



Universität Stuttgart

Projekt Data Science
Analyse von Mobilitätsdaten

Zwischenstand

Wintersemester 2023/24
Gruppe 02 – Ozan Tastekin

Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, führt eine erhöhte Anzahl von Fahrgästen zu einer Zunahme von Verspätungen.



An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.

Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.

1. Datenbereinigung Passagierdaten:

- 11% der Zeitstempel (Zeilen) fehlen
 - Es gab kein wirkliches Pattern, warum die Zeilen fehlen könnten
 - Außer, dass der öffentliche Verkehr zwischen 2 Uhr und 3 Uhr nicht fährt
 - Aber Zeitstempel fehlen durch den ganzen Tag und auch verteilt auf alle Wochentage
 - Keine Zeile hat 0 Ein- & Aussteiger
 - Störungen/Baustellen sind auf der Webseite angegeben
 - Annahme:
 - Fehlende Zeitstempel fehlen, weil es 0 Ein- & Aussteiger gab
 - Zeilen aufgefüllt

2. Datei erstellt für:

- Welcher Haltepunkt gehört zu welcher Haltestelle?
- Welche Linie fährt diesen Haltepunkt an?

Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.

3. Haltepunkte zuordnen

- Passagierdaten sind aufgeteilt in Ost und West
- Welcher Haltepunkt gehört zu Ost, welcher zu West?

4. Datentransformierung

- Zeitstempel auf das gleiche Format gebracht
- Verspätungen ausgerechnet
 - $\text{Verspätung} = \text{Abfahrtsverspätung} - \text{Ankunftsverspätung}$
- Auf 5 Minuten zusammengerechnet
- Tabellen verbunden (Passagierfrequenzen & Verspätungen)

5. Analyse

- Spearman-Rangkorrelation

Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.

Ergebnisse

Hypothese wird bei einem r Wert von $\geq 0,5$ angenommen

$0,5 < r < 0,7$ wird als „hohe positive Korrelation“ bezeichnet

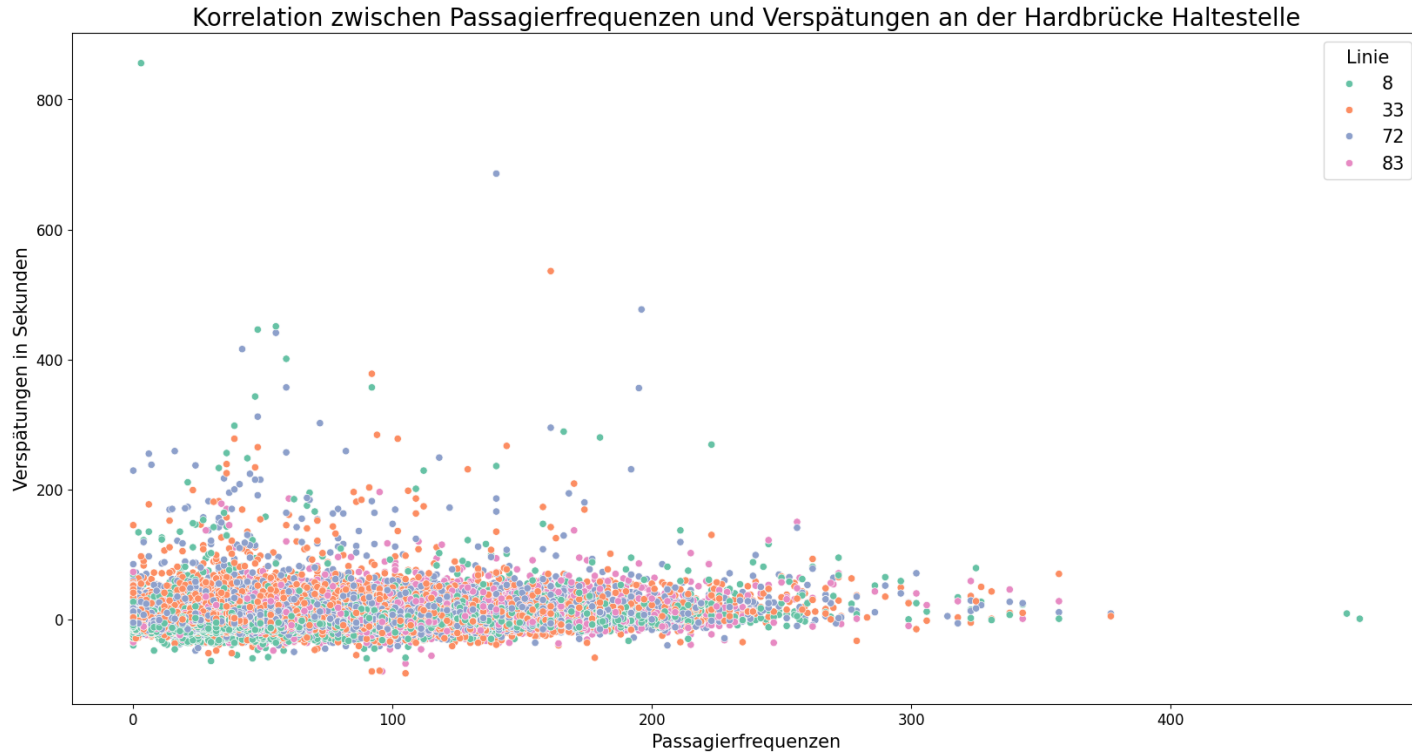
Spearman-Rangkorrelation hat $r = 0.1529132346606901$ ergeben

