

Leistungsnachweis

CAS Information Engineering

Modul „Scripting“

ZHAW School of Engineering  
Winterthur

Gruppe 3

André Meier  
Dominique Neff  
Kathrin Noser

29.11.2021

# Inhaltsverzeichnis

1	Background .....	1
2	Business Case.....	1
3	Daten.....	2
3.1	Polizeilich registrierte Verkehrsunfälle auf dem Stadtgebiet Zürich seit 2011 .....	2
3.2	Daten der Verkehrszählung zum motorisierten Individualverkehr .....	2
3.3	GeoJSON Daten .....	3
4	Übersicht Verarbeitungsschritte .....	4
5	Laden und Bereinigen der Daten .....	5
5.1	Polizeilich registrierte Verkehrsunfälle auf dem Stadtgebiet Zürich seit 2011 .....	5
5.2	Daten der Verkehrszählung zum motorisierten Individualverkehr .....	6
6	Analyse der Daten .....	8
7	Vertiefte Analyse der Daten / Business Case Analyse .....	11
7.1	Laden der Daten und zusätzliche Vorbereitungsschritte.....	11
7.2	Analyse der Daten .....	11
7.2.1	Cluster Plots mit Anzeige des Unfall-Schweregrads und Beteiligten.....	11
7.2.2	Heatmaps.....	14
7.2.3	Heatmaps mit zeitlicher Entwicklung .....	14
8	Ergebnisse .....	17
9	Probleme.....	18
10	Conclusion.....	18

# 1 Background

Im Rahmen dieser Arbeit geht es darum, die Verkehrsunfälle des Jahres 2020 in der Stadt Zürich zu analysieren. Die Stadt Zürich publiziert dazu auf ihrer Website umfangreiche statistische Rohdaten<sup>1</sup>, sowie jährlich aufbereitete Verkehrsstatistiken, die an Medienkonferenzen vorgestellt werden<sup>2</sup>.

In der Verkehrsstatistik für das Jahr 2020 werden verschiedene Faktoren aufgeführt, welche die Entwicklung der Zahlen massgeblich beeinflusst haben. Zum einen war das ganze Jahr durch die Corona Pandemie geprägt, dies insbesondere während des Lockdowns im Frühjahr. Bedingt durch Home Office waren deutlich weniger Fahrzeuge unterwegs, was nachvollziehbarerweise auch eine Reduktion der Verkehrsunfälle mit sich gebracht hat.

Eine spezielle Kategorie ist jedoch der Veloverkehr. Dieser hat während der Pandemie eher zugenommen, bedingt vielleicht unter anderem auch durch das schöne Wetter. Auch hier besteht eine Korrelation zwischen Anzahl benutzter Fahrzeuge und Unfallzahlen. Besonders stark angestiegen ist auch die Zahl der Unfälle mit Beteiligung von leistungsstarken E-Bikes. Zu guter Letzt ist aus nachvollziehbaren Gründen der Schweregrad der Unfälle bei den Velofahrern oft höher, da diese keine schützende Knautschzone um sich herum haben<sup>3</sup>.

Während also die Zahl der Verkehrsunfälle erfreulicherweise gesamthaft zurückgegangen ist nehmen die Unfälle mit Beteiligung von Velofahrern stetig zu und zwar schon seit mehreren Jahren. Insbesondere bei Unfällen mit Schwerverletzten sind beängstigend oft Velofahrer betroffen. Eine Bevölkerungsbefragung im Jahr 2015 zeigte zudem auf, dass diejenigen Personen, die innerhalb der Stadt Zürich mit dem Velo unterwegs sind, sich oft unsicher fühlen.

Der Stadtrat von Zürich hat daher schon im Jahr 2019 beschlossen, die Sicherheit des Veloverkehrs zu einem Strategie-Schwerpunkt zu ernennen. In diesem Zusammenhang ergreift er einerseits Massnahmen, alle Verkehrsteilnehmenden auf Unfall-vermeidendes Verhalten zu sensibilisieren, andererseits wird überprüft, inwiefern und wo Signalisierungen, auffällige Markierungen oder bauliche Massnahmen die Sicherheit der Verkehrsteilnehmer erhöhen können.<sup>4</sup>

## 2 Business Case

Anhand von Verkehrszählungen und von Verkehrsunfalldaten soll herausgefunden werden, wo die Hotspots für sicherheitsfördernde Massnahmen sind. Wir haben uns dabei auf Daten des Jahres 2020 beschränkt, um einerseits möglichst aktuelle Daten zu haben, andererseits aber auch ein vollständiges Jahr analysieren zu können.

Die Analyse der Unfallursache sowie weiterer Parameter wie beispielsweise Unfallzeitpunkt und Verkehrsaufkommen zu diesem Zeitpunkt soll in Verbindung mit den zugehörigen Unfallorten dazu dienen, an diesen Stellen präventive Massnahmen zu ergreifen.

---

<sup>1</sup> <https://data.stadt-zuerich.ch/>

<sup>2</sup> [https://www.stadt-zuerich.ch/pd/de/index/dav/themen\\_projekte/verkehrsunfallstatistik.html](https://www.stadt-zuerich.ch/pd/de/index/dav/themen_projekte/verkehrsunfallstatistik.html)

<sup>3</sup> <https://www.stadt-zuerich.ch/content/dam/stzh/pd/Deutsch/Dienstabteilung%20Verkehr/Publikationen%20und%20Bro schueren/Verkehrsunfallstatistik%202020.pdf>

<sup>4</sup> <https://www.stadt-zuerich.ch/sichervelofahren>

## 3 Daten

### 3.1 Polizeilich registrierte Verkehrsunfälle auf dem Stadtgebiet Zürich seit 2011

Dieser Datensatz liefert die polizeilich registrierten Strassenverkehrsunfälle mit Personen- und Sachschäden auf dem Stadtgebiet Zürich seit 2011.

Zu jedem Strassenverkehrsunfall sind der Unfallort, das Jahr, der Monat, der Wochentag, die Unfallstunde, die Strassenart, der Unfalltyp, die Unfallbeteiligung und die Unfallschwerekategorie verfügbar.

Als Strassenverkehrsunfall im Sinne der Strassenverkehrsunfall-Statistik zählen:

- ein unvorhergesehenes Ereignis auf einer öffentlichen Verkehrsfläche,
- welches in ursächlichem Zusammenhang mit dem Strassenverkehr und seinen Gefahren steht
- das einen Sachschaden und/oder Personenschaden zur Folge hat und
- an dem mindestens ein Fahrzeug oder ein fahrzeugähnliches Gerät beteiligt ist.

Planmässiges Handeln (z.B. Suizid- oder Tötungsabsicht) aller Beteiligten ist dabei ausgeschlossen.

Link: [https://data.stadt-zuerich.ch/dataset/sid\\_dav\\_strassenverkehrsunfallorte/download/RoadTrafficAccidentLocations.csv](https://data.stadt-zuerich.ch/dataset/sid_dav_strassenverkehrsunfallorte/download/RoadTrafficAccidentLocations.csv)

Weitere Informationen zu den Attributen des Datensatzes finden sich unter [https://data.stadt-zuerich.ch/dataset/sid\\_dav\\_strassenverkehrsunfallorte](https://data.stadt-zuerich.ch/dataset/sid_dav_strassenverkehrsunfallorte).

### 3.2 Daten der Verkehrszählung zum motorisierten Individualverkehr

Die hier verfügbaren Datensätze beinhalten die täglich aktualisierten Messwerte zum motorisierten Individualverkehr (MIV) in der Stadt Zürich, welche von den zahlreichen Zählstellen der Dienstabteilung Verkehr (DAV) gemessen wurden.

Zu einer Messstelle gehören mehrere Zählstellen, welche wiederum Messungen mit sogenannten Detektoren vornehmen. In den vorliegenden Datensätzen sind die Messwerte pro Zählstelle enthalten. Die genaue Lage der Detektoren einer Zählstelle kann einem zusätzlich verfügbaren detaillierten PDF-Situationsplan entnommen werden.

Die Daten sind verfügbar für die Jahre 2012 bis 2021, wobei jeweils die Daten eines Jahres in einer Datei enthalten sind.

Link: [https://data.stadt-zuerich.ch/dataset/sid\\_dav\\_verkehrszaehlung\\_miv\\_od2031/download/sid\\_dav\\_verkehrszaehlung\\_miv\\_OD2031\\_2020.csv](https://data.stadt-zuerich.ch/dataset/sid_dav_verkehrszaehlung_miv_od2031/download/sid_dav_verkehrszaehlung_miv_OD2031_2020.csv)

Weitere Informationen zu den Attributen des Datensatzes finden sich unter [https://data.stadt-zuerich.ch/dataset/sid\\_dav\\_verkehrszaehlung\\_miv\\_od2031](https://data.stadt-zuerich.ch/dataset/sid_dav_verkehrszaehlung_miv_od2031).

