

# Verkehrsanalyse mit Sensordaten: Einblicke aus Kalifornien

Bugra Görkem  
st183715@stud.uni-stuttgart.de  
Universität Stuttgart  
Stuttgart, Deutschland

Lucas Jester  
st183905@stud.uni-stuttgart.de  
Universität Stuttgart  
Stuttgart, Deutschland

Cem Sanlitürk  
st182238@stud.uni-stuttgart.de  
Universität Stuttgart  
Stuttgart, Deutschland

Taban Eshraghi Samani  
st185420@stud.uni-stuttgart.de  
Universität Stuttgart  
Stuttgart, Deutschland

## Zusammenfassung

Angesichts urbaner Mobilität und komplexer Verkehrssysteme gewinnen datenbasierte Analysen an Bedeutung[12]. Diese Arbeit präsentiert einen Data-Science-Ansatz, der auf Echtzeitdaten, aus dem Zeitraum 01.2023 - 12.2023, des California Performance Measurement Systems (PeMS) basiert, um die Dynamik des Verkehrsflusses in Kalifornien zu untersuchen. Zunächst wird das Themenfeld – geprägt von hoher Verkehrsdichte, variablen Umweltbedingungen und Unfallgeschehen – kurz eingeführt und die zentrale Fragestellung formuliert: Wie beeinflussen Unfälle und Umweltfaktoren die Wiederherstellung eines normalisierten Verkehrsflusses?

Nach einer detaillierten Beschreibung und Bereinigung der Datenbasis werden mithilfe von Hypothesentests (z.B. Mann-Whitney-U-Test) und prädiktiven Regressionsmodellen Zusammenhänge zwischen Unfalltypen, regionalen Unterschieden und Verkehrsverhalten analysiert. Die Ergebnisse liefern wertvolle Einblicke in die Komplexität städtischer und ländlicher Verkehrsdynamiken und tragen zur Optimierung von Verkehrsmanagement- und Sicherheitsstrategien bei.

## Keywords

Verkehrsaufkommen, Normalisierungsdauer, Unfallwahrscheinlichkeit, Regentag, Nichtregentag, Verkehrsanalyse, HOV-Spuren

## 1 Einführung

Trotz zahlreicher Initiativen zur Verbesserung der Verkehrssicherheit stehen kalifornische Autobahnen[10][13], insbesondere in Los Angeles, weiterhin vor großen Herausforderungen. Hohe Unfallzahlen, Staus und die Auswirkungen verschiedener Umweltfaktoren auf den Verkehrsfluss sind zentrale Themen in der Verkehrsforschung[14]. Aktuelle Daten zeigen, dass nicht einzelne Unfallursachen isoliert wirken, sondern ein vielschichtiges Zusammenspiel verschiedenster Einflussfaktoren – von Umweltbedingungen bis zu dynamischen Verkehrsmustern – das Unfallgeschehen prägen. Diese Arbeit verfolgt einen neuartigen Ansatz, der mittels umfangreicher Echtzeitdaten des *California Performance Measurement System* (PeMS)[4] kritische Parameter identifiziert und ihre Wechselwirkungen systematisch analysiert.

Diese Studie basiert auf einer umfassenden Analyse von Verkehrsdaten aus dem Jahr 2023, die aus dem PeMS stammen. Ziel ist es, über klassische Betrachtungen hinauszugehen und insbesondere die Auswirkungen von Umweltbedingungen, infrastrukturellen Gegebenheiten und temporären Verkehrsüberlastungen auf das

Unfallgeschehen zu untersuchen. Ein besonderer Fokus liegt auf der Frage, inwiefern extreme Witterungsbedingungen oder temporäre Verkehrsstörungen zu einer Veränderung des Fahrverhaltens führen und damit neue Risikoprofile entstehen lassen.

Um dies zu untersuchen, werden verschiedene statistische Verfahren angewendet, darunter auch nicht-parametrische Tests und Regressionsmodelle. Die Analyse soll helfen, neue Einsichten in die Verkehrsdynamik zu gewinnen und langfristig zur Optimierung von Sicherheitsmaßnahmen sowie Verkehrssteuerungsstrategien beizutragen.

## 1.1 Aufbau der Arbeit

Die Arbeit ist in mehrere, aufeinander aufbauende Abschnitte unterteilt, die eine systematische Analyse des Zusammenhangs zwischen Verkehrsunfällen und deren Einflussfaktoren ermöglichen. Zunächst gibt die Einführung 1 einen Überblick über das Thema, stellt die Forschungsfragen vor und erläutert die Motivation sowie die Relevanz der Analysen. Anschließend werden in den theoretischen Grundlagen 2 zentrale Konzepte und bisherige wissenschaftliche Erkenntnisse zu Verkehrsunfällen, Verkehrsflussmodellen und relevanten statistischen Verfahren dargestellt. In der darauf folgenden Datenakquise und Exploration 3 wird die Herkunft der verwendeten Datensätze erläutert, einschließlich der Sensordaten, Vorfallsdaten und ergänzender Wetterdaten. Die Datenbereinigung 4 beschreibt die durchgeführten Vorverarbeitungsschritte, darunter das Entfernen fehlerhafter oder unvollständiger Einträge sowie die Harmonisierung verschiedener Datenquellen. Im Abschnitt Datenanalyse 5 werden die gewählten statistischen Methoden und Modelle detailliert vorgestellt, mit denen die Hypothesen überprüft werden. Abschließend werden die Ergebnisse interpretiert und im Fazit zusammengefasst, wobei auch mögliche Limitationen sowie Ansätze für zukünftige Forschung aufgezeigt werden.

## 2 Grundlagen

### 2.1 Statistische Tests

**2.1.1 Mann-Whitney-U-Test.** Der Mann-Whitney-U-Test[18], ist ein nichtparametrischer statistischer Test, und somit anwendbar unabhängig der Verteilungseigenschaft der Stichproben. Er wird verwendet, um zu bestimmen, ob zwei unabhängige Stichproben aus Populationen mit identischer Verteilung stammen.

**2.1.2 Wilcoxon-Vorzeichen-Rang-Test.** Der Wilcoxon-Vorzeichen-Rang-Test [19] ist ein nichtparametrischer Test, der zur Überprüfung von Unterschieden zwischen **zwei abhängigen (gepaarten) Stichproben** verwendet wird.

**2.1.3 Cliff's delta.** Cliff's delta[5] ist ein robustes, nicht-parametrisches Maß für die Effektstärke, das den Unterschied zwischen zwei unabhängigen Stichproben quantifiziert, ohne Annahmen über deren Verteilung zu treffen. Der Wertebereich liegt zwischen  $-1$  und  $1$ , wobei ein Wert von  $0$  darauf hinweist, dass es keine systematischen Unterschiede zwischen den Gruppen gibt. Ein positiver Wert bedeutet, dass die Werte in der ersten Gruppe tendenziell höher sind als in der zweiten, während ein negativer Wert das Gegenteil anzeigt. Die Effektstärke wird anhand der folgenden Schwellenwerte interpretiert: Ein  $|\delta| < 0,147$  entspricht einem kleinen Effekt, Werte zwischen  $0,147$  und  $0,33$  deuten auf einen mittleren Effekt hin, und ein  $|\delta| \geq 0,33$  wird als großer Effekt gewertet.

**2.1.4 Shapiro-Wilk Test.** Der Shapiro-Wilk-Test[17] prüft die Nullhypothese, dass eine Stichprobe aus einer normalverteilten Grundgesamtheit stammt. Die Teststatistik  $W$  basiert auf der gewichteten Summe der quadrierten Abweichungen der geordneten Stichprobenwerte. Ein  $p$ -Wert kleiner als  $0,05$  deutet darauf hin, dass die Daten nicht normalverteilt sind.

**2.1.5 Lineare Regression.** Die Lineare Regression [8] ist ein statistisches Modell zur Untersuchung des Zusammenhangs zwischen einer abhängigen Variablen  $Y$  und einer oder mehreren unabhängigen Variablen  $X$ .

## 2.2 "California Department of Transportation"

Um eine einheitliche und konsistente Planung, den Bau und die Instandhaltung des staatlichen Autobahnnetzes in Kalifornien zu gewährleisten, ist das California Department of Transportation (Caltrans) <sup>1</sup> als zuständige Behörde verantwortlich. Die Expertise von Caltrans bietet im Bereich Sensornetzwerke und Verkehrsdatenanalyse wertvolle Einblicke für die Verkehrsanalyse mit Sensordaten. Zudem, stellt Caltrans Definitionen von Fachbegriffen bereit, auf die wir uns im Verlauf unseres Projekts beziehen werden.