Hypothese wird abgelehnt

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, führt eine erhöhte Anzahl von Fahrgästen zu **keiner** Zunahme von Verspätungen.

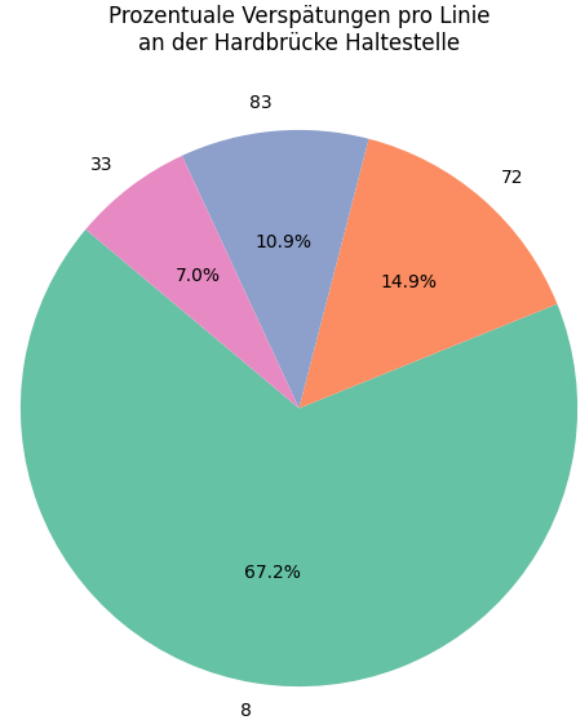
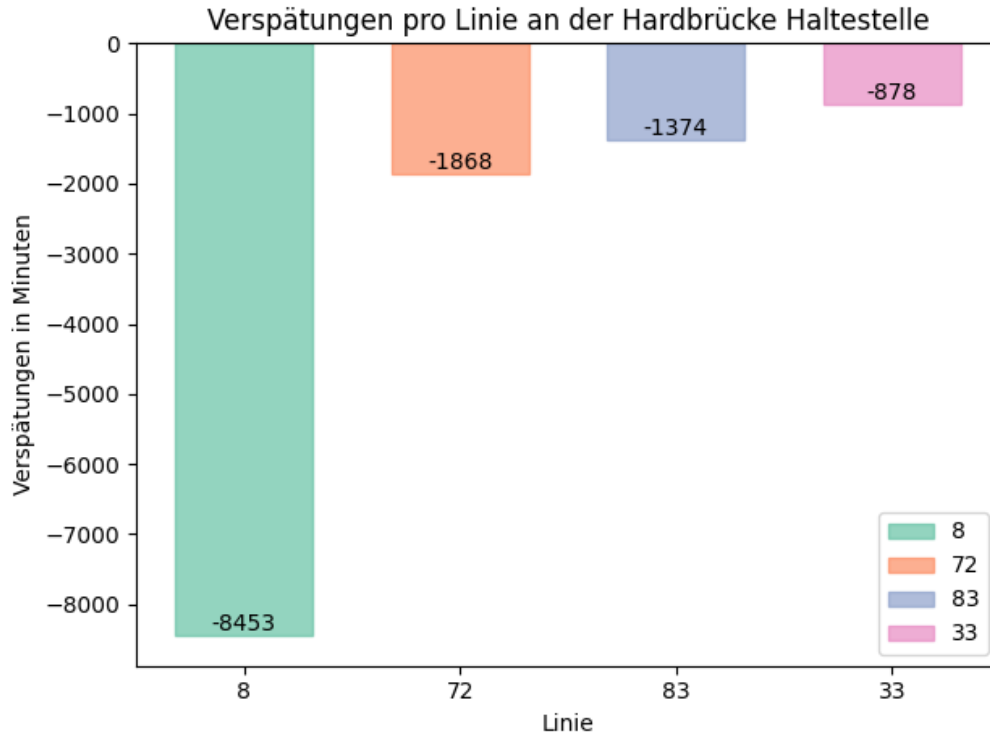
Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.



