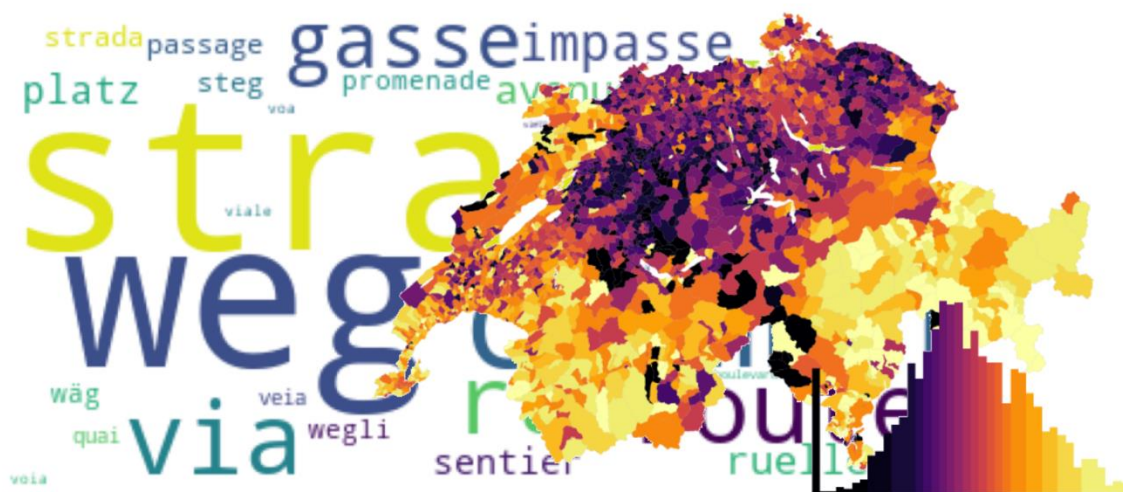


Spatial Data Mining mit Python: Möglichkeiten und Tools

Etymologie von Strassennamen der Schweiz



Autoren

Michael Bühler
Daniel Gmür

Betreuer

Martin Christen

Olten, 25. November 2022

Autoren

Michael Bühler

michael.buehler@bfs.admin.ch

Daniel Gmür

danielgmuer@gmail.com

Redlichkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der im Quellenverzeichnis angegebenen Quellen und Hilfsmittel angefertigt habe. Die wörtlich oder inhaltlich den im Quellenverzeichnis aufgeführten Quellen und Hilfsmitteln entnommenen Stellen sind in der Arbeit als Zitat bzw. Paraphrase kenntlich gemacht. Diese Studienarbeit ist noch nicht veröffentlicht worden. Sie ist somit weder anderen Interessierten zugänglich gemacht noch einer anderen Prüfungsbehörde vorgelegt worden.

Der vorliegende Bericht wurde von den Studierenden des CAS Spatial Data Analytics 2022 im Rahmen einer Zertifikatsarbeit erarbeitet.

Es muss an dieser Stelle darauf hingewiesen werden, dass die Arbeit nicht im Rahmen eines Auftragsverhältnisses erstellt wurde.

Weder die Autoren noch die Fachhochschule Nordwestschweiz können deshalb für Aktivitäten auf der Basis dieser Studierendenarbeit eine planerische Haftung übernehmen.

Zusammenfassung

Strassennamen bieten einen interessanten Einblick in eine Gesellschaft. Die Analyse der Strassennamen der Schweiz ist somit auch ein Blick in die Kultur und Geschichte dieses Landes.

Diese Arbeit untersucht die Strassennamen der Schweiz. Dafür wird der OGD-Datensatz «Amtliches Verzeichnis der Strassen» mit Daten aus der offenen, strukturierten Datenbank Wikidata und weiteren Datenquellen von Wikimedia abgeglichen. Die Auswertung erfolgte zu einem grossen Teil mittels den Pythonbibliotheken *pandas* und *geopandas* in JupyterLab Notebooks. Ergänzend wurde eine Programm-bibliothek zur *Names Entity Recognition* eingesetzt.

Es entstanden ein Datensatz zu den generischen Benennungen der Strassen und einer zu den spezifischen Benennungen. Beide Datensätze wurden mittels diverser Karten und Diagrammen illustriert.

Die Auswertungen der generischen Benennungen beschäftigt sich mit den oft vorkommenden Namenszusätzen wie *-strasse* oder *-gasse*. Es zeigte sich, dass über 55% aller Strassen den Term *strasse* oder *weg* aufweisen. Bei der Strassenlänge stellten sich *route* mit einer durchschnittlichen Länge von fast 1'200m und *strada* mit rund 1'100m an die Spitze. Unterschiede in den Sprachgebieten sind ebenfalls ersichtlich. Die Ostschweizerinnen geben den Strassennamen eher den Zusatz *Strasse*, die Nordwestschweizer benutzen hingegen mehr *Weg*, im Jura weisen sie oft ein *Rue* auf, während im Arc Lémanique *chemin* dominiert. Nicht sonderlich verwunderlich ist, dass auch innerhalb der Sprachgebiete kantonale Unterschiede festzustellen sind.

Bei den spezifischen Benennungen zeigte sich die grosse Herausforderung bei der Zuweisung an eine der neun definierten Kategorien. Viele Strassennamen sind nicht einem der zu vergleichenden Datenquellen zuzuweisen oder sind es mehrfach. Rund 60% aller Strassen der Romandie und der Deutschschweiz konnte mindestens einer Kategorie zugewiesen werden. Die Zuteilung wurde gegen Daten geprüft, die im kleinen Ausmass in OpenStreetMap verfügbar sind. Dabei hat sich gezeigt, dass etwa 50% der Namen richtig zugewiesen werden konnten. Bei der Auswertung der Zuweisungen zeigen sich bei den meisten Kategorien regionale Unterschiede. Diese folgen entweder den Sprachregionen, Stadt und Land oder dem Mittelland/Jura und Alpenregionen.

Vorwort

Seit tausenden von Jahren erleichtern sie den Menschen die Fortbewegung. Ohne Strassen wäre unsere Gesellschaft eine andere. Es könnte sein, dass deshalb aus Dankbarkeit an Strassen Namen vergeben wurden, wie das Menschen gerne mit Dingen tun, die sie mögen. Wahrscheinlich war es jedoch eine zweckmässige Massnahme, um im immer dichter werdenden Netz von Strassen nicht die Orientierung zu verlieren.

Wie alle Namen sagen auch Strassennamen etwas über die Namensgeber aus. Es wird deshalb davon ausgegangen, dass Strassennamen Rückschlüsse über die namensgebende Gesellschaft und deren Wertesystem zulassen. (Bancilhon, et al., 2021 S. 2)

Motivation

In der Schweiz gibt es unzählige Dorfstrassen, Bahnhofstrassen oder Hauptstrassen. Viele Strassen tragen Namen von geografischen Orten, die sie erschliessen, oder von bekannten Objekten, die daran liegen. Einige Strassen erhalten die Ehre, bewusst nach etwas oder sogar nach einer Person benannt zu werden.

Das Amtliche Verzeichnis der Strassen der Schweiz listet rund 221'000 Einträge mit entsprechendem Namen auf. Welche Namen wurden den Strassen vergeben? Lässt sich ein geografisches Muster erkennen? Diesen Fragen geht diese Zertifikatsarbeit nach.

Für die Antwortfindung spielen sowohl die Datenquellen als auch die Werkzeuge, mit welchen die Daten bearbeitet werden, eine wichtige Rolle. Ein Aspekt dieser Arbeit ist deshalb aufzuzeigen, welche Python Programmbibliotheken für die Datenaufbereitung und Auswertung verwendet werden können.

Betreuung

Diese Zertifikatsarbeit wurde an der Hochschule für Architektur, Bau und Geomatik der Fachhochschule Nordwestschweiz (FHNW) im CAS Spatial Data Analytics erstellt. Die Betreuung erfolgte durch Herrn Martin Christen der FHNW.

Inhaltsverzeichnis

	Zusammenfassung	1
	Vorwort	2
	Motivation	2
	Betreuung	2
1	Einleitung	7
1.1	Ausgangslage	7
1.2	Aufgabenanalyse	8
1.3	Zieldefinition	8
1.4	Zielpublikum	9
2	Vorgehen	10
2.1	Business understanding	10
2.2	Data understanding	10
2.2.1	Aufbau Odonym	10
2.2.2	Amtliches Verzeichnis der Strassen	11
2.2.3	Wikidata	13
2.2.4	Strassennamen Kanton Basel-Stadt	14
2.2.5	Strassennamenverzeichnis Stadt Zürich	15
2.2.6	OpenStreetMap	16
2.2.7	Kataloge	16
2.3	Data preparation	16
2.3.1	Data Cleaning	17
2.3.2	Extraktion generische Strassenbenennung	17
2.3.3	Extraktion spezifische Strassenbenennung	18
2.3.4	Verifikationsdatensatz	19
2.3.5	Weitere Verifikationsansätze	20
2.4	Modelling	21
2.4.1	Wahl und Definition der Kategorien	22
2.4.2	Wikidata	23
2.4.3	Zuordnen anhand von Katalogen	24
2.4.4	Named-Entity Recognition	24
2.4.5	Selbstkartierung	24
3	Resultate	25
3.1	Ablage GitHub	25
3.2	Verteilung und Statistiken	26
3.2.1	Generische Benennung	27
3.2.2	Spezifische Benennung generell	31
3.2.3	Ortschaften	35
3.2.4	Personen	36

3.2.5	Lebewesen	40
3.2.6	Berge	42
3.2.7	Gewässer	43
3.2.8	Berufe	45
3.3	Evaluation	46
4	Diskussion	48
4.1	Auswertung Strassennamen	48
4.2	Pandas / Geopandas	48
4.3	SPARQLWrapper	49
4.4	SpaCy	49
4.5	Pandarallel	50
4.6	Erstellung von Graphiken und Statistiken	50
5	Fazit und Ausblick	51
6	Glossar	52
6.1	Odonym	52
6.2	Etymologie	52
7	Quellenverzeichnis	53
8	Anhang	54
8.1	Übersicht Kapitel Notebook und Autor	54
8.2	Wikidataquery: Personen mit Bezug zur Schweiz	56
8.3	Wikidataquery: Strassen mit Statement benannt nach	56
8.4	Wikidataquery: Label von wikidataobjekten	56
8.5	Wikidataquery: Familienname von Personen	56
8.6	Wikidataquery: Label von allen Objekten	56
8.7	Wikidataquery: Abfrage über Suche	57

Abbildungsverzeichnis

Abbildung 1 The data mining life cycle (IBM Corporation, 2021)	10
Abbildung 2 Wortteile Odonym	11
Abbildung 3 Anzahl Strassen pro Kanton	12
Abbildung 4 Beispiele Wikidataobjekte mit «instance of» Statements	13
Abbildung 5 Ablauf der Zuweisung einer Kategorie an ein Odonym	13
Abbildung 6 Strassenobjekte in Wikidata mit Statement «named after» und einer geografischen Koordinate	14
Abbildung 7 Darstellung Zusammenhang spezifische Strassenbenennungen zu Kategorie	14
Abbildung 8 Übersicht der OSM-Strassen mit Informationen zur Etymologie	16
Abbildung 9 Übersicht der Aufbereitung der Strassennamen des Amtlichen Verzeichnis der Strassen	17
Abbildung 10 Übersicht des Modells zur Kategorisierung der spezifischen Strassenbenennungen	21
Abbildung 11 Visualisierung aller Strassen des Amtliches Verzeichnis der Strassen	25
Abbildung 12 Verteilung der amtlichen Strassen	26
Abbildung 13 Dichte des Netzes der benannten Strassen	27
Abbildung 14 Häufigkeit der generischen Strassenbezeichnungen	28
Abbildung 15 Durchschnittliche Strassenlänge nach generischer Bezeichnung	28
Abbildung 16 Top generische Strassenbezeichnung pro Gemeinde	29
Abbildung 17 Relative Häufigkeit der generischen Strassenbezeichnung pro Kanton	30
Abbildung 18 Visualisierung der generischen Strassenbezeichnung der Stadt Genf	30
Abbildung 19 Visualisierung der generischen Strassenbezeichnung der Stadt Zürich	31
Abbildung 20 Top spezifische Strassenbezeichnung pro Gemeinde	32
Abbildung 21 Häufigkeit der spezifischen Strassenbezeichnungen	32
Abbildung 22 Relative Häufigkeit der Namenskategorie pro Kanton	33
Abbildung 23 Anteil nicht klassifizierte Strassennamen pro Gemeinde	34
Abbildung 24 Verteilung der Strassen bekannt nach bekannten Ortschaften	35
Abbildung 25 Anteil der nach Ortschaften benannten Strassen pro Gemeinde	36
Abbildung 26 Anteil der nach Personen benannten Strassennamen pro Gemeinde	37
Abbildung 27 LISA-Cluster der nach Personen benannten Strassennamen pro Gemeinde	38
Abbildung 28 Nach Personen benannte Strassen in Zürich nach Geschlecht	39
Abbildung 29 Anteil Strassennamen benannt nach Lebewesen pro Gemeinde	40