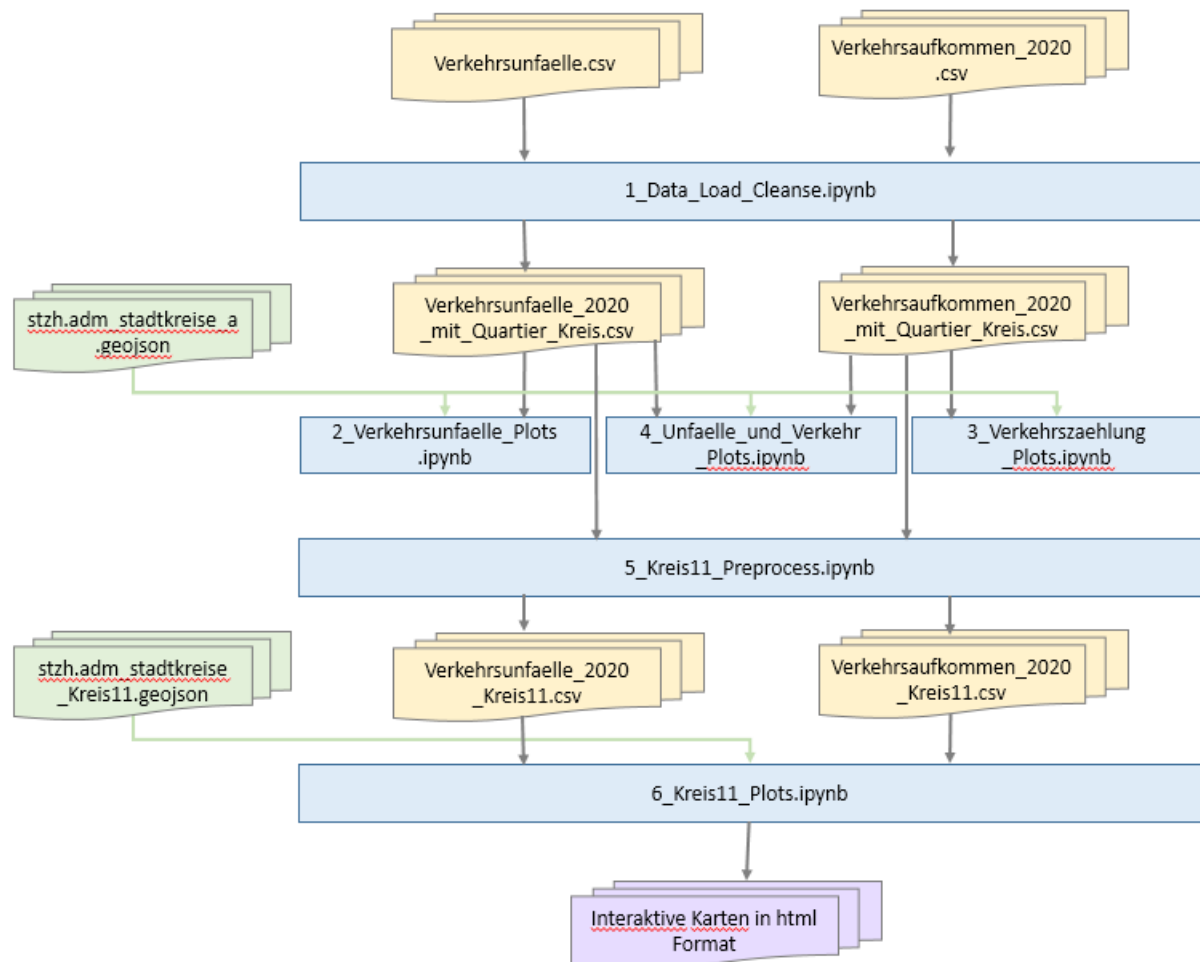
### 3.3 GeoJSON Daten

Um die Daten auf den Karten nachvollziehbar den einzelnen Stadtkreisen zuordnen zu können, resp. Unterschiede zwischen den Stadtkreisen anschaulich präsentieren zu können, wurden entsprechende GeoJSON Dateien verwendet.

**Dominique: hast Du mir hier noch die Links dazu?**

## 4 Übersicht Verarbeitungsschritte

Die einzelnen Verarbeitungsschritte hängen wie nachfolgend dargestellt miteinander zusammen.



In einem ersten Schritt werden die Daten in pandas DataFrames geladen und bereinigt, sowie anschliessend in neuen .csv Dateien abgespeichert (s. Kapitel 5). Anschliessend werden in den Notebooks 2 bis 4 die Daten analysiert und auf verschiedene Arten und unter Verwendung von folium Libraries visualisiert (Kapitel 6). Aufgrund der Erkenntnisse aus diesem Schritt wird ein Stadtkreis ausgewählt, dessen Daten genauer analysiert werden sollen.

Die Daten des ausgewählten Stadtkreises durchlaufen nochmals ein Preprocessing, bevor sie anschliessend zu guter Letzt im Detail analysiert und visuell dargestellt werden (Kapitel 7).

Die Idee, hier nochmals einen separaten Preprocessing Schritt einzufügen, beruht auf dem Bestreben, zeitintensive Verarbeitungsschritte nur für die Daten des ausgewählten Stadtkreises durchführen zu müssen.

Die .csv Dateien sind im Verzeichnis /data zu finden, ebenso die .geojson Dateien (in separaten Unterverzeichnissen). Die resultierenden .html Karten werden im Verzeichnis /results abgelegt.

## 5 Laden und Bereinigen der Daten

Die in diesem Kapitel beschriebenen Schritte sind im Jupyter Notebook „1\_Data\_Load\_Cleanse“ implementiert.

### 5.1 Polizeilich registrierte Verkehrsunfälle auf dem Stadtgebiet Zürich seit 2011

Die Datei RoadTrafficAccidentLocations.csv<sup>5</sup> (bei uns abgelegt als /data/verkehrsunfaelle.csv) umfasst die Daten sämtlicher Verkehrsunfälle der Jahre 2011 bis 2021 auf dem Gebiet der Stadt Zürich. Die Daten werden direkt in ein pandas DataFrame eingelesen und anschliessend einem Cleansing unterzogen.

Folgende Bereinigungs-schritte werden durchgeführt:

Löschen von nicht benötigten Spalten. So sind z.B. verschiedene Bezeichnungen je in einer Spalte sowohl in Deutsch, Französisch, Italienisch und Englisch enthalten. Hier reicht es, die Spalten mit den deutschen Bezeichnungen beizubehalten, die übrigen werden gelöscht.

Löschen sämtlicher Daten, die nicht das Jahr 2020 betreffen.

Überprüfen, ob Null Values vorhanden sind, was glücklicherweise nicht der Fall ist.

Geodaten transformieren: Die in der Datei enthaltenen geografischen Koordinaten sind im LV95 Format. Diese müssen ins WGS84 Format übersetzt werden, damit sie später auf einer Karte angezeigt werden können. Dies wird mit der Library pyproj durchgeführt.

Reverse Geocoding: Um die Daten einem bestimmten Stadtkreis zuordnen zu können muss ein Reverse Geocoding betrieben werden. Dies wird mittels Nominatim aus der Library geopy.geocoders durchgeführt. Aufgrund der Datenmenge (5291 Unfälle) dauert dieser Vorgang recht lange. Zudem werden nicht alle Adressen im selben Format retourniert. Manche Adressen haben ein Feld für die Hausnummer, andere nicht, wieder andere haben dafür ein zusätzliches Feld für eine Gebäudebezeichnung. Einige Werte müssen daher anschliessend mithilfe einer Lookup Tabelle nachträglich korrigiert werden.

Da der obige Schritt sehr zeitaufwändig ist werden die Daten anschliessend ein erstes Mal zwischengespeichert.

Kategorisierung in Jahreszeiten: Die Daten werden aufgrund der Monatsangabe im Datum den vier Jahreszeiten zugeordnet. Hierbei werden jedoch nicht die genauen Jahreszeiten gemäss Kalender verwendet, bei denen die Jahreszeiten inmitten einzelner Monate ändern, sondern es werden jeweils drei ganze Monate einer Jahreszeit zugeordnet.

Kategorisierung in Tageszeit-Abschnitte: Um die gewünschten Abfragen zu vereinfachen wird den Daten hier schon eine Tageszeit-Kategorie zugeordnet. Wir haben uns für folgende Kategorien entschieden:

- 0-5h: Nacht
- 5-9h: Rushhour Morgen
- 9-12h: Morgen
- 12-15h: Nachmittag
- 15-19h: Rushhour Abend
- 19-24h: Abend

Diese Kategorien sind bewusst nicht alle gleich lang gewählt worden, weil bei einer Aufteilung in gleich grosse „bins“ schlussendlich in jeder Kategorie „von allem ein bisschen etwas“

---

<sup>5</sup> Link siehe Kapitel 3

vorhanden ist und keine klaren Effekte mehr herausgelesen werden können. Diejenigen Kategorien mit dem grössten Verkehrsaufkommen (und damit auch mit den meisten Unfällen) sind dabei am kürzesten. Ganz korrekt müsste man wohl die Frequenzen noch so bearbeiten, dass sie proportional zur Anzahl Stunden korrigiert werden. Wir haben darauf der Einfachheit halber aber verzichtet weil die Effekte auch so sehr schön sichtbar sind.

