Visual Data Analysis

UNIVERSITY
OF SKÖVDE

# CRIME PATTERNS IN LA WITH AIRBNB DATA

KIM WURSTER

GROUP 7: PRERNA, SHIVAM NAYAR, ANDREAS ALGBORG

# Table of Contents

# 1. Introduction

## 1.1 One research question

The research question guiding this study is: **How do crime patterns vary across time variables (weekdays, time of day) and location variables (area names, premises), and how can this information help to inform safer Airbnb bookings?** This question is multivariable and can be effectively answered using dashboards, which allow users to explore key safety-related insights without overwhelming them with excessive details.

## 1.2 Motivation and aim

The motivation behind this study stems from the need for safer Airbnb bookings in a large and diverse city like Los Angeles. A family of four, planning their vacation, wants to identify the safest Airbnb locations. The challenge is to present crime-related data in a manner that is easy to interpret and supports informed decision-making. The chosen visualization, a dashboard, is essential for addressing the research question of identifying the safest Airbnb areas in Los Angeles for a family visit. Given the city's size and complexity, a dashboard is necessary to present large amounts of data in an accessible way. Dashboards provide clarity by structuring information logically while allowing interactivity, addressing the challenge of balancing depth and simplicity (Sarikaya et al., 2019). This approach to build dashboards balances depth and simplicity, allowing users to explore key safety-related insights without overwhelming them with excessive details.

One key challenge is fitting all relevant data into a single visualization without overwhelming the user. Previous research highlights that dashboards must cater to different decision-making levels, balancing storytelling with data density (Bach et al., 2022). By leveraging pre-attentive processing techniques such as colour and spatial positioning (Janes et al., 2013), the dashboard effectively directs user attention to critical safety indicators.

Inspired by storytelling dashboard techniques (Lavalle et al., 2025), this design ensures structured, goal-oriented navigation. While limitations include data aggregation and the exclusive focus on Airbnb's, the approach significantly improves decision-making by reducing misinterpretation and enhancing usability. This paper helped me understand how to create storytelling dashboards that are easy to understand and intuitive.

# 2. Data

The dataset used for this study consists of two primary sources: Los Angeles Crime Data and Los Angeles Airbnb Listings. The crime dataset includes detailed records of reported crimes from 2020 to the present, covering factors such as crime types, locations, times, and associated premises (Isha, 2025). This dataset provides valuable insights into crime trends and patterns over time. The Airbnb dataset, on the other hand, contains extensive information on available listings across Los Angeles, including location details, pricing, and property characteristics (Batiz, 2025). These datasets were merged based on geographic crime areas using the k-Nearest Neighbours (kNN) algorithm, which classifies Airbnb locations based on crime data locations. The kNN algorithm is like a voting system where the majority class label determines the class of the new data point when taking its k – nearest neighbours into consideration (Shafi, 2023).

The primary features used in this study include Airbnb location, weekday, crime type, premises, weapons used, and time of day. By integrating these attributes, the analysis aims to identify trends in crime occurrences relative to Airbnb locations, helping users determine which areas are safer for short-term stays. The use of kNN for classification allows for efficient grouping of crime patterns within specific geographic regions. This approach ensures that Airbnb listings can be assessed based on their proximity to high-risk areas, enhancing the overall relevance of the visualization.

# 3. Evaluation

The evaluation of our visualizations was conducted using the Stopwatch Method, incorporating three structured questions alongside additional feedback from users. This approach ensured that both qualitative and quantitative aspects of the visualization's effectiveness were captured. To assess them, a user evaluation with ten participants was conducted. The evaluation aimed to determine whether the

dashboard successfully conveyed crime patterns and enabled users to make informed Airbnb booking decisions. The Stopwatch Method was used, where participants were given three questions to answer while their task completion time and accuracy were measured. In addition to structured questions, users were encouraged to provide open-ended feedback regarding the clarity, usability, and overall effectiveness of the visualization.

One of the key factors in evaluating visualizations is their efficiency, which is typically measured through task completion time and correctness. A faster interpretation time indicates a well-structured and intuitive visualization (Nazemi et al., 2015). The Stopwatch Method measures the task completion and the evaluator made sure the question was answered correctly, therefore the best structured and intuitive visualization got the highest score in the evaluation. Asking the participants for feedback on the visualizations is important to detect potential issues in the visualizations and possibly change them to be more intuitive and easier to interpret.

