Berlin Bicycle Theft Analysis — Final Project Report

Project GitHub: https://github.com/moe221/Bicycle-theft-in-Berlin

1 Introduction to Data, Goals, and Tasks

For this project, I worked with a dataset of reported bicycle thefts in Berlin published and maintained by the Berlin police department, updated daily and spanning from 2024 to Jun 2025. It includes key attributes like theft time, location, type of bike, and coordinates. Drawing from early-stage task elicitation, I identified three main questions (**Tasks**) to guide my design:

- Where are the hotspots? (Task 1)
  - **Why**: To identify high-risk areas for bike theft in order to inform prevention or route planning
  - **How**: By ranking or filtering areas based on bike theft counts or rates
  - **What**: Spatial patterns or clusters of high theft activity
  - **Where**: Borough-level bike theft data in Berlin
  - **When**: During the initial exploration phase to scope areas of interest
  - **Who**: Citizens or bicycle owners seeking to assess personal risk

- When do thefts peak in those hotspots? (Task 2)
  - **Why**: To understand when bike theft risk is highest in hotspot areas
  - **How**: By examining bike theft frequency across different time periods
  - **What**: Temporal trends or peaks in theft occurrences
  - **Where**: Bike theft incidents in identified high-theft boroughs
  - **When**: After identifying hotspot areas, to further investigate patterns and risk factors
  - **Who**: Bicycle owners planning safer usage times, or local authorities considering patrol schedules

- What is the safest parking location near a given address? (Task 3)
    - **Why**: To choose a bike parking location with lower theft risk near a planned destination
    - **How**: By comparing bike theft counts within walking distance of a destination, possibly using a map or proximity filter
    - **What**: Spatial variation in theft risk at a fine-grained level (e.g. by street or neighborhood block)
    - **Where**: Bike theft incidents within a 500m radius of the user's destination
    - **When**: During trip planning or upon arrival at the destination
    - **Who:** Individual bicycle users planning secure parking, or app users receiving safety recommendations

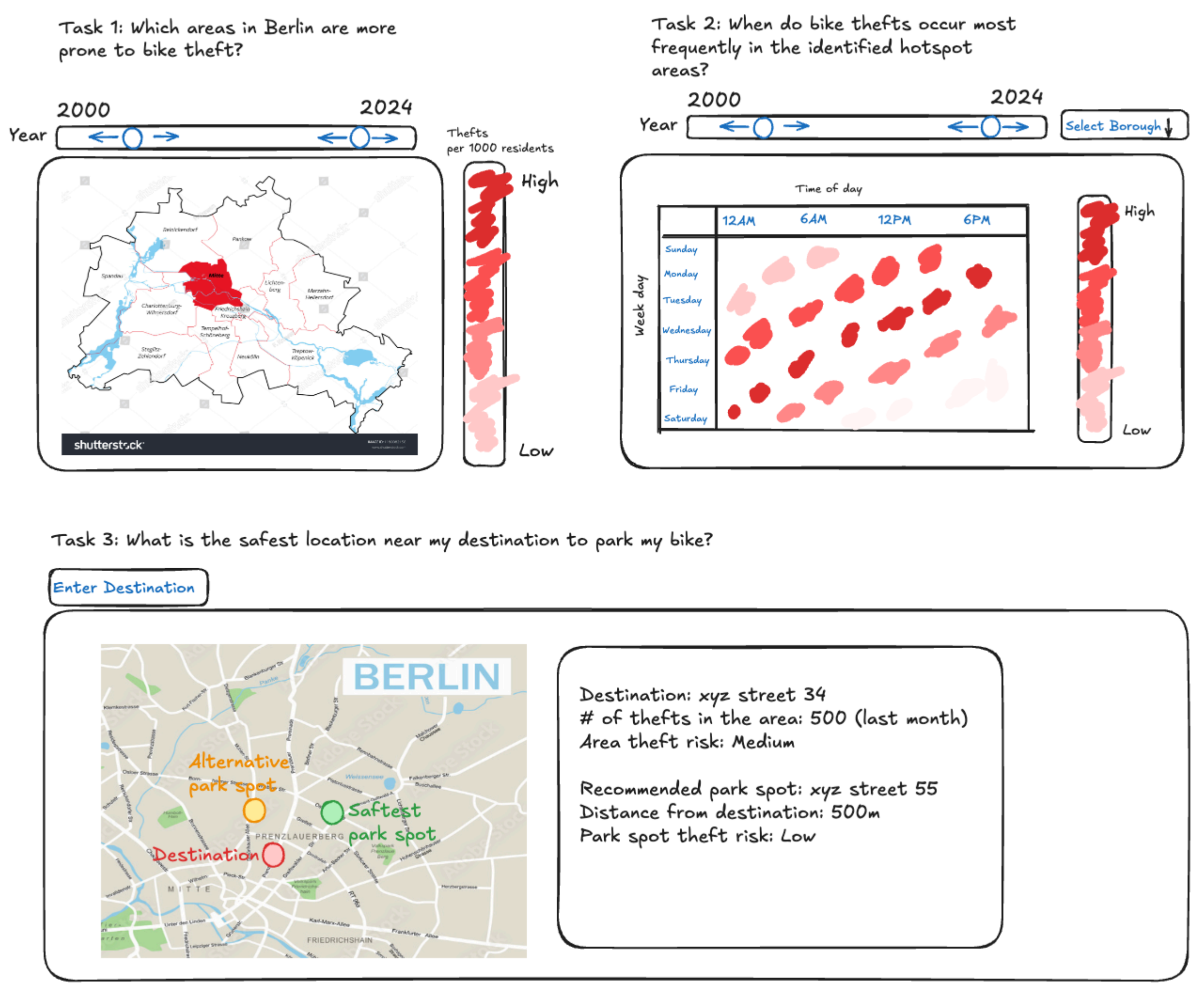
These questions directly shaped the design and functionality of my dashboard, which combines a population-normalized choropleth map, a time-based heat-map, and an interactive parking recommendation tool. Each of these supports a task and reflects a core use case: **exploration**, **planning**, and **decision-making**.