## 3 Datenakquise und -exploration

Die verwendeten Daten stammen ursprünglich aus dem PeMS, einem Verkehrs- und Analysetool von Caltrans. Die Originaldaten wurden im Rahmen des Projekts "XAI-Traffic"[9] verarbeitet und in einem strukturierten Datensatz auf Kaggle<sup>2</sup>. Im Einzelnen beinhaltet der Datensatz folgende Informationen:

### 3.1 Vorfallsmeldungen (incidents.csv)

Die Datei incidents.csv enthält Informationen zu insgesamt 476.766 Verkehrsvorfällen auf kalifornischen Straßen im Zeitraum von 01.2023 bis 12.2023. Jeder Vorfalleintrag umfasst Angaben zur

- Incident Id – Eindeutige ID des Vorfalls
- Duration (mins) – Dauer des Vorfalls in Minuten

<sup>1</sup><https://dot.ca.gov/-/media/dot-media/programs/environmental-analysis/documents/env/glossary-terms-a11y.pdf>

<sup>2</sup><https://www.kaggle.com/datasets/gpxlcj/xtraffic> veröffentlicht, der insgesamt 17 Dateien umfasst

- Fwy – Autobahn, auf der sich der Vorfall ereignet hat (z. B. „I10-E“)
- Abs PM – Absolute Postmile (Meilenmarkierung auf der Autobahn)
- AREA – Region des Vorfalls (z. B. „Central LA“)
- LOCATION – Detaillierte Standortangabe
- DESCRIPTION – Beschreibung des Vorfalls (z. B. Unfall, Hindernis)
- type – Typ des Vorfalls 1 (z.B. Gefahrenstelle, Unfall, Wetterereignisse)
- dt – Datum und Uhrzeit des Vorfalls
- date – Datum des Vorfalls
- time – Uhrzeit des Vorfalls

Beachte, dass die Absolute Postmile (Abs PM) die fortlaufende Distanz in Meilen entlang einer Autobahn angibt, gemessen vom definierten Nullpunkt der Route (meistens Staatsgrenzen, Küstenlinien, Kreuzungen mit anderen Autobahnen oder wichtigen Verkehrsknotenpunkten). Sie wächst kontinuierlich von West nach Ost bzw. Süd nach Nord und bleibt auch bei Streckenänderungen als einheitliche Referenz bestehen. In Systemen wie PeMS dient sie zur präzisen Lokalisierung und Verknüpfung von Vorfällen mit Verkehrssensordaten.

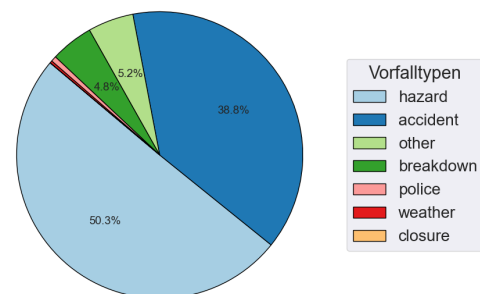


Abbildung 1: Verteilung der gemeldeten Vorfälle

### 3.2 Metadaten der Sensoren (sensor\_meta\_feature.csv)

Diese Datei enthält Informationen zu 16.972 Sensoren, die über das gesamte kalifornische Straßennetz wie in Abbildung 2 dargestellt verteilt sind.

Die bereitgestellte Sensormetadatei ist ein von den Wissenschaftlern des XTraffic-Projekts aus dem PeMS extrahierter und nach strengen Kriterien gefilterter Datensatz. Sie enthält nur Sensoren mit mindestens 50% gültigen Messwerten über das gesamte Jahr. Dies gewährleistet eine hohe Datenqualität. Zusätzlich werden irrelevante oder redundante Attribute entfernt, wodurch der Datensatz eine optimierte Grundlage für die Analyse von Verkehrsströmen und Vorfällen bietet. Sie enthält 26 Meta-Features zu Verkehrssensoren, die in drei Kategorien unterteilt sind. Diese Merkmale beschreiben physische, funktionale und geografische Eigenschaften der Sensorstandorte und sind essenziell für die Analyse von Verkehrsfluss und -vorfällen.

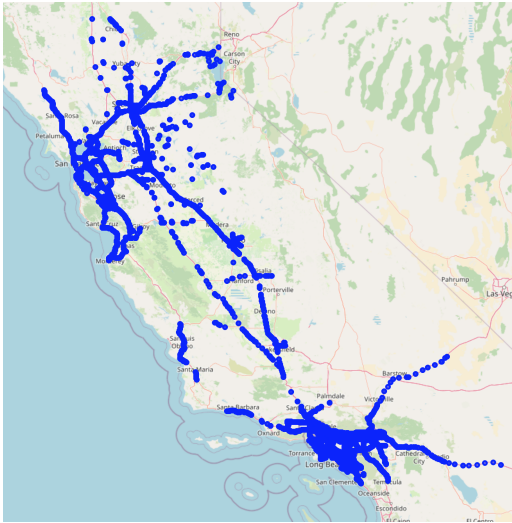


Abbildung 2: Sensorverteilung in Kalifornien

1. Allgemeine Sensorinformationen, wie z.B. Eindeutige Sensor-ID, zugehörige Autobahn und Verwaltungsregion (District, County, City), sowie ABS PM.
2. physikalische und infrastrukturelle Eigenschaften der Fahrbahn, wie z.B. Fahrspur-, Straßenbreite sowie Existenz von *high-occupancy vehicle* (HOV) Spuren, Straßenkategorie und zugelassene Höchstgeschwindigkeit
3. Geografische Merkmale, wie z.B. Koordinaten (Längen- und Breitengrade), sowie Topografie des Umfelds

Beachte dabei, dass die vorgefilterten Daten trotzdem einer gründlichen Überprüfung unterzogen werden muss, um ihre Qualität sicherzustellen.

### 3.3 Distanzmatrix (dis\_matrix.npy)

Die Distanzmatrix beschreibt die kürzeste Strecke in Kilometern zwischen verschiedenen Sensorpunkten. Dabei wird nicht die Luftlinie, sondern die tatsächlich befahrbare Straßenroute berücksichtigt, weshalb die Matrix nicht symmetrisch ist.

Eine Überprüfung der ursprünglichen Daten ergibt, dass 99,59% der Werte fehlerhaft sind (negative Distanzen). Aus diesem Grund werden alle Distanzen mithilfe der OSM API neu berechnet.

### 3.4 Zeitreihendaten (p01\_done.npy bis p12\_done.npy)

Die Zeitreihendaten liegen als dreidimensionale Matrizen vor, die für jeden Monat in separaten Dateien (p01\_done.npy bis p12\_done.npy) gespeichert sind. Wie in Tabelle 1 gezeigt, umfassen diese Matrizen für ihren respektiven Monat eine Anzahl an Messungen (fünf-minütige Messungen) für den ganzen Monat, was in der ersten Dimension gespeichert ist. Darum variieren auch diese Werte je nach Monat, z.B. Januar in Abbildung 3 hat 8.928 Einträge während der Februar nur 8.064 Werte hat. In jeder dieser Messung befinden sich die Messwerte der 16.972 Sensoren (zweite Dimension) für drei verschiedene Messtypen (Dimension 3): Anzahl der

passierten Fahrzeuge (*Verkehrsaufkommen*), die Belegungsrate und die Geschwindigkeit.

Monat	5 Minutenintervalle	Sensoren	Messtypen
Januar	8928	16972	3
Februar	8064	16972	3
März	8928	16972	3
April	8640	16972	3
Mai	8928	16972	3
Juni	8640	16972	3
Juli	8928	16972	3
August	8928	16972	3
September	8640	16972	3
Oktober	8928	16972	3
November	8640	16972	3
Dezember	8928	16972	3

Tabelle 1: Tabelle der Sensordaten von Januar bis Dezember

Januar-Matrix:

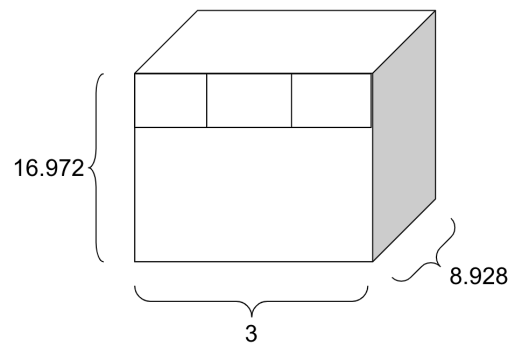


Abbildung 3: 3D Matrix für Januar

### 3.5 Nodeliste (node\_order.npy)

Die Datei node\_order.npy dient dazu, die Metasensordaten den entsprechenden Sensoren zuzuordnen. Der Index dieser Datei entspricht dem Index in den Zeitreihendaten (p-Dateien), während der gespeicherte Wert die zugehörige station\_id aus der Datei sensor\_meta\_feature.csv repräsentiert. Somit ergibt sich ein funktionales und korrektes Mapping zwischen den Metasensordaten und den Zeitreihendaten.

