



► Data Science Challenge AXA Data Innovation Lab

Thomas Dygutsch

18. Dez. 2025

Aufgabe

- ▶ Der Fahrradverleih **CitiBike** vermietet in New York über 12.000 Fahrräder an 750 Verleihstationen. Somit ist CitiBike eine echte Alternative zu den herkömmlichen Transportmitteln, wie z.B. U-Bahn oder Taxi. CitiBike stellt die durch den Verleih gesammelten Daten der Öffentlichkeit zur Verfügung (s. bspw. „2023-citibike-tripdata.zip“).
- ▶ Deine Aufgabe als Data Scientist ist es, CitiBike dabei zu helfen diese Daten wertstiftend zu nutzen, beispielsweise indem du für CitiBike Kooperationsmöglichkeiten mit einer Versicherung (und/oder umgekehrt) skizzierst. Dazu kannst du zusätzlich die öffentlich zugängigen Daten des NYPD zu **Verkehrsunfällen** nutzen.

Agenda

- ▶ Datenangleichung und -bereinigung
 - ▶ Citibike
 - ▶ NYPD
- ▶ Datenanalyse
 - ▶ Geografisch - Unfallrisiko nach NYC-Stadtgebiet
 - ▶ Zeitlich - Unfallrisiko nach Wochentag
- ▶ Kooperationsmöglichkeiten mit Versicherung

Datenbereinigung Citibike

- ▶ Viele unterschiedliche Datensätze
 - ▶ Zwei verschiedene Spaltenbenennungen (2014-2019 und 2020-heute)
 - ▶ Spezifische Spaltennamen für einen Jahresbereich rausfiltern:
 - ▶ 2014-2019: tripduration, bikeid, gender, birth year
 - ▶ ab 2020-2025: ride_id
 - ▶ Usertype Anpassung → `replace({'Subscriber': 'member', 'Customer': 'casual'})`
 - ▶ E-Bikes 2019 eingeführt → davor nur „Classic Bikes“
 - ▶ DTYPE für alle Spalten angeglichen → Station Ids als String (z.B.SYS016)

