesting that the spatial patterns of perceived safety in Airbnb reviews do not strongly align with the spatial distribution of reported crime. These findings reinforce the notion that subjective safety perception and officially reported incidents may reflect different dimensions of urban experience.

5. Conclusions

In this thesis, the potential of crowdsourced data was explored by using Airbnb reviews as a case study to assess urban safety perception at the neighbourhood level. The aim was to evaluate whether these reviews could reliably capture meaningful insights about perceived safety in the surrounding environment, and how such perceptions vary across space. The work addressed three main research questions and focused on Stockholm as the case study city.

First, to address whether machine learning models can accurately classify user neighbourhood-related reviews, unsupervised clustering techniques such as K-Means and BIRCH were applied. However, these models showed low Silhouette Scores and overlapping group characteristics, indicating poor cluster separability. As a result, a heuristic filtering method using semantic similarity with TF-IDF and SBERT embeddings was implemented to extract a cleaner subset of neighbourhood-related reviews, ensuring greater relevance for subsequent analysis.

Second, to understand how sentiments vary across neighbourhood topics, three sentiment classification methods were applied: VADER, TextBlob, and SBERT-based clustering. The results showed that the majority of reviews expressed positive sentiments, although the strength and degree of positivity varied by model. VADER and TextBlob often classified reviews as strongly positive, while SBERT revealed more balanced sentiment groupings. Using multiple sentiment models allowed for a broader comparison of how different approaches interpret review tone and structure, helping to capture variations in how safety perceptions are expressed.

Lastly, the spatial distribution of sentiments was analysed using a hexagonal grid framework and compared with police report data. A Spearman correlation analysis was conducted to evaluate the relationship between perceived and reported safety. The results showed weak and statistically insignificant correlations, suggesting that while user reviews could offer different subjective impressions of neighbourhoods, they do not strongly align with the spatial patterns of reported crime. This highlights the value of using crowdsourced reviews to complement, but not replace, traditional safety data. It is also important to note that the analysis was based on entire reviews, rather than isolating only the segments directly referring to the neighbourhood, which may have introduced noise into the sentiment classification.

Overall, this research demonstrated that crowdsourced review data holds potential for capturing safety perception in urban environments, particularly when filtered and analysed with NLP techniques. While not without limitations, these reviews offer a scalable and accessible way to gather localised insights on how neighbourhoods are experienced by visitors. The methodology presented can be adapted and extended to other cities or datasets, providing a foundation for future work in perception-driven urban analytics and safe route planning.

Future work could focus on refining the review filtering process by isolating only the specific segments that refer to the surrounding environment, rather than analysing entire comments. Additionally, combining Airbnb reviews with other perception-based data sources, such as street-level imagery or social media, could offer a more comprehensive view of urban safety perception. These insights could also support broader applications in urban planning. For example, incorporating neighbourhood perception data into safe route planning tools could help tailor mobility solutions based on users' safety preferences and contextual urban features.

References

- Abubakar, H. D., Umar, M., & Bakale, M. A. (2022). Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec. *Sule Lamido University Journal of Science & Technology*. <https://doi.org/10.56471/slujst.v4i.266>
- Bonta, V., Kumares, N., & Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology*. <https://doi.org/10.51983/ajcst-2019.8.S2.2037>
- Brodsky, I. (2018). *H3: Uber's hexagonal hierarchical spatial index*. <https://www.uber.com/en-EE/blog/h3/>
- City of Stockholm. (2012). *Urban mobility strategy*. <https://start.stockholm/globalassets/start/om-stockholms-stad/politik-och-demokrati/styrdokument/urban-mobility-strategy.pdf>
- Culotta, A., Jin, G. Z., Sun, Y., & Wagman, L. (2023). *Safety reviews on Airbnb: An information tale*. <https://www.nber.org/papers/w31855>
- Falini, A. (2022). A review on the selection criteria for the truncated SVD in Data Science applications. *Journal of Computational Mathematics and Data Science*. <https://doi.org/10.1016/j.jcmds.2022.100064>
- Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). A survey on text classification algorithms: From text to predictions. *Information*. <https://doi.org/10.3390/info13020083>
- GeeksforGeeks. (2020). *ML*. <https://www.geeksforgeeks.org/ml-birch-clustering/>
- GeeksforGeeks. (2024). *Python – Sentiment Analysis using VADER*. <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>
- Hargrave, M. (2025). *Crowdsourcing*. <https://www.investopedia.com/terms/c/crowdsourcing.asp>
- Huang, X., Wang, S., Lu, T., Liu, Y., & Serrano-Estrada, L. (2024). Crowdsourced geospatial data is reshaping urban sciences. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2024.103687>
- Inside Airbnb. *Get the data*. <https://insideairbnb.com/get-the-data/>
- Jacoby, R., Bou-Zeid, E., Pigliautile, I., & Pisello, A. L. (2024). Crowdsourced data as a strategic approach to include the human dimension in outdoor environmental quality assessments. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2024.175925>
- Khadka, N. (2020). *Truncated SVD*. <https://dataaspirant.com/truncated-svd/>

- Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*. <https://doi.org/10.3390/info10040150>
- Ladyous, A. K. (2021). *Sentiment analysis using VADER*. <https://akladyous.medium.com/sentiment-analysis-using-vader-c56bcffe6f24>
- Laerd Statistics. (n.d.). *Spearman's rank-order correlation*. <https://statistics.laerd.com/statistical-guides/spearman-rank-order-correlation-statistical-guide.php>
- Lagutina, K., Larionov, V. ., Petryakov, V. ., Lagutina, N. ., Paramonov, I. ., & Shchitov, I. (2019). *Sentiment classification into three classes applying multinomial Bayes algorithm, N-grams, and thesaurus*. <https://doi.org/10.23919/FRUCT.2019.8711968>
- Liu, Y., Whitfield, C., Zhang, T., Hauser, A., Reynolds, T., & Anwar, M. (2021). Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Information Science and Systems*. <https://doi.org/10.1007/s13755-021-00158-4>
- Loria, S. (n.d.). *Quickstart—TextBlob 0.15.3 documentation*. <https://textblob.readthedocs.io/en/dev/quickstart.html>
- Lovmar, L., Ahlford, A., Jonsson, M., & Syvänen, A.-C. (2005). Silhouette scores for assessment of SNP genotype clusters. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-6-35>
- Marqo. (n.d.). *Introduction to Sentence Transformers*. <https://www.marqo.ai/course/introduction-to-sentence-transformers#:~:text=Sentence%20transformers%20address%20this%20limitation,accurate%20and%20meaningful%20text%20analysis>.
- Niu, H., & Silva, E. A. (2020). Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. *Journal of Urban Planning and Development*. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000566](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566)
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. <https://doi.org/10.1561/1500000011>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., & P. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.

- Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*. <https://doi.org/10.1007/s13278-023-01030-x>
- Qin, T., Van de Weghe, N., & Huang, H. (2025). The role of urban built environments in the perception of safety: A systematic review. *Environmental Impact Assessment Review*. <https://doi.org/10.1016/j.eiar.2025.107915>
- Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using Siamese BERT-networks*. <https://doi.org/10.18653/v1/D19-1410>
- Sampath, M., & Vignesh, S. (2024). *Text clustering for topic identification: A TF-IDF and K-means approach applied to the 20 Newsgroups dataset*. <https://easychair.org/publications/preprint/C6Ft>
- Scribbr. (n.d.). *What are the assumptions of the Pearson correlation coefficient?* <https://www.scribbr.com/frequently-asked-questions/assumptions-of-pearson-correlation-coefficient/#:~:text=These%20are%20the%20assumptions%20your,Your%20data%20have%20no%20outliers>
- Shah, P. (2024). *Ways to overcome the issues of sparse matrix in NLP*. <https://pankti0919.medium.com/ways-to-overcome-the-issues-of-sparse-matrix-in-nlp-0b135bf677ca>
- Shandeep. (2019). *CountVectorizer vs. TfidfVectorizer*. <https://medium.com/@shandeep92/countvectorizer-vs-tfidfvectorizer-cf62d0a54fa4>
- Shuyo, N. (n.d.). *Langdetect*.
- Sinha, A. (2020). *Why do we use hexagons (and not squares) to aggregate location data?* <https://hackernoon.com/why-do-we-use-hexagons-and-not-squares-to-aggregate-location-data-ouc23y35>
- Youvan, D. C. (2024). *Understanding Sentiment Analysis with VADER: A Comprehensive Overview and Application*. https://www.researchgate.net/publication/381650914_Understanding_Sentiment_Analysis_with_VADER_A_Comprehensive_Overview_and_Application
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). *BIRCH: An efficient data clustering method for very large databases*. <https://doi.org/10.1145/233269.233324>