## 4 Datenbereinigung

### 4.1 Vorfallsmeldungen (incidents.csv)

Bei der Datenprüfung der Vorfallsdatei fallen einige Inkonsistenzen auf. So weisen einige Vorfälle negative Unfallzeiten zwischen -1 und -105 Minuten auf. Da diese nur 0,0005% der Daten ausmachen, werden sie als fehlerhaft eingestuft und entfernt. Anderweitige fehlerhafte Daten werden ebenfalls entfernt. Darunter fallen unter anderem fehlende Angaben zur Lokalisierung des Vorfalls.

Da es sich bei den Daten um Einträge von mehreren verschiedenen Institutionen handelt, wie z.B. *California Highway Patrol* (CHP), der *Freeway Service Patrol* oder anderen Leitstellen der kalifornischen Autobahnen, wiederholen sich manche Vorfälle. Diese erscheinen auf den ersten Blick jedoch verschieden, denn sie besitzen eine andere Identifikationsnummer und unterscheiden sich in anderen Informationen zu den Vorfällen. Dennoch beschreiben sie den gleichen Vorfall, sind aber mit unterschiedlichen Zeitstempeln, oder verschiedenen Referenzhighways gelistet. Durch einen geeigneten Algorithmus werden diese sich wiederholenden Einträge identifiziert und zu einem einzigen neuen zusammengefasst. Dies erfolgt beispielsweise über die Mittelwertbildung der Dauer des Vorfalls und die Zusammenfassung der anderen Referenzwerte.

## 4.2 Metadaten der Sensoren (`sensor_meta_feature.csv`)

Obwohl die Datenqualität insgesamt hoch ist, treten dennoch Unvollständigkeiten auf. Beispielsweise fehlen bei insgesamt 21 Sensoren die exakten Positionen, also die Längen- und Breitengrade des Sensors. Da jedoch weitere Beschreibungen wie z.B. die Standortbeschreibungen, der Straßenname ist, können die Sensoren zunächst gewissen Regionen zugewiesen werden. Die exakte Standortbestimmung wird rekursiv mithilfe der ABS Pm durchgeführt.

Darüber hinaus wird bei 833 Sensoren Angaben zu infrastrukturellen Merkmalen wie Straßenbreite oder Spurbreite vermisst. Diese Lücken werden in erster Linie durch die Betrachtung der Sensoreigenschaften des gegenüberliegenden Sensors ergänzt. Wenn dies nicht möglich ist, werden die fehlenden Daten durch ein Hybridansatz vervollständigt. Dieser Hybridansatz beinhaltet den Einsatz externer Geodaten (von Caltrans), da sie nicht alle Merkmale abdeckt implementieren wir eine Nachbarschafts-Interpolation. Damit ist gemeint, dass benachbarte Sensoren, welche auf der selben Strecke liegen, sehr ähnliche Bedingungen der Umgebung aufweisen. Bei der Wahl der Nachbarn werden alle vollständigen Sensoren auf der selben Fahrbahnstrecke betrachtet und für numerischen Werte (z.B. Straßenbreite) der Mittelwert übernommen, während bei nicht numerischen Werten wie Oberflächenbeschaffenheit der Modus übernommen wird.

Ferner fehlen anfangs bei knapp 20% der Sensoren die Angabe zur zugehörigen Stadt. Eine Überprüfung ergibt, dass einige Sensoren tatsächlich innerhalb von Städten liegen. Auch hier wird wie im Schritt zuvor, die Regionen der Sensoren mit der fehlenden Zugehörigkeit zu Städten untersucht und mit externen Datenquellen erweitert. Hierfür werden die Dienste von OpenStreetMaps (OSM) genutzt, um somit die Zugehörigkeit der Städte durchführen und validieren zu können. Somit ist der fehlende Anteil auf 10% gesunken.

## 4.3 Zeitreihendaten (`p01_done.npy` bis `p12_done.npy`)

Um eine fundierte Analyse durchführen zu können, ist es entscheidend, die Qualität der Zeitreihendaten sicherzustellen.

Die Zeitreihendaten werden zunächst in jeder Dimension einzeln betrachtet, um die Datenqualität zu verbessern. Zu Beginn werden doppelte und sich wiederholende Messwerte aus den Zeitreihendaten entfernt. Zudem werden Sensoren identifiziert, die im Laufe eines Monats oder sogar eines ganzen Jahres keine Messungen

liefern. Es stellt sich heraus, dass diese fehlerhaften Sensoren nicht konstant waren, was zu einer variierenden Anzahl funktionierender Sensoren über die Monate führt (siehe Tabelle 2).

Um eine konsistente Datenbasis zu gewährleisten, wird für jeden Monat eine separate Zuordnungstabelle (`node_order.npy`) erstellt, die die korrekte Zuweisung der Sensordaten zu den jeweiligen Sensor-IDs sicherstellt.

Monat	#Sensoren Dimension 1	#Sensoren Dimension 2	#Sensoren Dimension 3
Januar	12.602	11.626	8.816
Februar	12.674	11.693	8.844
März	12.731	11.721	8.854
April	12.623	11.646	8.830
Mai	12.490	11.549	8.744
Juni	12.281	11.364	8.617
Juli	12.037	11.089	8.572
August	11.725	10.729	8.205
September	12.361	11.359	8.639
Oktober	12.372	11.383	8.639
November	12.324	11.338	8.618
Dezember	12.437	11.437	8.712

Tabelle 2: Anzahl der verbleibenden Sensoren

Für weiterführende Analysen werden nur die Sensoren berücksichtigt, die über das gesamte Jahr hinweg kontinuierliche Messwerte liefern. Dies gewährleistet eine zuverlässigere Vergleichbarkeit und vermeidet Verzerrungen durch zeitweise ausgefallene Sensoren. Nach dieser Selektion blieben 9.927 Sensoren für die Messung der passierten Fahrzeuge, 9.275 Sensoren für die Belegungsrate und 7.315 Sensoren für die Geschwindigkeitsmessung übrig.

## 5 Datenanalyse

### 5.1 Formulierung der Hypothesen

Im Folgenden werden inhaltliche, zeitliche und geografische Zusammenhänge untersucht. Die formulierten Hypothesen lauten:

- **Hypothese 1. —Regen und Unfallwahrscheinlichkeit**  
Es gibt keinen signifikanten Unterschied in der Unfallwahrscheinlichkeit zwischen Regen- und Nichtregentagen auf Freeways in Los Angeles.
- **Hypothese 2. —Hauptverkehrszeiten und Unfälle**  
Auf der Interstate 710 unterscheidet sich die Unfallwahrscheinlichkeit während den Hauptverkehrszeiten nicht signifikant von der außerhalb der Hauptverkehrszeiten.
- **Hypothese 3. —Feiertage und Verkehrsaufkommen**  
Das durchschnittliche Verkehrsaufkommen auf Interstates in Kalifornien unterscheidet sich an Feiertagen und Wochenenden nicht signifikant von dem an Werktagen.
- **Hypothese 4. —Unfall- und Staudauer**  
Es besteht kein signifikanter Zusammenhang zwischen der Falldauer und der Normalisierungsdauer.
- **Hypothese 5. —Einfluss auf HOV-Spuren**  
Man stellt keinen Signifikanten Unterschied im Verkehrsfluss auf HOV-Spur nach einem Unfall.

Hierbei wird je das in der Praxis gängige Signifikanzniveau von 5% gefordert[6].

## 5.2 Hypothese 1. – Regen und Unfallwahrscheinlichkeit

Es gibt keinen signifikanten Unterschied in der Unfallwahrscheinlichkeit zwischen Regen- und Nichtregentagen auf Freeways in Los Angeles.

**5.2.1 Motivation.** Die Untersuchung des Zusammenhangs zwischen Niederschlag und Unfallwahrscheinlichkeit ist von großer Bedeutung für die Verkehrssicherheit und Verkehrssteuerung. Regen verändert die Straßenverhältnisse, kann die Bremswege verlängern, die Sichtverhältnisse verschlechtern und das Fahrverhalten der Verkehrsteilnehmer beeinflussen. Frühere Studien[1] haben gezeigt, dass Regen das Unfallrisiko erhöhen kann, jedoch variiert dieser Effekt stark in Abhängigkeit von der Verkehrsdichte, Infrastruktur und Fahrweise.

