



Universität Stuttgart

**Projekt Data Science**  
**Analyse von Mobilitätsdaten**

# Hypothesen

Wintersemester 2023/24  
Gruppe 02 – Ozan Tastekin & Tony Klasan

# Letztes Mal bei...

## Datenbereinigung

### Datenbereinigung – Spalten Umbenennen & Entfernen



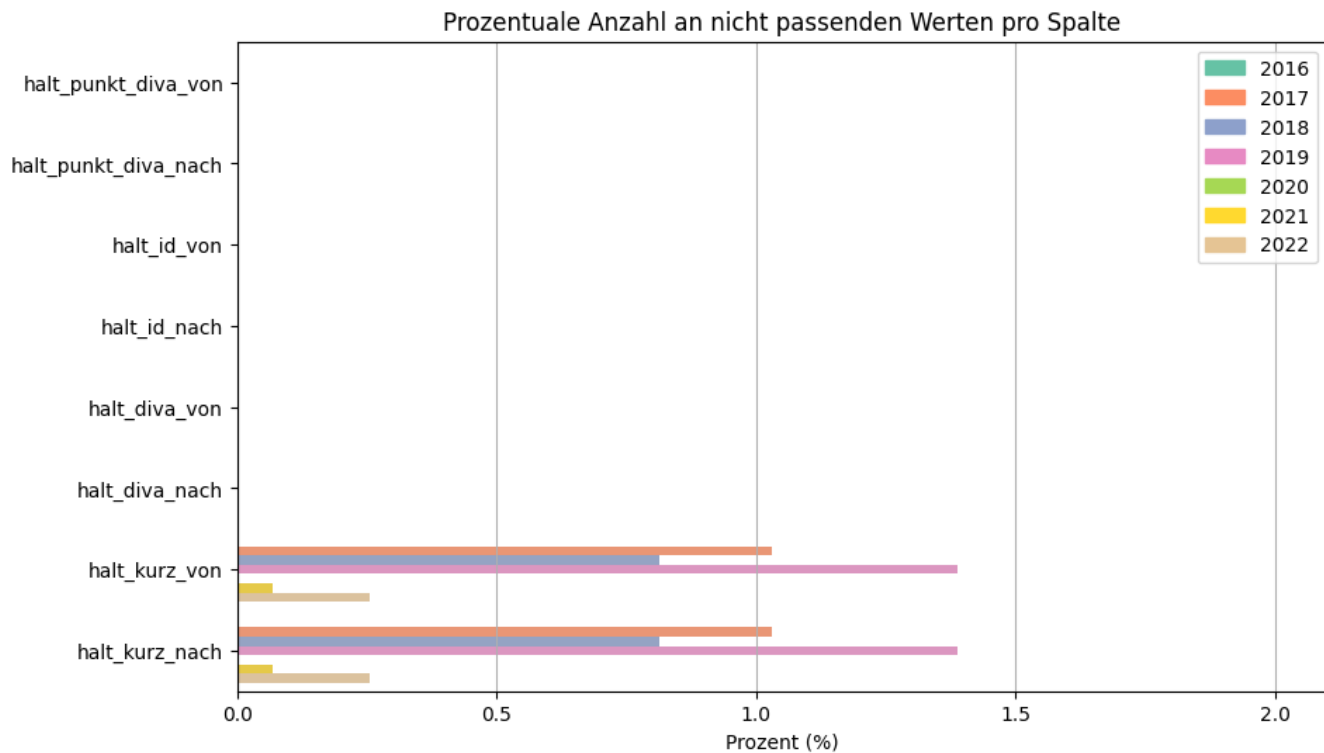
Fahrzeiten		
<u>linie</u>	<u>soll ab von</u>	<u>fahrweg id</u>
<u>richtung</u>	<u>ist ab von</u>	<u>fw no</u>
<u>betriebsdatum</u>	<u>seq nach</u>	<u>fw typ</u>
<u>fahrzeug</u>	<u>halt diva nach</u>	<u>fw kurz</u>
<u>kurs</u>	<u>halt punkt diva nach</u>	<u>fw lang</u>
<u>seq von</u>	<u>halt kurz nach1</u>	<u>umlauf von</u>
<u>halt diva von</u>	<u>datum nach</u>	<u>halt id von</u>
<u>halt punkt diva von</u>	<u>soll an nach</u>	<u>halt id nach</u>
<u>halt kurz von1</u>	<u>ist an nach1</u>	<u>halt punkt id von</u>
<u>datum von</u>	<u>soll ab nach</u>	<u>halt punkt id nach</u>
<u>soll an von</u>	<u>ist ab nach</u>	
<u>ist an von</u>	<u>fahrt id</u>	