Appendix

6.1. Mapping of Original Crime Types to Grouped Categories

Main Category	Crime Type
Violent Crime and Personal Threats	Assault
	Aggravated Assault
	Murder/Manslaughter
	Attempted Murder/Manslaughter
	Robbery
	Armed Robbery
	Attempted Robbery
	Violence/Threat Against Officer
	Rape
	Attempted Rape
	Shooting
	Shooting, Suspected
	Other Robbery
	Weapons Act
	Unlawful Deprivation of Liberty/Kidnapping
	Unlawful Deprivation of Liberty
	Incitement Against a Group of People
	Harassment/Annoyance
Public Disorder and Disturbance	Bomb Threat
	bomb threat
	Fight
	Intoxication/LOB (Law on Public Order)
	Police Operation/Command
	Morality Crime
	Found Dead
Theft and Property Crime	harassment/annoyance
	Misconduct by Child/Youth
	Theft
	Attempted Theft
	Theft/Burglary
	Burglary
	Attempted Burglary
	Motor Vehicle Theft
	Motor Vehicle, Found Stolen
	Handling Stolen Goods
	Crime Against Elderly
Forgery Crime	
Unlawful Intrusion/Violation of Domestic Peace	

	Petty Theft
	Violation of domestic peace
Traffic Incidents and Road Safety	Traffic Accident
	Traffic Accident, Personal Injury
	Single Vehicle Accident
	Traffic Accident, Hit and Run
	Traffic Accident, Wildlife
	Drunk Driving
	Traffic Offense
	Traffic Obstruction
	Traffic Control
	Illness/Accident
	Causing Bodily Harm
	Unauthorized Driving
	Suspicious Activity and Potential Hazards
Sabotage Against Emergency Services	
Explosion	
Detonation	
Warning Alarm/Crash	
Assault Alarm	
Fraud	
Vandalism	
Unlawful Entry	
Unlawful Threat	
Drug Crime	
Burglary Alarm	
Spread of Infection/Chemicals	
Fire and Rescue	Fire
	Fire Automatic Alarm
	Rescue Operation
	Mountain Rescue
Animal-Related Incidents	Animal
	Traffic Accident, Wildlife
	Animal Injured/Taken Into Custody
Missing/Wanted Persons	Missing Person
	Wanted Person
Other/Miscellaneous	Other
	Person/Vehicle Check
	Public Order Act
	Environmental Crime
	Alcohol Law
	Economic Crime
	Aviation Law

	Maritime Law
	Knife Law
	Protection Act
	Workplace Accident
	Found Property
	Foreigners Act
	Dog and Cat Act
	Daily Summary
	Morning Summary
	Evening Summary
	Weekly Summary
	Night Summary
	Evening and Night Summary
	Update
	Abuse of documents
	the dog and cat act
	border control
	theft, petty

6.2. Full Text of Sample Airbnb Reviews

Review	Full comment
1	big recommendation easy get centre city metro appartement clean cozy includes needed staying day super friendly host fast answering always check everything alright
2	stayed david olivias wonderful apartment month really enjoyed stay location perfect walkable close grocery store restaurant bakery coffee shop major transit hub apartment set back street quiet there pay parking lot front convenient carthe apartment lovely clean bright comfortable kitchen well set everything need cookingdining master bedroom quiet bed comfortable david olivia delightful show around orient home incredibly helpful responsive question came cant say enough positive look forward next stay return
3	spacious lot morning sun kitchen large well equipped block major shopping center grocery store public transportation lot restaurant appreciate host flexibility checkout