## 4. Visualization comparison and evaluation results

The selection process, comparison of the four visualizations, and evaluation results are presented in Table 1.

### 4.1 Justification of the final selected visualization technique

The selected visualization stands out due to several key factors. First, it includes a clear legend, which reduces cognitive load and minimizes potential misinterpretations. A well-designed dashboard must prioritize simplicity, avoiding unnecessary clutter while maintaining readability and ease of interpretation (Mokkup, 2023). Compared to other dashboards evaluated, the lack of a legend in Dashboards 1 and 4 resulted in confusion from participants.

Additionally, the chosen visualization follows a structured and consistent layout. Consistency in design enhances usability and allows users to navigate the dashboard intuitively (Mokkup, 2023). This structural coherence makes the visualization more accessible to diverse audiences.

Another major strength of the selected visualization is its effective use of colour. The choice of a monochromatic or complementary colour scheme aids in differentiation and visual clarity. Blue, symbolizing trust and clarity, is particularly suitable for conveying stable and reliable data (Fresh BI, 2025). Additionally, using visual cues like colour intensity and size variations helps direct attention to significant data points, making interpretation more efficient (Mokkup, 2023). In contrast, Dashboard 3 struggled with distinguishing between areas within similar colour ranges, leading to misinterpretation of regional differences.

While alternative visualizations, such as radar charts and complex multi-variable dashboards, were considered, they presented notable limitations. For example, Dashboard 3's radar chart appeared cluttered, making it difficult for users to extract meaningful insights efficiently. Similarly, Dashboard 4's reliance on multiple colour schemes and excessive visual elements led to confusion, contradicting the principle of simplicity in dashboard design (Mokkup, 2023). By contrast, the selected visualization ensures that complexity does not come at the expense of usability.

Furthermore, the selected visualization effectively shows a mix of numerical and text-based data. Text elements provide context, while numerical values are visualized in a way that enhances pattern recognition without overwhelming the user. This combination ensures a holistic representation of the dataset while maintaining clarity and interpretability.

## 5. Discussion

The evaluation of the visualization techniques revealed several key insights regarding the effectiveness of different design choices. A consistent colour scheme and simple visualizations contributed to faster response times, enabling users to interpret data quickly and efficiently. However, a recurring issue in user feedback was the absence of legends, which led to difficulties in understanding certain elements of the visualization. Additionally, while easy-to-interpret visualizations were beneficial, they often lacked depth. To address this, simple dashboards need to incorporate additional dimensions or interactivity to provide a more comprehensive analysis.

The factors that contribute to easily understandable visualizations include the use of a consistent colour scheme, a limited number of variables per visualization, simpler chart types, and the inclusion of clear headings and legends. These elements enhance user comprehension and reduce cognitive load, making it easier to extract insights from the data.

One of the key takeaways from the evaluation process is the importance of selecting appropriate charts and graphics based on the average visualization literacy of the target audience. The colour scheme should accurately represent relationships in the data, ensuring that visuals using the same variables maintain similar colour encoding for consistency. Furthermore, users tend to subconsciously memorize answers to specific questions, which can influence their ability to analyse new data objectively. Open-ended questions also posed challenges, as users often found them confusing or difficult to answer. To improve evaluation effectiveness, the Stopwatch Method should be combined with other evaluation techniques to yield more significant results. Interactivity is another crucial aspect, as it allows for deeper exploration and interpretation of data.

The strengths of the final visualization include a simple and structured design with clear headings, effective colour coding, and a well-defined legend. The use of colour luminance facilitated pattern recognition, and the dashboard successfully showcased multiple dimensions. However, some weaknesses were identified, such as the presence of two similar tables, which caused confusion among users. Additionally, certain pieces of information presented in text format were not as intuitive as graphical representations. The dashboard was also limited in the number of dimensions it could display at once, restricting the depth of analysis. Most notably, the lack of interactivity hindered users from engaging more deeply with the data.

The dashboard was specifically designed to address the challenges identified in the research question. The need to fit all relevant information into a single visualization was met through the use of a dashboard format, which allows for the structured presentation of complex data. The focus on Airbnb locations ensured that the visualization remained relevant to the target audience, as they were only interested in areas with available Airbnb listings. Although crime data was distributed generically across different areas, it still provided valuable insights when combined with other relevant features, making it applicable to a broader decision-making process.