Haltepunkte
<u>halt punkt id</u>
<u>halt punkt diva</u>
<u>halt id</u>
<u>GPS Latitude</u>
<u>GPS Longitude</u>
<u>GPS Bearing</u>
<u>halt punkt ist aktiv</u>

Haltestellen
<u>halt id</u>
<u>halt diva</u>
<u>halt kurz</u>
<u>halt lang</u>
<u>halt ist aktiv</u>

# Letztes Mal bei...

## Datenbereinigung



Hypothese 01 - Ozan

**Heißen mehr Passagiere auch automatisch  
mehr Verspätungen?**

**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.**

## Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022
- Passagierfrequenzen an Haltestelle Hardbrücke<sup>[2]</sup>
  - Ein- und Aussteiger an 2 Gleisen (Hin- und Rückrichtung)  
=> Buslinien 33, 72, 83 und Tramlinie 8 davon betroffen
  - Gleise sind auf separaten Bahnsteigen
  - Alle **5 Minuten** gezählt
  - 2020-2023

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

[2] [https://data.stadt-zuerich.ch/dataset/vbz\\_frequenzen\\_hardbruecke](https://data.stadt-zuerich.ch/dataset/vbz_frequenzen_hardbruecke)

## Spearman-Rangkorrelation

Weil beide Variablen nicht Normalverteilt sind

## Variablen & grober Ablauf

### **Verspätungen an Hst. & Passagierfrequenz pro 5 Min**

- Nicht sicher wegen Ablauf, Problem:  
Linien sind nicht genau bei 5 Minuten Marke an Haltestelle  
=> Welche Linie bekommt wie viele Ein- und Aussteiger?
- Korrelationskoeffizient  $r$  wird ausgerechnet

## Annahmekriterium

Hypothese wird bei einem  $r$  Wert von  $\geq 0,6$  angenommen

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

## Hypothese 02 - Tony

**Gab es Fahrplanänderungen, die eine Linie entlasten sollte, wenn ja, haben sich die Verspätungen der Linie auch verbessert?**

**Durch die Fahrplanänderung am 09. Dezember 2018 haben sich die Verspätungen bei den Linien 744 und 743 von Montag bis Freitag verringert.**

## Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022
- Fahrplanwechsel Dezember 2018<sup>[2]</sup>
  - Zur Mittagszeit fährt mit der Linie 744 ein neuer halbstündiger Bus
  - Linie 743 entlastet durch Linie 745
    - 745 fährt länger

## Annahmekriterium

Statistischer Test

Wert zwischen 0 und 1  $\Rightarrow$  Annahmewert  $\geq 0,05$

## Erklärung

- $p < 0,01$  – starke Evidenz gegen die Nullhypothese
- $p > 0,10$  – keine Evidenz gegen die Nullhypothese

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

[2] [https://www.stadt-zuerich.ch/vbz/de/index/die\\_vbz/medien/medienmitteilungen/fahrplanwechsel18.html](https://www.stadt-zuerich.ch/vbz/de/index/die_vbz/medien/medienmitteilungen/fahrplanwechsel18.html)

Die Hypothese wird angenommen wenn der p-Wert mehr als 0,05 ist, was eine starke Evidenz aufweist

## Hypothese 03 - Ozan

**Beeinflusst der Regen die Verspätungen der  
Busse, sodass sie mehr Verspätungen  
haben?**



## Im Jahr 2022 verzeichneten Busse an den Haltestellen in Stunden mit mehr als 10 Minuten Regen überdurchschnittlich hohe Verspätungen.

### Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022
- Stündliche Wetterdaten in Zürich<sup>[2]</sup>
  - 3 Standorte der Messung in Zürich
  - Temperatur, Regendauer, Luftdruck, Luftfeuchtigkeit, etc.
  - 2000-2023

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

[2] [https://data.stadt-zuerich.ch/dataset/ugz\\_meteorodaten\\_stundenmittelwerte](https://data.stadt-zuerich.ch/dataset/ugz_meteorodaten_stundenmittelwerte)

### Mann-Whitney-U-Test

Weil beide Variablen nicht Normalverteilt sind

### Variablen & grober Ablauf

#### **Verspätungen pro Stunde & Regendauer pro Stunde**

- Verspätungen der Busse pro Stunde aufsummieren
- In eine der beiden Kategorien klassifizieren (Regen)
- p-Wert ausrechnen

### Annahmekriterium

Hypothese wird bei einem p Wert von  $\leq 0,05$  angenommen

## Hypothese 04 - Tony

**Gibt es während den Zeiten, in denen die  
meisten zur Arbeit gehen, mehr  
Verspätungen?**

**In 2021 haben Busse zwischen 6 und 9 Uhr höhere Verspätungen, als in den Zeiten zwischen 10 und 13 Uhr.**

## Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

## Annahmekriterium

Statistischer Test

Wert zwischen 0 und 1 => Annahmewert  $\geq 0,05$

## Erklärung

- $p < 0,01$  – starke Evidenz gegen die Nullhypothese
- $p > 0,10$  – keine Evidenz gegen die Nullhypothese

