# Usage Patterns of Bicycle Counting Stations in Heidelberg and the Influence of External Factors

**Julian Jurcevic**[*1] **Martin Eichler**[*2] **Simon Rappenecker**[*3] **Tarik Eker**[*4]


## Abstract

- Desribe topic
- Data source (just heidelberg) and external Factors
- sentence in one method
- Results briefly


## 1. Introduction

- Short introduction, but more motivational, bicycle traffic more important over the years...
- Name what is interesting and why it matters
- Shortly introduce the data, and the main method

## 2. Data

The analysis relies on a combination of traffic, weather, and calendar datasets. In the following, we describe the acquisition, processing, and quality assessment of the cyclist data.

### 2.1. Cyclist Counting Network

We analyze continuous hourly cyclist counts from 15 permanent monitoring stations in Heidelberg. The dataset covers the period from May 2014 to the end of 2024. The data is physically collected using **Eco-Counter** systems equipped with **ZELT inductive loops** (Eco-Counter, 2025). These loops are embedded in the pavement and differentiate between cyclists and motorized vehicles based on electromagnetic signatures. Most stations record traffic separately for each direction ("Inbound" vs. "Outbound"), which allows for a detailed analysis of commuting flows.

**Acquisition and Geography.** We retrieved the data from the **MobiData BW** (Nahverkehrsgesellschaft Baden-Württemberg mbH (NVBW), 2025) open data platform. Using a custom Python script, we programmatically fetched and merged the raw CSV files, applying a spatial filter to extract only stations within the administrative domain of Heidelberg.(Maybe rein) Although the raw data includes meteorological fields, we excluded them due to sensor inconsistencies and relied on official weather records instead.

Heidelberg's central topography is predominantly flat, which facilitates a high modal share of cycling. The counting stations are located at critical bottlenecks and key corridors (see Figure 1). The most important sensors are situated on the Neckar bridges, specifically the **Ernst-Walz-Brücke** and the **Theodor-Heuss-Brücke**, in 1 marked red. These locations capture the main traffic volume commuting between the university campus (*Im Neuenheimer Feld*) and the city center.

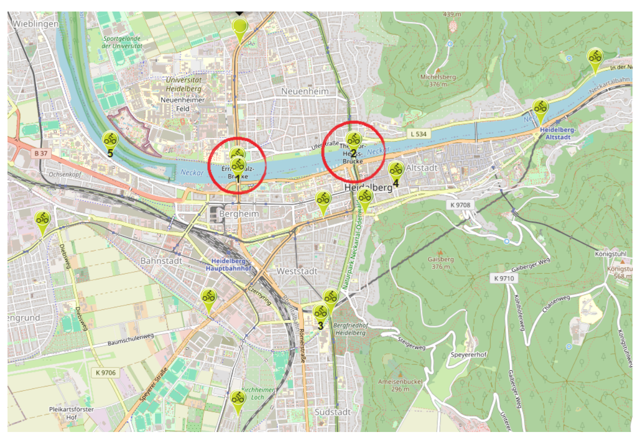


*Figure 1.* Geographical distribution of bicycle counting stations in Heidelberg. The core network along the Neckar river captures the majority of commuter traffic, while peripheral stations monitor local routes.

---

[*]Equal contribution [1]Matrikelnummer 7167089, MSc Computer Science [2]Matrikelnummer 6009076, MSc Computer Science [3]Matrikelnummer 6324777, MSc Machine Learning [4]Matrikelnummer 5668988, MSc Computer Science. Correspondence to: JJ <7167089julian-steffan.jurcevic@student.uni-tuebingen.de>, ME <martin.eichler@student.uni-tuebingen.de>, SR <simon.rappenecker@student.uni-tuebingen.de>, TE <tarik.eker@student.uni-tuebingen.de>.

*Project report for the "Data Literacy" course at the University of Tübingen, Winter 2025/26 (Module ML4201).* Style template based on the ICML style files 2025. 

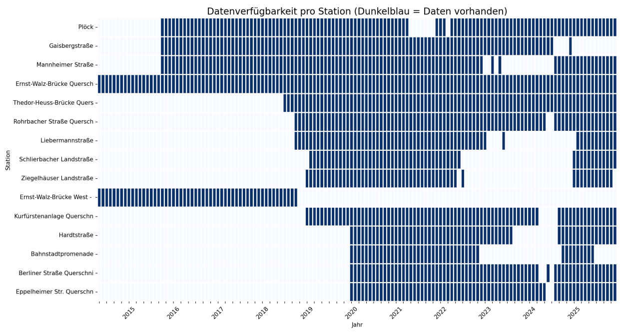


*Figure 2.* Data quality heatmap for all counting stations over the analysis period. Blue cells indicate available data, while white cells represent missing hours or false-zero days.

**Data Sanity.** To ensure the validity of our results, we performed a longitudinal quality check. We checked for two main issues: temporal gaps (missing hours) and "false zeros" (days with a total count of zero, indicating sensor failure). The data quality varies significantly between stations. Figure 2 shows a visualization of data availability over the entire 11-year period. Blue cells indicate valid data, while white cells represent either missing hours or false-zero days.