Nach dem erfolgreichen Durchlaufen aller obigen Schritte wird der entstandene DataFrame in der Datei `/data/verkehrsunfaelle_mit_quartier_kreis.csv` abgelegt.

## 5.2 Daten der Verkehrszählung zum motorisierten Individualverkehr

Die Datei `sid_dav_verkehrszaehlung_miv_OD2031_2020.csv`<sup>6</sup> (bei uns abgelegt als `/data/-verkehrszaehlung_2020.csv`) umfasst die Daten von 94 Messstationen auf dem Gebiet der Stadt Zürich, wobei für jede Messstation stündlich ein Wert enthalten ist. Die Daten werden auch hier direkt in ein pandas DataFrame eingelesen und anschliessend einem Cleansing unterzogen.

Folgende Bereinigungs-schritte werden durchgeführt:

Überprüfung auf Null Values: In 63'615 Datenzeilen (von insgesamt knapp 1.6 Mio. Zeilen) sind Null Values enthalten, immer in der Spalte „AnzFahrzeuge“. Es handelt sich hier also um fehlende Messwerte.

Fehlende Messwerte sind in der Datei zum Teil schon imputiert, jedoch nur für einzelne fehlende Stundenwerte. Die vorgenommene Imputation basiert gemäss Dokumentation<sup>7</sup> auf den Messwerten zum gleichen Zeitpunkt des gleichen Wochentags aus den letzten 5 Jahren. Dazu wird ein Faktor zum jeweiligen Wert der vorher gemessenen Stunde berechnet.

Wenn eine Zählstelle länger ausgefallen ist, sind die entsprechenden Werte jeweils mit „AnzFahrzeugeStatus“ = „Fehlend“ gekennzeichnet. Diese Zeilen werden mittels `fillna(0)` behandelt.

Datumsformat umwandeln und aufsplitten: Die Messzeiten sind in einem recht unbrauchbaren Format abgelegt (`dtype('O')`) und müssen erst einmal mittels `pd.to_datetime` in ein vernünftiges Format umgewandelt werden. Anschliessend werden Jahr, Monat und Wochentag aus dem Datum extrahiert und in separaten Spalten abgelegt, wobei Monat und Wochentag zusätzlich auch noch als String abgelegt werden (also z.B. „Januar“ statt „1“).

Die Kategorisierung in Jahreszeit und Tageszeit-Abschnitt wird analog zum Vorgehen im Datensatz Verkehrsunfälle durchgeführt (s.o.).

Die Transformation der Geodaten sowie das Reverse Geocoding wird ebenfalls analog zu den Verkehrsdaten durchgeführt. Allerdings werden hierfür die Daten erst pro Messstellen-ID, Messdatum und Tageszeit-Kategorie gruppiert, wobei die Anzahl Fahrzeuge pro Gruppe summiert wird. Anschliessend werden die geografischen Informationen pro Messstelle einmal nachgeschaut. Der Datensatz umfasst insgesamt 94 Messstellen, jedoch rund 1.6 Mio. einzelne Zeilen mit Messdaten. Diese einzeln nachzuschauen wäre kaum praktisch durchführbar.

Kategorisierung des Verkehrsaufkommens: Den schon aggregierten Daten wird nun noch eine Kategorie bezüglich des Verkehrsaufkommens zugeordnet. Hierbei werden drei gleich grosse „bins“ dazu verwendet, die Labels „Gering“, „Mittel“ und „Hoch“ entsprechend zuzuweisen.

---

<sup>6</sup> Link siehe Kapitel 3

<sup>7</sup> [https://data.stadt-zuerich.ch/dataset/sid\\_dav\\_verkehrszaehlung\\_miv\\_od2031](https://data.stadt-zuerich.ch/dataset/sid_dav_verkehrszaehlung_miv_od2031).



Nach dem erfolgreichen Durchlaufen aller obigen Schritte wird auch der hier entstandene DataFrame gespeichert, diesmal in der Datei `/data/verkehrsaufkommen_mit_quartier_kreis.csv`.

## 6 Analyse der Daten

Die in Kapitel 5 aufbereiteten Daten werden anschliessend analysiert und in verschiedenen Grafiken visuell dargestellt. Dies geschieht in drei separaten verschiedenen Notebooks: 2\_Verkehrsunfaelle\_Plots.ipynb für die Unfalldaten, 3\_Verkehrszaehlung\_Plots.ipynb für die Daten zum Verkehrsaufkommen sowie 4\_Unfaelle\_und\_Zaehlung\_Plots.ipynb.

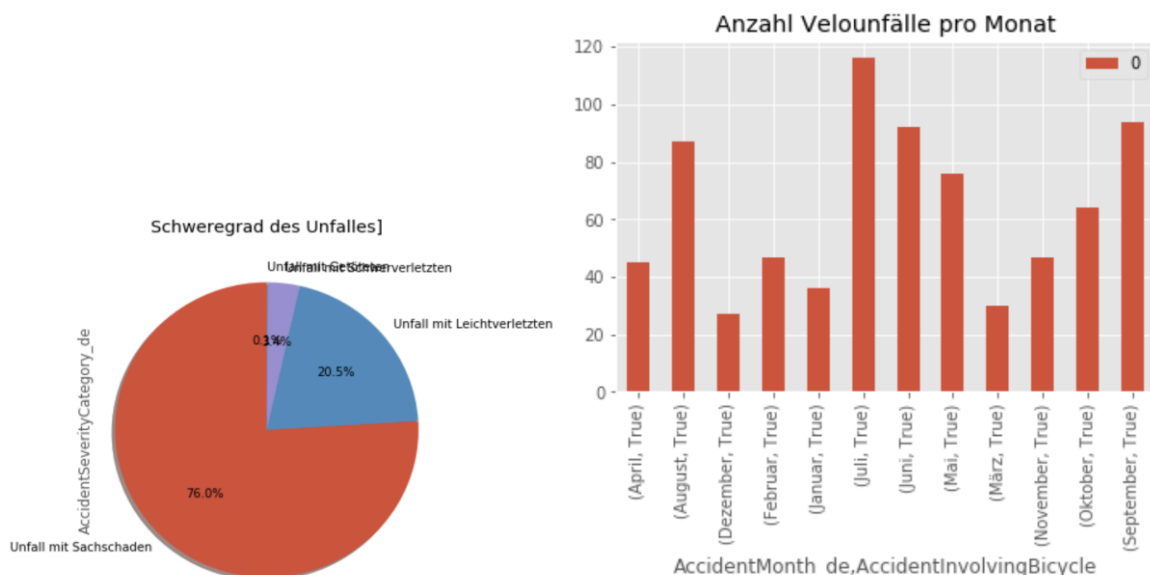
Hierbei kommen verschiedene Techniken zum Zug, wie das Erstellen von Statistiken, Pie- und Bar Charts, oder die Anzeige von Markern und Chloropleth Grafiken auf einer Karte mit Hilfe der folium Library.

All die Grafiken und Tabellen hier aufzuführen würde den Rahmen dieses Reports sprengen. Es werden daher im Folgenden nur einzelne Beispiele aufgeführt. Für Details verweisen wir auf die Jupyter Notebooks.