To avoid clutter, I aggregated data and prioritized only information directly relevant to the stated goals. This helped maintain clarity while keeping the interface focused.

2 Visualization Implementation

Low-fidelity Prototype

To kick off the design process, I created a low-fidelity prototype using Excalidraw. This early sketch helped me quickly lay out the structure of the dashboard, define the key views for each task, and experiment with layout ideas without getting bogged down in visuals or implementation. It also served as a reference point for feedback and guided the transition to high-fidelity development by clarifying the user flow and necessary interactions.

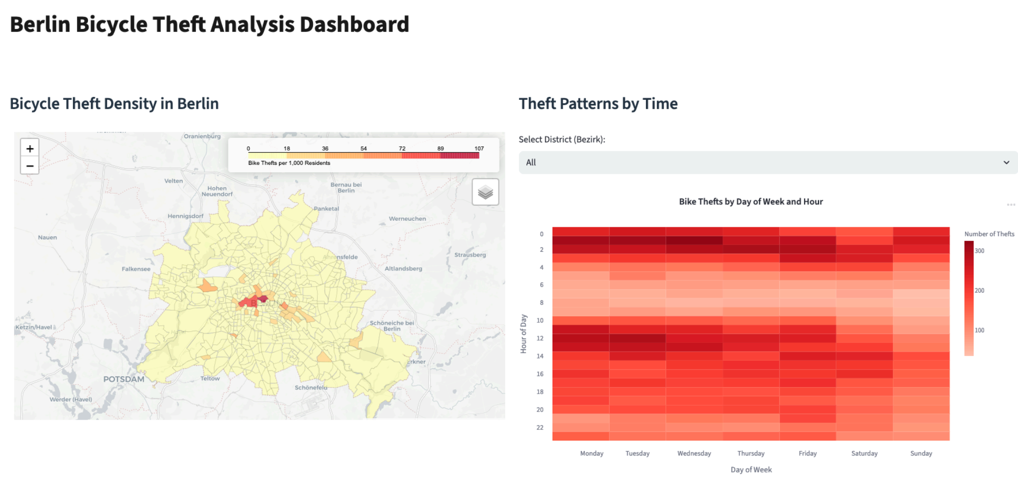


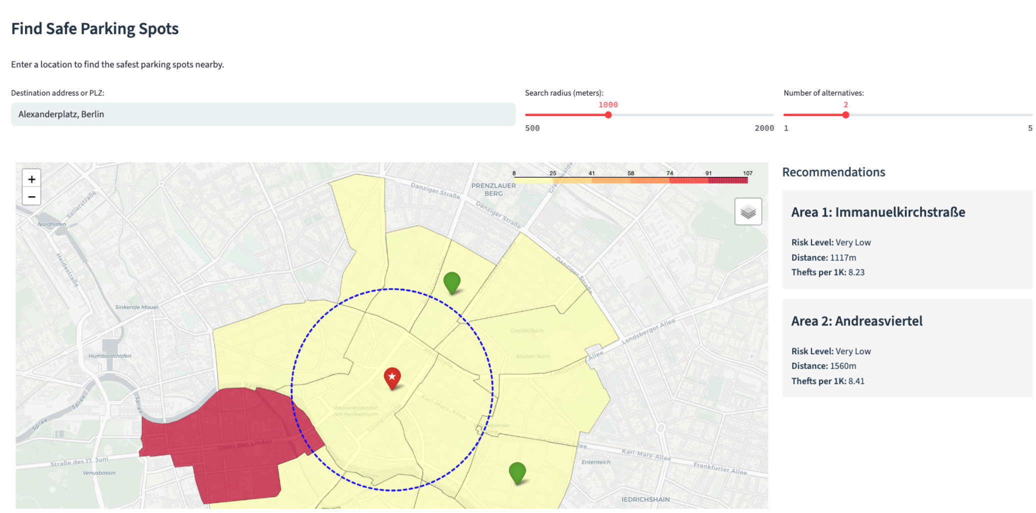
Platform & Deployment (High Fidelity Version)