Die Hypothese wird angenommen wenn der p-Wert mehr als 0,05 ist, was eine starke Evidenz aufweist

Hypothese 05 - Ozan

**Gibt es weniger Ein- und Aussteiger, wenn es regnet?**

**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.**

## Welche Daten werden benötigt?

- Stündliche Wetterdaten in Zürich<sup>[1]</sup>
  - 3 Standorte der Messung in Zürich
  - Temperatur, Regendauer, Luftdruck, Luftfeuchtigkeit, etc.
  - 2000-2023
- Passagierfrequenzen an Haltestelle Hardbrücke<sup>[2]</sup>
  - Ein- und Aussteiger an 2 Gleisen (Hin- und Rückrichtung)  
=> Buslinien 33, 72, 83 und Tramlinie 8 davon betroffen
  - Alle 5 Minuten gezählt
  - 2020-2023

[1] [https://data.stadt-zuerich.ch/dataset/ugz\\_meteodaten\\_stundenmittelwerte](https://data.stadt-zuerich.ch/dataset/ugz_meteodaten_stundenmittelwerte)

[2] [https://data.stadt-zuerich.ch/dataset/vbz\\_frequenzen\\_hardbruecke](https://data.stadt-zuerich.ch/dataset/vbz_frequenzen_hardbruecke)

## Mann-Whitney-U-Test

Weil beide Variablen nicht Normalverteilt sind

## Variablen & grober Ablauf

### **Regendauer pro Stunde & Passagierfrequenz pro Stunde**

- Ein- und Aussteiger pro Stunde aufsummieren
- In eine der beiden Kategorien klassifizieren (Regen)
- p-Wert ausrechnen

## Annahmekriterium

Hypothese wird bei einem p Wert von  $\leq 0,05$  angenommen

## Hypothese 06 - Tony

**Gab es Maßnahmen wegen Covid-19, die den öffentlichen Verkehr betrafen und wie haben die sich ausgewirkt auf die Verspätungen?**

**Durch die Covid-19 Maßnahmen, die den öffentlichen Verkehr betrafen, vom 19.März 2020 bis 27.April 2020 hatten Trams überdurchschnittlich niedrige Verspätungen.**

## Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022
- Covid Pandemiemaßnahmen Schweiz<sup>[2]</sup>

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

[2] [https://de.wikipedia.org/wiki/COVID-19-Pandemie\\_in\\_der\\_Schweiz](https://de.wikipedia.org/wiki/COVID-19-Pandemie_in_der_Schweiz)

## Annahmekriterium

Regressionsanalyse

Wert zwischen 0 und 1 => Annahmewert  $\geq 0,05$

## Erklärung

- $p < 0,01$  – starke Evidenz gegen die Nullhypothese
- $p > 0,10$  – keine Evidenz gegen die Nullhypothese

Die Hypothese wird angenommen wenn der p-Wert mehr als 0,05 ist, was eine starke Evidenz aufweist

Hypothese 07 - Ozan

**Haben Busse oder Trams mehr  
Verspätungen?**



**An Silvester 2021 hatten Trams mehr Verspätungen als Busse. ODER**  
**In 2021, am Züricher Hauptbahnhof haben Busse mehr Verspätungen als Trams.**

Welche Daten werden benötigt?

- Fahrzeiten SOLL und IST Vergleich in Zürich<sup>[1]</sup>
  - Basisdatensatz
  - 2016-2022

[1] <https://data.europa.eu/data/datasets/878a98b8-4973-4d76-858e-eddd88652d9f-stadt-zurich>

Mann-Whitney-U-Test

Weil Variable nicht Normalverteilt ist

Variable & grober Ablauf

**Verspätungen der Busse & Trams**

- Verspätungen aufsummieren
- In eine der beiden Kategorien klassifizieren (Tram/Bus)
- p-Wert ausrechnen

Annahmekriterium

Hypothese wird bei einem p Wert von  $\leq 0,05$  angenommen

# Zeitaufwand

## Hypothesenaufstellung

- Ozan: ~94 Stunden + ~ 20 Stunden von vorherigen Meilensteinen, die ausgelassen wurden
  - Feedback bearbeitet
  - Daten gesucht
  - Hypothesen erstellt
  - Präsentationsfolien erstellt
- Tony: ~32 Stunden
  - Hypothesen erstellt
- Hypothesenaufstellung: ~126 Stunden + ~ 20 Stunden
- Insgesamt: Ozan: ~252 Stunden Tony: ~94 Stunden