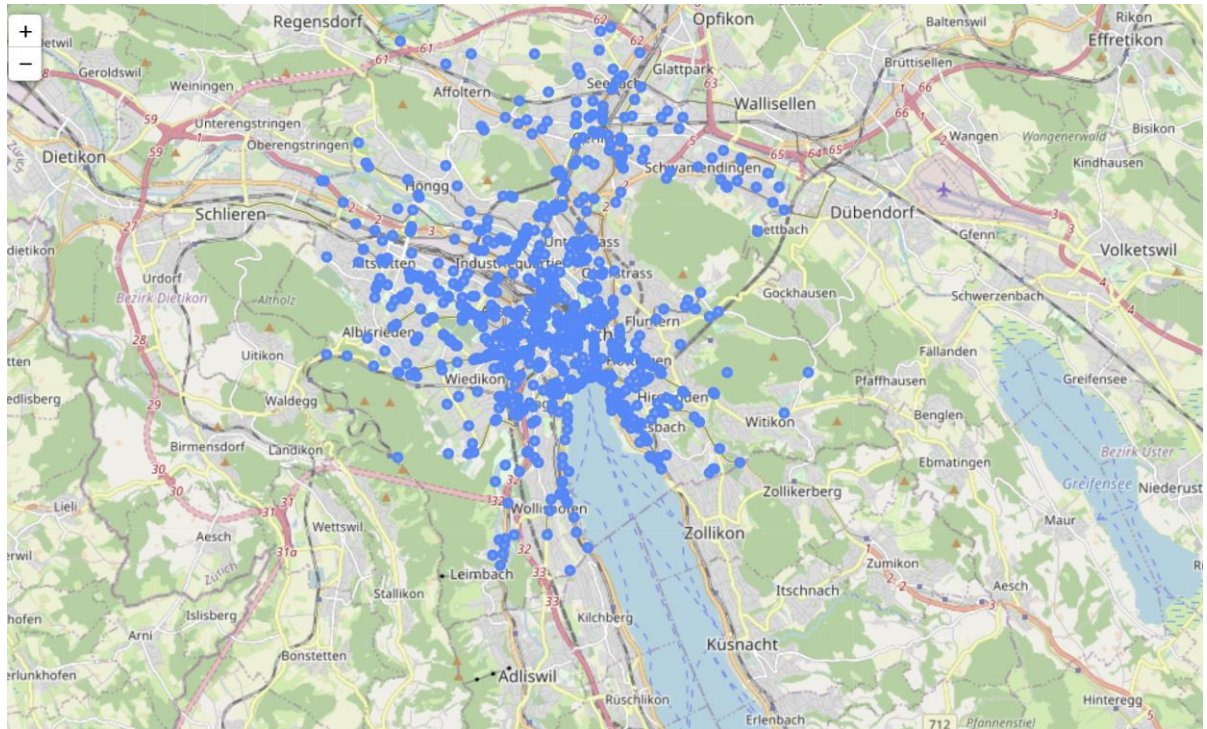
### Erstellen von Statistiken

AccidentSeverityCategory_de	
Unfall mit Sachschaden	4019
Unfall mit Leichtverletzten	1087
Unfall mit Schwerverletzten	180
Unfall mit Getöteten	5

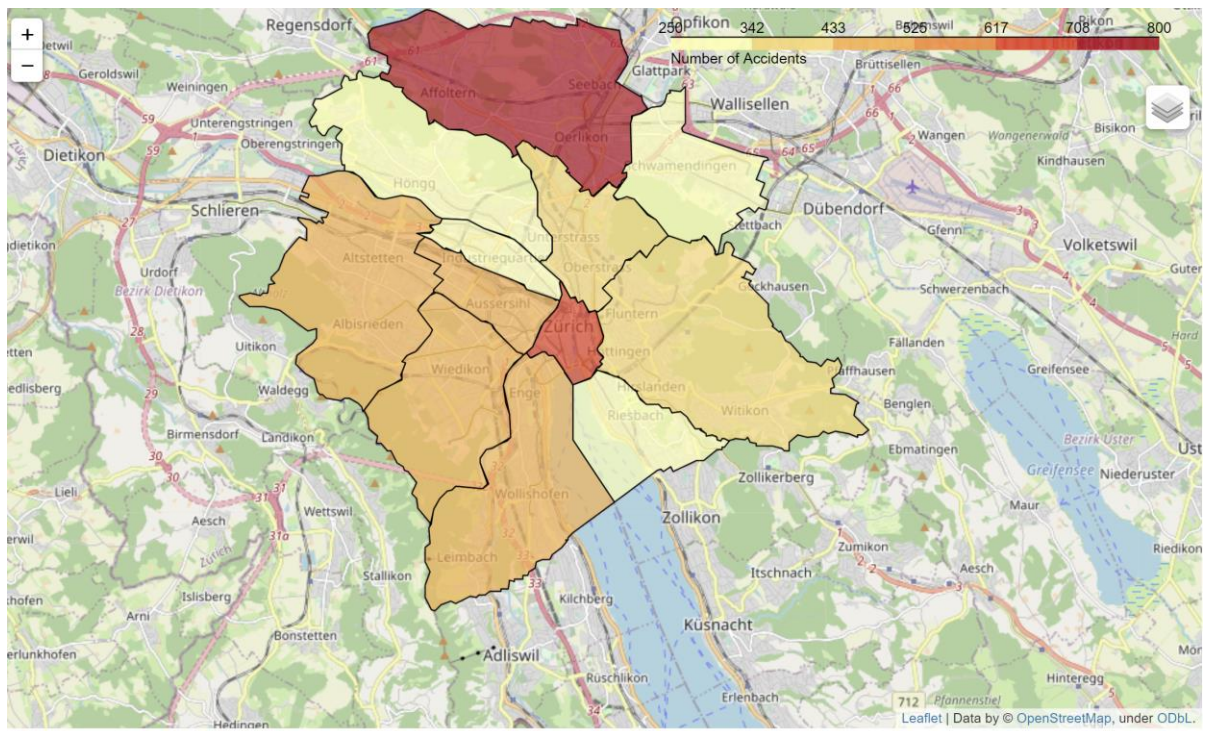
### Pie- und Bar Charts



## Anzeige von Markern auf einer Karte mit Hilfe der folium Library Velounfälle Stadt Zürich

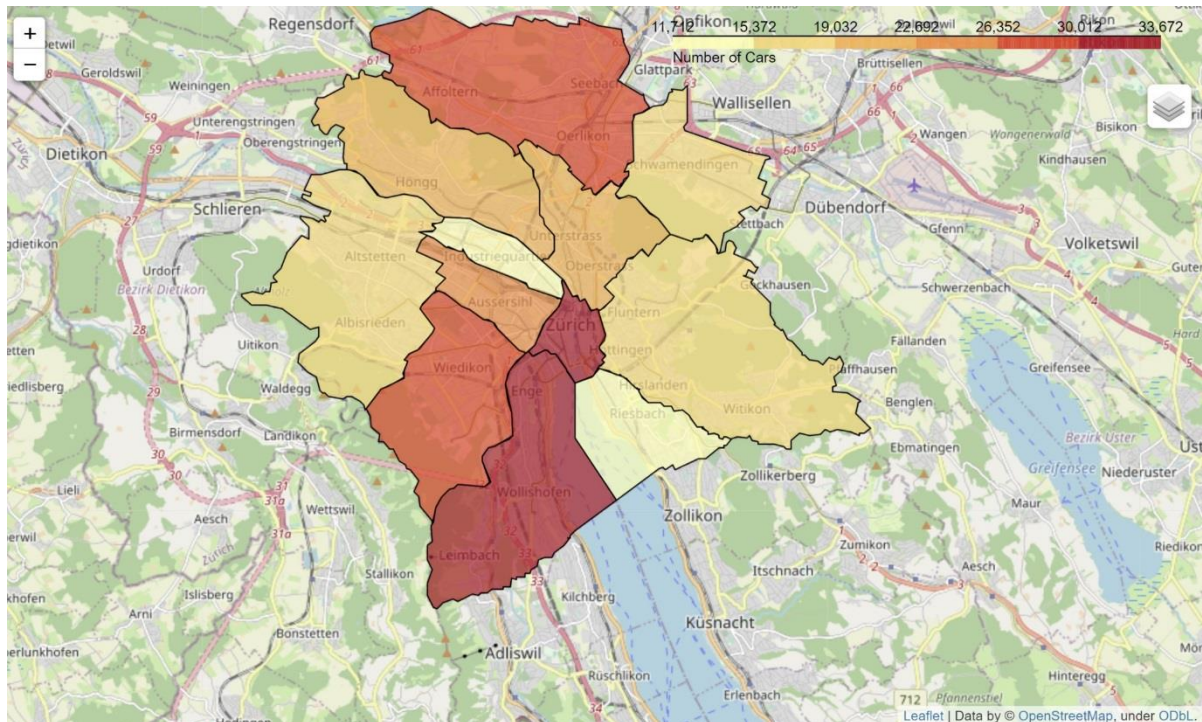


## Chloropleth Grafiken mit Hilfe der folium Library Anzahl Unfälle nach Stadtkreis





## Chloropleth Grafiken mit Hilfe der folium Library Verkehrsaufkommen nach Stadtkreis



Zusammengefasst lassen sich folgende **Hauptaussagen** machen:

- Der weitaus grösste Teil der Unfälle zieht nur Sachschaden nach sich (4'019). Unfälle mit Leichtverletzten sind knapp vier Mal seltener (1'087), Unfälle mit Schwerverletzten nochmals gut fünf Mal seltener (180). Es gab 5 Unfälle mit Todesfolge.
- Die häufigste Unfallart ist „Schleuder- oder Selbstunfall“, gefolgt von „Parkierunfall“ und „Auffahrunfall“ (sowie weiteren, selteneren Unfallursachen).
- In den Sommermonaten gibt es am meisten Unfälle, am wenigsten im Frühling (wobei dies ein Corona Effekt sein dürfte).
- An Wochentagen gibt es mehr Unfälle als am Wochenende, am allerwenigsten am Sonntag, am meisten am Mittwoch.
- Die meisten Unfälle erfolgen in der Rushhour am Abend. Erstaunlicherweise gibt es aber am Nachmittag (12-15h) immer noch mehr Unfälle als in der Rushhour am Morgen (5-9h), obwohl letztere Kategorie eine Stunde länger dauert.
- Im Kreis 11 treten am meisten Unfälle auf, gefolgt von der Altstadt (obwohl dies ein sehr kleiner Kreis ist) und zwar in allen Jahreszeiten. Bei den Velo-, Motorrad- und Fussgängerunfällen schwingt jedoch die Altstadt obenaus.
- Die Zählstellen für das Verkehrsaufkommen sind nicht gleichmässig über das Stadtgebiet verteilt. Im Kreis Altstadt hat es am meisten davon.
- In den Kreisen 1 und 2 ist das Verkehrsaufkommen am höchsten.
- Die Häufigkeit von Unfällen ist klar korreliert mit dem Verkehrsaufkommen.