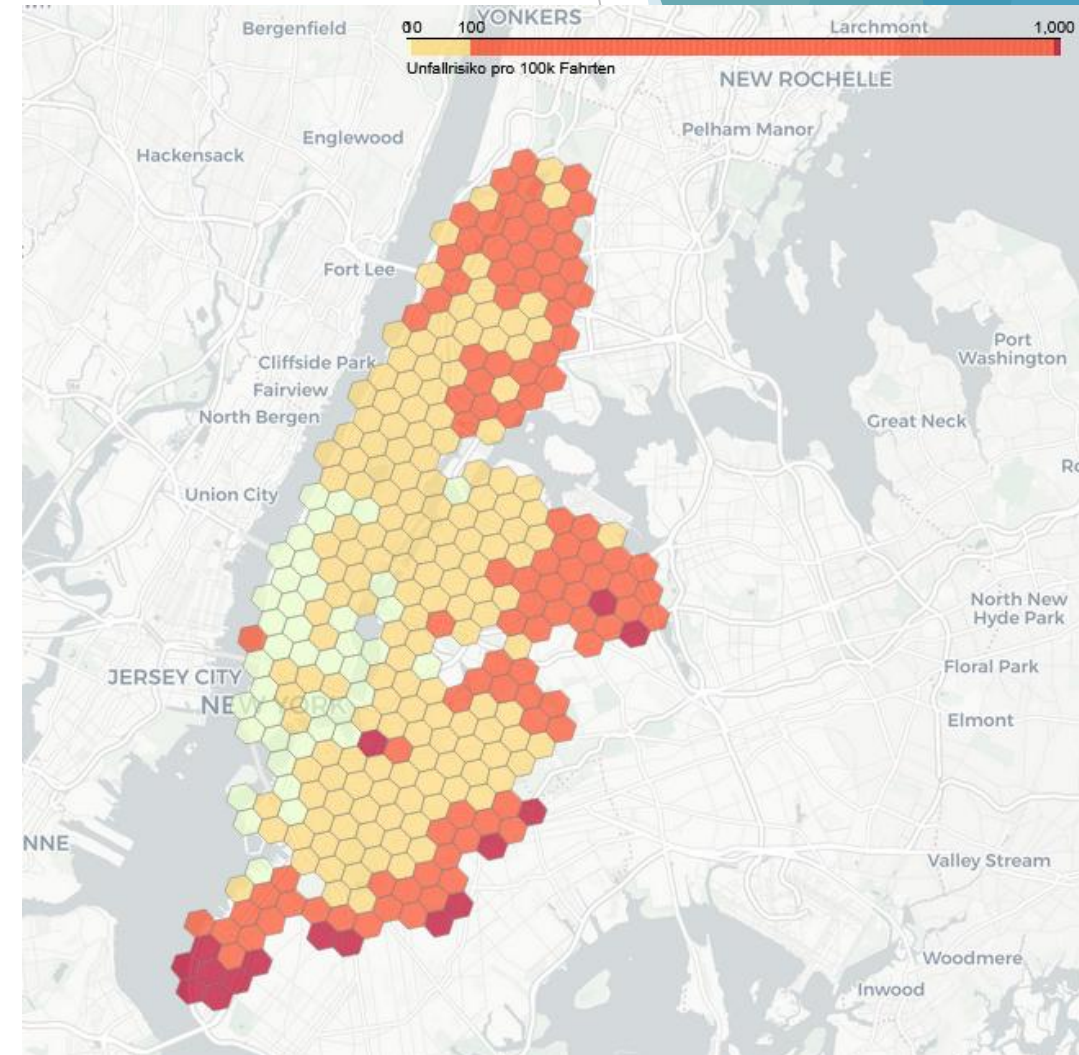
Datenbereinigung NYPD

- ▶ Ein einzelner Datensatz
 - ▶ Zeit / Datum konvertieren
 - ▶ Filtern des Datensatz nach Fahrradbeteiligung 'NUMBER OF CYCLIST INJURED', 'NUMBER OF CYCLIST KILLED' oder VEHICLE TYPE CODE X enthält Angaben über Fahrradbeteiligung

Datenanalyse

Geografisch - Unfallrisiko nach NYC-Stadtgebiet

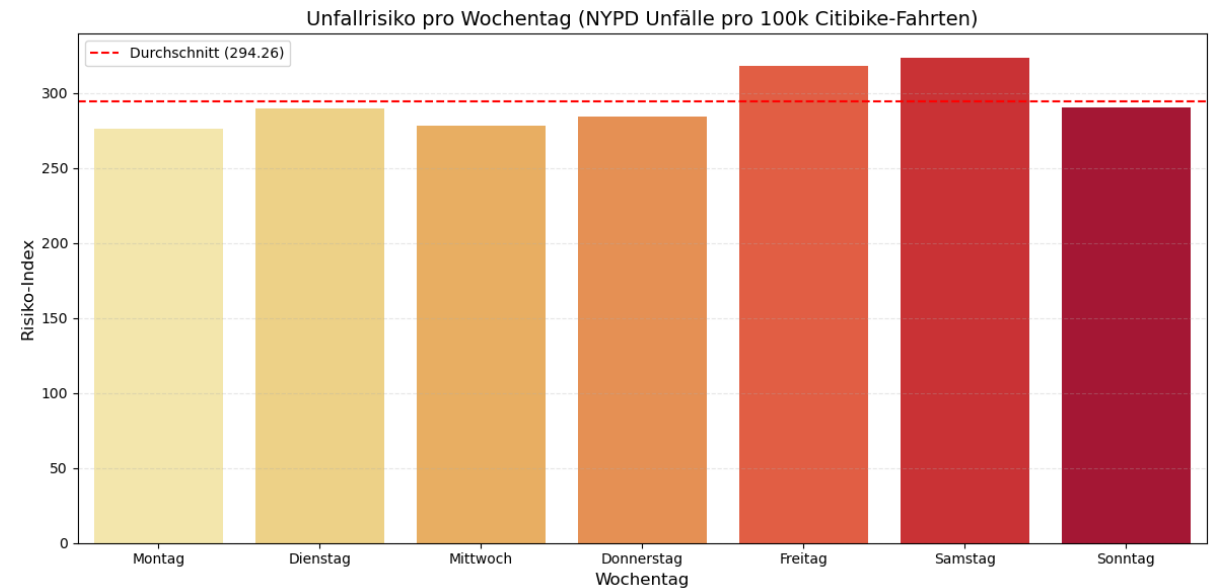
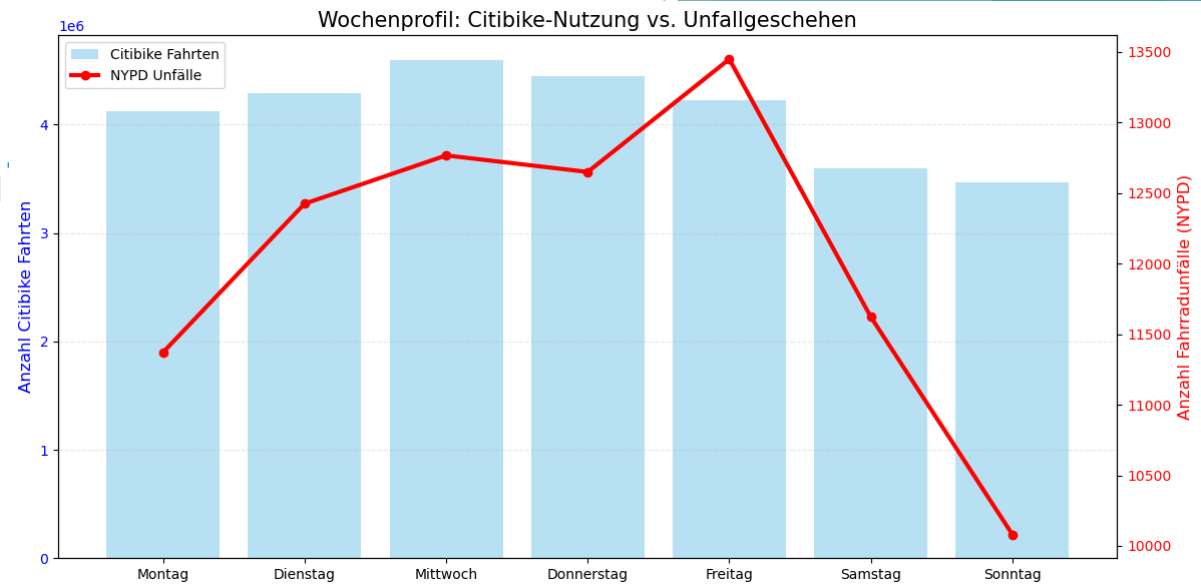
- ▶ Das Unfallrisiko (Unfälle je 100k Fahrten) ist in den Randgebieten deutlich höher als im Zentrum.
- ▶ Annahme: Die Infrastruktur (Fahrradwege, Ampeln etc.) ist hier weniger gut ausgebaut.



