
Abstract Template for Data Literacy 2023/24

Deutsche Bahn - Predicting and Visualising the One Exception to the German Punctuality Rule

Jiha Kim ^{* 1} Kim Zierahn ^{* 2} Sumitrra Bala Subramaniam ^{* 3} Frederik Panse ^{* 4}

Abstract

The reputation of Deutsche Bahn (DB) raises questions regarding its reliability in maintaining schedules. This project aims to predict the punctuality of trains operated by DB with respect to both time and geographical locations. We are planning to use data provided by the [Deutsche Bahn](#) in 2016 and the corresponding [directory of operating sites](#). We came across this dataset on recorded train schedules and delays in 2016 on Deutsche Bahn's website as some of our friends and other members of the community in Tübingen are affected by the train delays almost on a day-to-day basis. We are planning to do a linear regression, predicting the minutes of delay. Our features will include the coordinates, the month, the (week)day and the number of trains. We are planning to do a visualization in form of a map, depicting the delay in Germany. We are expecting to identify and predict operating locations or regions with high or low delay minutes.

1. Contribution Plan

We are planning to work as a team on this project in all phases. However, Kim Zierahn is going to do the data cleaning and EDA. Jiha Kim and Sumitrra Bala Subramaniam are going to do the predictive model. Frederik Panse is going to do the data visualization. All authors will jointly write the text of the report.

^{*}Equal contribution ¹Matrikelnummer 6640082, jiha.kim@student.uni-tuebingen.de, MSc Quantitative Data Science Methods ²Matrikelnummer 6635183, kim-isabella.zierahn@student.uni-tuebingen.de, MSc Quantitative Data Science Methods ³Matrikelnummer 6642271, sumitrra.bala-subramaniam@student.uni-tuebingen.de, MSc Quantitative Data Science Methods ⁴Matrikelnummer 5810899, frederik.panse@student.uni-tuebingen.de, MSc Quantitative Data Science Methods.