Da im Kreis 11 am meisten Verkehrsunfälle aufgetreten sind (absolute Werte, nicht proportional zur Fläche des Kreises) haben wir beschlossen, diesen Kreis anschliessend noch vertiefter zu analysieren. Die Resultate dazu finden sich in Kapitel 7.

## 7 Vertiefte Analyse der Daten / Business Case Analyse

### 7.1 Laden der Daten und zusätzliche Vorbereitungsschritte

Im Notebook `5_Kreis11_Load_Preprocess.ipynb` werden die im Schritt 1 (`1_Data_Load_Cleanse.ipynb`) aufbereiteten Dateien für die Unfall- und Verkehrsaufkommens-Daten geladen, auf die Daten des Kreis 11 reduziert und anschliessend in den neuen Dateien `verkehrsunfaelle_2020_kreis11.csv` und `verkehrszaehlung_2020_kreis11.csv` für den nachfolgenden Schritt 6 bereitgelegt.

Die Idee, hier nochmals einen separaten Preprocessing Schritt einzufügen beruht auf dem Bestreben, zeitintensive Verarbeitungsschritte nur für die Daten des ausgewählten Stadtkreises durchführen zu müssen. Aktuell wird jedoch nur ein einziger zusätzlicher Preprocessing Schritt durchgeführt, der darin besteht, jedem Verkehrsunfall das passende Quartier zuzuweisen. (Dies ist allerdings hauptsächlich darauf zurückzuführen, dass die beiden Dateien durch verschiedene Personen geschrieben worden sind. Beide Schritte liessen sich ohne Mehraufwand im gleichen Schritt durchführen (der Kreis muss sowieso nachgeschaut werden, das Quartier könnte bei der gleichen Anfrage mitangefragt werden). Es ist aber durchaus denkbar, dass es bei weiteren Auswertungen Sinn machen würde, zusätzlich benötigte Vorverarbeitungsschritte hier unterzubringen.

### 7.2 Analyse der Daten

Die in diesem Kapitel beschriebenen Schritte sind im Jupyter Notebook „`6_Kreis11_Plots.ipynb`“ implementiert.

#### 7.2.1 Cluster Plots mit Anzeige des Unfall-Schweregrads und Beteiligten

Als haben wir eine Karte erzeugt, in der alle Unfälle je nach Zoom Level entweder einzeln als Marker oder gruppiert in Clustern angezeigt werden. Wird auf ein Cluster draufgeklickt so geht dieser auf und es wird der nächsthöhere Detaillierungsgrad angezeigt.

Die Marker für die einzelnen Unfälle zeigen folgende Informationen an:

- Position: Ort des Unfalls
- Farbe des Markers: Schweregrad des Unfalls (hellblau: nur Sachschaden, hellrot: Leichtverletzte, dunkelrot: Schwerverletzte, schwarz: Tote)
- Icon des Markers: Falls Fussgänger, Motorrad- oder Velofahrer am Unfall beteiligt waren so wird dies mit einem entsprechenden Icon angezeigt.
- Tooltip: Wenn sich die Maus über einem Marker befindet werden zusätzliche Informationen zum betreffenden Unfall angezeigt (Art des Unfalls, Wochentag, Monat und Tageszeit-Kategorie).

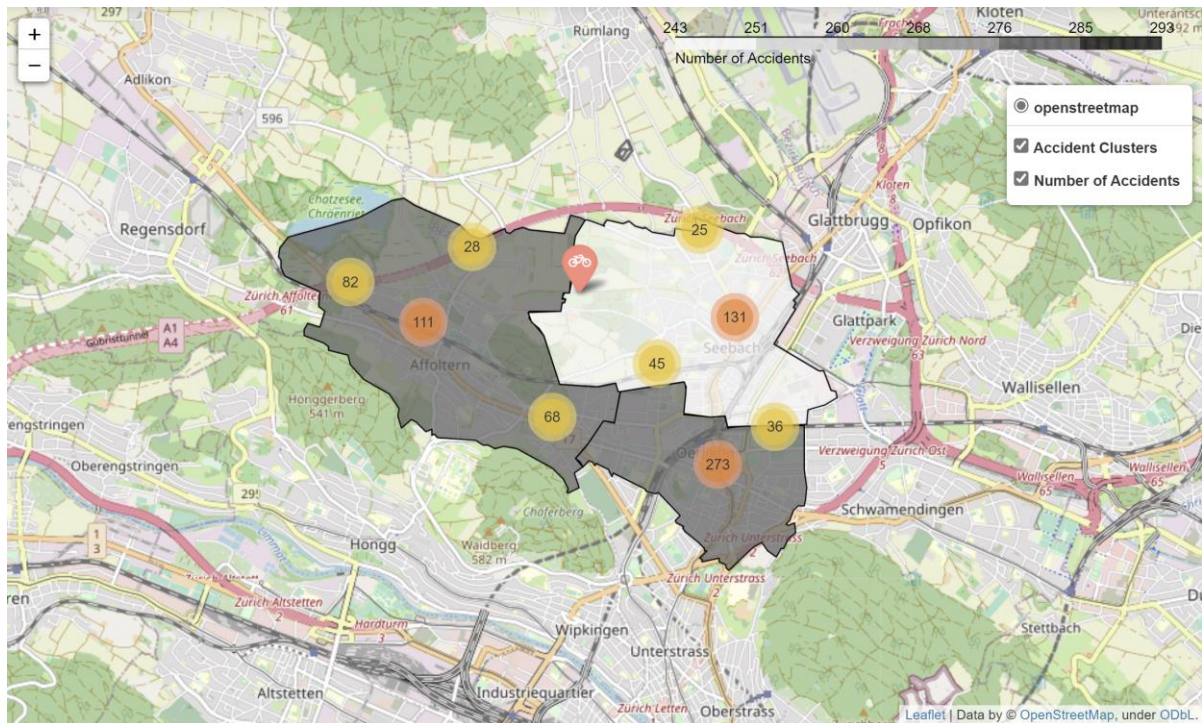
Im Hintergrund sind zudem immer noch die Angaben zur Unfalldichte pro Quartier in Form eines Chloropleth Plots ersichtlich (je dunkler desto dichter). Dieser Layer kann aber auch mittels LayerControl weg-selektiert werden.

Es wird eine Funktion zur Verfügung gestellt, die eine beliebige Teilmenge der Unfalldaten entgegennimmt, z.B. alle Velounfälle, alle Unfälle im Winter, alle Unfälle während der Rush-hour am Abend etc. (oder auch alle Daten für den Kreis 11) und diese auf einer Karte darstellt. Die resultierende Karte wird einerseits auf dem Bildschirm angezeigt und andererseits zusätzlich als Datei abgespeichert

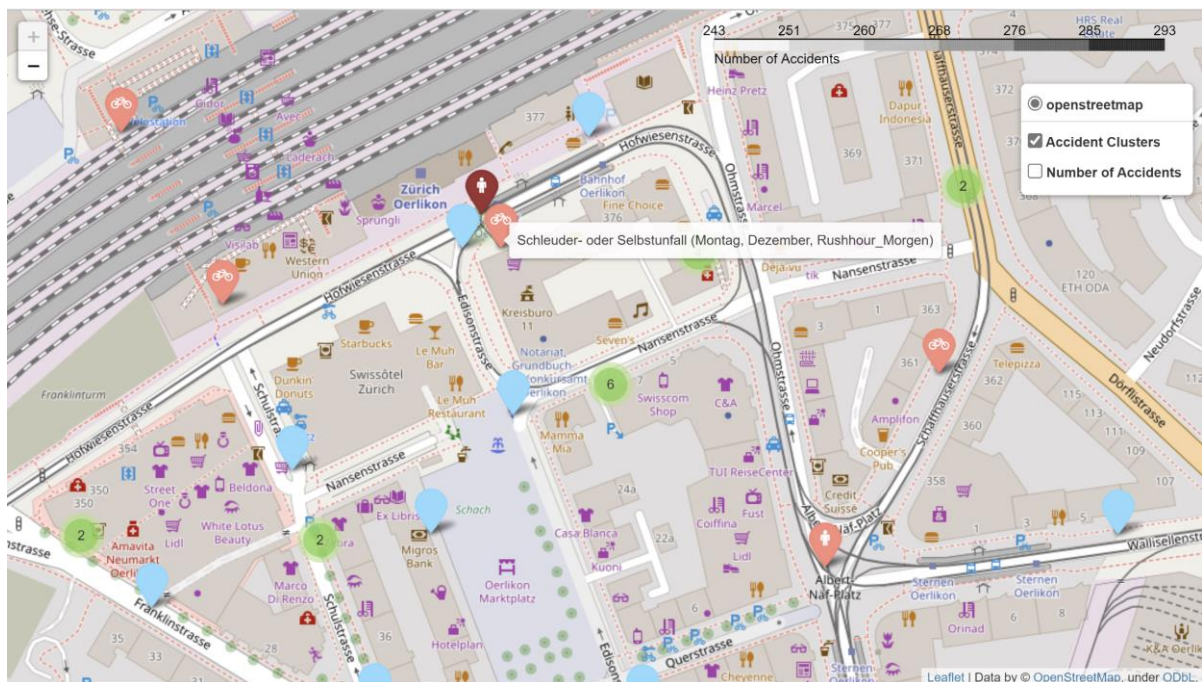
Grundsätzlich können der Funktion auch Daten aus anderen Kreisen übergeben werden, auch diese werden korrekt auf der Karte angezeigt. Einzig die Hintergrundinformation