6.3. Enlarged Spatial Sentiment Maps

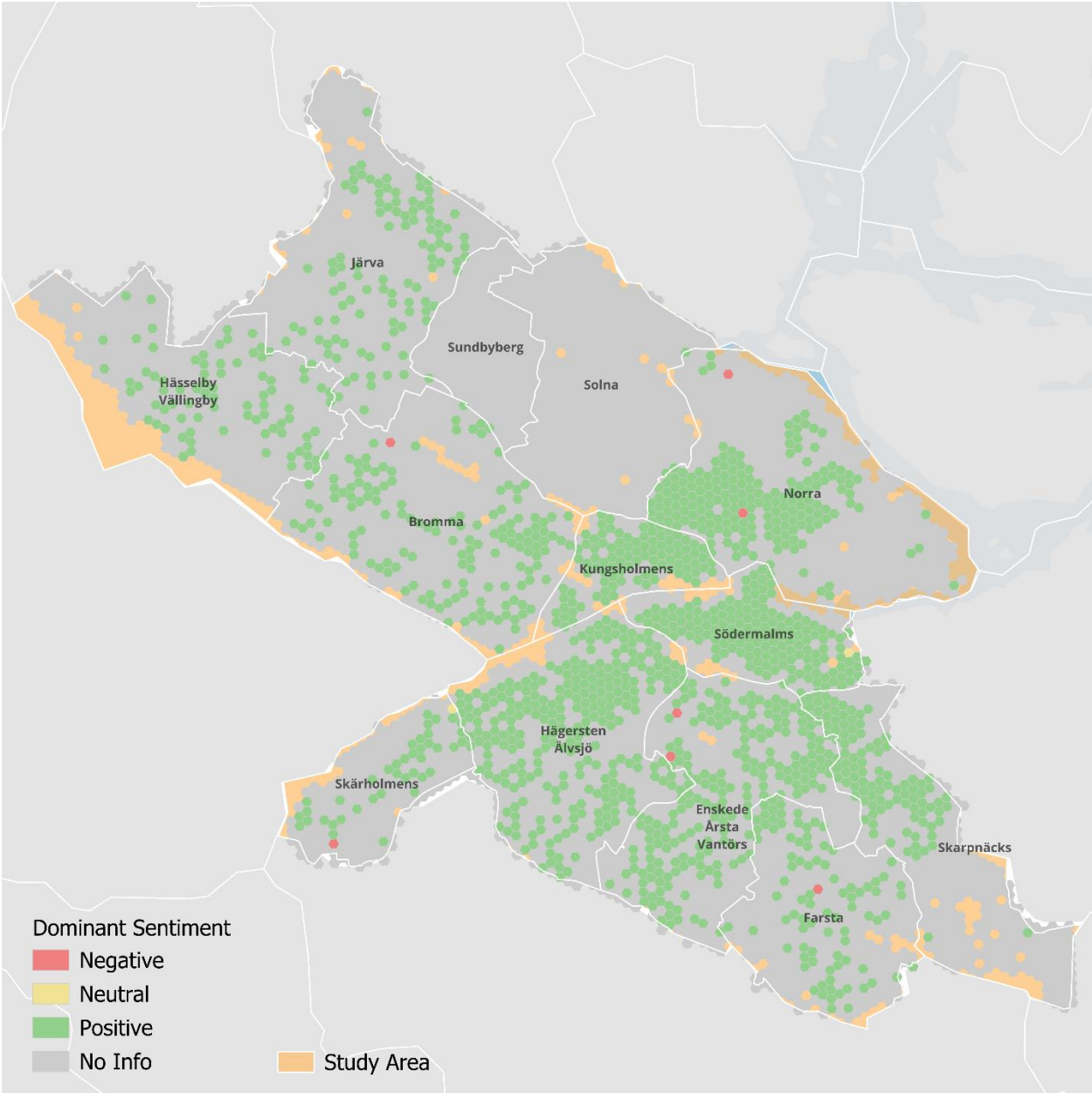


Figure 31. Enlarged VADER Sentiment Map

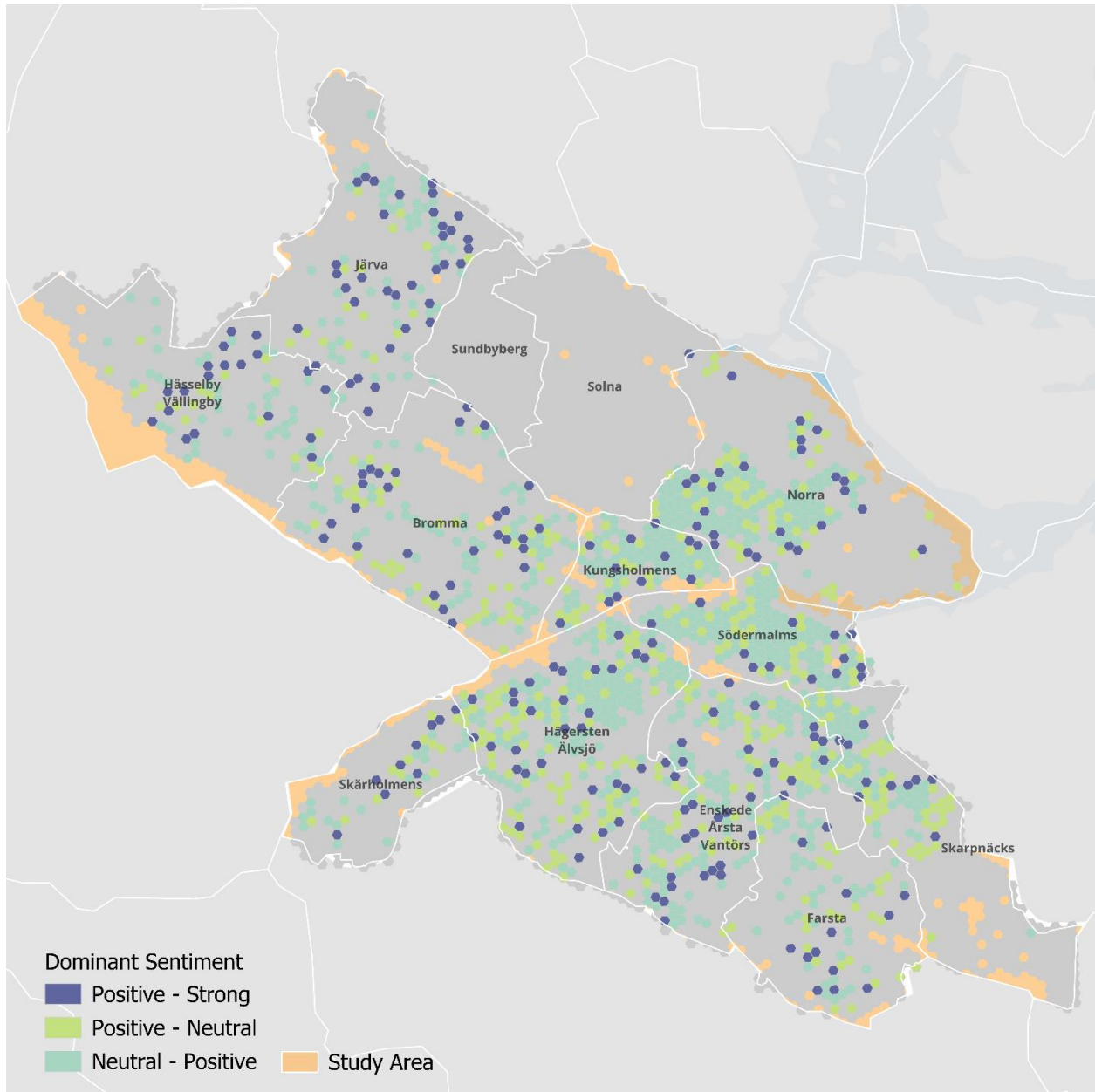


Figure 32. Enlarged SBERT Sentiment Map

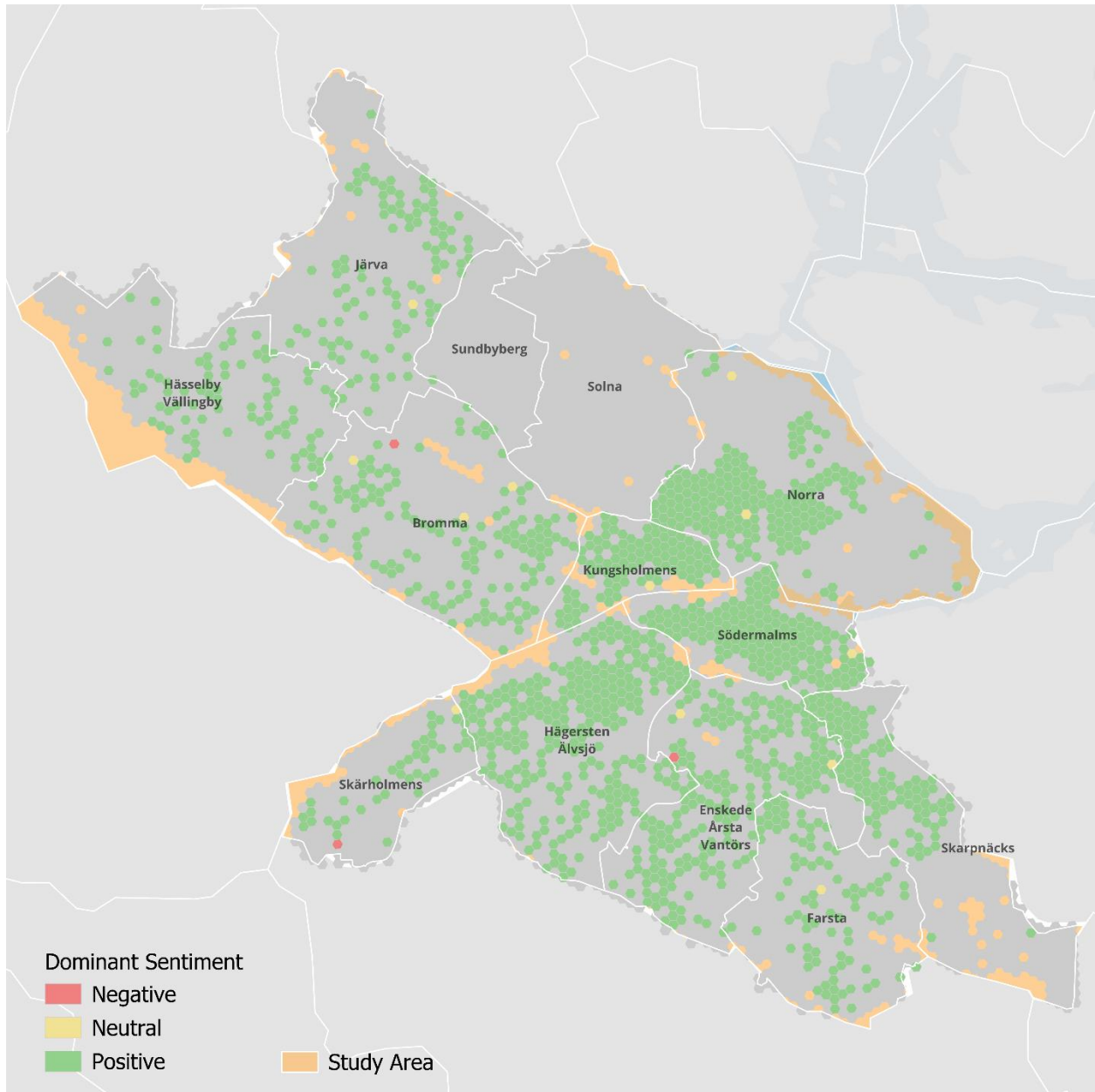


Figure 33. Enlarged TextBlob Sentiment Map

