Exposure-Normalized Bicycle Crash Risk Along Berlin Routes

Eric Berger [*1] **Edward Eichhorn** [*2] **Liaisan Faidrakhmanova** [*3] **Luise Grasl** [*4] **Tobias Schnarr** [*5]

Abstract

We plan to investigate route-level bicycle crash risk in Berlin by addressing a key limitation in current safety analyses: raw crash counts don't distinguish between dangerous roads and simply busy roads. Our approach will combine police-recorded crash data from the German Unfallatlas with recent city-wide traffic volume estimates to calculate a normalized risk metric at the street-segment level. The core idea is to identify which segments have high crash rates per cyclist rather than just high absolute crash numbers. We will then explore how to aggregate these segment-level risks into route-level scores that could inform safer cycling navigation. Building on this, we aim to score a user-specified route and, where possible, recommend alternative routes with lower estimated risk while maintaining comparable convenience (e.g., distance or travel time).

1. Introduction

Motivate the problem, situation or topic you decided to work on. Describe why it matters (is it of societal, economic, scientific value?). Outline the rest of the paper (use references, e.g. to Section 2: What kind of data you are working with, how you analyse it, and what kind of conclusion you reached. The point of the introduction is to make the reader want to read the rest of the paper.

[*]Equal contribution [1]Matrikelnummer 12345678, MSc Machine Learning [2]Matrikelnummer 12345678, MSc Computer Science [3]Matrikelnummer 12345678, MSc Media Informatics [4]Matrikelnummer 12345678, MSc Medical Informatics [5]Matrikelnummer 12345678, MSc Medical Informatics. Correspondence to: Initials1 <first1.last1@uni-tuebingen.de>, Initials2 <first2.last2@uni-tuebingen.de>, Initials3 <first3.last3@uni-tuebingen.de>, Initials4 <first4.last4@uni-tuebingen.de>, Initials5 <first4.last4@uni-tuebingen.de>.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2025/26 (Module ML4201).

2. Data and Methods

In this section, describe *what you did*. Roughly speaking, explain what data you worked with, how or from where it was collected, it's structure and size. Explain your analysis, and any specific choices you made in it. Depending on the nature of your project, you may focus more or less on certain aspects. If you collected data yourself, explain the collection process in detail. If you downloaded data from the net, show an exploratory analysis that builds intuition for the data, and shows that you know the data well. If you are doing a custom analysis, explain how it works and why it is the right choice. If you are using a standard tool, it may still help to briefly outline it. Cite relevant works. You can use the `\citep` and `\citet` commands for this purpose (MacKay, 2003).

3. Results

In this section outline your results. At this point, you are just stating the outcome of your analysis. You can highlight important aspects ("we observe a significantly higher value of x over y"), but leave interpretation and opinion to the next section. This section absoultely *must* include at least two figures.

4. Discussion & Conclusion

Use this section to briefly summarize the entire text. Highlight limitations and problems, but also make clear statements where they are possible and supported by the analysis.

Contribution Statement

Explain here, in one sentence per person, what each group member contributed. For example, you could write: Max Mustermann collected and prepared data. Gabi Musterfrau and John Doe performed the data analysis. Jane Doe produced visualizations. All authors will jointly wrote the text of the report. Note that you, as a group, a collectively responsible for the report. Your contributions should be roughly equal in amount and difficulty.

Notes

Your entire report has a **hard page limit of 4 pages** excluding references and the contribution statement. (I.e. any pages beyond page 4 must only contain the contribution statement and references). Appendices are *not* possible. But you can put additional material, like interactive visualizations or videos, on a githunb repo (use links in your pdf to refer to them). Each report has to contain **at least three plots or visualizations**, and **cite at least two references**. More details about how to prepare the report, inclucing how to produce plots, cite correctly, and how to ideally structure your github repo, will be discussed in the lecture, where a rubric for the evaluation will also be provided.

References

MacKay, D. J. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.