Abbildung 30 LISA-Cluster der nach Lebewesen benannten Strassennamen pro Gemeinde	41
Abbildung 31 Verteilung der Strassen benannt nach den bekanntesten Bergen	42
Abbildung 32 LISA-Cluster der nach Bergen benannten Strassennamen pro Gemeinde	43
Abbildung 33 Verteilung der Strassen benannt nach den bekanntesten Flüssen	44
Abbildung 34 LISA-Cluster der nach Gewässern benannten Strassennamen pro Gemeinde	44
Abbildung 35 Verteilung der Strassen benannt nach Berufen	45
Abbildung 36 Konfusionsmatrix zur Evaluation des Modells der Kategorisierung der spezifischen Strassenbenennung	46

Tabellenverzeichnis

Tabelle 1 Beispiele Aufbau von Odonymen	11
Tabelle 2 Informationen über Amtliches Verzeichnis der Strassen	12
Tabelle 3 Informationen über Wikidata	13
Tabelle 4 Informationen zum Datensatz Strassennamen des Kanton BS	15
Tabelle 5 Informationen zum Datensatz Strassennamenverzeichnis Stadt Zürich	15
Tabelle 6 Auszug auf den bereinigten Strassennamen	17
Tabelle 7 Auszug auf den extrahierten generischen Bezeichnungen	18
Tabelle 8 Auszug auf den extrahierten spezifischen Bezeichnungen	19
Tabelle 9 Auszug auf den Resultaten der Lemmatisierung	19
Tabelle 10 Übersicht Kategorien der spezifischen Strassenbenennungen	23
Tabelle 11 Kennzahlen zur Performance des Modells zur Vorhersage des spezifischen Strassenkategorie	47

1 Einleitung

Strassennamen spiegeln die Kultur eines Landes wider und können als Stellvertreter für die historischen und kulturellen Merkmale des Landes verwendet werden.¹ (Bancilhon, et al., 2021 S. 14)

Anhand dieses Zitats ist es denkbar, dass auch die Strassennamen der Schweiz Rückschlüsse auf das Wertesystem der Gesellschaft in der Schweiz zulassen. Ein Versuch der Kategorisierung und Auswertung der Strassennamen bietet sich deshalb an.

1.1 Ausgangslage

Im März 2020 lancierte *Open Knowledge Belgium* die Plattform equalstreetnames.eu. Das Ziel dieses Projekts ist aufzuzeigen, wie viele Strassen in Brüssel nach Frauen und Männern benannt sind. Dafür wurden Datenobjekte von Wikidata mit den Strassengeometrien in OpenStreetMap verknüpft. Diese Arbeiten wurden manuell in einem Crowdsourcing-Projekt durchgeführt.² Unterdessen sind auf der Plattform equalstreetnames.eu auch die Städte Basel, Bern, Lausanne, St. Gallen und Zürich ersichtlich.³ Wobei die Daten der Stadt Basel noch nicht vollständig aufbereitet sind.⁴

Die Studie «Streetonomics: Quantifying Culture Using Street Names» (Bancilhon, et al., 2021) hat im Jahr 2021 die Vergabe von Strassennamen in London, Paris, Wien und New York untersucht. Dabei wurde, wie in equalstreetnames.eu auch, die Verteilung der Strassennamen an männliche und weibliche Personen beschrieben. Zudem wurde die Daten so aufbereitet, dass ersichtlich ist, welcher Nationalität die Personen angehörten und welchem Beruf sie nachgingen. Eine grafische Übersicht ist unter folgendem Link abrufbar: <http://social-dynamics.net/streetonomics/>⁵.

Beide Projekte betrachten nur Strassennamen, die nach Personen benannt sind und auch nur im städtischen Gebiet. Es gibt zwar in OpenStreetMap weitere Objekte, die mit dem entsprechenden Schlüssel `name:etymology:wikidata`⁶ getagt sind. In der Schweiz ist dieser Tag jedoch in keiner Region flächendeckend erfasst und bis jetzt gab es auch keine Bestrebungen, diesen systematisch zu erfassen. Eine Übersicht über die Elemente in OpenStreetMap mit einem bestehenden `name:etymology:wikidata` Schlüssel ist etwa auf der Webseite der Open Etymology Map⁷ ersichtlich.

Grundsätzlich stellt sich die Frage, woher weiss man überhaupt nach welchem Objekt eine Strasse benannt ist. In der Schweiz ist das Bundesamt für Landestopografie swisstopo verpflichtet, das Amtliche Verzeichnis der Strassen zu führen (GeoNV S. Art 26a Abs 1). Die effektive Strassenbenennung ist jedoch in der Hoheit der Kantone (GeoNV S. Art 26 Abs 1f). In vielen Kantonen sind die Gemeinden für die Festlegung der Strassennamen zuständig, z.B. im Kanton Zürich (KVAV-ZH S. Art 33 Abs 1c) oder Kanton Bern (SV-BE S. Art 4 Abs 1). Somit enthält zwar das Amtliche Verzeichnis der Strassen sämtliche Strassennamen der Schweiz, die Bedeutung des Namens muss jedoch bei den einzelnen Gemeinden abgeklärt werden.

Zwei Gemeinden der Schweiz haben die Benennung der Strassen als Open Government Data veröffentlicht. Dies sind:

¹ Im Original: «street names reflect a country's culture and can be used as proxies for the country's historical and cultural characteristics»

² <https://github.com/EqualStreetNames/equalstreetnames-brussels>. Abgerufen am 06. November 2022

³ <https://github.com/EqualStreetNames/equalstreetnames>. Abgerufen am 11. November 2022

⁴ Stand 11. November 2022

⁵ Abgerufen am 25. November 2022

⁶ <https://wiki.openstreetmap.org/wiki/Key:name:etymology:wikidata>. Abgerufen am 11. November 2022

⁷ <https://etymology.dsantini.it>. Abgerufen am 11. November 2022

- Stadt Basel. Datensatz: «Strassennamen»⁸. Dieser deckt zwar den ganzen Kanton Basel-Stadt ab, es sind allerdings nur auf dem Gebiet der Stadt Basel Erklärungen zu den Strassennamen enthalten.
- Stadt Zürich. Datensatz: «Strassennamenverzeichnis»⁹. Die Informationen sind auch in der «Strassennamen-Datenbank der Stadt Zürich»¹⁰ abrufbar.

Weiter unterhält der Kanton Genf die Webapplikation «Noms géographiques»¹¹. Die Lizenz dieser Daten ist jedoch unklar. Zudem sind die Daten nicht über eine Programmierschnittstelle (API) beziehbar.

In diversen Medien wurde das Thema der Strassennamen ebenfalls aufgegriffen. Meistens handelte es sich dabei um die Betrachtung einzelner Strassennamen, wie zum Beispiel in einer Sendung von Radio SRF 1 über die angeblich verrücktesten Strassennamen (Kohler, Barbara; Schweizer Radio und Fernsehen, 2016). Eine ausführlichere Analyse hat watson durchgeführt (Senn, Lea; Fehr, Reto; watson % FixxPunkt AG, 2022). Darin wird etwa aufgeführt, welche die häufigsten Strassennamen oder die häufigsten Endungen der Strassennamen sind. Zusätzlich wird auch die Anzahl der Strassennamen nach Kategorien wie Orte, Berge, Bäume, Tiere usw. aufgelistet. Die Statistiken betrachten dabei vor allem die Anzahl der entsprechenden Strassennamen und nehmen nur in einem begrenzten Umfang eine geografische Analyse vor.

1.2 Aufgabenanalyse

Es gibt zu den Strassen der Schweiz nur sehr begrenzt systematische Auswertungen zur Herkunft des Namens. Diese beschränken sich auf die fünf Städte, welche in Kapitel 1.1 «Ausgangslage» beschrieben sind. Dabei werden jedoch bei allen nur Strassennamen mit Bezug zu Personen betrachtet. Es stellt sich somit die Frage, ob eine Auswertung über die ganze Schweiz möglich ist, und welche Werkzeuge diese am besten unterstützen.

Eine Schwierigkeit liegt darin, dass zu den Namen in den allermeisten Fällen keine gesicherten Informationen über deren Herkunft vorliegen. Somit muss aus den Namen eine Vermutung abgeleitet werden.

Die OGD-Datensätze der beiden Städte Basel und Zürich, sowie die in OpenStreet-Map bereits aufbereiteten Strassen, können für eine Verifikation der Auswertung in Betracht gezogen werden.

1.3 Zieldefinition

Die Strassennamen aus dem Amtliches Verzeichnis der Strassen der Schweiz werden analysiert. Die Analyse soll eine Auswertung der Strassennamen im Allgemeinen beinhalten. Diese gibt Auskunft über die generische Benennung der Strassennamen wie *strasse*, *weg*, *rue*, *via*, usw. in Bezug auf die Lage und Länge der Strassen.

Ausserdem soll die mögliche Herkunft der Namen im Speziellen analysiert werden. Als Definition der Herkunft werden sinnvolle Kategorien gebildet, wie zum Beispiel Personen, Berge oder Gewässer. Diese werden ebenfalls in Bezug zur Lage ausgewiesen.

Es gilt zudem herauszufinden, welche Pythonbibliotheken eine solche Auswertung möglichst effektiv unterstützen.

⁸ <https://data.bs.ch/explore/dataset/100189/information/>. Abgerufen am 12. November 2022

⁹ https://data.stadt-zuerich.ch/dataset/geo_strassennamenverzeichnis. Abgerufen am 12. November 2022

¹⁰ stadt-zuerich.ch/strassennamen-datenbank. Abgerufen am 12. November 2022

¹¹ <https://ge.ch/noms-geographiques/>. Abgerufen am 12. November 2022

1.4 Zielpublikum

Diese Arbeit demonstriert den Umgang von geografischen Daten mit Hilfe von Pythonbibliotheken. Die Daten werden explorativ begutachtet und ausgewertet. Somit richtet sie sich in erster Linie an Fachleute, die sich mit Spatial Data Mining oder der Auswertung geografischer Daten auseinandersetzen. Weiter werden Strassennamen ausgewertet, deren Resultate für alle Personen interessant sein können, da sie sich mit der Herkunft von Strassennamen und deren räumlichen Auswertung in der Schweiz beschäftigen.

2 Vorgehen

Um die benötigten Informationen aus den unterschiedlichen Datensätzen in einem strukturierten Vorgehen gewinnen zu können, orientiert sich der Ablauf dieser Arbeit dem *Cross-Industry Standard Process for Data Mining (CRISP-DM)* Modell. Das Modell beschreibt ein Vorgehen in sechs Phasen (IBM Corporation, 2021), deren Ablauf in der Abbildung 1 dargestellt ist.