Gerade in einer Metropole wie Los Angeles, die für ihre hohe Verkehrsdichte bekannt ist, stellt sich die Frage, ob Regen einen signifikanten Einfluss auf das Unfallgeschehen hat oder ob die Verkehrsinfrastruktur und das Fahrverhalten diesen Effekt abschwächen. Die gewonnenen Erkenntnisse könnten dazu beitragen, präventive Maßnahmen wie gezieltere Warnsysteme, Geschwindigkeitsanpassungen oder infrastrukturelle Verbesserungen zu entwickeln, um die Verkehrssicherheit bei Regenwetter zu erhöhen.

**5.2.2 Ansatz.** Das Ziel dieser Hypothese ist es, zu untersuchen, ob Niederschlag in einer Metropole wie Los Angeles einen signifikanten Einfluss auf die Unfallwahrscheinlichkeit hat. Dafür werden die Vorfalldaten aus incidents.csv herangezogen, wobei nur Unfälle innerhalb von Los Angeles betrachtet werden. Die sensor\_meta\_feature.csv liefert dazu die Geokoordinaten der Sensoren, während die p01\_done bis p12\_done numpy-Dateien genutzt werden, um das tägliche Verkehrsaufkommen zu ermitteln. Angereicherte Daten:

Um eine präzisere Analyse zu ermöglichen, wird der Datensatz um Niederschlagsinformationen angereichert. Mithilfe der Visual Crossing Wetter-API werden für alle Verkehrssensoren in Los Angeles tägliche Niederschlagsmessungen basierend auf ihren Koordinaten abgerufen.

Auf dieser Grundlage wird das Stadtgebiet in Regen- und Nichtregengebiete unterteilt. Ein Bereich wird als Regentag definiert, wenn dort mindestens 0,1 mm Niederschlag gemessen wird (Fläche von 1 Quadratmeter innerhalb eines Tages 0,1 Millimeter Wasserhöhe) – basierend auf der Definition des Deutschen Wetterdienstes (DWD)[7] und der Schweizerischen Eidgenossenschaft[1], die diesen Schwellenwert als bedeutend für die Unfallwahrscheinlichkeit betrachten.

In Abbildung 4 sieht man eine Beispielkategorisation für den 08.03.2024.

Anschließend wird für beide Kategorien (Regen- und Nichtregengebiete) die Unfallwahrscheinlichkeit berechnet für jeden Tag. Dafür werden den Unfällen anhand ihrer Absoluten Postmile (Abs PM) die nächstgelegenen Sensoren zugeordnet, die entweder im Regen- oder im Nichtregengebiet liegen. Die Unfallwahrscheinlichkeit für einen Bereich wird als Verhältnis der Anzahl der dort registrierten Unfälle zur Gesamtzahl der in diesem Bereich gezählten Fahrzeuge berechnet.

$$\text{UnfallwahrscheinlichkeitRegenTag}_x : U_{wRx} = \frac{UR_x}{TR_x}$$

$$\text{UnfallwahrscheinlichkeitNichtRegenTag}_x : U_{wNRx} = \frac{UNR_x}{TNR_x}$$

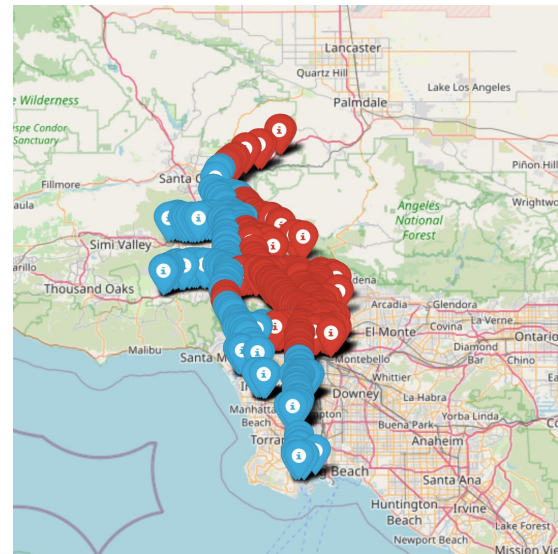
Dabei sei  $U_{wRx}$  die Unfallwahrscheinlichkeit im Regnbereich an Tag  $x$  und  $UR_x$  Anzahl Unfälle- und  $TR_x$  Anzahl Autos im Regnbereich für Tag  $x$ . Analog gilt das gleiche für  $UNR_x$  und  $TNR_x$ .

Das Resultat sind zwei Mengen

- RegenW: { $U_{wR1}, U_{wR2}, \dots, U_{wR365}$ }
- NichtregenW: { $U_{wNR1}, U_{wNR2}, \dots, U_{wNR365}$ }

Unfallwahrscheinlichkeiten die gleich 0 sind werden von den Mengen entfernt. Damit ergeben sich zwei unabhängige Stichproben und übrig bleibt nur noch das Durchführen eines passenden Tests wofür hier der Mann-Whitney U Test sehr geeignet ist. Das liegt daran, dass dieser ein nicht-parametrisierter Test ist, d.h. wir können es anwenden ohne Rücksicht auf die Verteilung der Daten zu nehmen. Zudem führen wir den Test zweiseitig durch und berechnen den p-Wert. Wir wählen ein Signifikanzniveau von  $\alpha = 0,05\%$ .

Außerdem wird neben dem Mann-Whitney-U-Test auch noch Cliff's Delta berechnet, um die Effektgröße zu bestimmen.



**Abbildung 4: Regenbereich Blau | Nichtregengebiet Rot für den 08.03.2024**

**5.2.3 Akzeptanzkriterium.** Das Akzeptanzkriterium für die Hypothese „Regen und Unfallwahrscheinlichkeit“ sieht vor, dass bei einem durch den Mann-Whitney-U-Test ermittelten p-Wert unter 0,05 die Nullhypothese abgelehnt werden kann, was auf einen signifikanten Unterschied in der Unfallwahrscheinlichkeit zwischen Regen- und Nichtregentagen hinweist. Liegt der p-Wert hingegen über diesem Schwellenwert, kann die Nullhypothese beibehalten werden, sodass kein signifikanter Einfluss von Regen angenommen wird. Ergänzend wird mittels Cliff's Delta  $|\delta|$  die Effektstärke bewertet. Ein Wert oberhalb des definierten Mindestniveaus (0,147)



deutet auf mindestens einen mittleren Effekt hin. Ist sowohl  $p < 0.05$  und  $|\delta| > 0.147$  dann wird  $H_0$  abgelehnt, sonst angenommen.

**5.2.4 Ergebnisse und Fazit.** Die Analyse des Einflusses von Regen auf die Unfallwahrscheinlichkeit in Los Angeles ergab einen p-Wert von  $1.70 \times 10^{-11}$ , der weit unter dem Signifikanzniveau von 0,05 liegt. Dies bedeutet, dass die Nullhypothese, die von keinem Unterschied zwischen Regen- und Nichtregentagen ausgeht, abgelehnt werden kann. Statistisch gesehen gibt es also einen signifikanten Zusammenhang zwischen Niederschlag und Unfallwahrscheinlichkeit. Zusätzlich zeigt die Berechnung von Cliff's Delta (0.6909) eine starke Effektstärke, was darauf hinweist, dass Regen einen deutlichen Einfluss auf das Unfallgeschehen hat. Dies bestätigt, dass Unfälle an Regentagen signifikant häufiger auftreten als an trockenen Tagen. Diese Erkenntnisse unterstreichen die Notwendigkeit gezielter Maßnahmen zur Verbesserung der Verkehrssicherheit bei Regen, etwa durch angepasste Tempolimits, optimierte Warnsysteme oder verbesserte Straßenentwässerung. Zukünftige Untersuchungen könnten sich auf die genauen Mechanismen der Unfallentstehung konzentrieren, etwa die Rolle von Starkregen, Sichtverhältnissen oder Fahrbahnbeschaffenheit, um präzisere Präventionsstrategien zu entwickeln.

### 5.3 Hypothese 2. – Hauptverkehrszeiten und Unfälle

Auf der Interstate 710 unterscheidet sich die Unfallwahrscheinlichkeit während den Hauptverkehrszeiten nicht signifikant von der außerhalb der Hauptverkehrszeiten.

**5.3.1 Motivation.** In der vorliegenden Studie wird untersucht, ob die Unfallwahrscheinlichkeit auf der Interstate 710 während der Hauptverkehrszeiten signifikant höher ist als während der Nicht Hauptverkehrszeiten. Die Interstate 710, die entlang des LA River verläuft und als wichtige Verkehrsader Ost- und Süd-Los Angeles miteinander verbindet, spielt eine zentrale Rolle im regionalen Verkehrsfluss. Daraus leitet sich die Hypothese ab, dass die erhöhte Verkehrsbelastung in den definierten Hauptverkehrszeiten zu einer gesteigerten Unfallwahrscheinlichkeit führt.