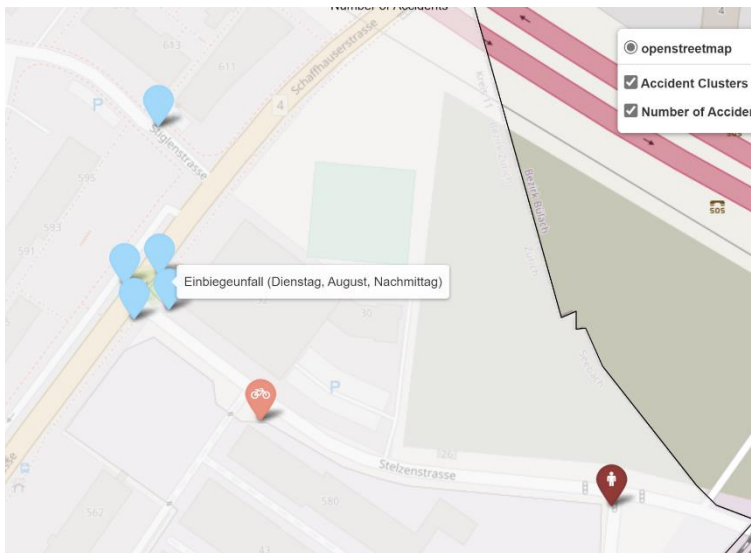
bezüglich der allgemeinen Unfalldichte pro Quartier (Chloropleth Layer) wird unabhängig von den Daten immer nur für den Kreis 11 angezeigt. Dies könnte aber mit wenig Aufwand auch noch integriert werden.



Einstiegslevel: hauptsächlich Cluster ersichtlich mit Gesamtanzahl Unfälle pro Region



Höherer Zoomlevel: mehr Details, hier rund um den Bahnhof Oerlikon, Chloropleth Layer deaktiviert



Schaffhauserstrasse

Abbiegeunfälle ohne  
Fahrradbeteiligung

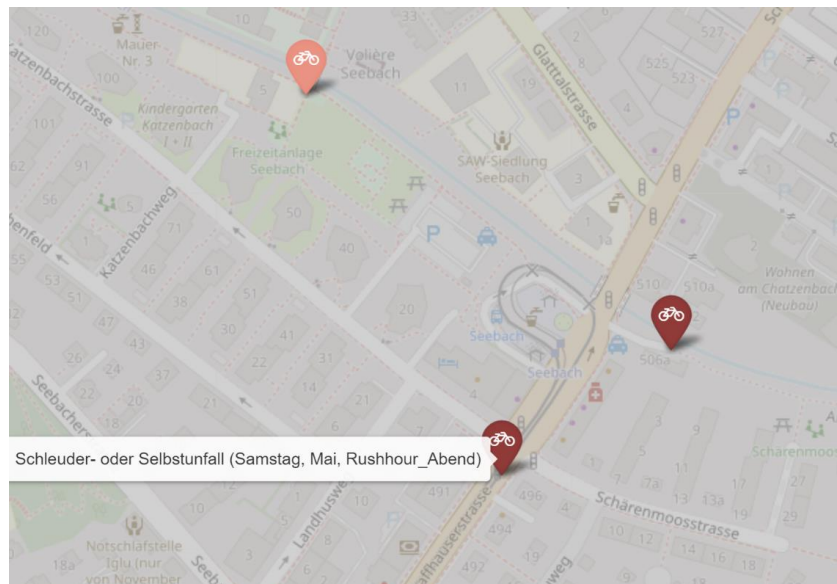
nur Sachschaden

Schaffhauserstrasse,  
etwas weiter oben:

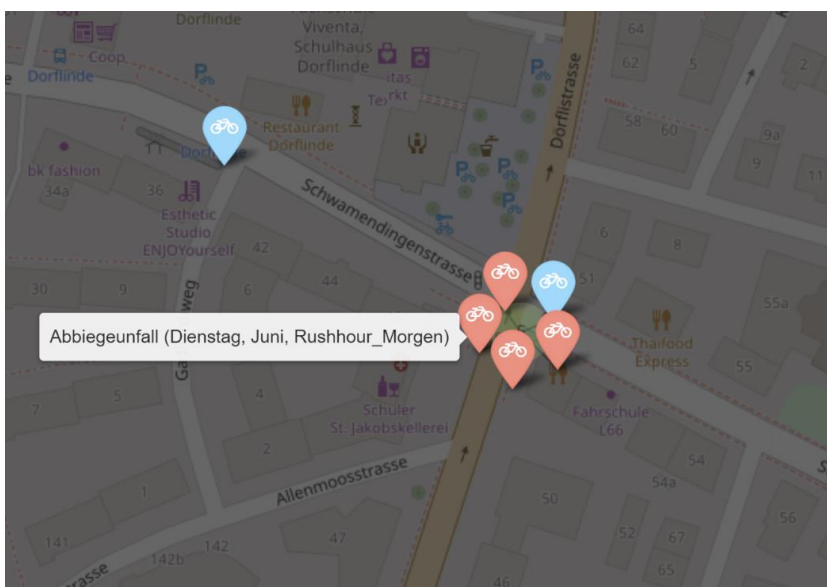
2 Schleuder- oder  
Selbst-unfälle mit  
Schwerverletzten

beide in der Rushhour  
am Abend

beide am Wochenende



Schleuder- oder Selbstunfall (Samstag, Mai, Rushhour\_Abend)



Abbiegeunfall (Dienstag, Juni, Rushhour\_Morgen)

Oerlikon, Kreuzung  
Schwamendingenstrasse  
/ Dörflistrasse

Cluster von Velounfällen

allesamt Abbiegeunfälle  
unter der Woche

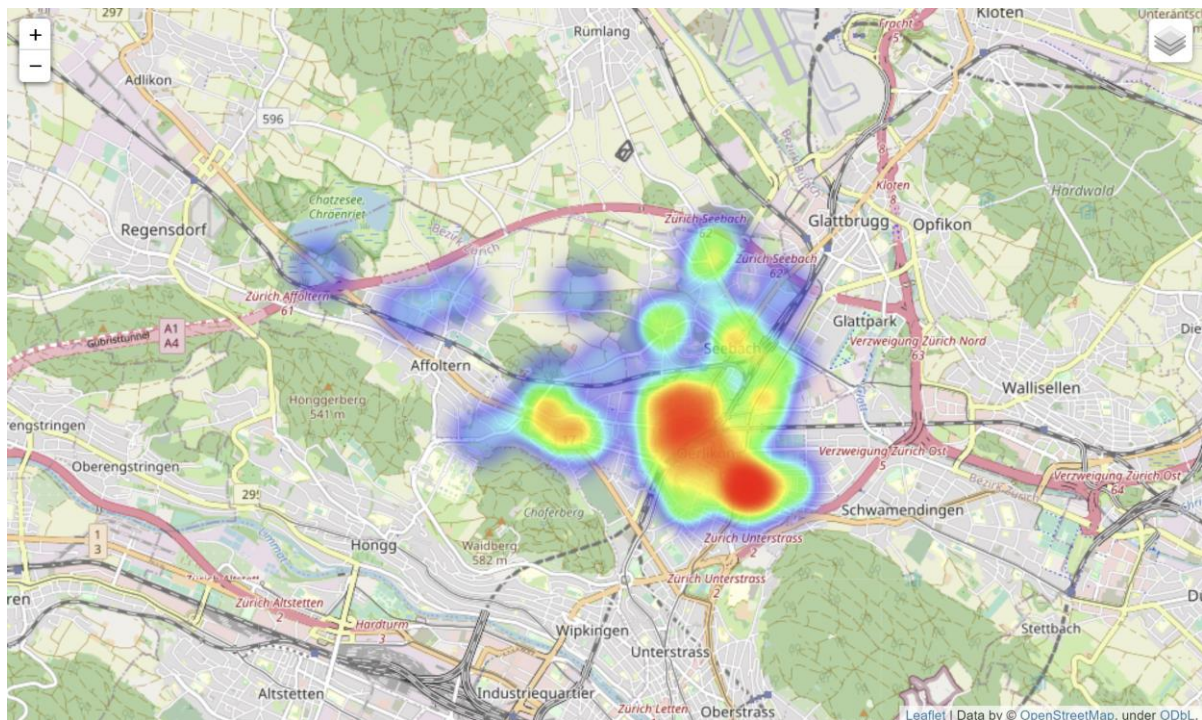
3x in der Rushhour am  
Abend, 1x in der  
Rushhour am Morgen  
und einmal am Morgen