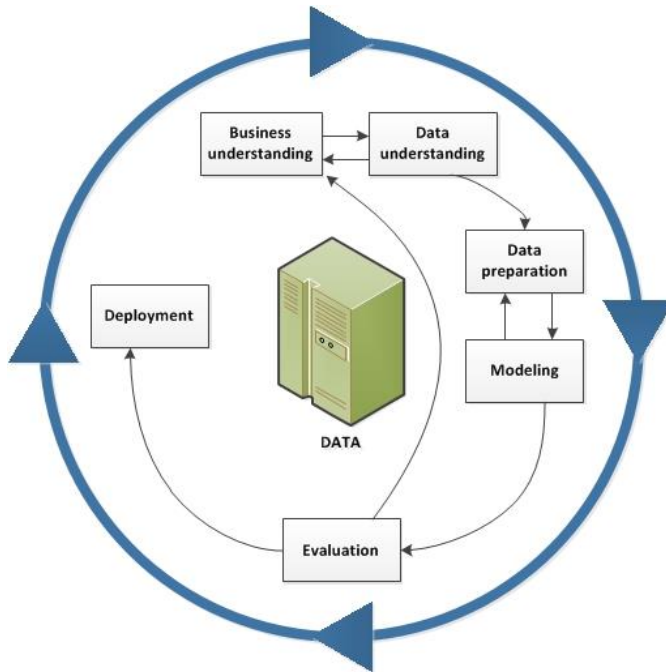


Abbildung 1 The data mining life cycle (IBM Corporation, 2021)

Die folgenden Kapitel sind analog diesen Phasen gegliedert, obwohl der effektive Ablauf auch iterative Schritte enthält.

2.1 Business understanding

Eine Situationsanalyse sowie ein Projektplan wurden vorgängig zu den Hauptarbeiten dieser Arbeit in einer Projektskizze (Bühler, et al., 2022) aufgezeigt. Die Ziele der Datenanalyse sind im Kapitel 1.2 «Aufgabenanalyse» dargestellt.

Zu den Risiken im Business-Kontext kann der unsichere Zeitaufwand für ein einigermaßen aussagekräftiges Resultat genannt werden. Es liess sich vor der ersten Analyse der Daten kaum abschätzen, wie gut eine automatische Analyse der Strassennamen funktionieren könnte. So kann etwa gemäss (Brunner, et al., 2008 S. 25) die Mehrdeutigkeiten nur schon von Toponymen bei über 50% liegen. Dies lässt für Strassennamen auf eine tiefe Quote bei der Bestimmung der Herkunft des Namens deuten.

2.2 Data understanding

Die in Kapitel 1.1 «Ausgangslage» aufgelisteten Datensätze weisen sehr viele Daten auf. In den folgenden Kapiteln ist eine Übersicht über die Datensätze sowie deren Datenmodelle und Daten enthalten.

2.2.1 Aufbau Odonym

Odonym ist der Fachbegriff für Strassennamen. Beide Begriffe werden hier gleichwertig verwendet. Odonyme sind meist aus mehreren Wortteilen aufgebaut. Nachfolgend sind die, in dieser Arbeit verwendeten, Definitionen dieser Wortteile beschrieben.

Viele Odonyme in der Schweiz bestehen aus einer spezifischen und generischen Strassenbenennung. Am Beispiel von «Riggenbachstrasse»:

- Riggenbach = spezifische Strassenbenennung
- strasse = generische Strassenbenennung

Zusätzlich können Präpositionen, Stopwörter und spezielle Zeichen in einem Odonym enthalten sein.

Als Präpositionen sind Zusätze wie etwa *Alt, Hinter, Gross, Mont* usw. zu verstehen

Stopwörter sind Füllwörter wie *am, Im, Bei, de, aux, sur, della* usw.

Zu den speziellen Zeichen zählen etwa Bindestriche, Apostroph oder der Punkt, welcher oft bei der Abkürzung von *St.* vorkommt. Ausserdem wurden bei den Odonymen gewisse diakritische Zeichen entfernt, um den Abgleich über die verschiedenen Datensätze zu vereinfachen. Dadurch entstanden aus *ï* ein *i* oder aus *è* ein *e*.

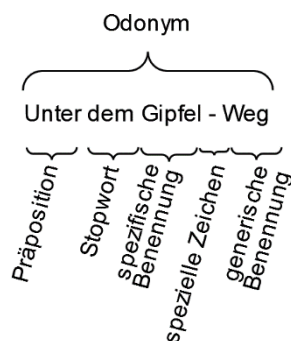


Abbildung 2 Wortteile Odonym

Folgend eine paar Beispiele:

Odonym	spezifisch	generisch	Präposition	Stopwort
Riggenbachstrasse	Riggenbach	strasse	--	--
St. Hilarienweg	Hilarien	Weg	Sankt	--
Veia San Martin	Martin	Veia	Sankt	--
Impasse Saint-Sébastien	Sebastien	Impasse	Sankt	--
Sous-l'Eglise	Eglise	--	Unter	l
Unterer Scharanserweg	Scharanser	weg	Unter	--
Chemin du Grand-Bois	Bois	chemin	gross	du

Tabelle 1 Beispiele Aufbau von Odonymen

2.2.2 Amtliches Verzeichnis der Strassen

Das amtliche Verzeichnis der Strassen enthält alle Strassen, Plätze, benannten Gebiete, Passagen etc., die Bestandteil einer oder mehrerer offizieller Gebäudeadressen sind, und/oder die auf einem Stadtplan oder einem Geoportal erscheinen. Das Verzeichnis der Strassen basiert ursprünglich auf den Daten der amtlichen Vermessung. (Bundesamt für Landestopografie swisstopo, 2021 S. 4)

Das «Amtliches Verzeichnis der Strassen» dient als Grundlage. Daraus werden die Strassennamen extrahiert, welche für die Analyse verwendet werden.

Anzahl Datensätze	Ca. 221'000 ¹² . Davon analysiert: ca. 172'000
Bezug	https://www.cadaastre.ch/de/services/service/register/street.html ¹³
Format	Unter anderem: csv und gdb
Lizenz	Nutzungsbedingungen für kostenlose Geodaten und Geodienste (OGD) von swisstopo

Tabelle 2 Informationen über Amtliches Verzeichnis der Strassen

Folgende Attribute wurde verarbeitet:

- STR_ESID: eindeutiger Identifikator
- COM_NAME: zugehöriger Gemeindennamen
- COM_CANTON: zugehöriger Kantonsname (Abkürzung)
- STR_TYPE: Typ der Strasse
- STN_LABEL: Strassenname
- STR_EASTING und STR_NORTHING: Punktkoordinate

Eine Auswertung über die Anzahl Strassen pro Kanton ist in Abbildung 3 dargestellt. Diese zeigt die Anzahl Strassen pro Kanton des kompletten Datensatzes. In den beiden bevölkerungsreichsten Kantonen Bern und Zürich liegen die meisten Strassen. Eher überraschend kommt auf Rang drei St. Gallen, welcher rund 40% mehr Strassen als der nächstfolgende Kantone Aargau aufweist. Der Grund für diese Tatsache ist im Kapitel 3.2 beschrieben. Interessant scheint auch, dass in den fünf Kantonen mit den meisten Strassen mehr Strassen liegen als in den restlichen 21.

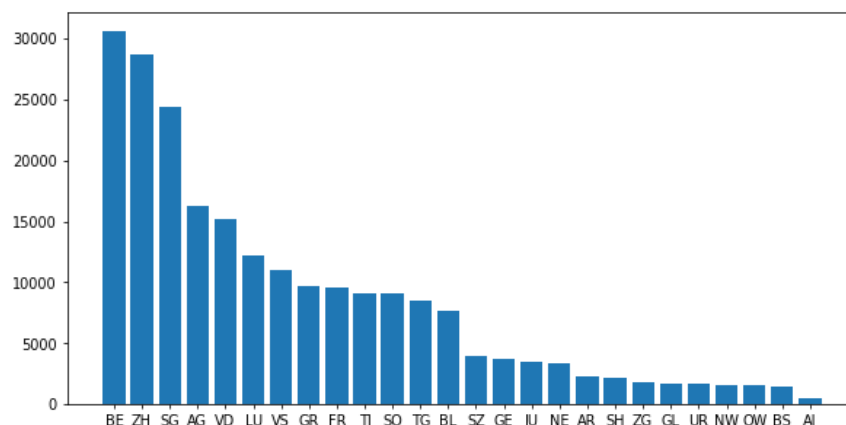


Abbildung 3 Anzahl Strassen pro Kanton

Um die Analyse aus zeitlichen Gründen möglichst einfach zu gestalten, wurde nicht der komplette Datensatz verwendet, sondern nur Elemente vom Typ STR_TYPE == 'Street'. Somit reduziert sich der effektiv analysierte Datensatz auf rund 172'000 Elemente.

¹² Stand: November 2022

¹³ Abgerufen am 18. November 2022

2.2.3 Wikidata

Wikidata ist eine freie, gemeinsame, mehrsprachige und sekundäre Datenbank zur Sammlung strukturierter Daten.¹⁴

Wikidata ist eine grosse, strukturierte Datenbank mit dem Schwerpunkt enzyklopädischen Inhalt.¹⁵ Zu Wikidata gehört neben der Datenbank der «Wikidata Query Service» (WDQS), ein SPARQL-Endpunkt für Abfragen der Datenbank¹⁶. Der Endpunkt ist über eine API automatisiert abfragbar.

Anzahl Datensätze	Ca. 100Mio ¹⁷ . Davon ca. 42'000 Personen mit Bezug zur Schweiz ¹⁸ .
Bezug	https://query.wikidata.org/ ¹⁹
Format	Webservice
Lizenz	Creative Commons Public Domain Dedication 1.0

Tabelle 3 Informationen über Wikidata

Strukturierte Datenbanken sind so aufgebaut, dass sie sowohl für Menschen wie Maschinen einfach zu verarbeiten sind. Das Datenmodell besteht im Wesentlichen aus Objekten, die über Statements beschrieben sind. Viele Objekte weisen über ein Statement *instance of* aus. Einige Beispiele:

Objekt Label	Objektnummer	Instance of Label	Instance of
Rigi	Q15291	mountain	Q8502
Birke	Q25243	taxon	Q16521
Niklaus Riggenbach	Q115633	human	Q5
Olten	Q68129	municipality of Switzerland	Q70208

Abbildung 4 Beispiele Wikidataobjekte mit «instance of» Statements

Mit Hilfe von Wikidata können somit die aus den Odonymen extrahierten spezifischen Strassenbenennungen einer Kategorie, einer Instance, zugewiesen werden. Der Ablauf dazu ist in Abbildung 5 dargestellt.

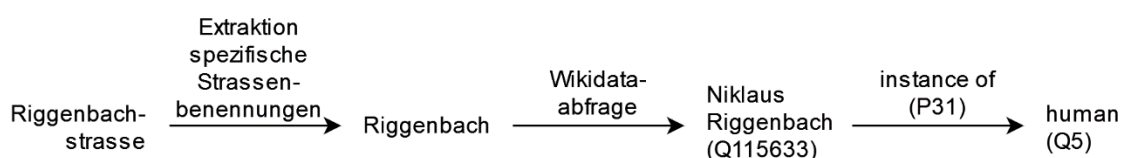


Abbildung 5 Ablauf der Zuweisung einer Kategorie an ein Odonym

Eine Herausforderung ist die Begrenzung der Abfragen über die API, welche auf 60sec Prozessierungszeit pro Minute limitiert ist.²⁰ Eine Abfrage aller 172'000 spezifischen Strassenbenennungen würde deshalb mindestens 48h dauern. Ein Problem

¹⁴ <https://www.wikidata.org/wiki/Wikidata:Introduction/de>. Abgerufen am 18. November 2022

¹⁵ https://www.wikidata.org/wiki/Help:About_data/de. Abgerufen am 18. November 2022

¹⁶ https://www.mediawiki.org/wiki/Wikidata_Query_Service. Abgerufen am 18. November 2022

¹⁷ <https://www.wikidata.org/wiki/Special:Statistics>. Abgerufen am 18. November 2022

¹⁸ Personen mit Geburts- oder Todesort in der Schweiz. Query ersichtlich im Anhang «Wikidataquery: Personen mit Bezug zur Schweiz». Abgerufen am 18. November 2022

¹⁹ Abgerufen am 18. November 2022

²⁰ https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual#Query_limits. Abgerufen am 18. November 2022

stellen auch die Datensätze insgesamt dar. So können spezifische Strassenbenennungen nur einem Wikidataobjekt zugewiesen werden, wenn dieses auch in Wikidata erfasst ist. Erste explorative Analysen haben gezeigt, dass zum Beispiel die Personen, welche namensgebend für die Daniel Fechter-Promenade und Emil Angst-Strasse in Basel stehen, nicht in Wikidata vorhanden sind.²¹

In Wikidata sind bereits rund 2'500 Strassen der Schweiz erfasst. Davon weisen 641 über ein Statement *named after* auf.²² Wovon mehr als 600 über eine Koordinate verfügen. Diese Strassen sind in Abbildung 6 rot markiert.