I used Python with Altair for static plots, Folium for the interactive map, and Streamlit to build a web-based dashboard. I unfortunately did not deploy the dashboard as it seemed out of scope for this course project; however, **the full code is on [GitHub] with instructions to run it locally**, so I welcome you to clone the repo and follow the instructions to run the dashboard locally using streamlit.

Screenshots of Interactive Dashboard

Choropleth Map of Theft Rate (Neighbourhoods, normalized), Heat-map of Thefts by Day & Hour **and** Safe-Parking Finder Interface





(Screenshots and source code are available on the project GitHub.)

3 Design Rationale

Throughout this project, I followed the principles: "overview first, zoom and filter, details on demand". The choropleth map provides a quick overview of risk areas across Berlin. Users can then zoom into specific neighborhoods and filter by time using the heat-map. For detailed planning, the parking finder shows exact zones to consider within a given radius of a destination.

I used the Five Design Sheets method to iterate through both low-fidelity and high-fidelity prototypes. This helped me stay grounded in user needs while evolving the design rapidly.

I also relied on design principles emphasized in the course: insight-driven evaluation, avoiding data overload, and encoding clarity. Instead of packing the interface with raw data, I made intentional design choices such as aggregating thefts by hour and normalizing rates to support insight generation and storytelling.

4 Evaluation Approach & Procedure

I conducted three interviews with participants recruited through friends and family. Each session lasted about 25 minutes and followed a structured process:

- **Context Interview:** I began by asking how they currently choose where to park their bike.
- **Think-Aloud:** Participants performed two tasks:
  - finding the top three theft areas and
  - identifying the peak time for thefts in Kreuzberg while talking me through their thinking process.
- **Scenario Walk-Through:** For the third task, they entered a real destination and selected a recommended parking spot.
- **Semi-Structured Wrap-Up:** I collected feedback on perceived usefulness, trust in the system, and any difficulties encountered.
- **Insight Capture:** Finally, participants listed up to five surprising or interesting findings they discovered through the tool.

Measures & Success Criteria

- **Task Completion:** Over 95% of tasks were completed without assistance.
- **Insight Richness:** Each participant generated more than three insights.
- **Trust Signals:** Participants shared thoughts indicating they would actually use the dashboard in real-life.

5 Evaluation Results & Insights

Key Qualitative Findings

Participants provided some actionable feedback:

- They asked for more granular parking recommendations ideally street pins instead of general blocks. → This is a limitation from my data so it might not be possibly at this time.
- A couple noted they wanted walking-time labels to help weigh distance to safest parking area.
- All expressed a desire for more annotations such as CCTV presence or how recently thefts had occurred.
- Clearer legends and redundant encodings (color + number) were appreciated and reduced confusion.

6 Summary & Future Improvements

What Worked Well

The population-normalized theft map effectively highlighted areas of relative risk, which users found useful. The heat-map clearly showed peak times of theft activity, and the parking recommendation tool helped users quickly understand and compare their options. For longer-term parking, in particular, participants trusted the guidance provided.

Areas to Improve

1. Users want recommendations at the street level.
2. Display walking distances and time to make proximity trade-offs clearer.
3. Add short annotations to explain why a zone is safer (low incident history or surveillance).
4. Let users filter results based on a mix of distance and theft history.
5. Provide tooltips or a quick intro guide to improve feature discoverability.
6. Deploy dashboard and make it accessible online

7 Conclusion

This project shows how interactive recommendation tools, can help Berlin cyclists to make data-driven parking decisions. My qualitative evaluation with semistructured interviews and think-aloud studies showed both strong usability and specific areas for improvement. These methods aligned with the emphasis on insight generation and formative feedback.

Moving forward, refining data granularity and improving data transparency could make the dashboard into a reliable decision-making tool.

Project GitHub: https://github.com/moe221/Bicycle-theft-in-Berlin

Data Sources:

- **Bicycle Theft data**
- **Shape Files - Berlin**
- **Population Data - Berlin**