### 7.2.2 Heatmaps

Ebenfalls ein sehr gutes Mittel, sich schnell einen Überblick zu verschaffen über die Anzahl der Unfälle ist eine HeatMap. Diese bietet zwar nicht so viele Möglichkeiten, zusätzliche Informationen zu den Unfallorten verfügbar zu machen wie die Cluster Maps, dafür sind sie beim Herein- und Herauszoomen wesentlich einfacher zu handhaben, indem sich stets alle Punkte auf dem gleichen Zoom-Level befinden, wohingegen bei der Clustermaps immer nur einzelne Cluster aufgehen und andere zum Teil gleichzeitig wieder zugehen wenn man einen neuen Cluster öffnet.

Auch hier wurde wieder dasselbe Vorgehen gewählt, dass eine Funktion für das Erstellen der Karte zur Verfügung gestellt wird, die dann mit beliebigen Teilmengen des Datensatzes aufgerufen werden kann.



Heatmap für Unfälle mit Velobeteiligung, es sind zwei klare Hotspots um den Bahnhof Oerlikon erkennbar.

### 7.2.3 Heatmaps mit zeitlicher Entwicklung

Sehr spannend ist des Weiteren die Möglichkeit, eine HeatMap mit einer zeitlichen Entwicklung zu erzeugen. Hierbei kann der Benutzer durch die einzelnen Zeitpunkte wandern und verfolgen, wie sich die Unfallschwerpunkte im Laufe der Zeit verändern.

Wie immer kann die Auswertung über beliebige Untermenge der Daten vorgenommen werden. In der entstandenen Karte kann durch Klicken auf die einfachen Vorwärts- resp. Rückwärtspfeile in der Kontrollleiste unten links der Verlauf der Unfallorte und -häufigkeiten automatisch für alle Zeitscheiben hintereinander wiedergegeben werden. Wenn man das Tempo der Abfolge selber steuern möchte ist dies mittels Klicken auf die doppelten Pfeile ebenfalls möglich. In der Mitte der Kontrollleiste wird zudem der Name der aktuell dargestellten Zeitscheibe angezeigt.

Als erstes wurde versucht, die Daten im Verlauf des Jahres darzustellen, also pro Monat. Dieser Versuch erwies sich aber als nicht besonders spannend. Die Verteilung der Unfallorte über die einzelnen Monate erscheint sehr heterogen, es erschliessen sich keine grossen

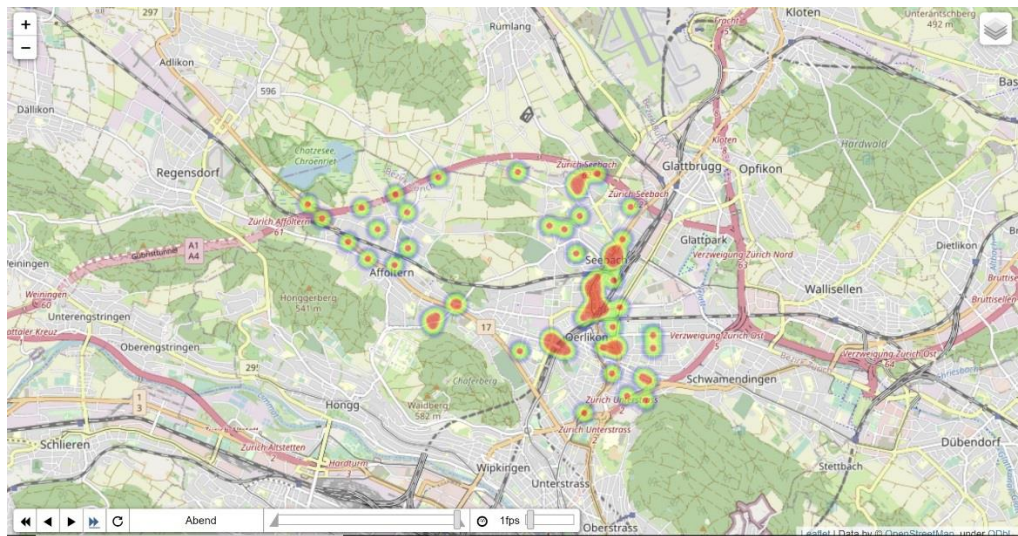
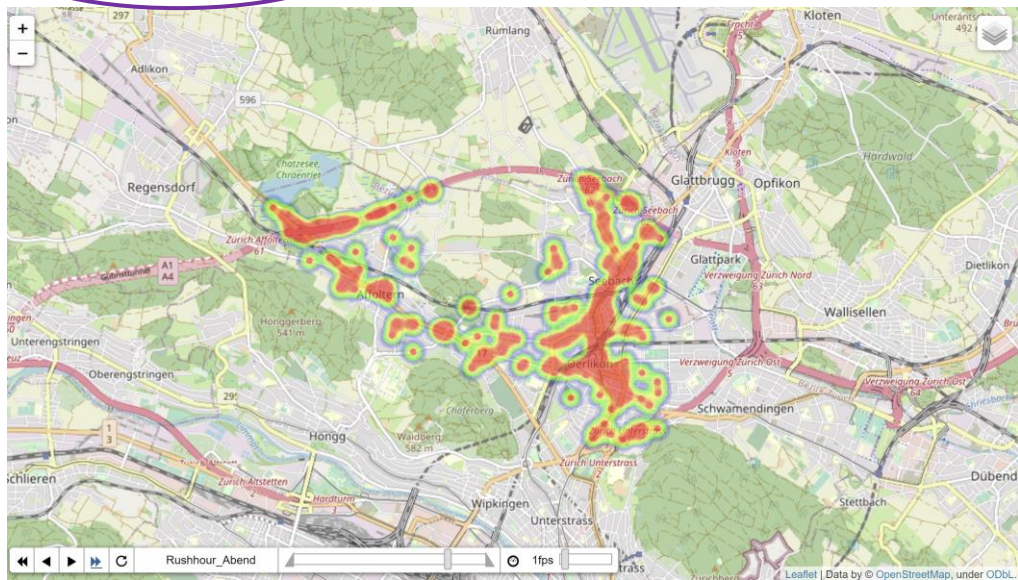
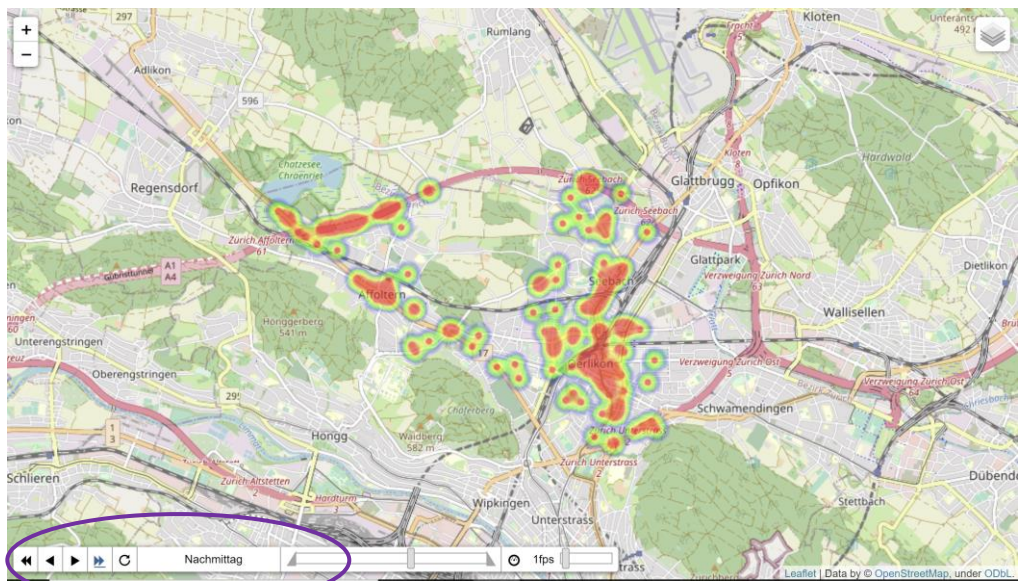


sinnvollen Zusammenhänge. Die Unfälle trugen sich in den verschiedenen Monaten auf (zum Teil erstaunlich stark) disjunkten Verkehrsachsen zu, was die Vermutung aufkommen lässt, dass dies stark mit aktuell vorhandenen Baustellen oder Strassensperrungen zu tun hat. Dies kann natürlich ebenfalls von Interesse sein, aber nur wenn man zusätzlich die Daten zur Verfügung hat, wann wo Bauarbeiten, Strassensperrungen oder sonstige besondere Vorkommnisse stattgefunden haben.