Abbildung 6 Strassenobjekte in Wikidata mit Statement «named after» und einer geografischen Koordinate

Der Zusammenhang eines Strassenobjekts zu einer namensgebenden Kategorie über *named after* und *instance of* ist in Abbildung 7 ersichtlich.



Abbildung 7 Darstellung Zusammenhang spezifische Strassenbenennungen zu Kategorie

Da die Strassenobjekte in Wikidata bis jetzt keine ESID aufweisen, ist ein Abgleich mit den Elementen im Amtlichen Verzeichnis der Strassen nicht ohne weiteres möglich. Zudem sind fast ausschliesslich Objekte in den Städten Genf und Zürich vorhanden, welche zu einem grossen Teil nach Personen benannt sind. Es wurde deshalb darauf verzichtet, diese Daten etwa für eine Verifikation zu benutzen, da sie nur eine sehr spezifische Aussage zulassen.

2.2.4 Strassennamen Kanton Basel-Stadt

Der Datensatz beinhaltet die Namen aller Strassen im Kanton Basel-Stadt sowie kurze Erläuterungen zur Bedeutung der Strassennamen der Stadt Basel. Die Felder zu den Erklärungen sind ausserhalb des Stadtgebietes leer.²³

²¹ Stand 18. November 2022

²² Query ersichtlich im Anhang «Wikidataquery: Strassen mit Statement benannt nach Wikidataquery: Personen mit Bezug zur Schweiz». Abgerufen am 18. November 2022

²³ <https://data.bs.ch/explore/dataset/100189/information/>. Abgerufen am 19. November 2022

Anzahl Datensätze	1442 ²⁴
Bezug	https://data.bs.ch/explore/dataset/100189/ ²⁵
Format	Unter anderem: csv
Lizenz	CC BY 3.0 CH + OpenStreetMap

Tabelle 4 Informationen zum Datensatz Strassennamen des Kanton BS

Das Verzeichnis der Strassennamen des Kanton Basel-Stadt ist einer der wenigen OGD-Datensätze in der Schweiz, welche Informationen zu den Strassennamen beinhaltet (siehe auch Kapitel 1.1 Ausgangslage).

Die Informationen über die Strassennamen sind in den zwei Feldern *Erklärung erste Zeile* und *Erklärung zweite Zeile* in Prosa enthalten. Eine automatisierte Zuweisung an ein namengebendes Objekt ist deshalb nicht einfach zu bewerkstelligen. Der Datensatz wurde dennoch für eine Aufbereitung als Verifikationsdatensatz in Betracht gezogen und entsprechend aufbereitet.

2.2.5 Strassennamenverzeichnis Stadt Zürich

Die Strassennamen der Stadt Zürich werden in einer laufend aktualisierten Datenbank geführt.

Anzahl Datensätze	2546 ²⁶
Bezug	https://data.stadt-zuerich.ch/dataset/geo_strassennamen-verzeichnis/ ²⁷
Format	Unter anderem: json und wfs
Lizenz	Creative Commons CCZero

Tabelle 5 Informationen zum Datensatz Strassennamenverzeichnis Stadt Zürich

Es sind alle der über 2'500 offiziell benannte Strassen aufgeführt. Interessant ist insbesondere, dass dieser Datensatz neben den Strassennamen auch Erläuterungen zu ihrer Herkunft und Bedeutung enthält. Dieser Datensatz wurde in dieser Arbeit zur Erstellung eines Verifikationsdatensatzes verwendet.

²⁴ Abgerufen am 18. November 2022

²⁵ Abgerufen am 23. November 2022

²⁶ Abgerufen am 18. November 2022

²⁷ Abgerufen am 23. November 2022

2.2.6 OpenStreetMap

Zurzeit besitzen rund 4'300 Schweizer Strassen in OpenStreetMap (OSM) den Wiki-data-Schlüssel *name:etymology:wikidata*. Dieser Schlüssel enthält Informationen zur Namensgebung der Strasse oder eben deren Etymologie. Die OSM-Daten wurden via Overpass API²⁸ bezogen. Abbildung 8 zeigt die räumliche Verteilung dieser Strassen. Der Grossteil befindet sich in den grossen Städten Zürich, Basel, Lausanne, Bern, Luzern und St. Gallen sowie in der Region Yverdon.

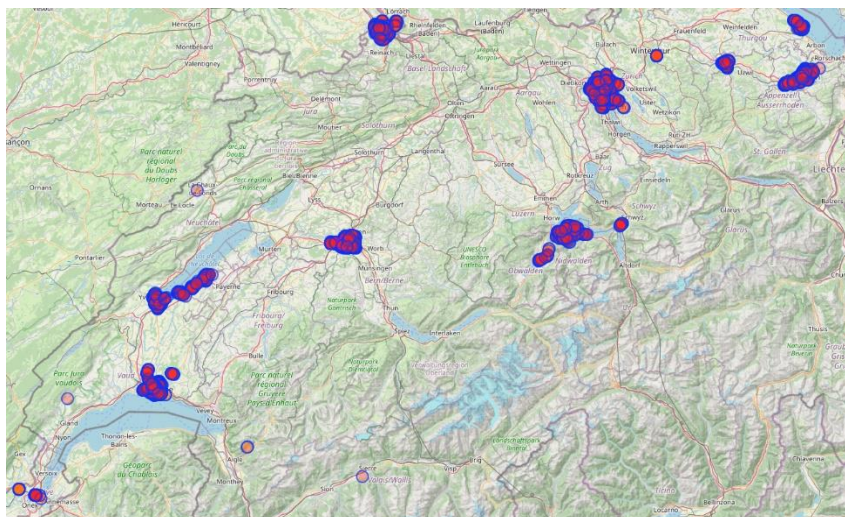


Abbildung 8 Übersicht der OSM-Strassen mit Informationen zur Etymologie

2.2.7 Kataloge

In dieser Arbeit wurden mehrere Übersichtslisten von unterschiedlichen Quellen verwendet. Es handelt sich dabei um Übersichtslisten zu den Themen Tiere²⁹, Pflanzen³⁰ und Berufen³¹ des freien Wörterbuchs Wiktionary, um Daten zum Gewässernetz der Schweiz des Bundesamts für Umwelt³² und um Listen zu den Schweizer Bergen³³ und Pässen³⁴ der freien Enzyklopädie Wikipedia.

2.3 Data preparation

Bei den meisten Strassennamen kann zwischen einer generischen Benennung, wie *Strasse*, *Weg* oder *Avenue* und einer spezifischen Benennung, wie *Anton von Blarer*, *Nelken*, oder *Kirch* unterschieden werden. Da in dieser Arbeit die generischen sowie die spezifischen Benennungen analysiert werden, werden die Strassennamen in einen generischen und einen spezifischen Teil aufgeteilt (Abbildung 9). Dabei werden von den Strassennamen zuerst die generischen Benennungen, anhand einer Liste der gängigsten Benennungen, extrahiert (z.B. *Ruelle* aus *Ruelle de l'Eglise*). Vom verbleibenden Teil, z.B. *de l'Eglise*, werden möglichst viele nicht relevante Wörter oder Zeichen entfernt, damit nur noch die spezifische Strassenbenennung, z.B. *Eglise* übrig

²⁸ <https://overpass-turbo.osm.ch/>. Abgerufen am 12. Oktober 2022

²⁹ <https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Tiere>. Abgerufen am 13. Oktober 2022

³⁰ <https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Pflanzen>. Abgerufen am 13. Oktober 2022

³¹ <https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Berufe>. Abgerufen am 13. Oktober 2022

³² <https://www.bafu.admin.ch/bafu/de/home/themen/wasser/zustand/karten/karten-und-abgeleitete-daten/gewaessernetz-der-schweiz.html>. Abgerufen am 13. Oktober 2022

³³ https://de.wikipedia.org/wiki/Liste_von_Bergen_in_der_Schweiz. Abgerufen am 13. Oktober 2022

³⁴ https://de.wikipedia.org/wiki/Liste_der_P%C3%A4sse_in_der_Schweiz. Abgerufen am 13. Oktober 2022

bleibt. Die generische Benennung wird ohne Weiterverarbeitung analysiert. Die spezifische Benennung wird zur Zuordnung eines Strassennamens zu einer Kategorie, z.B. Person, Lebewesen oder Gewässer, verwendet (vgl. 2.4 Modelling).

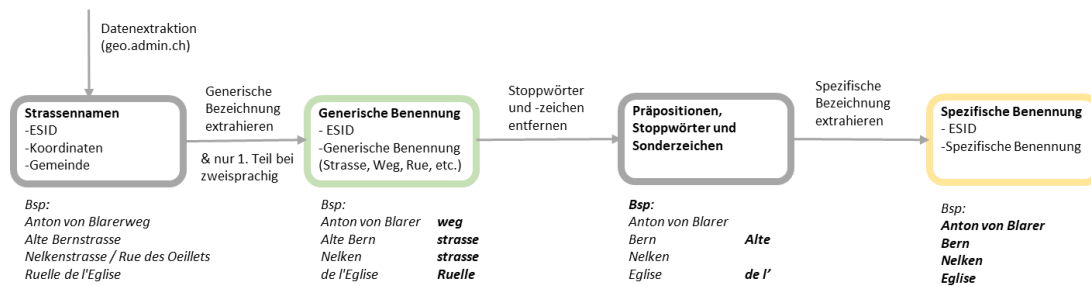


Abbildung 9 Übersicht der Aufbereitung der Strassennamen des Amtlichen Verzeichnis der Strassen

2.3.1 Data Cleaning

Vom gesamten Datensatz der Amtlichen Strassen (vgl. 2.2.2) wurden die Areale und Plätze entfernt, u.a. da erstere oft keine generische Strassenbezeichnung besitzen und die Geometrie im Gegensatz zu den Strassen Polygone sind. Zudem enthält der Datensatz rund 100 Strassen ohne Koordinaten, welche ebenfalls entfernt wurden. Bei den Strassennamen, Attribut *STN_LABEL*, wurden ebenfalls alle Leerschläge am Anfang und Ende des Namens entfernt. Ausserdem wurden in den Strassennamen alle Informationen nach einer Klammer oder einem Schrägstrich entfernt. Letzterer Schritt wurde durchgeführt, um mehrsprachige Strassennamen auf eine Sprache zu reduzieren und Zusatzinformationen in Klammern zu entfernen. Ein Beispiel zeigt Tabelle 6, wobei *STN_LABEL_NO_BI* die bereinigte Spalte ist.