6.4. Enlarged Police Report Category Maps

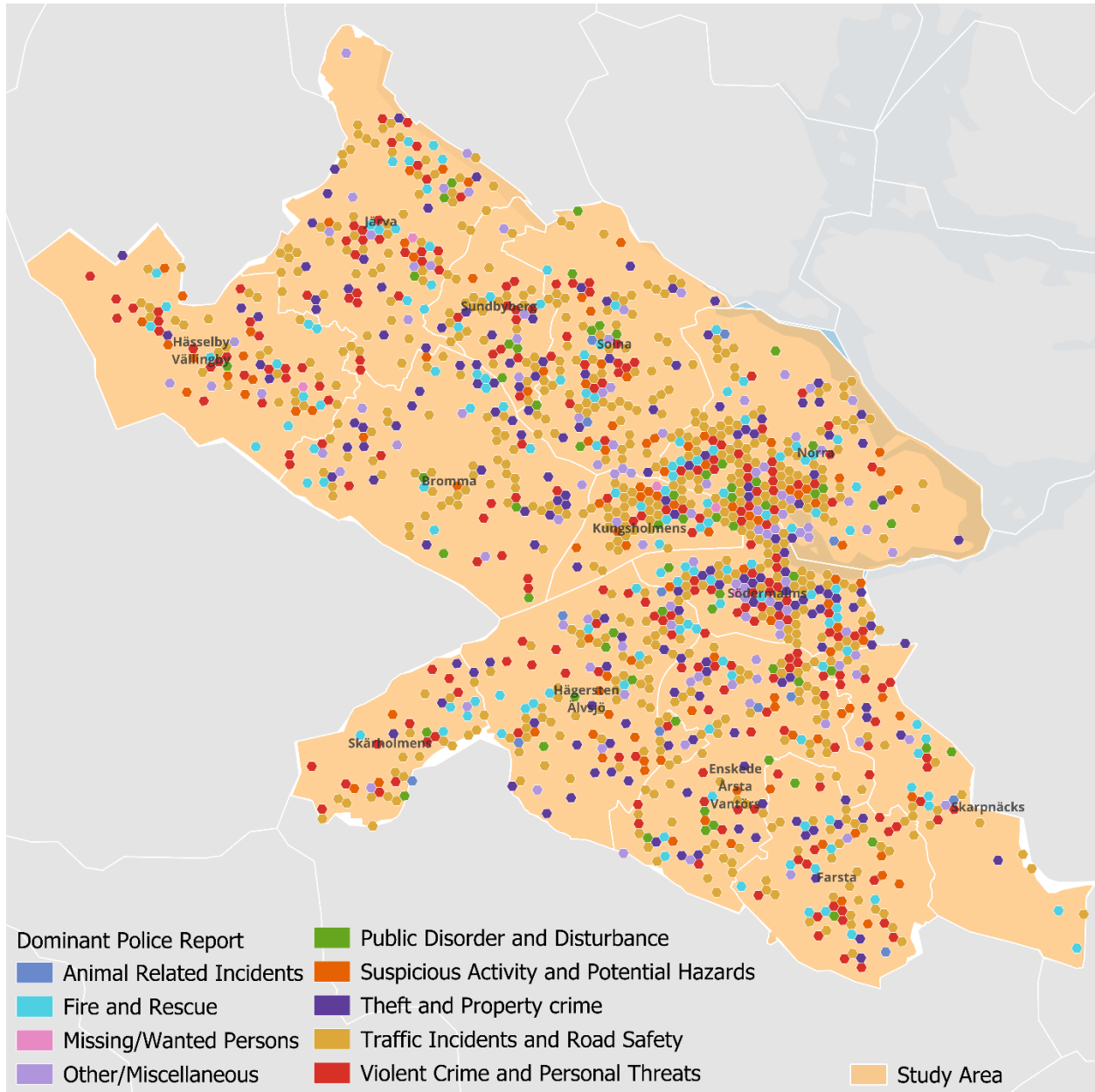


Figure 34. Enlarged Most Common Police Report Category per Hexagon

License

Non-exclusive licence to reproduce thesis and make thesis public

I, María Alejandra Rodríguez Sánchez,
(*author's name*)

grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

Crowdsourced Perception of Neighbourhood Safety: A Case Study Using Airbnb Reviews ,

(*title of thesis*)

supervised by Victor Henrique Cabral Pinheiro, Amnir Hadachi, and Katarzyna ;
Sowinska

(*supervisor's name*)

1. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;
2. am aware of the fact that the author retains the rights specified in points 1 and 2;
3. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

María Alejandra Rodríguez Sánchez

15 May 2025