Als nächstes wurde daher eine HeatMap Serie entwickelt, welche die Daten für die verschiedenen Tageszeit-Kategorien aufzeigt. Natürlich kann auch in dieser Karte hinein- und herausgezoomt werden.

Diese Karte zeigt ein deutlich intuitiveres Bild als die Darstellung über die Monate hinweg. Es sind klare Unfallschwerpunkte auszumachen, etwa in der Umgebung des Bahnhofs Oerlikon. Diese bewegen sich von Kategorie zu Kategorie nur wenig, blähen sich aber während der Stosszeiten deutlich auf.

Dieselbe Karte lässt sich natürlich auch nur für die Velounfälle erstellen: Hier sieht man wieder etwas mehr Bewegung als bei der Sicht auf alle Unfälle. Einerseits hat dies sicher damit zu tun, dass es weniger Velo-Unfälle gibt als Unfälle überhaupt und somit hier noch mehr statistische Schwankungen zu erwarten sind. Andererseits ist es aber auch gut möglich, dass Velo-Aktivitäten je nach Tageszeit an unterschiedlichen Orten stattfinden. So bewegen sich die Velofahrer in Affoltern offenbar am Vormittag deutlich weniger (oder aber zumindest vorsichtiger) als am Nachmittag. Diese Tatsache liesse sich unter Umständen nutzen, indem zum Beispiel überprüft werden könnte, ob die Verkehrsströme je nach Tageszeit etwas anders gelenkt werden könnten um kritische Gebiete zu entlasten.



Snapshots für die drei Zeitscheiben «Nachmittag», «Rushhour Abend» und «Abend»



## 8 Ergebnisse

Aus Platzgründen können in diesem Report nicht alle Karten und schon gar nicht alle einzelnen Folgerungen daraus dargestellt werden. Es stehen jedoch alle Karten als .html Dateien im /results Folder bereit, um interaktiv darin zu forschen. Hat man andere Interessensschwerpunkte, so können mit dem Jupyter Notebook auf sehr einfache Art neue Karten mit einer anderen Datenmenge erzeugt werden. Es reicht, das Selektionskriterium für den gewünschten Datensatz zu formulieren und die entsprechende Funktion aufzurufen. Noch komfortabler wäre es, wenn hierfür ein kleines GUI erstellt würde, was den Rahmen dieser Arbeit aber definitiv sprengen würde.

Die ClusterMap Karten sind ein sehr gutes Tool, um gezielt nach Unfallclustern zu suchen und sich schnell einen Überblick zu verschaffen, ob die Unfälle einem Muster folgen oder tatsächlich zufällig verstreut sind. So ist z.B. das oben gezeigte Beispiel des Velounfallclusters an der Kreuzung Schwamendingerstrasse / Dörflistrasse ein klares Beispiel für eine Stelle, an der Massnahmen ergriffen werden sollten.

Tatsächlich hat dies auch die Stadt Zürich erkannt und hat noch im Jahr 2020 an jener Stelle eine sogenannte Trottoirnase eingebaut, die verhindern soll, dass beim Rechtsabbiegen Velofahrer übersehen werden.<sup>8</sup>

### Dörflistrasse



Beim Rechtsabbiegen das Velo übersehen – das war die Gefahr an der Kreuzung Dörfli-/Schwamendingenstrasse. Aus diesem Grund haben wir eine «Trottoirnase» gebaut. Sie ist rechts unten im Nachher-Bild zu erkennen und hilft, das Rechtsabbiegen zu vereinfachen und Unfälle zu verhindern.

Ebenso fällt beim Anschauen der Karte mit lediglich Velounfällen auf Level der einzelnen Marker auf, dass der Anteil an roten Markern, also an Unfällen mit Verletzten, sehr viel höher ist als auf der Karte mit allen Unfällen. Es macht also sehr viel Sinn, den Velounfällen eine besondere Aufmerksamkeit zu schenken.

Die HeatMaps über Zeitabschnitte wiederum sind ein nützliches Tool, um Verkehrsströme zu analysieren. (Also eigentlich hier Unfallströme, Verkehrsströme und Unfallströme sind ja aber, wie wir schon in der Grundlagenanalyse gesehen hatten, stark gekoppelt.)

Schon in der Grundanalyse (4\_Unfälle\_und\_Verkehr\_Plots.ipynb) konnte gezeigt werden, dass die Höhe des Verkehrsaufkommens und die Anzahl der Unfälle sehr stark gekoppelt sind. Ebenso fällt beim Analysieren der Unfallcluster auf, dass etliche dieser Cluster zu Stosszeiten entstanden sind. Nebst der Entschärfung von kritischen Stellen ist es somit ebenfalls empfehlenswert, die Verkehrsströme zu analysieren und diese besser zu verteilen.

---

<sup>8</sup> <https://www.stadt-zuerich.ch/site/velo/de/index/erfolgreich-umgesetzt.html>

## 9 Probleme

Schwieriger als erwartet hat sich die Verbindung der Unfalldaten mit den Daten aus den Verkehrszählungen erwiesen. Dies hauptsächlich deshalb, weil bei den Unfalldaten kein exaktes Datum hinterlegt ist, sondern lediglich Monat, Wochentag (also beispielsweise „Dienstag“) und Uhrzeit. Dies verunmöglicht ein genaues Matching von Unfallzeitpunkt und Verkehrsaufkommen zu diesem Zeitpunkt, es kann lediglich mit Mittelwerten gearbeitet werden. Für einen Unfall an einem Mittwoch im Juli um 17h also mit dem Mittelwert aller Verkehrsmessungen von Mittwochen im Juli, in der Tageszeit-Kategorie „Rushhour Abend“.

Ebenso schwierig ist die Wahl, welche Zählstellen denn für die Ermittlung dieser Mittelwerte beigezogen werden sollen. Eine Option wäre, mittels Distanzberechnung herauszufinden, welche Zählstelle am nächsten beim Unfallort liegt. Befindet sich die Zählstelle aber beispielsweise an einer Autobahn und der Unfall findet auf einem gleich daneben vorbeiführenden Veloweg statt, so hat das Verkehrsaufkommen auf der Autobahn vermutlich wenig Einfluss auf den Unfall gehabt.

Eine weitere Alternative wäre möglicherweise das Aufsummieren aller Fahrzeuge pro Quartier für einen gegebenen Tagesabschnitt. Auch dies hat jedoch seine Tücken, denn die Zählstellen sind sehr unregelmässig verteilt. Es müsste also ein Kontrollfaktor angewendet werden für die Anzahl der Zählstellen im betrachteten Gebiet und für dessen Grösse damit nicht einfach das Quartier mit den meisten Zählstellen als das verkehrsreichste dargestellt wird.

Somit haben wir uns im Rahmen dieser Arbeit im Hinblick auf das Verkehrsaufkommen darauf beschränkt, Choropleth Karten zu erstellen, die aufzeigen, welche Stadtkreise wann ein besonders hohes Verkehrsaufkommen aufweisen. Allerdings haben wir darauf verzichtet, einen Kontrollfaktor anzuwenden. Zudem konnten wir zeigen, dass Verkehrsaufkommen und Unfalldichte sehr stark gekoppelt sind, so dass ähnlich relevante Informationen auch aus den Angaben zur Tageszeit Kategorie sowie den HeatMaps bezüglich der Unfalldichte erhalten werden können. Sehr viel nützliche Information kann auf jeden Fall schon aus den Verkehrsunfalldaten alleine extrahiert werden. Denkbar wäre, beim Vorliegen von bestimmten Unfallclustern spezifisch zu überprüfen, ob eine passende Zählstation in der Nähe liegt und deren Daten dann beizuziehen.

## 10 Conclusion

Das Erarbeiten der obigen Tools und Analysen hat uns grossen Spass gemacht und wir haben viel gelernt dabei.

Im Kapitel Ergebnisse haben wir einige mögliche Erkenntnisse aufgeführt, man könnte aber auch noch deutlich mehr Zeit investieren um eine umfangreichere (Rang-)Liste zu erstellen mit sinnvollen Massnahmen, sowohl in Bezug auf die Entschärfung von Unfall Hotspots, wie auch auf die Lenkung der Verkehrsströme.

Sicherheit beim Velofahren ist ein Strategie-Schwerpunkt des Zürcher Stadtrats<sup>9</sup>. Angesichts des viel grösseren Anteils an Verletzten gegenüber anderen Unfällen ist dies sicher eine gute Entscheidung.

---

<sup>9</sup> <https://www.stadt-zuerich.ch/sichervelofahren>