STR_ESID	STN_LABEL	COM_FOSNR	COM_NAME	COM_CANTON	STR_OFFICIAL	STR_EASTING	STR_NORTHING	STN_LABEL_NO_BI
10022105	Chandosselweg / Chemin de Chandossel	2250	Courgevaux	FR	True	2574274.0	1194118.0	Chandosselweg
10244635	Figgione (nucleo)	5072	Faido	TI	True	2706220.0	1148322.0	Figgione

Tabelle 6 Auszug auf den bereinigten Strassennamen

2.3.2 Extraktion generische Strassenbenennung

Der Extraktion der generischen Strassenbenennung geht die manuelle Erstellung einer Liste der gängigsten generischen Strassenbenennungen der Schweiz voraus. Dazu wurden die Daten durchsichtet sowie eine ausführlichere Internetsuche durchgeführt. Anschliessend wurde getestet, wie jede dieser generischen Benennungen am besten gefunden werden kann. Die jeweiligen Testresultate wurden durch die räumliche Darstellung auf einer Schweizer Karte und durch Begutachten der gefundenen Strassennamen geprüft und wo nötig angepasst. Dabei mussten auch diverse reguläre Ausdrücke *Regex* verwendet werden. Z.B. der generische Strassenterm *rue* kann am besten via *rue\$* gefunden werden, da ohne das *\$*-Zeichen fälschlicherweise viele deutschsprachige Strassennamen wie z.B. *Chlei Brruederhölzli* gefunden werden. Testresultate die weniger als fünf fehlerhafte Strassennamen enthielten wurden in der Regel vernachlässigt.

Anschliessend wurde die Extraktion der generischen Benennung, mittels der *String-Contains-Methode* durchgeführt. Dabei wurde darauf geachtet, dass generische Bezeichnungen, welche ein Substring einer anderen generischen Bezeichnung sind, z.B. *weg* von *wegli* nicht doppelt extrahiert werden. Die Spalte *STR_TERMS* der Tabelle 7 zeigt einige Beispiele der Resultate dieser Extraktion

STR_ESID	STN_LABEL	STR_TERMS
10166111	Platzpromenade	promenade
10009703	In der Au	NaN
10098076	Rue du Sentier	sentier
10242098	14-15-Weg	weg
10062278	J.-C.-Heer-Strasse	strasse

Tabelle 7 Auszug auf den extrahierten generischen Bezeichnungen

2.3.3 Extraktion spezifische Strassenbenennung

Die Information der Strassennamen, welche nach der Extraktion der generischen Benennung noch übrigbleibt, wurde nun so weiterverarbeitet, dass möglichst nur noch die spezifische Bezeichnung dasteht. Die italienischen und romanischen Strassennamen werden für die Extraktion und alle darauffolgenden Teilprozesse nicht mehr verwendet. Zum einen wurden, bis auf die deutschen Umlaute, alle diakritischen Zeichen entfernt oder mit dem jeweiligen *Standardbuchstaben* ersetzt. Stoppwörter und Präpositionen wurden ebenfalls entfernt. Stoppwörter sind Wörter, welche häufig auftreten und gewöhnlich keine Relevanz für die spätere Kategorisierung der spezifischen Strassenbezeichnungen haben, respektive diese erschweren oder verhindern. Unter anderem wurden diverse Artikel wie *'der'* oder *'la'*, Präpositionen wie *'auf'*, oder *'Della'* oder auch Adjektive wie *'ancienne'* oder *'alt'* vom ursprünglichen Strassennamen entfernt. Einige Stoppwörter wie z.B. *'san'* wurden pro Strassenname gespeichert, da sie für die Kategorisierung der spezifischen Bezeichnungen wieder verwendet werden, jedoch für den String-Vergleich u.a. mit Wikidata entfernt wurden. Dabei wurde möglichst darauf geachtet, dass deutsche Stoppwörter nicht grössere Mengen an französischen spezifischen Benennungen entfernen und umgekehrt genauso. Z.B. der Strassenname *In der Au* wurde komplett entfernt. Es handelt sich dabei jedoch um Einzelfälle und sehr kurze Strings sind ausserdem häufig sehr ungeeignet für den späteren String-Vergleich. Tabelle 8 zeigt einige Beispiele dieses Prozesses.

STR_ESID	STN_LABEL	STR_TERMS	STN_LABEL_FINAL
10142970	Meisenstrasse	strasse	Meisen
10248748	Florence Schelling-Weg	weg	Florence Schelling
10154128	Alter Paradiesweg	weg	Paradies
10016257	Impasse de la Gare	impasse	Gare
10121463	Rue de la Tuilière	rue	Tuiliere
10023110	Z.I. d'In-Riaux	NaN	Z I Riaux
10087133	Avenue Jean-Marie-Musy	avenue	Jean Marie Musy
10150667	Veia da Reschen	veia	Reschen
10106374	Lischenweg / Chemin des Fléoles	weg	Lischen
10189084	Chräi	NaN	Chräi
10116516	Gässli	gässli	

Tabelle 8 Auszug auf den extrahierten spezifischen Bezeichnungen

Die Verwendung der Natural Language Processing NLP Bibliothek spaCy für die Extraktion der spezifischen Benennungen wurde getestet. Zum einen die Verwendung der darin enthaltenen Stoppwortlisten für diverse Sprachen, zum anderen die Lemmatisierung. Schliesslich wurde beides nicht in den Workflow integriert, da die Resultate der Kategorisierung nicht verbessert werden konnten. Durch die Lemmatisierung werden einige spezifische Strassennamen optimiert, wie z.B. *Häusern* zu *Haus*, andere wiederum verschlechtert oder verfälscht wie *Eisen* zu *Ei*. Tabelle 1 zeigt einige Beispiele der Lemmatisierung mittels SpaCy.

STR_ESID	STN_LABEL_FINAL	SPACY_DE_LEMMA
10140563	Emanuelen	Emanuel
10169078	Kelten	Kelt
10142970	Meisen	Meisen
10103851	Ringlichen	Ringliche
10108780	Eisen	Ei
10209758	Sonnhalde	Sonnhald
10013619	Moselen	Mosel
10116917	Häusern	Haus
10064127	Alpbachsäge	Alpbachsag
10119756	Truffersbrunnen	Truffersbrunn
10009440	Ins	in

Tabelle 9 Auszug auf den Resultaten der Lemmatisierung

2.3.4 Verifikationsdatensatz

Wie in Kapitel 2.1 «Business understanding» erwähnt, ist eine Analyse von Odonymen nicht einfach und fehlerbehaftet. Um eine Aussage über die Genauigkeit der prozessierten Daten machen zu können, wird ein Verifikationsdatensatz aus Strassen einschliesslich bekannter Etymologie benötigt. Dieser Verifikationsdatensatz wird auf Basis der OSM-Strassen mit Etymologie Tag (vgl. 2.2.6) und des Strassennamenverzeichnis der Stadt Zürich (vgl. 2.2.5) erstellt. Unser Verifikationsdatensatz, auch *Ground Truth* genannt, enthält 1'227 Einträge und wird folgendermassen erstellt.

Erstens wird für alle 4'300 Schweizer OSM-Strassen, welche den Etymologie-Tag besitzen, die mit ihnen verknüpfte Wikidata-Instanz bezogen (Gurtovoy, et al., 2022). Durch diesen, in Kapitel 2.2.3 beschriebenen, Mining Prozess erhalten wir für alle

4'300 Strassennamen etymologische Informationen, durch welche die Strassennamen in eine unserer spezifischen Benennungskategorien zugeordnet werden können. Die Zuordnung wird manuell durchgeführt. So werden z.B. alle Strassen mit der Wikidata-Instanz *Q16251* = *taxon* der Kategorie *Lebewesen* zugewiesen und entsprechen die Wikidata-Instanzen *Q11879590* = *female given name* der Kategorie *Person*. Nach dem Bereinigen der Daten, u.a. entfernen von Duplikaten, werden sie via String-Vergleich der Strassennamen mit den, von uns analysierten, Strassen des amtlichen Verzeichnisses verknüpft. Schlussendlich erhalten wir so 946 einzigartige Einträge.

Zweitens haben wir 281 zusätzliche Einträge durch Interpretation der Erläuterungen der Strassen des Strassennamenverzeichnis der Stadt Zürich generiert. So haben wir z.B. der Strasse mit der Erläuterung «*Das Land zwischen dem Seeufer und dem Bleicherweg wurde seit dem Mittelalter als städtische Leinwand-Bleichen genutzt. Hier wurden die Leinentücher ausgelegt, um in der Sonne gebleicht zu werden*» die Kategorie *Beruf* zugewiesen. Dabei haben wir darauf geachtet, dass wir nur Erläuterungen manuell klassifizieren, die nicht bereits im OSM-Teil enthalten sind und mit den Strassen des amtlichen Verzeichnisses verknüpft werden können.

2.3.5 Weitere Verifikationsansätze

Im Kanton Basel-Stadt sind in OpenStreetMap 102 Strassen erfasst, die nach Personen benannt sind.³⁵ Der Datensatz ist allerdings nicht vollständig, vergleiche Kapitel 1.1 «Ausgangslage». Es wurde deshalb versucht, die Daten in OpenStreetMap mit dem Datensatz der Strassennamen, vergleiche Kapitel 2.2.4 «Strassennamen Kanton Basel-Stadt», zu ergänzen. Aus den Feldern *Erklärung erste Zeile* und *Erklärung zweite Zeile* wurden 147 Strassen mit einem möglichen Bezug zu einer Person extrahiert. Die mögliche Datensätze wurden danach zusätzlich mit Wikidata abgeglichen, um das Resultat möglichst zu verbessern. Die Abfrage wurde zweistufig durchgeführt. Im ersten Schritt fand ein Vergleich des vorgängig extrahierten spezifischen Strassenbenennungen gegen das Label des Wikidataobjekts statt, um die Vor- und Nachnamen zu erhalten³⁶. Im zweiten Schritt wurde angenommen, dass die spezifische Strassenbenennung nur aus dem Nachnamen bestand und somit gegen das Statement *family name* (P734) verglichen³⁷.

Aus den 147 Strassen wurden schliesslich 103 identifiziert, die nach einer Person benannt sein könnten. Eine visuelle Überprüfung ergab einige Fehler zweiter Art (Falsch-Negativ). Zum Beispiel wurden folgende Strassen nicht erkannt:

- C.F. Meyer-Strasse. C.F. in Wikidata ausgeschrieben.
- General Guisan-Strasse. *General* kein Teil des Namens.
- Katja Wulff-Anlage. Generische Strassenbenennung *Anlage* wird durch einen Fehler in der Prozessierung nicht korrekt verarbeitet.
- Henric Petri-Strasse. Kein Eintrag in Wikidata.

Da die 103 erkannten Strassen keinen Mehrwert gegenüber den 102 bereits in OpenStreetMap erfassten Angaben brachten, wurde die erstellte Liste nicht weiterverwendet.

In einem weiteren Ansatz wurden versucht die spezifische Strassenbenennungen aus dem Amtlichen Verzeichnis der Strassen mit dem Wikidata Query Service zu analysieren. Da die Datenmenge des Verzeichnisses sehr umfangreich ist, wurden nur Strassen im Kanton Basel-Stadt ausgewertet, was 1'439 Strassen entspricht.

³⁵ <https://basel.equalstreetnames.eu>. Abgerufen am 19. November 2022

³⁶ Query ersichtlich im Anhang 8.4 «Wikidataquery: Label von wikidataobjekten»

³⁷ Query ersichtlich im Anhang 8.5 «Wikidataquery: Familienname von Personen»

Die Abfrage bestand darin, die spezifischen Strassenbenennungen gegen das deutschsprachige Label des Wikidataobjekts zu vergleichen.³⁸ Beim Abgleich wurden zu 789 spezifische Strassenbenennungen ein Wikidataobjekt gefunden. Davon wurden 62 als Menschen erkannt. Dies entspricht wahrscheinlich weniger als 50% aller nach Personen benannten Strassen in Basel, da davon auszugehen ist, dass zu den 102 Objekten in OpenStreetMap noch weitere bestehen. Bei den Resultaten zeigten sich zudem weitere Probleme bei der Abfrage gegen das Wikidata-Label:

- Plural werden nicht erkannt. Zum Beispiel: Nachtkerzenweg (Nachtkerzen), Backstubenweg. (Backstuben)
- Baslerstrasse (Basler), Zürcherstrasse (Zürcher), Bernerstrasse (Berner) usw.
- Kein Wikidataobjekt vorhanden: Weinlagerstrasse (Weinlager),
- Mehrdeutigkeiten:
 - Reichensteinerstrasse, Reichenstein ist eine Rebsorte und eine Burgruine
 - Spechtweg: Specht ist auch ein Familienname
 - Jurastrasse: Jura ist eine Periode der Erdgeschichte, ein Gebirge, ein Kanton, usw.