Hypothese 01

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, gibt es einen positiven linearen Zusammenhang zwischen Fahrgästen und Verspätungen.



Hypothese 02

An der Hardbrücke-Haltestelle, auf den Linien 33, 72, 83 und 8 im Jahr 2022, zeigen Stunden mit weniger als 10 Minuten Regen eine erhöhte Anzahl von Fahrgästen auf.



In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.

Hypothese 02

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.

1. Datenbereinigung Wetterdaten

- Mussten nicht bereinigt werden

Version 1: Maximale Regendauer

2. Datentransformierung

- Zeitstempel auf das gleiche Format gebracht
- Verspätungen ausgerechnet
 - $\text{Verspätung} = \text{Abfahrtsverspätung} - \text{Ankunftsverspätung}$
 - Auf 1 Stunde zusammengerechnet
- **Maximale Regendauer der 3 Stationen genommen**
- Tabellen verbunden

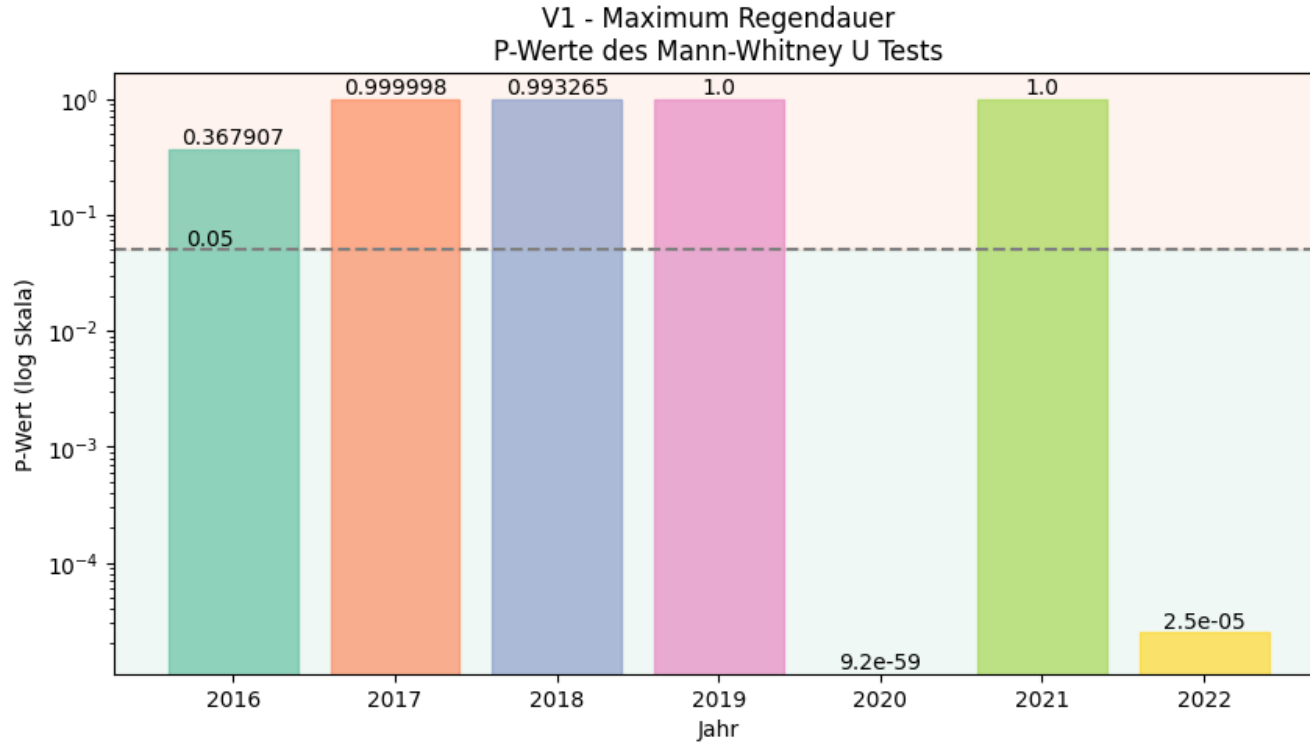
Version 2: Nächste Station

2. Datentransformierung

- Zeitstempel auf das gleiche Format gebracht
- Verspätungen ausgerechnet
 - $\text{Verspätung} = \text{Abfahrtsverspätung} - \text{Ankunftsverspätung}$
 - Auf 1 Stunde zusammengerechnet
- **Nächste Wetterstation zur Haltestelle ausgerechnet und Regendauer genommen**
- Haversine Formel zur berechnung von Distanz auf einer Kugel
- Tabellen verbunden

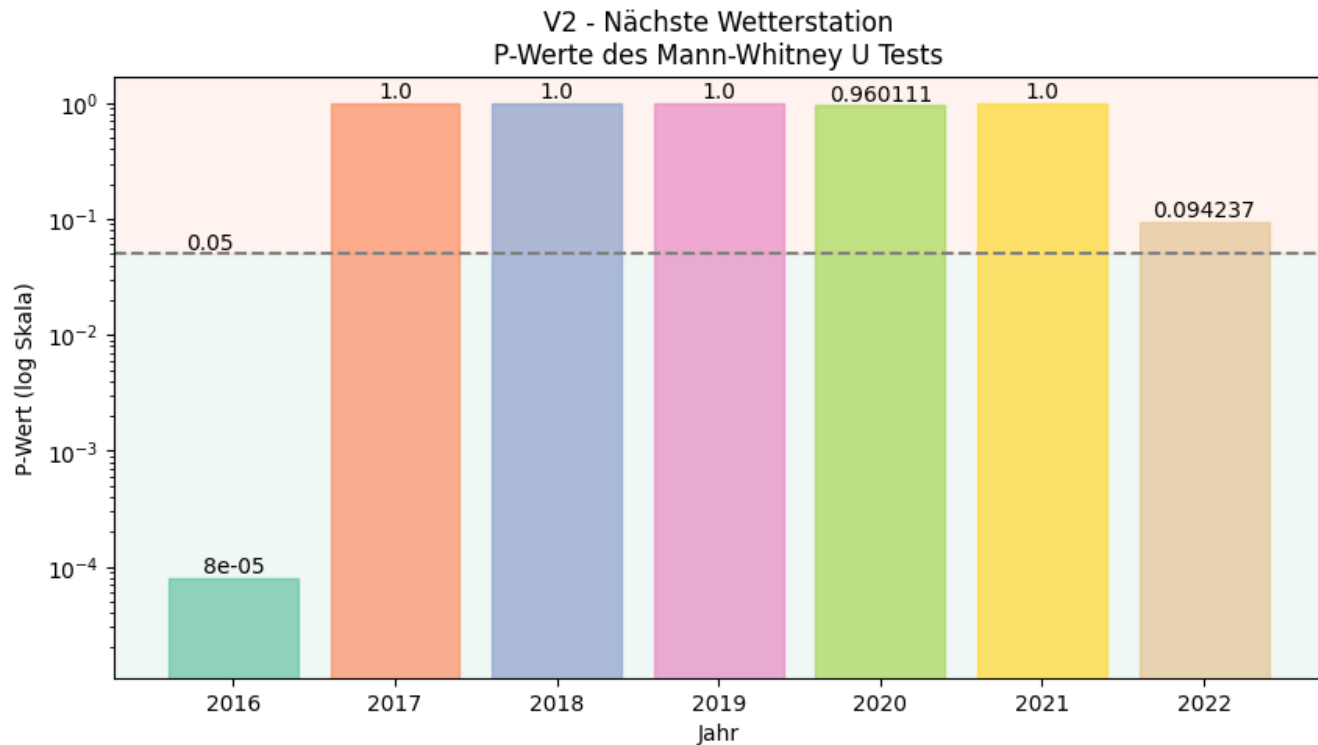
Hypothese 02 – V1

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



Hypothese 02 – V2

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



Hypothese 02

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.