In scenarios involving larger datasets, the use of interactivity becomes even more critical. An interactive dashboard allows users to filter, zoom, and explore different layers of data, revealing deeper insights that would not be immediately apparent in a static visualization. The dashboard design is well-suited for handling large amounts of data, as it organizes information in a structured and accessible manner.

The overall design of the visualization is justified by several factors. The use of a clear and structured colour scheme, predominantly blue, contributed to the readability and user-friendliness of the dashboard. Visual clarity was ensured through the use of well-defined headings, a legend, and distinct colour encoding, all of which guided users toward key insights. The dashboard's ability to present a large volume of information in an organized manner further reinforced its effectiveness as a data exploration tool.

# 6. References

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2022). *Dashboard design patterns.* IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/tvcg.2022.3209448

Batiz, O. (2025). *Los Angeles Airbnb Listings*. Kaggle. https://www.kaggle.com/datasets/oscarbatiz/los-angeles-airbnb-listings

FreshBI. (2025). *Color Theory in Dashboard Design*. https://freshbi.com/blogs/color-theory-in-dashboard-design/

Isha (2025). *Crime Data from 2020 to Present*. Kaggle.https://www.kaggle.com/datasets/ishajangir/crime-data

Janes, A., Sillitti, A., & Succi, G. (2013). *Effective dashboard design*. Cutter IT Journal, 26, 17–24.

Lavalle, A., Maté, A., Santos, M. Y., Guimarães, P., Trujillo, J., & Santos, A. (2025). *A methodology for the systematic design of storytelling dashboards applied to industry 4.0*. Data amp; Knowledge Engineering, 156, 102410. https://doi.org/10.1016/j.datak.2025.102410

Mokkup (2023). *8 Essential Dashboard Design Principles for Effective Data Visualization.* Medium. https://medium.com/@mokkup/8-essential-dashboard-design-principles-for-effective-data-visualization-40653c5fd135

Nazemi, K., Burkhardt, D., Hoppe, D., Nazemi, M., & Kohlhammer, J. (2015). Web-based evaluation of information visualization. Procedia Manufacturing, 3, 5527–5534. https://doi.org/10.1016/j.promfg.2015.07.718

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). *What do we talk about when we talk about dashboards?* IEEE Transactions on Visualization and Computer Graphics, 25 (1), 682–692. https://doi.org/10.1109/tvcg.2018.2864903

Shafi, A (2023). *K-Nearest Neighbors (KNN) Classification with scikit-learn.* Datacamp. https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn

# Appendix

Interview Questions:

1. What is the most dangerous time to visit LA in terms of time of the day and weekday?
2. Which areas that have decent amount of Airbnbs would you avoid living in given the crimes?
3. What are the most dangerous public places to visit in LA?
4. Additional comments about the visualization?

.

*Table 1- Comparison of selected visualization techniques.*

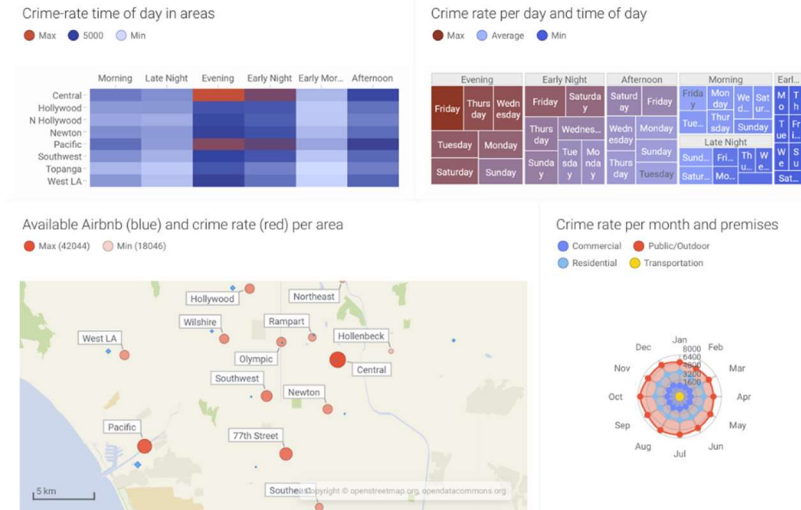| | Vis 1: Dashboard 1 (Evaluators' Decision) | Vis 2: Dashboard 2 (My selection) |
|---|---|---|
| | <br>Fig 1. Dashboard 1 | <br>Fig 2. Dashboard 2 |
| What | Categorical attribute, quantitative, ordered and spatial data. | static table with a lot of categorical attributes (Premis, Time in the day, Weekday, Location), and some quantitative attributes (Crime count, Airbnb count) about crime data in LA and Airbnb listings |
| Why | Communicating to the target audience dangerous timings, areas and places to be in LA. Dashboard is perfect for multivariable questions. | give tourists information on how to be safe in LA especially focusing on weekday, time in the day and location, Dashboard is perfect for multivariable questions |
| How | Highlight tables (using intensity), map (using scaled colours and marker size) horizontal bar chart and treemap (using brightness) | Dashboard: Symbol Map (Colour Luminance, Position & Size), Tree Map (Size, Position & Color Luminance), Highlight tables (Colour Luminance, Position) |
| Evaluation result | Fast response times and consistent answers, Consistent color scheme & simple visualizations = faster response time, questions were answered very fast since easy to understand charts, Common Feedback: Absence of clear legends | Dashboard is easy to understand and people are able to answer the questions fast, Consistent color scheme & simple visualizations = faster response time, Dashboard was very simple and therefore doesn't show complex correlations between features |
| Advantage (briefly) | Simplicity, minimalistic design, easy to read & well used channels. | simple structure (with headings) and colour coding, clear legend, easy pattern recognition through colour luminance, showcasing multiple dimensions |
| Disadvantage (briefly) | Low number of variables and lack of complexity, Simple dashboards must include more dimensions to add visual depth | two similar tables are confusing, some information in text format (not intuitive), limited amount of dimesions on the dashbord (5 in total), Simple dashboards must include more dimensions to add visual depth |