Um eine guten Verifikationsdatensatz zu gewinnen, hätte die Wikidata-Abfrage verbessert werden müssen. Darauf wurde aus zeitlichen Gründen verzichtet.

2.4 Modelling

In diesem Kapitel wird beschrieben, wie die in 2.3.3 extrahierten spezifischen Strassenbenennungen des amtlichen Verzeichnisses, automatisch kategorisiert werden. Die verwendeten Kategorien werden im Unterkapitel 2.4.1 beschrieben. Um die Modellierung etwas zu vereinfachen haben wir uns für folgende Prämisse entschieden: *Mehrfach vorhandene Spezifische Benennungen gehören derselben Kategorie an.* Dies ist selbstverständlich nicht immer der Fall, z.B. kann *Kirch* einmal nach einer Person und einmal nach einer Kirche benannt sein.

Abbildung 10 bietet eine Übersicht, über das verwendete Modell zur Kategorisierung, hier teilweise auch Klassifikation genannt, der spezifischen Strassenbenennungen. Alle spezifischen Benennungen werden folglich parallel durch vier unterschiedliche Zu-

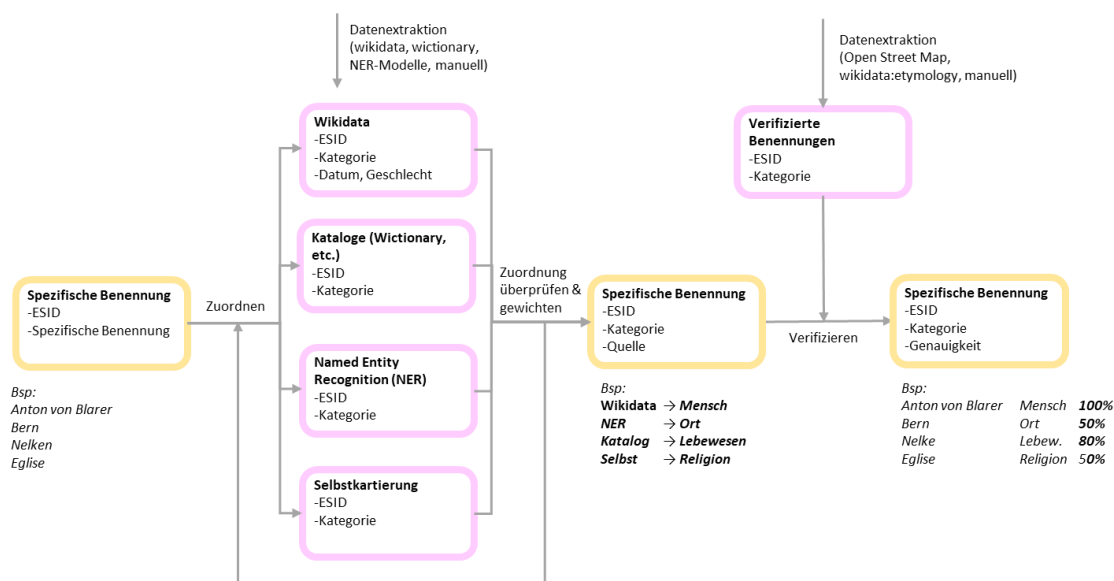


Abbildung 10 Übersicht des Modells zur Kategorisierung der spezifischen Strassenbenennungen

³⁸ Query ersichtlich im Anhang 8.6 «Wikidataquery: Label von allen Objekten»

ordnungsarten, wie z.B. durch Named Entity Recognition (NER), einer Kategorie zugeordnet. Diese vier Zuordnungsarten werden in den Unterkapiteln 2.4.2 bis 2.4.5 beschrieben. Alle vier Zuordnungsarten wurden jeweils iterativ optimiert. D.h. die jeweiligen Resultate wurden mehrfach manuell validiert, z.B. durch die stichprobenweise oder spezifische Suche nach Falschzuordnungen und anschliessend die Zuordnung angepasst. Z.B. durch Anpassung der Wikidata Query, durch Komplettierung von Katalogen oder auch durch Anpassung der vorangehenden *Data preparation* (2.3). Durch diesen iterativen Validierungsprozess konnte zusätzlich die Genauigkeit der einzelnen Zuordnungsarten grob quantifiziert werden. Diese Informationen nutzten wir, um die spezifische Benennung schliesslich einer finalen Kategorie zuzuordnen. Bei nur einem oder keinem Resultat war die finale Kategorie eindeutig. Bei mehreren Resultaten wurde jeweils das Resultat der *genauesten* Zuordnungsart verwendet. Dies sind, nach absteigender Genauigkeit geordnet, folgende Zuordnungsarten:

- Extraktion spezifische Strassenbenennung Religion;
- Katalog Gewässer;
- Katalog Berg;
- Selbstkartierung Natur;
- Selbstkartierung Religion;
- Selbstkartierung Siedlung;
- Katalog Lebewesen;
- Wikidata Person;
- Named Entity Recognition Person;
- Katalog Beruf;
- Katalog Ortschaft.

So wird z.B. eine Strasse deren spezifische Bezeichnung der Kategorie *Ortschaft* und der Kategorie *Gewässer* zugeordnet wird schliesslich der Kategorie *Gewässer* zugeordnet. Schliesslich wird diese finale Zuordnung durch einen unabhängig erstellten Test- oder Verifikationsdatensatz überprüft, respektive die Genauigkeit berechnet (vgl. 3.3).

2.4.1 Wahl und Definition der Kategorien

Die Auswahl der schliesslich 9 Kategorien für Strassennamen (Berg, Beruf, Gewässer, Lebewesen, Natur, Ortschaft, Person, Religion, Siedlung und Andere) war ein unschweifiger Prozess, auf welchen mehrere Faktoren einwirkten. In der frühen Phase der Auswahlfindung lenkten uns insbesondere einige Produkte von Datenjournalisten³⁹ und einige wissenschaftliche Publikationen (Bancilhon, et al., 2021), (Gurtovoy, et al., 2022)). Später haben wir Erkenntnisse der ersten quantitativen Resultate wie z.B. die häufigsten spezifischen Bezeichnungen (z.B. *Dorf* 904-mal, *Kirch* 761-mal, *Birken* 406-mal) hinzugezogen. Zudem haben wir übergeordnet versucht, nicht zu kleine Kategorien zu definieren und Überlappungen zu minimieren. Nachfolgende Tabelle bietet eine grobe Beschreibung dieser 9 Kategorien, oder 10, wenn man die nicht kategorisierbaren Benennungen als eigene Klasse betrachtet.

³⁹ <https://www.watson.ch/wissen/schweiz/133417259-das-sind-die-witzigsten-strassennamen-der-schweiz>. Abgerufen am 13. Oktober 2022

⁴⁰ <https://www.zeit.de/2018/05/strassenverzeichnis-strassennamen-deutschland>. Abgerufen am 10. Oktober 2022

Berg	Gipfel, Pässe, Gebirge, Hügel
Beruf	Historische und aktuelle Berufsgattungen
Gewässer	Flüsse, Seen und Bäch
Lebewesen	Pflanzen und Tiere
Natur	Natürliche Artefakte wie ein Wald oder eine Halde
Ortschaft	Ortschaften, Länder, Städte, Gemeinden
Person	Personennamen, Familiennamen
Religion	Religiöse Einrichtungen, <i>Götter</i> , <i>Heilige</i>
Siedlung	Artefakte von Menschen erstellt wie eine Treppe oder eine Schule

Tabelle 10 Übersicht Kategorien der spezifischen Strassenbenennungen

2.4.2 Wikidata

Um Personen in den spezifischen Benennungen zu finden, wurde auf die Abfrage von Wikidata zurückgegriffen. Es schien als eine der zuverlässigsten Möglichkeiten die Personennamen zu extrahieren, auch wenn beim Aufbau des Verifikationsdatensatzes die Erfolgsquote wahrscheinlich weniger als 50% war.⁴¹ Als Ergänzung zur Auswertung mittels NER⁴², schien die Methode mittels Wikidata eine gute Ergänzung.

Für die Abfrage wurde zwei Queries verwendet. Zum einen war dies eine leicht angepasste Abfrage über das Wikidata Label,⁴³ welche nur Objekte von der Instanz Mensch (wdt:P31 = wd:Q5), plus einige zusätzliche Informationen wie das Geschlecht, zurück gibt.

Zusätzlich wurde eine Abfrage über die Wikidatasuche durchgeführt.⁴⁴ Dies hat zum Vorteil, dass Abfragen wie *General Guisan* ebenfalls zum gewünschten Wikidataobjekt führen. Der Nachteil ist allerdings, dass eine grosse Zahl Objekte als Ergebnisse zurückkommt. In dieser Abfrage wurde jeweils das erste Objekt verwendet und alle anderen verworfen, was sicher verbesserungswürdig ist.

Mit der Abfrage über die Wikidatasuche wurden 13'705 Strassen mit Personenbezug gefunden. Bei der Abfrage über das Label 8'046.

Zu erwähnen ist, dass die Abfragen über sämtliche rund 178'000 bei beiden Queries über 72h dauerte. Probleme verursachten wiederkehrende Ausfälle beim Endpunkt. In den ersten Versuchen musste nach einem Unterbruch die Abfrage manuell neu gestartet werden. Dies wurde verbessert in dem der Code aus dem Notebook in ein Pythonskript⁴⁵ überführt wurde, welches in einem Dockercontainer mit der Option – restart=always auf einem Server lief.

⁴¹ Vergleiche Kapitel 2.3.5 «Weitere Verifikationsansätze»

⁴² Vergleiche Kapitel 2.4.4 «Named-Entity Recognition»

⁴³ Query ersichtlich im Anhang 8.4 «Wikidataquery: Label von wikidataobjekten»

⁴⁴ Query ersichtlich im Anhang 8.7 «Wikidataquery: Abfrage über Suche»

⁴⁵ mapping wikisearch person.py. Extrakt aus mapping wiki person.ipynb

2.4.3 Zuordnen anhand von Katalogen

Die im Kapitel 2.2.7 beschriebenen Kataloge wurden zuerst, wo nicht auf Deutsch und Französisch vorhanden mittels *deepl.com* übersetzt und anschliessend bereinigt. Falls für eine Kategorie mehrere Kataloge verwendet wurden, wurden diese zusammengefügt wie z.B. die Kataloge der Schweizer Pässe und Gipfel zur Kategorie *Berge*. Danach wurde ein String-Vergleich (*match*) für jede Kategorie mit den spezifischen Benennungen durchgeführt. Die spaCy Funktionalität *similarity* wurde als Alternative getestet, ist jedoch hier nicht geeignet. Einige wenige Einträge der Kataloge wurden manuell geändert. Es handelt sich dabei insbesondere um Strings mit 3 oder weniger Zeichen, welche fälschlicherweise mit viel zu vielen spezifischen Strassenbenennungen matchen wie z.B. *Ins* oder *Ober*.

2.4.4 Named-Entity Recognition

Die im Python Modul Spacy integrierte *Natural language processing*-Funktionalität Named-Entity Recognition (NER) ermöglicht die automatische Identifikation und Klassifikation von Eigennamen. So erkennen die von uns getesteten deutschen und französischen Sprachmodelle *de_core_new_md* und *fr_core_news_md* im amtlichen Verzeichnis der Strassen Eigennamen der Kategorien *PER* – *person*, *LOC* – *location*, *ORG* – *organization* und *MISC* – *miscellaneous*. Einzig die so erkannten Personen (*PER*) entsprechen einer von uns gewählten Kategorie. Die anderen NER Kategorien *LOC*, *ORG* und *MISC* wurden nicht weiter verwendet.

2.4.5 Selbstkartierung

Für die Kategorien *Natur*, *Religion* und *Siedlung* konnten keinen passenden Zuordnungsinformationen aufgefunden werden, weshalb wir für diese manuell eine Liste zusammenstellten. Die Liste der Siedlungsnamen wie *bahn*, *reservoir*, *musee* oder *chateau* wurde insbesondere durch die Attribute des topografischen Landschaftsmodells der Schweiz⁴⁶ ergänzt. Die Listen *Natur* und *Religion* wurden hauptsächlich durch Betrachten von regional gleichmässig verteilten Stichproben erstellt.