Ergebnisse

Hypothese wird bei einem p Wert von $\leq 0,05$
bei mindestens 4 der 7 Jahre angenommen

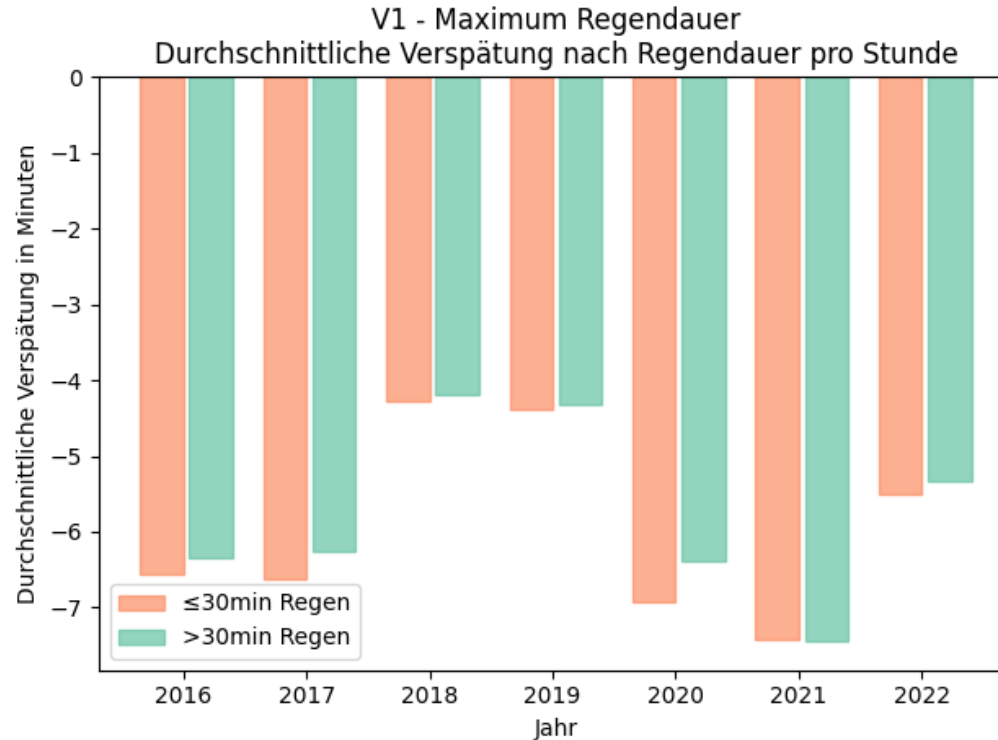
Mann-Whitney-U-Test hat bei beiden Versionen nicht
mind. 4 von 7 Jahre $p \leq 0,05$ ausgegeben

Hypothese wird abgelehnt

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen **nicht** überdurchschnittlich hohe Verspätungen.