- **High Reliability:** The core network is extremely stable. The *Ernst-Walz-Brücke* recorded only 12 missing hours over the entire 11-year period and had zero false-zero days. This makes it an ideal baseline for long-term trend analysis.
- **Sensor Failures:** In contrast, peripheral stations like *Mannheimer Straße* (621 zero-days) or *Schlierbacher Landstraße* (large data gaps) showed significant hardware issues. Consequently, we excluded these unstable stations from our longitudinal analysis to avoid statistical bias. (noch nd sicher)

- Weather data,
- Why we need this, Station data is bad, describing fetching, here we name the source
- OPTIONAL: Give comparision plot between temperatures and bad metrics as rain PLOT: WEATHERDATA

- Accident data? Not used, but need to include it

- Holiday data,
- We we need it, describe source, also name that there were some errors in the dataset, we have to be careful, ...

## 3. Method

Previous works often rely on predefined rules to categorize bicycle traffic (Miranda-Moreno et al., 2013). However, such an approach may fail to capture hybrid usage patterns. We propose a more data-driven, implicit approach. By extracting features that quantify the 'shape' of daily, weekly, and seasonal profiles, we map each station into a multi-dimensional feature space. This allows us to use unsupervised k-means clustering (Lloyd, 1982) to discover different station types.

In a subsequent step in chapter 4, we assign one of the following labels to each cluster (cf. Fig. ):

Figure Patterns, Indices

- *Utilitarian*: Traffic showing strong commuting patterns, with two distinct peaks in traffic volume during the morning and afternoon hours.

- *Recreational*: Traffic dominated by leisure activities. These locations exhibit a single peak in traffic volume around midday or early afternoon.

- *Mixed*: Locations that serve a dual purpose, showing characteristics of both utilitarian and recreational usage.

To characterise the counting stations, we derive three distinct features that capture traffic patterns across different timescales: daily, weekly and seasonal to exploit the distinct pattern visible in Figure xyz .

Figure pattern

### Double Peak Index (DPI)

Weekday hourly profiles of some stations exhibit a double-peak structure, typically associated with morning and evening commuting. The DPI quantifies this behaviour by identifying dominant morning (5–10 h) and evening (14–20 h) peaks and relating their magnitudes to the average midday level (8–14 h). Stations with clear, balanced commuting peaks yield high DPI values, whereas flat or single-peak profiles result in low scores.

Formally, let $p_m$ and $p_e$ denote the magnitudes of the morning and evening peaks at hours $h_m$ and $h_e$, and let $m$ be the average midday traffic level. The DPI is defined as

$$\mathrm{DPI} = \max(S \cdot Y \cdot D,\, 0)$$

where $S$, $Y$ and $D$ models strength, symmetry and distance, respectively:

$$S = \frac{(p_m - m) + (p_e - m)}{2}$$
$$Y = 1 - \frac{|p_m - p_e|}{\max(p_m, p_e)}$$
$$D = \min\left(\frac{|h_e - h_m|}{10}, 1\right)$$

**Weekend Shape Difference (WSD)**

Differences between weekday and weekend hourly traffic patterns provide an additional discriminator between usage types. To capture this effect, we compare the shape of the weekday and weekend hourly profiles. Let $\mathbf{p}^{wd}$ and $\mathbf{p}^{we}$ denote the weekday and weekend hourly profiles, normalised to sum to one.

$$\text{WSD} = \left\| \frac{\mathbf{p}^{wd}}{\sum_h p_h^{wd}} - \frac{\mathbf{p}^{we}}{\sum_h p_h^{we}} \right\|_2$$

**Seasonal Drop Index (SDI)**

Finally, we consider long-term patterns. Seasonality provides a discriminator between leisure and utilitarian-oriented stations. The SDI quantifies the relative decline between high and low-usage months. Let $I_m$ denote the monthly index values of a station. Using upper and lower quantiles to ensure robustness against outliers

$$q_{90} = \text{quantile}_{0.9}(I_m)$$
$$q_{10} = \text{quantile}_{0.1}(I_m)$$

the SDI is defined as

$$\text{SDI} = \frac{q_{90} - q_{10}}{q_{90}}$$

High values indicate strong seasonal variation, whereas low values correspond to relatively stable, year-round usage.

## 4. Results

- describe temporal shift, describe why this is expected and why, PLOT: city,
- describe holidays impact, describe behaviour (similar as weekend) PLOT: show shift
- include also plot of different public holidays (my own one), this is just to have a funny fact (Vater Tag) PLOT: funny plot, na werden zu viele plots
- describe weather impact, obvious PLOT: weather

## 5. Conclusion

- summarize
- limitations
- problems
- statements that can be made

## Contribution Statement

Add this, see original template

## References


Eco-Counter. Zelt evo – inductive loop for mixed traffic. https://de.eco-counter.com/losungen/produkte/zelt-evo-selective, 2025. Accessed: 2025-12-27.

Lloyd, S. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

Miranda-Moreno, L. F., Nosal, T., Schneider, R. J., and Proulx, F. Classification of bicycle traffic patterns in five north american cities. *Transportation Research Record*, 2339(1):68–79, 2013. doi: 10.3141/2339-08. URL https://doi.org/10.3141/2339-08.

Nahverkehrsgesellschaft Baden-Württemberg mbH (NVBW). Bicycle counting data repository (api v2). https://mobidata-bw.de/fahrradzaehldaten/v2/, 2025. Accessed: 2025-12-27.