**5.3.2 Ansatz.** Als Hauptverkehrszeiten werden hier ausschließlich, nach dem Traffic Monitoring Guide, die Werktage betrachtet, und zwar von 6:00-10:00 Uhr morgens und 15:00-19:00 abends. Alle anderen Zeiten dazwischen bilden die Nicht - Hauptverkehrszeiten. Da Wochenenden und Feiertage andere Hauptverkehrszeiten haben, werden diese im Folgenden nicht mehr betrachtet. Weiterhin wird die Unfallwahrscheinlichkeit folgendermaßen definiert:

$$P = \frac{U_T}{V_T}$$

Wobei  $U$ : die Anzahl an Unfällen auf der Interstate 710 im Zeitfenster  $T$  ist.

Und  $V$ : das Verkehrsaufkommen des nächstgelegenen Sensors zum Unfall im Zeitfenster  $T$  ist. Dabei werden hier täglich jeweils fünf Zeitfenster betrachtet, zwei für die Hauptverkehrszeiten (6:00–10:00 Uhr und 15:00–19:00 Uhr) und drei für die Nicht-Hauptverkehrszeiten (0:00–6:00 Uhr, 10:00–15:00 Uhr und 19:00–0:00 Uhr). Diese Unterteilung geschieht für jeden Werktag.

Unfälle, die beispielsweise um 9:55 Uhr auftreten, könnten durch den dadurch verursachten Stau das Verkehrsaufkommen des unmittelbar folgenden Zeitfensters beeinflussen. Da dies allerdings, nach Prüfung der Vorfalldaten, nur selten bei Unfällen auf der Interstate 710 der Fall ist, kann dies vernachlässigt werden. Somit beschränkt sich die Analyse der Hypothese nur auf die festgesetzten Zeitgrenzen, da dieses Vorgehen eine konsistente und reproduzierbare Bewertung der Unfallwahrscheinlichkeiten ermöglicht.

Zur Überprüfung der Hypothese, dass die Unfallwahrscheinlichkeit auf der Interstate 710 während der Hauptverkehrszeiten höher ist als in den Nicht-Hauptverkehrszeiten, wird ein mehrstufiger Ansatz verfolgt. Zunächst werden alle gemessenen Unfälle auf der Interstate 710 für das gesamte Beobachtungsjahr aus den Vorfalldaten gefiltert und für jeden Werktag in die oben genannten fünf Zeitfenster unterteilt. Dies geschieht basierend auf dem Datum und den Zeitstempeln der Vorfalldaten. Anschließend wird mit Hilfe der Sensor Sensor Meta Daten für jeden Unfall der nächstgelegene Sensor über "Abs PM" ermittelt. Über die Sensordaten wird nun die vom nächstgelegenen Sensor gemessene Autoanzahl für das gesamte betrachtete Zeitfenster kalkuliert. Dieser Prozess wird für jeden Unfall durchgeführt, um schließlich mithilfe der bereits definierten Formel für jeden Unfall eine Unfallwahrscheinlichkeit zu berechnen.

Zuletzt werden die ermittelten Unfallwahrscheinlichkeiten in zwei Gruppen unterteilt: eine Gruppe umfasst die Wahrscheinlichkeiten der Hauptverkehrszeiten, die andere die der Nicht Hauptverkehrszeiten. Zur statistischen Überprüfung der Hypothese wird ein Mann-Whitney-U-Test auf diese beiden Gruppen angewendet bei einem Signifikanzniveau von  $\alpha = 0,05$ . Das Ergebnis des Tests liefert ein  $p$ , welches schließlich als Entscheidungsgrundlage dient, ob die Hypothese verworfen oder angenommen wird.

**5.3.3 Akzeptanzkriterium.** Das Akzeptanzkriterium für die Hypothesenprüfung basiert auf dem p-Wert. Wird der p-Wert kleiner als 0,05, wird die Nullhypothese ( $H_0$ ) abgelehnt. Liegt der p-Wert jedoch bei 0,05 oder darüber, wird die Nullhypothese beibehalten.

**5.3.4 Ergebnisse und Fazit.** Beim Analysieren der Hypothese ergab der Mann-Whitney-U-Test einen p-Wert von ungefähr  $p \approx 0,96$ . Da dieser deutlich über dem Signifikanzniveau von 0,05 liegt, kann kein statistisch signifikanter Unterschied zwischen den beiden Zeiträumen festgestellt werden. Die Nullhypothese wird also nicht verworfen.

Im ganzen Jahr wurden an Werktagen insgesamt 2.385 Unfälle auf der Interstate 710 erfasst, von denen 1.173 in den definierten Hauptverkehrszeiten (06:00–10:00 und 15:00–19:00 Uhr) und 1.211 in den Nicht-Hauptverkehrszeiten (0:00–6:00 Uhr, 10:00–15:00 Uhr und 19:00–0:00 Uhr). Die Abbildung 5 zeigt die genaue Verteilung der Unfälle an Werktagen im Jahr 2023 auf die 5 festgelegten Zeitfenster.

Das Ergebnis verdeutlicht, dass die Definition der Unfallwahrscheinlichkeit eine zentrale Rolle spielt. Betrachtet man lediglich die absolute Anzahl der Unfälle, so ereignen sich geringfügig mehr Unfälle außerhalb der Hauptverkehrszeiten. Allerdings erfasst diese Betrachtung nicht die zeitliche Dimension der Verkehrsdynamik.

Ein genauerer Blick auf die Zeiträume zeigt, dass die Hauptverkehrszeiten insgesamt 8 Stunden eines Werktages umfassen, während die Nicht-Hauptverkehrszeiten mit 16 Stunden doppelt so

lang sind. Daraus folgt, dass sich während der Hauptverkehrszeiten pro Stunde nahezu doppelt so viele Unfälle ereignen wie in den übrigen Stunden des Tages.

Trotz dieses offensichtlichen Unterschieds konnte der statistische Test keinen signifikanten Unterschied in der Unfallwahrscheinlichkeit feststellen. Dies lässt sich durch das stark erhöhte Verkehrsaufkommen während der Hauptverkehrszeiten erklären, das zwar die Anzahl der Unfälle pro Zeiteinheit erhöht, aber gleichzeitig die Wahrscheinlichkeit eines Unfalls pro Fahrzeug möglicherweise konstant hält. Diese Ergebnisse legen nahe, dass nicht nur die absolute Unfallanzahl, sondern auch die Verkehrsdichte und andere Einflussfaktoren bei der Bewertung der Verkehrssicherheit berücksichtigt werden sollen.

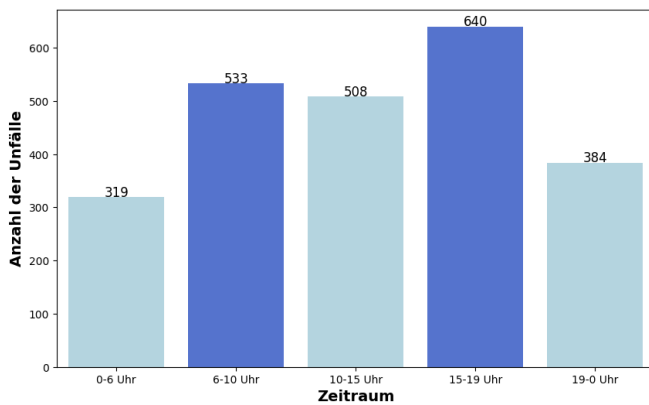


Abbildung 5: Verteilung der Unfälle an Werktagen auf der I710 im Jahr 2023

## 5.4 Hypothese 3. – Feiertage und Verkehrsaufkommen

Das durchschnittliche Verkehrsaufkommen auf Interstates in Kalifornien unterscheidet sich an Wochenenden und Feiertagen nicht signifikant von dem an Werktagen.

**5.4.1 Motivation.** Die Hypothese basiert auf der Annahme, dass das Verkehrsaufkommen in Kalifornien durch eine Vielzahl von Faktoren beeinflusst wird, die sowohl an Werktagen als auch an Wochenenden und Feiertagen konstant bleiben könnten. Studien zeigen, dass Kalifornien aufgrund der starken Abhängigkeit vom Individualverkehr ein durchgehend hohes Verkehrsaufkommen aufweist [3]. Insbesondere in Ballungsräumen wie Los Angeles und der Bay Area bleibt der Verkehr auch an Wochenenden hoch, da Freizeitaktivitäten und Tourismus eine bedeutende Rolle spielen [11].