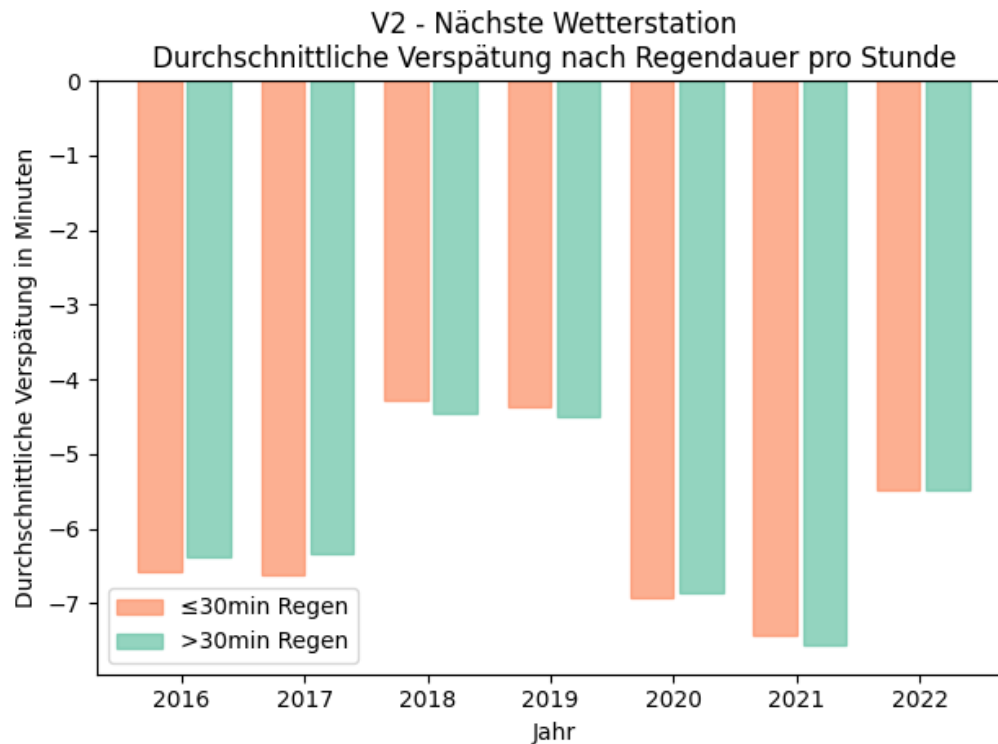
Hypothese 02 – V1

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



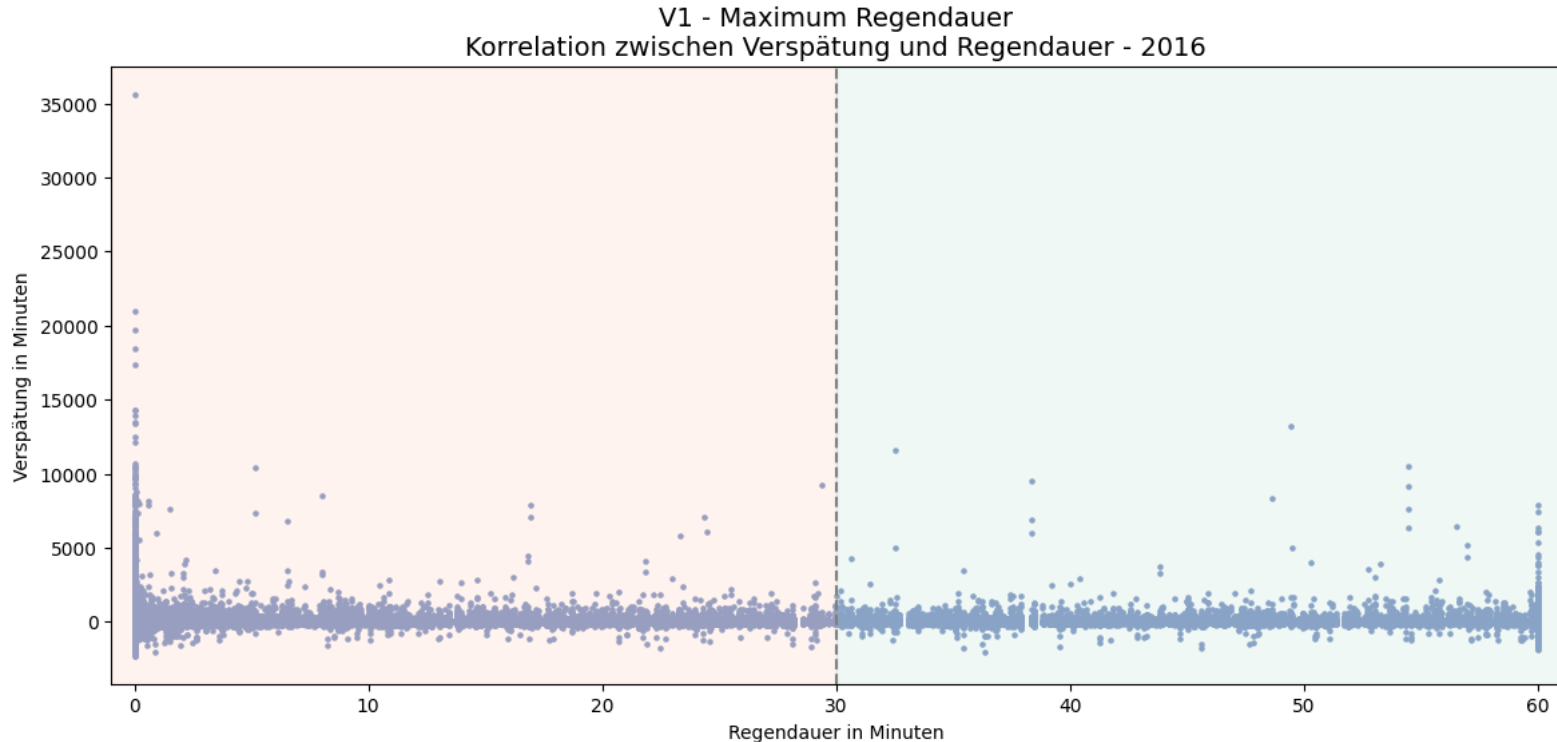
Hypothese 02 – V2

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