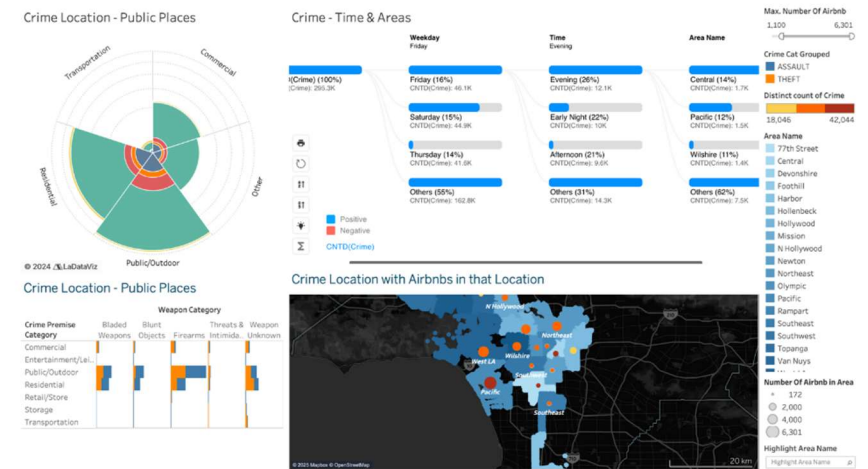| | Vis 3: Dashboard 3 | Vis 4: Dashboard 4 |
|---|---|---|
| | Fig 3. Dashboard 3 | Fig 4. Dashboard 4 |
| What | Categorical attribute, quantitative-, ordered- and Spatial Data | Data of crime & Airbnbs in LA, across various attributes - categorical (Location, Time, Crime & Weapon category) & numerical (Crime count & Airbnb counts) |
| Why | To discover and compare crime rate across different area and time, Dashboard is perfect for multivariable questions | To help tourists in LA to find crime hotspots - basis the intensity of crime (to enable safer stay in LA), Dashboard is perfect for multivariable questions |
| How | With the use of color intensity, size, position, color hue, length and spatial region | Polar Area Chart, Drill-down Tree, Map, Gantt Bar Chart - Colors, Size |
| Evaluation result | Color intensity could have been higher and radar chart was hard to read, Common Feedback: Absence of clear legends | Dashboard was moderately easy to interpret, as interactive features of the Dashboard required time, but the information was clear and enough for them to answer the questions |
| Advantage (briefly) | Make it easy to identify crime hotspots across different times and locations | Ability to capture multiple variables at once. interactivity, easy outlier detection, clarity of data presented |
| Disadvantage (briefly) | Radar chart could appear cluttered. Hard to distinguish between areas within the same color range | Complex visualisations, difference in colour schemes, 1 visual lacked legends, colour and shapes as paramteres - confusing |