⁴⁶ <https://www.swisstopo.admin.ch/de/wissen-fakten/topografisches-landschaftsmodell.html>. Abgerufen am 14. November 2022

3 Resultate

In dem folgenden Kapitel werden die, aus Sicht der Autoren, interessantesten Erkenntnisse dieser Arbeit detaillierter ausgeführt.

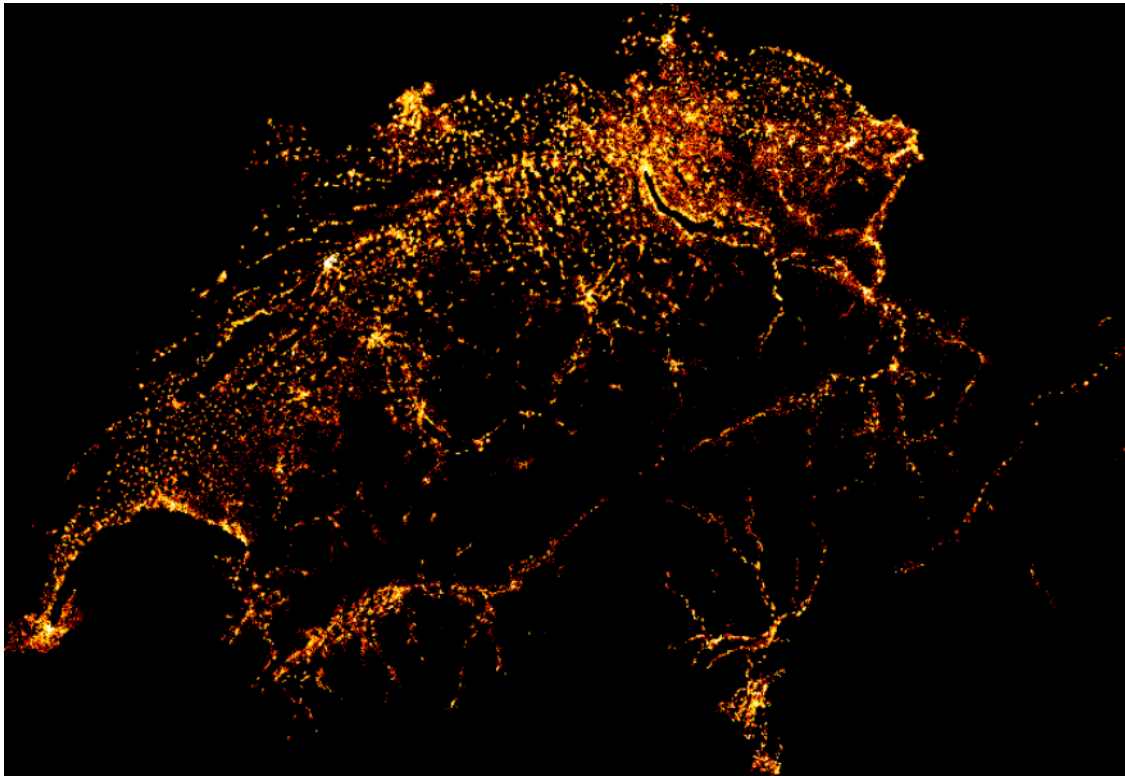


Abbildung 11 Visualisierung aller Strassen des Amtliches Verzeichnis der Strassen

3.1 Ablage GitHub

Sämtlicher erstellter Programmecode, inklusive den Daten, ist auf GitHub in folgendem Repository verfügbar: <https://github.com/CaptainInler/cassda-zertifikatsarbeit>. Die Ablage entspricht dem Aufbau dieses Dokuments welcher wiederum den CRISP-Prozess, wie in Kapitel 2 «Vorgehen» beschrieben, angelehnt ist. Der Zusammenhang der Kapitel dieser Arbeit mit den JupyterLab Notebooks ist im Anhang 8.1 ersichtlich.

3.2 Verteilung und Statistiken

Am häufigsten wird der Name Dorfstrasse verwendet (637-mal), viele gibt es nur ein einziges Mal, wie den Elefantenweg. Wie bereits in der Motivation breiter ausgeführt ist keiner dieser Namen zufällig gewählt. Die hier präsentierten Ergebnisse fokussieren, gemäss Fragestellung (vgl. 1.3) insbesondere auf die räumliche, regionale Verteilung der Strassennamen und liefern so Informationen zu aktuellen Gegebenheiten, welche auf historischen Prozessen basieren. Ein erster Blick auf die Verteilung aller Strassen (Abbildung 11) zeigt, dass die Verteilung der Strassen, wenig erstaunlich, der Bevölkerungsverteilung nachkommt. Die Visualisierung der Kerndichte (Abbildung 12) zeigt denn auch deutlich, dass wir in der bevölkerungsreichsten Agglomeration,

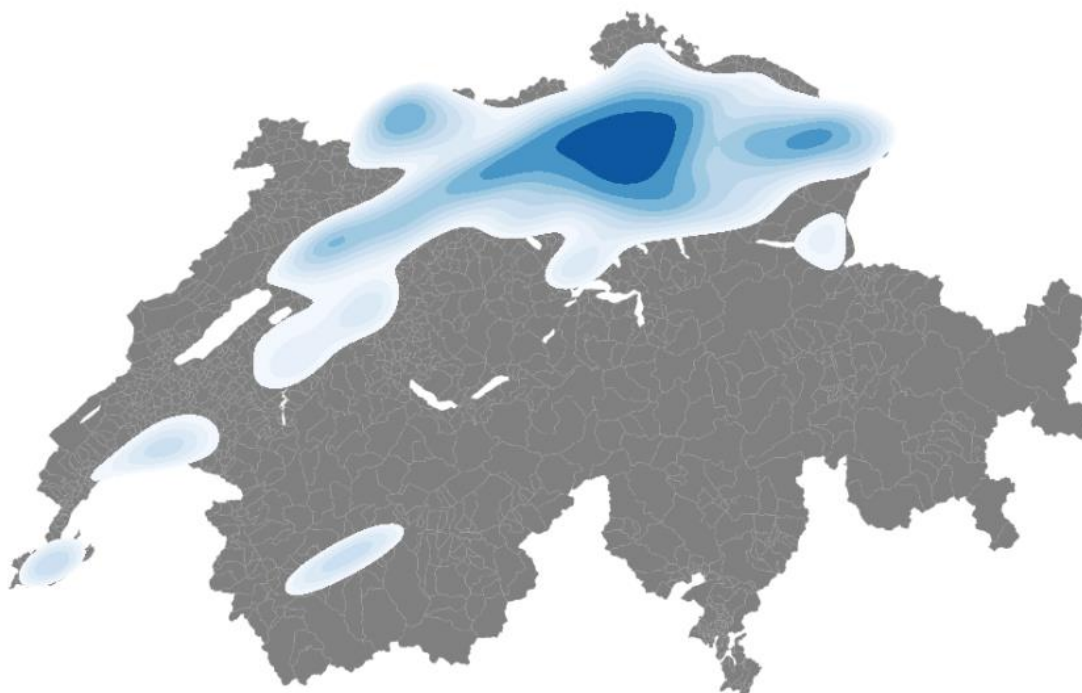


Abbildung 12 Verteilung der amtlichen Strassen

Diese Karte zeigt die Kerndichteschätzung (Kernel Density Estimation KDE) der amtlichen Strassen der Schweiz (Strassen wurden durch Punkte abgebildet). Je dunkler blau, desto häufiger gibt es in diesem Gebiet amtliche Strassen. Es wurden die Default-Parameter von Seaborn (seaborn.kdeplot) verwendet (bis auf tresh = 0.4).

der Agglomeration Zürich, die höchste Dichte an amtlichen Strassen vorfinden. Neben Zürich stechen auch die Agglomerationen der weiteren Schweizer Grossstädte hervor. Dabei spiegelt die Strassendichte jedoch nicht überall proportional die Bevölkerungsdichte wider. Die Agglomerationen St. Gallen, Sargans und Sion haben eine verhältnismässig hohe Strassendichte. Die Agglomerationen Genf, Lausanne und Bern eine geringe. Dieser Sachverhalt lässt sich durch den Vergleich mit der Statistik der Strassenlängen (Bundesamt für Statistik) erklären. Die Statistik der Strassenlänge bietet einen Überblick über das Schweizer Strassennetz, welches für den Motorfahrzeugverkehr geöffnet ist und beinhaltet alle befahrbaren Strassen mit einer Fahrbahnbreite von zumindest 2.80 Metern. Damit ist in diesem Datensatz eine gewisse schweizweite

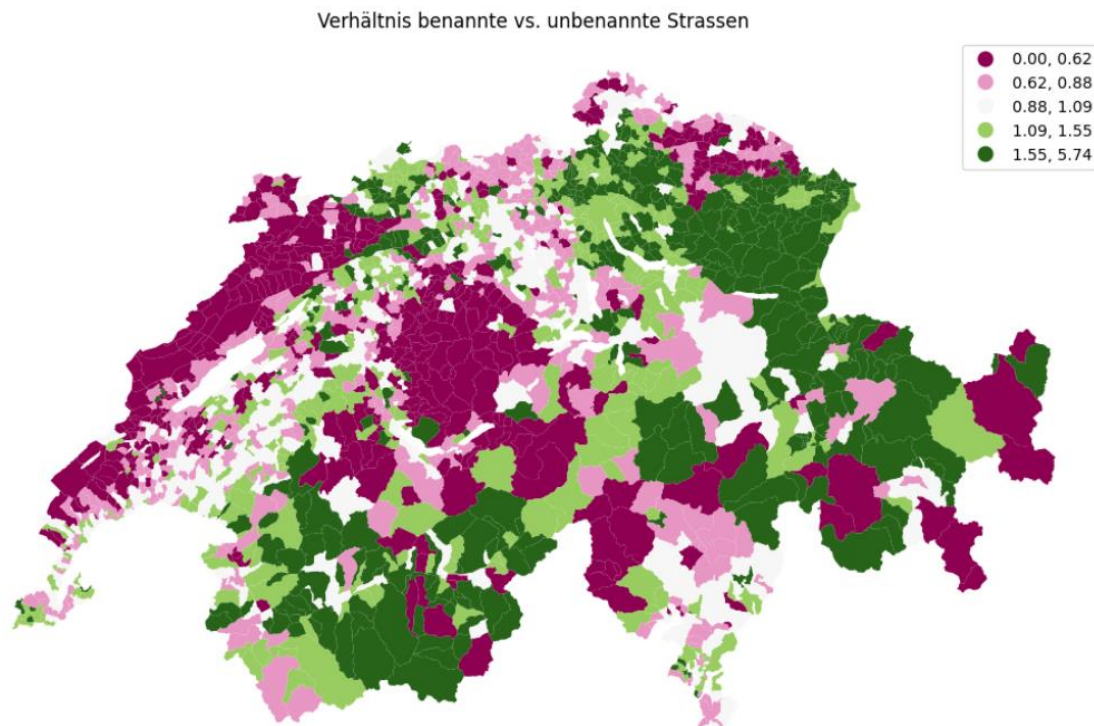


Abbildung 13 Dichte des Netzes der benannten Strassen

Diese Abbildung zeigt das Verhältnis zwischen der Länge der benannten Strassen (aus dem Amtlichen Verzeichnis der Strassen) und der Länge der insgesamt vorhandenen Strassen. In den roten Gemeinden sind deshalb tendenziell weniger der vorhandenen Strassen benannt als in den grünen Gemeinden.