Darüber hinaus legen Untersuchungen nahe, dass Feiertage und Wochenenden in den USA oft mit Freizeit- und Urlaubsreisen verbunden sind, was das Verkehrsaufkommen auf Interstates erhöhen kann [2]. Diese Mobilitätsmuster könnten dazu führen, dass sich das durchschnittliche Verkehrsaufkommen an diesen Tagen nicht signifikant von dem an Werktagen unterscheidet. Eine Analyse solcher

Muster ist nicht nur für die Verkehrsplanung von Bedeutung, sondern auch für die Entwicklung von Modellen zur Vorhersage und Optimierung von Verkehrsflüssen

**5.4.2 Ansatz.** Um die Hypothese zu überprüfen, wird ein systematischer Ansatz gewählt, der die Verkehrsdaten nach Tagestypen kategorisiert und vergleicht. Die Analyse konzentriert sich ausschließlich auf die Hauptfahrspuren (Mainline) der Interstates in Kalifornien, um eine konsistente Datenbasis zu gewährleisten.

Zunächst werden die Sensordaten nach ihrer Zugehörigkeit zu den verschiedenen Interstates gruppiert. Hierfür wurden aus den Metadaten der Sensoren diejenigen identifiziert, die sich auf den Hauptfahrspuren der Interstates befinden. Anschließend wurden die Zeitreihendaten um Datums- und Zeitinformationen ergänzt.

Für diese Kategorisierung werden die gesetzlichen kalifornischen Feiertage identifiziert. Für jeden Interstate werden die durchschnittlichen Verkehrsaufkommen pro Tag berechnet, indem die Messwerte aller zugehörigen Sensoren aggregiert werden.

Das Verkehrsaufkommen wird als Anzahl der Fahrzeuge definiert, die einen Sensor in einem bestimmten Zeitintervall passieren. Die Berechnung des Verkehrsaufkommens  $V$  erfolgt durch die Anzahl der Autos  $A$  dividiert durch die betrachtete Zeit  $T$ .

Die Formel hierfür lautet:

$$\text{Verkehrsaufkommen} := V = \frac{A}{T}$$

Die Tage werden in drei Kategorien unterteilt: Werktage (Montag bis Freitag, ausgenommen Feiertage), Wochenenden (Samstag und Sonntag) und Feiertage (offizielle Feiertage in Kalifornien).

Für jeden Interstate werden die durchschnittlichen Verkehrsaufkommen pro Tag berechnet, indem die Messwerte aller zugehörigen Sensoren aggregiert werden. Diese Werte werden dann nach Tagestyp gruppiert, um zwei Hauptgruppen zu bilden: Das durchschnittliche Verkehrsaufkommen an Werktagen und das durchschnittliche Verkehrsaufkommen an Wochenenden und Feiertagen.

Die Daten werden nach Interstates gruppiert, wobei für jede Interstate die verfügbaren Sensoren identifiziert werden. Für jede dieser Gruppen werden die Daten nach Tageskategorien (Werktag, Wochenende, Feiertag) aufgeteilt und anschließend aggregiert, um das durchschnittliche Verkehrsaufkommen zu berechnen.

Zur statistischen Überprüfung der Hypothese wird der Mann-Whitney-U-Test angewendet, da dieser keine Normalverteilung der Daten voraussetzt. Der Test wird zweiseitig durchgeführt, um festzustellen, ob ein signifikanter Unterschied zwischen den beiden Gruppen besteht.

**5.4.3 Akzeptanzkriterium.** Der Mann-Whitney-U Test wird mit einem Signifikanzniveau von 0,05 durchgeführt. Sollte der p-Wert des Tests kleiner als 0,05 sein, wird die Hypothese abgelehnt, was darauf hindeutet, dass ein signifikanter Unterschied im Verkehrsaufkommen zwischen den verschiedenen Tagestypen besteht.

Im Gegensatz dazu wird die Nullhypothese angenommen, wenn der p-Wert gleich oder größer als 0,05 ist, was darauf hinweist, dass es keinen signifikanten Unterschied im durchschnittlichen Verkehrsaufkommen zwischen Feiertagen, Wochenenden und Werktagen gibt.

**5.4.4 Ergebnisse und Fazit.** Die Analyse des Verkehrsaufkommens auf kalifornischen Interstates ergab einen p-Wert von  $\approx 3,176 \times$

$10^{-27}$ , der deutlich unter dem Signifikanzniveau von 0,05 liegt. Dies bedeutet, dass die Nullhypothese abgelehnt werden muss. Entgegen der ursprünglichen Annahme unterscheidet sich das durchschnittliche Verkehrsaufkommen an Wochenenden und Feiertagen signifikant von dem an Werktagen.

Die Ergebnisse zeigen, dass trotz der hohen Mobilität in Kalifornien und der vielfältigen Nutzung der Straßen zu verschiedenen Zeiten, deutliche Unterschiede im Verkehrsaufkommen zwischen Werktagen und Wochenenden/Feiertagen bestehen. Diese Unterschiede könnten auf verschiedene Faktoren zurückzuführen sein, wie beispielsweise unterschiedliche Reisezwecke (Pendeln vs. Freizeitaktivitäten), veränderte Reiseziele oder unterschiedliche Tageszeiten, zu denen Fahrten unternommen werden.

Diese Erkenntnisse haben wichtige Implikationen für die Verkehrsplanung und -steuerung in Kalifornien. Sie deuten darauf hin, dass Verkehrsmanagementstrategien an die spezifischen Muster verschiedener Tagestypen angepasst werden sollten, um eine optimale Verkehrsflusssteuerung zu gewährleisten. Beispielsweise könnten an Werktagen andere Maßnahmen zur Stauvermeidung erforderlich sein als an Wochenenden oder Feiertagen.

Zukünftige Forschungen könnten sich auf die genauen Unterschiede in den Verkehrsmustern konzentrieren und untersuchen, wie sich diese auf verschiedenen Streckenabschnitten und zu verschiedenen Tageszeiten manifestieren. Auch die Analyse saisonaler Schwankungen oder der Einfluss spezifischer Feiertage auf das Verkehrsaufkommen könnte wertvolle Einblicke liefern.

## 5.5 Hypothese 4. – Unfall- und Normalisierungsdauer

Es besteht kein signifikanter Zusammenhang zwischen der Unfalldauer und der Normalisierungsdauer.

**5.5.1 Motivation.** Die Normalisierungsdauer, also die Dauer, bis sich der Verkehr nach einem Unfall wieder normalisiert, ist ein zentraler Faktor für die Verkehrseffizienz und das Störungsmanagement. Eine längere Normalisierungszeit kann zu erheblichen Stau Problemen, wirtschaftlichen Verlusten und sekundären Unfällen führen. Daher ist es entscheidend zu verstehen, welche Faktoren diese Dauer beeinflussen.

Ein naheliegender Einflussfaktor ist die Unfalldauer selbst. Je länger ein Unfall anhält, desto größer ist die potenzielle Störung im Verkehrsfluss. Dabei könnte ein linearer Zusammenhang bestehen: Eine längere Unfallzeit führt zu einer entsprechend längeren Erholungsphase des Verkehrs. Diese Hypothese zu überprüfen ist essenziell, um Verkehrsprognosen zu verbessern und effizientere Räumstrategien zu entwickeln.

Durch ein besseres Verständnis dieses Zusammenhangs lassen sich gezielte Maßnahmen ableiten, um die Wiederherstellung des Verkehrsflusses zu beschleunigen. Dazu zählen beispielsweise optimierte Einsatzstrategien für Räumdienste, frühzeitige Umleitungen oder adaptive Verkehrssteuerungen, die helfen können, Verzögerungen zu minimieren und den Verkehrsfluss schneller zu stabilisieren.