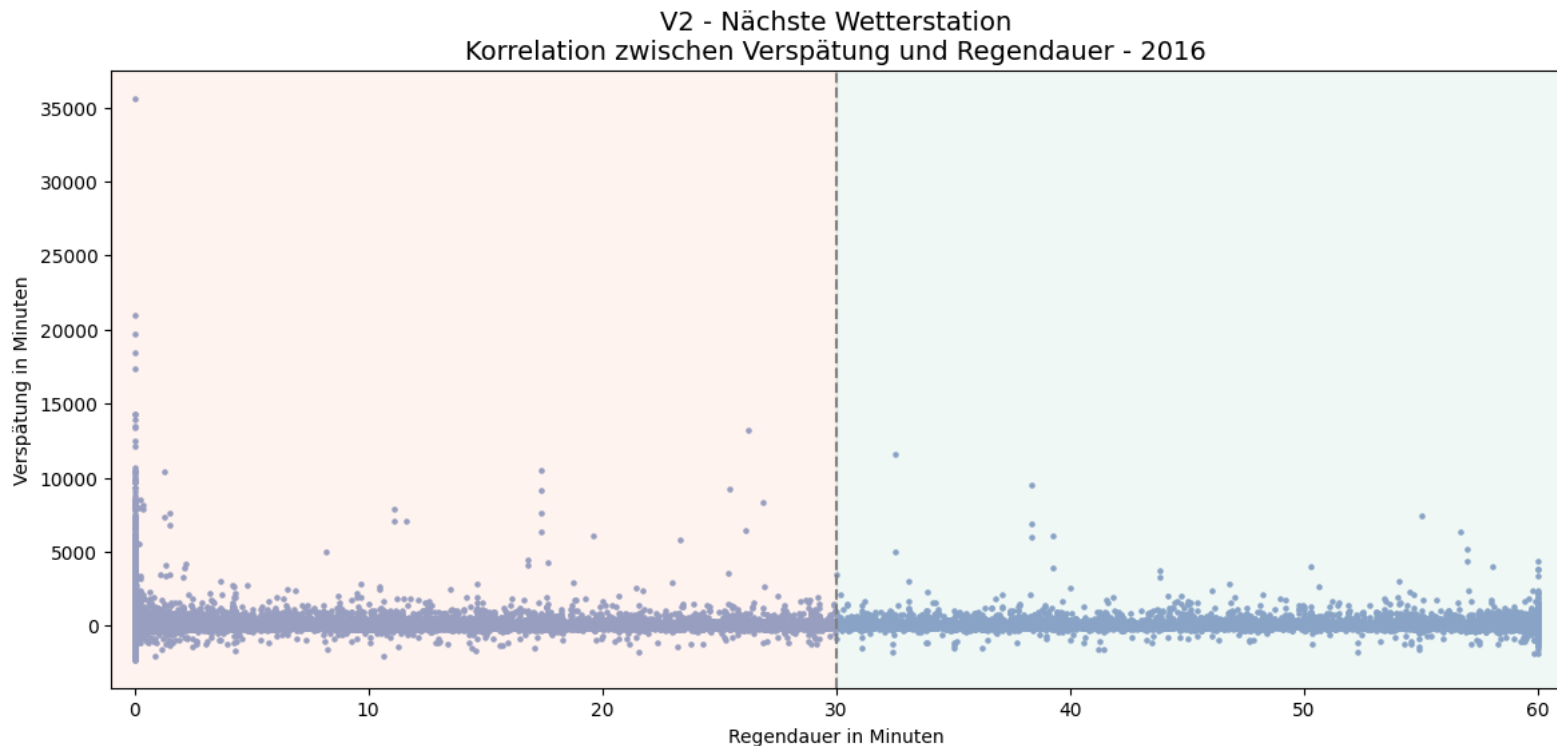
Hypothese 02 – V1

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



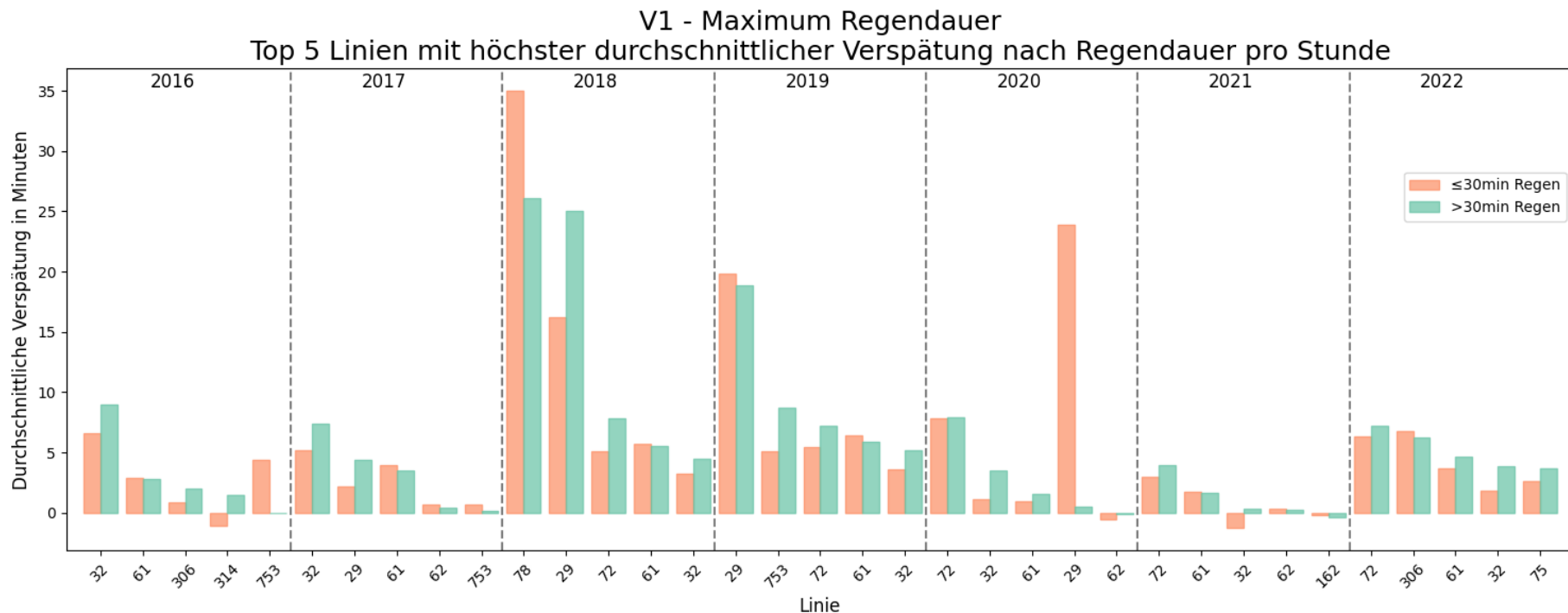
Hypothese 02 – V2

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



Hypothese 02 – V1

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.



Hypothese 02 – V2

In der Mehrheit der Jahre von 2016 bis 2022 verzeichneten Busse in Stunden mit mehr als 30 Minuten Regen überdurchschnittlich hohe Verspätungen.