Datenanalyse

Zeitlich - Unfallrisiko nach Wochen

- ▶ Von Mo-Fr werden mehr Fahrräder genutzt
- ▶ Das Unfallrisiko steigt zum Wochenende (Fr,Sa).
- ▶ Annahme: Hier spielen wahrscheinlich späte Fahrten mitten in der Nacht und Alkoholeinfluss eine Rolle.
- ▶ Weiterführende Analyse: Verlauf des Unfallrisikos über die Uhrzeit wäre interessant um These zu bestätigen.



Kooperationsmöglichkeiten

- ▶ **Versicherung pro Fahrt (Laufkunden):** Eine Versicherung könnte eine Option anbieten, die für jede einzelne CitiBike-Fahrt abgeschlossen wird. Die Prämie könnte dynamisch an das identifizierte Risikoprofil der Fahrt (Route, Tageszeit) angepasst werden.
- ▶ **Abonnement-Versicherung (Member) :** Für CitiBike-Member könnte eine individuelle Versicherung angeboten je nach Risikoprofil (Routen, Fahrtzeiten).
- ▶ **Risikobasierte Prämienmodelle (Laufkunden/Member):** Die Versicherung und/oder Citibike könnte Prämien anbieten z.B. Bonuspunkte oder Rabatte für die Nutzung von als sicher eingestuften Routen oder zu risikoarmen Zeiten.

Kooperationsmöglichkeiten

- ▶ **Anonymisierte Daten für die Versicherung:** CitiBike könnte anonymisierte Nutzungsdaten mit der Versicherung teilen, um deren Risikomodelle zu verfeinern und präzisere, wettbewerbsfähigere Versicherungsprodukte zu entwickeln.
- ▶ **Unfall-Einblicke für CitiBike:** Die Versicherung könnte detaillierte, anonymisierte Einblicke in die Art und Ursachen von Unfällen, die Radfahrer betreffen, an CitiBike weitergeben. Dies würde CitiBike helfen, die Sicherheit seines Dienstes und seiner Fahrräder weiter zu optimieren.