**5.5.2 Ansatz.** Für die Überprüfung dieser Hypothese werden sämtliche Vorfalldaten aus der Datei `incidents.csv` verwendet. Die Datei `sensor_meta_feature.csv` liefert die entsprechenden Geokoordinaten

der Sensoren sowie deren Station ID. Zur Ermittlung der Geschwindigkeit zu den jeweiligen Zeitstempeln werden die in den `p01_done` bis `p01_done` gespeicherten Messwerte herangezogen. Um die Dauer bis zur Wiederherstellung des normalen Verkehrsflusses nach einem Unfall zu bestimmen, wird die Zeit gemessen, die vergeht, bis die Geschwindigkeit auf ein als normal definiertes Niveau zurückkehrt. Dieser Normalzustand wird anhand von 5-Minuten-Intervallen ermittelt, indem für jeden Sensor die charakteristische Geschwindigkeit zu einem bestimmten Zeitpunkt bestimmt wird. Die Berechnung erfolgt durch eine historische Referenzbildung: Für einen gegebenen Unfallzeitpunkt (z. B. 15:00 Uhr) wird der nächstgelegene Sensor in Fahrtrichtung betrachtet, und der Messwert  $w_i$  für das nächste Zeitintervall (15:05 Uhr) wird herangezogen. Zur Festlegung des Normalbereichs werden alle Geschwindigkeitsmessungen desselben Sensors zur selben Uhrzeit über das gesamte Jahr hinweg aggregiert. Der Mittelwert  $M_{wi}$  dieser Messwerte repräsentiert die charakteristische Geschwindigkeit, ergänzt durch die Varianz  $V_{wi}$ , um Schwankungen im Verkehrsfluss zu berücksichtigen. Ein Messwert wird als nicht normalisiert betrachtet, wenn  $w_i < M_{wi} - V_{wi}$ . In diesem Fall wird der Normalzustand als noch nicht wiederhergestellt angesehen und die Überprüfung wird für das nächste Zeitintervall fortgesetzt. Sobald ein Messwert  $w_i \geq M_{wi} - V_{wi}$  erreicht wird, gilt der Verkehrsfluss als normalisiert, und die Zeitspanne bis zu diesem Zeitpunkt wird als Normalisierungsdauer definiert. Zur Überprüfung des Zusammenhangs zwischen Unfalldauer und Normalisierungsdauer wird eine lineare Regression durchgeführt. Das Signifikanzniveau wird standardmäßig auf  $\alpha = 0,05\%$  gesetzt, und zusätzlich erfolgt eine Residuenanalyse sowie eine Homoskedastizitätsprüfung, um die Gültigkeit des Modells sicherzustellen.

**5.5.3 Akzeptanzkriterium.** Das Akzeptanzkriterium für die Hypothese basiert auf einer linearen Regressionsanalyse. Wenn der p-Wert  $< 0,05$  ist, deutet das daraufhin, dass ein signifikanter linearer Zusammenhang zwischen der Unfalldauer und der Normalisierungsdauer besteht. Zusätzlich dazu wird eine Residuenanalyse sowie eine Homoskedastizitätsprüfung gemacht, um sicherzugehen, dass der Zusammenhang wirklich linear ist und nicht durch andere Faktoren verzerrt wird. Erst wenn die Residuen normalverteilt sind, keine sichtbare Struktur haben (Residuenanalyse) und die Varianz der Residuen konstant sind (Homoskedastizität) dürfen wir  $H_0$  ablehnen. Sollte der p-Wert  $> 0,05$  sein, wird die Nullhypothese angenommen.

**5.5.4 Ergebnisse und Fazit.** Der berechnete p-Wert von 0,598 liegt deutlich über der typischen Signifikanzschwelle von 0,05, wodurch die Nullhypothese, dass kein signifikanter linearer Zusammenhang zwischen Unfalldauer und Normalisierungsdauer besteht, nicht verworfen werden kann. Das Streudiagramm 6 zeigt eine starke Streuung der Daten, insbesondere bei kürzeren Unfalldauern, und bestätigt diesen Befund. Die lineare Anpassung (rote Linie) verdeutlicht zudem, dass keine klare lineare Tendenz erkennbar ist. Dies bedeutet, dass die Dauer eines Unfalls nicht zwangsläufig einen direkten Einfluss darauf hat, wie lange es dauert, bis sich der Verkehrsfluss normalisiert.



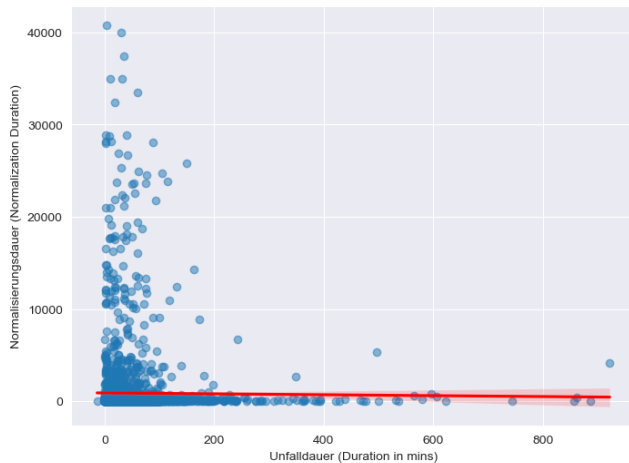


Abbildung 6: Streudiagramm für Normalisierungsdauer und Unfalldauer

## 5.6 Hypothese 5. – Verkehrsfluss bei Unfällen auf HOV Spuren

Man stellt keinen signifikanten Unterschied im Verkehrsfluss auf HOV-Spur nach einem Unfall fest.

**5.6.1 Motivation.** Die Analyse des Verkehrsflusses auf HOV-Spuren im Zusammenhang mit Unfällen auf der Hauptfahrbahn basiert auf der Annahme, dass Fahrer bei plötzlicher Staubbildung nach einem Unfall[16][15] verstärkt auf alternative Fahrspuren ausweichen, um Verzögerungen zu minimieren. Da HOV-Spuren nur für eine begrenzte Gruppe von Verkehrsteilnehmern zugänglich sind, stellen sie eine besonders attraktive Option dar. Dennoch liegt der Hypothese zugrunde, dass sich der Verkehrsfluss auf den HOV-Spuren nach einem Unfall trotz dieses potenziellen Anreizes nicht signifikant verändert.

**5.6.2 Ansatz.** Um diese Fragestellung zu untersuchen, wird ein methodisch präziser Zugang gewählt. Hierfür kommen ausschließlich HOV-Sensordaten zum Einsatz. Für jeden Unfall wird der nächstgelegene Sensor ermittelt, der sich auf derselben Straßenseite befindet und innerhalb eines maximalen Radius von 100 Metern liegt. Diese räumliche Begrenzung stellt sicher, dass nur direkt betroffene Messwerte in die Analyse einfließen. Zusätzlich wird ein festes Beobachtungszeitfenster von 15 Minuten vor und 15 Minuten nach dem Unfall definiert, um den kurzfristigen Einfluss des Unfalls auf den Verkehrsfluss zu erfassen.

Um den Unterschied im Verkehrsfluss vor und nach dem Unfall zu quantifizieren, wird ein gepaarter Wilcoxon-Test[19] angewandt. Dieser nicht-parametrische Test prüft, ob der Pre-Unfall Verkehrsfluss signifikant höher ist als der Vor-Unfall-Wert. Die Entscheidung für den Wilcoxon-Test erfolgt, da es sich um gepaarte Beobachtungen handelt und der Test robust gegenüber Ausreißern ist.

**5.6.3 Akzeptanzkriterium.** Die Hypothese gilt als bestätigt, wenn der Wilcoxon-Test für das Merkmal *Verkehrsaufkommen* keinen signifikanten Unterschied zwischen den Messwerten vor und nach dem Unfall aufzeigt ( $p\text{-Wert} \geq 0.05$ ).

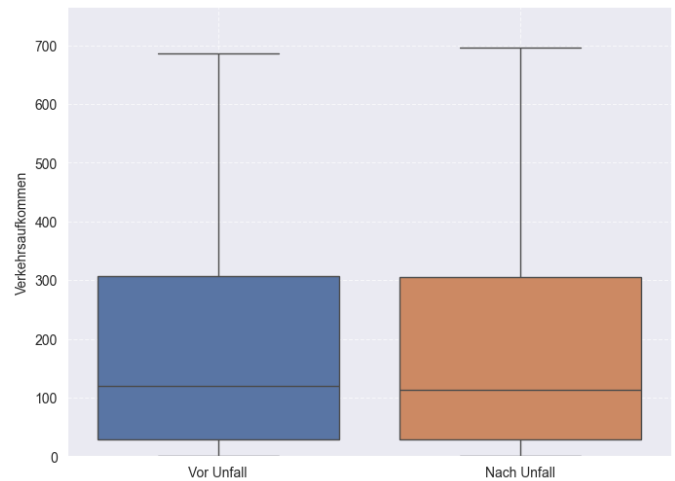


Abbildung 7: Verkehrsaufkommen vor und nach dem Zwischenfall.

**5.6.4 Ergebnis und Fazit.** Abbildung 7 zeigt das Verkehrsaufkommen auf der HOV-Spur im definierten Beobachtungszeitraum von 15 Minuten vor bis 15 Minuten nach dem Unfall. Aus den Daten geht hervor, dass es nach dem Unfall zu gewissen Schwankungen im Verkehrsaufkommen kommt, jedoch keine eindeutige, visuell auffällige Veränderung erkennbar ist. Während in einigen Fällen ein leichter Anstieg oder Rückgang des Verkehrsflusses sichtbar wird, bleibt das Gesamtbild relativ stabil. Dies deutet darauf hin, dass der Unfall auf der Mainline keine unmittelbaren, drastischen Auswirkungen auf die Nutzung der HOV-Spuren hatte.

Um diese visuelle Einschätzung quantitativ zu untermauern, wird ein gepaarter Wilcoxon-Test auf knapp 20.000 Messwerte angewendet. Dieser ergibt einen  $p$ -Wert von unter 0.0002, was darauf hindeutet, dass die Unterschiede im Verkehrsfluss vor und nach dem Unfall nicht zufällig sind, sondern statistisch signifikant voneinander abweichen. Das bedeutet, dass sich der Verkehrsfluss auf den HOV-Spuren durch den Unfall messbar verändert hat, auch wenn dies in der grafischen Darstellung nicht unmittelbar ersichtlich ist.

Trotz dieser statistischen Signifikanz zeigt der berechnete Cliff's Delta-Wert, der nahe 0 liegt, dass die praktische Relevanz dieses Effekts sehr gering ist. Anders ausgedrückt existieren zwar Unterschiede im Verkehrsaufkommen, ihre tatsächliche Wirkung ist jedoch minimal. Eine mögliche Erklärung hierfür könnte sein, dass sich die Fahrer auf den HOV-Spuren nur geringfügig anpassen oder dass sich der Einfluss der Mainline-Störung nicht direkt auf diese speziellen Fahrspuren überträgt.

Zukünftige Analysen könnten sich intensiver damit befassen, wie sich ein Unfall auf der Mainline in Situationen mit Stau auswirkt. Eine Überprüfung unserer Daten zeigt, dass nach einem Unfall nicht zwingend ein Stau entsteht: Tatsächlich tritt nur in 5% der Fälle überhaupt ein Stau auf.

## 6 Schlussfolgerung

Die vorliegende Studie liefert wertvolle Einblicke in die komplexen Zusammenhänge zwischen Umweltfaktoren, Unfallereignissen und

Verkehrsfluss auf kalifornischen Autobahnen. Im Rahmen der Datenanalyse wurden mehrere Hypothesen überprüft, die u.a. den Einfluss von Niederschlag auf die Unfallwahrscheinlichkeit, die Unterschiede zwischen Hauptverkehrszeiten und Nebenzeiten sowie die Variation des Verkehrsaufkommens an Werktagen, Wochenenden und Feiertagen untersuchten. Während sich insbesondere bei der Analyse des Regeneinflusses und der Tagestypen signifikante Effekte zeigten, konnte der vermutete lineare Zusammenhang zwischen Unfalldauer und der Dauer der Wiederherstellung eines normalen Verkehrsflusses nicht bestätigt werden. Zudem ergab die Untersuchung des Verkehrsflusses auf HOV-Spuren zwar statistisch signifikante Differenzen vor und nach Unfallereignissen, jedoch wies der berechnete Effekt – trotz eines signifikanten p-Werts – eine nur minimale praktische Relevanz auf.

Ein zentraler Kritikpunkt an dem verwendeten Datensatz zeigt sich in der Tatsache, dass lediglich in etwa 5% der Unfälle ein signifikanter Stau festgestellt wurde. Dieses Ergebnis steht in klarem Widerspruch zu zahlreichen Studien[15][16], die von einem deutlich häufigeren und gravierenderen Zusammenhang zwischen Unfällen und der Entstehung von Verkehrsstaus berichten.

Trotz dieser Limitationen zeigt die Analyse, dass Umweltfaktoren wie Niederschlag einen signifikanten Einfluss auf das Unfallgeschehen haben und dass das Verkehrsaufkommen je nach Tagestyp (Werktag vs. Wochenende/Feiertag) variierende Muster aufweist. Diese Erkenntnisse unterstreichen die Notwendigkeit einer differenzierten Verkehrssteuerung und präventiver Maßnahmen, etwa durch adaptive Tempolimits und gezielte Warnsysteme bei ungünstigen Witterungsbedingungen.

## Literaturverzeichnis

- [1] Autor(en). 2025. Einflüsse des Wetters auf das Straßenunfallgeschehen. [file:///Users/lucasjester/Downloads/einfluesse\\_des\\_wettersaufdasstrassenunfallgeschehen%20\(2\).pdf](file:///Users/lucasjester/Downloads/einfluesse_des_wettersaufdasstrassenunfallgeschehen%20(2).pdf) Zugriff am 1. März 2025.
- [2] Christopher J. Bacon, A. Abdel-Rahim, Kevin Chang, Kushalkumar G. Patel, and Ben Sonner. 2016. Intersection Treatment Alternatives to Address Holiday Traffic Congestion and Delays. (2016).
- [3] R. Cervero and Peter Bosselmann. 1994. An Evaluation of the Market Potential for Transit-Oriented Development Using Visual Simulation Techniques. (1994).
- [4] Tom Choe, Alexander Skabardonis, and Pravin Varaiya. 2002. FREEWAY PERFORMANCE MEASUREMENT SYSTEM (PeMS): AN OPERATIONAL ANALYSIS TOOL. (January 2002). [https://ntlrepository.blob.core.windows.net/lib/34000/34000/34065/10.1.1.112.4428\\_1\\_.pdf](https://ntlrepository.blob.core.windows.net/lib/34000/34000/34065/10.1.1.112.4428_1_.pdf)
- [5] Norman Cliff. 1993. Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions. *Psychological Bulletin* 114 (11 1993), 494–509. doi:10.1037/0033-2909.114.3.494
- [6] M. Cowles and C. Davis. 1982. On the Origins of the .05 Level of Statistical Significance. *American Psychologist* 37 (1982), 553–558. doi:10.1037/0003-066X.37.5.553
- [7] Deutscher Wetterdienst. 2025. Definition Regentag. <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv3=102196&lv2=102134> Zugriff am 1. März 2025.
- [8] David A. Freedman. 2009. Statistical Models: Theory and Practice. (2009). doi:10.1017/CBO9780511815867
- [9] Xiaochuan Gou, Ziyue Li, Tian Lan, Junpeng Lin, Zhishuai Li, Bingyu Zhao, Chen Zhang, Di Wang, and Xiangliang Zhang. 2024. XTraffic: A Dataset Where Traffic Meets Incidents with Explainability and More. arXiv:2407.11477 [cs.LG] <https://arxiv.org/abs/2407.11477>
- [10] J. Kaplan. 1950. EFFECTIVE SAFETY-COUNCIL PROGRAMS. (1950).
- [11] Junghwan Kim and M. Kwan. 2018. Beyond Commuting: Ignoring Individuals' Activity-Travel Patterns May Lead to Inaccurate Assessments of Their Exposure to Traffic Congestion. *International Journal of Environmental Research and Public Health* 16 (2018). doi:10.3390/ijerph16010089
- [12] Yisheng Lv, Y. Duan, Wenwen Kang, Z. Li, and Feiyue Wang. 2015. Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems* 16 (2015), 865–873. doi:10.1109/TITS.2014.2345663
- [13] Gustavo Marfia, M. Roccetti, A. Amoroso, and G. Pau. 2013. Safe Driving in LA: Report from the Greatest Intervehicular Accident Detection Test Ever. *IEEE Transactions on Vehicular Technology* 62 (2013), 522–535. doi:10.1109/TVT.2012.2226484
- [14] Viacheslav V. Morozov, Artur I. Petrov, Vladimir Shepelev, and Mohammed Balfaqih. 2024. Ideology of Urban Road Transport Chaos and Accident Risk Management for Sustainable Transport Systems. *Sustainability* (2024). doi:10.3390/su16062596
- [15] T. Nagatani. 1993. Effect of traffic accident on jamming transition in traffic-flow model. *Journal of Physics A* 26 (1993). doi:10.1088/0305-4470/26/19/008
- [16] Fu Peng-bo and M. Key. 2010. Study of the Control Measures for Traffic Jam Caused by Accidents on Expressway. *Transactions of Beijing Institute of Technology* (2010).
- [17] S. S. Shapiro and M. B. Wilk. 1965. An analysis of variance test for normality: Complete samples. *Biometrika* 52 (1965), 591–611. doi:10.1093/biomet/52.3-4.591
- [18] David J. Sheskin. 2004. Handbook of Parametric and Nonparametric Statistical Procedures. (2004).
- [19] Frank Wilcoxon. 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1, 6 (1945), 80–83. doi:10.2307/3001968