Homogenität gegeben. Wir haben nun pro Gemeinde die Länge der in der Strassenlängenstatistik enthaltenen Strassen mit der Länge der benannten Strassen, also dem von uns analysierten Datensatz des amtlichen Strassenverzeichnisses, verglichen und visualisiert (Abbildung 13). Die Visualisierung zeigt starke regionale Unterschiede. So tragen im Jura und im Emmental verhältnismässig viel weniger Strassen einen Namen als z.B. im Kanton St. Gallen. Die genaue Analyse einiger zufälliger Standorte in den genannten Regionen zeigt denn auch, dass im Kanton St. Gallen viele Feldwege im Amtlichen Verzeichnis der Strassen enthalten und somit benannt sind. Im Jura hingegen sind weitgehend nur die Strassen innerhalb von Ortschaften benannt.

3.2.1 Generische Benennung

Die beiden am häufigsten verwendeten generischen Bezeichnungen für Odonyme sind, wenig überraschend, *strasse* mit fast 50'000 Erscheinungen und *weg* mit 43'000. Das macht einen Anteil von fast 55% (Abbildung 14). Dahinter bricht die Verteilung zusammen und es folgt mit der dritthäufigsten Bezeichnung *chemin*, knapp 15'000, die häufigste der Romandie. Die häufigste Bezeichnung der italienisch- und romanischsprachigen Schweiz liegt mit über 6'600 Benennungen schweizweit auf Platz 5. Auf den ersten zehn Plätze schaffen es auch die zwei Verkleinerungsformen *gässli* mit über 1000 Benennungen und *ruelle* mit über 700. Wesentlich geringer, als eine der eben genannten generischen Strassenbezeichnungen anzutreffen, ist die Wahrscheinlichkeit über eine *piazza* (41), eine *boulevard* (34) oder eine *vietta* (26) zu gehen. Rund 33'000 Strassennamen besitzen keine generische Strassenbezeichnung, dies macht einen Anteil von etwas über 19%. Dazu zählen z.B. Strassennamen wie *Kolonie*, *Zum Kuckuck* oder *Chueschwanz*, aber auch sehr seltene Verkleinerungsformen wie *wääg* oder Landschaftsbezeichnungen wie *Allmeind*, *Linthgut* oder *Im Steig*.

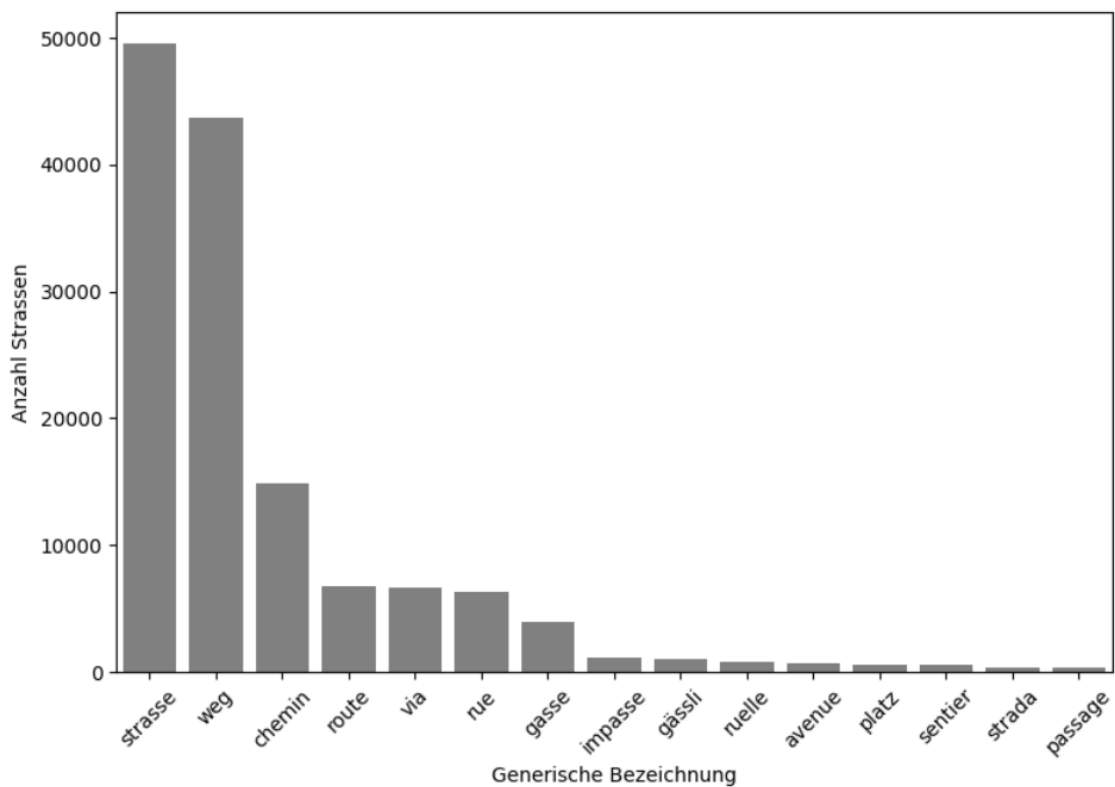


Abbildung 14 Häufigkeit der generischen Strassenbezeichnungen

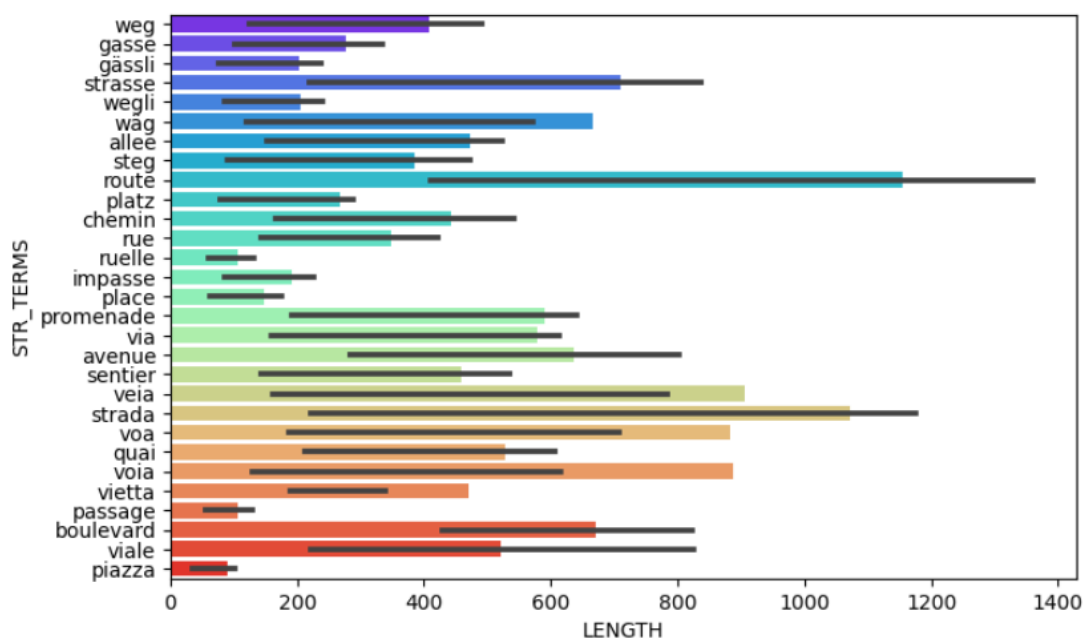


Abbildung 15 Durchschnittliche Strassenlänge nach generischer Bezeichnung

Dieses Balkendiagramm zeigt die durchschnittliche Länge (in Metern) der generischen Bezeichnungen der amtlichen Strassen der Schweiz. Der Fehlerbalken (schwarzer Strich) stellt das Konfidenzintervall mit Konfidenzniveau 50 % dar. So ist z.B. eine amtliche Strasse mit der generischen Bezeichnung «route» durchschnittlich 1155m lang und das Konfidenzintervall enthält es den wahren Wert mit 50 % Wahrscheinlichkeit.

Die durchschnittliche Länge der generischen Strassenbezeichnungen unterscheidet sich erheblich (Abbildung 15). Eine *route* ist mit 1155m durchschnittlich am längsten, wobei zu beachten ist, dass die eigentliche Länge stark variiert. So ist eine zufällig aufgefundene *route* mit 50% Wahrscheinlichkeit zwischen rund 400m und 1'350m lang, respektive zu 50% kürzer als 400m oder länger als 1'350m. Um einiges kürzer als die *routes* sind die *ruelles*, *passages* und *piazzas* mit durchschnittlich rund 100m. Die Verkleinerungsformen *gässli*, *wegli* oder *ruelle* sind, wenig erstaunlich, alle durchschnittlich kürzer als ihre Substantive.

Die Visualisierung der häufigsten generischen Strassenbezeichnung pro Gemeinde zeigt erwartungsgemäss die Verteilung der vier Schweizer Sprachregionen (Abbildung 16). In der Deutschschweiz teilen sich die *strassen* und *wege* den Topplatz in den allermeisten Gemeinden unter sich auf. Dabei hat die Bezeichnung *weg* in Richtung Romandie immer häufiger den Topplatz inne und umgekehrt in östlicher Richtung die *strassen*, mit Ausnahme der Region Chur-Sargans. In der Romandie sind die *chemins* als Topplatzierung am meisten verbreitet, wobei insbesondere zwei Regionen davon abweichen. Erstens ist der Jurabogen erkennbar, in welchem häufig *rue* den Topplatz belegt. Zweitens sind im Kanton Fribourg die *route* und interessanterweise auch die *impasse* (Sackgasse) häufig als Topplatz vorhanden. In der italienisch- und romansprachigen Schweiz ist die *via* klar am meisten die häufigste generische Strassenbezeichnung pro Gemeinde. Die grau gefärbten Gemeinden haben keine amtlichen Strassen oder nach unserer Definition keine amtlichen Strassen mit einer generischen Bezeichnung.

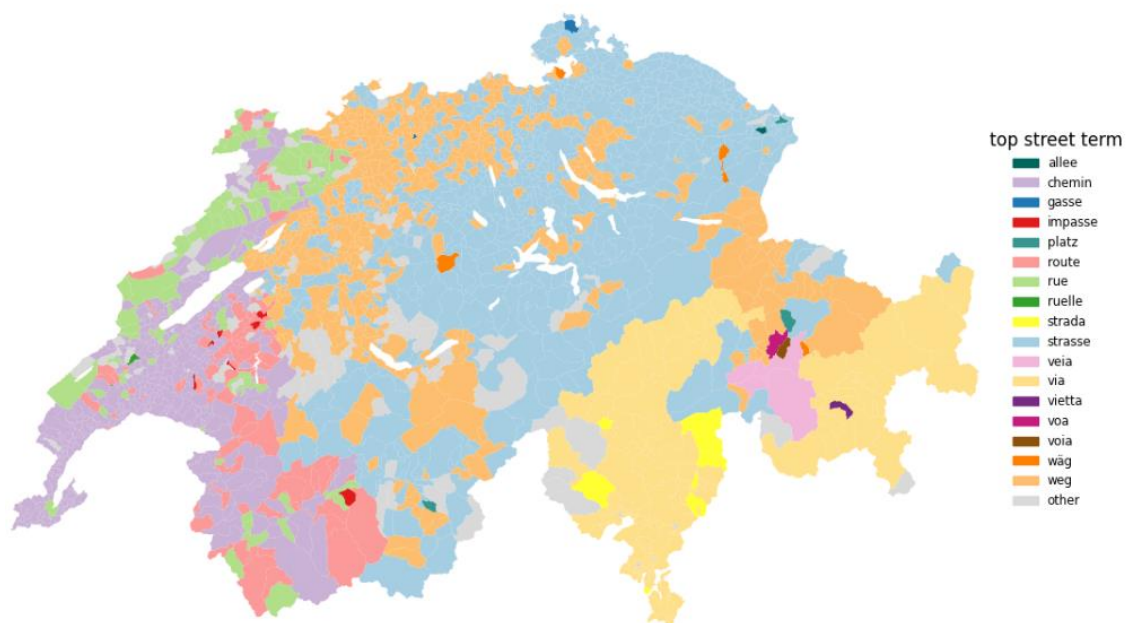


Abbildung 16 Top generische Strassenbezeichnung pro Gemeinde

Diese Karte zeigt die häufigste, generische Bezeichnung der amtlichen Strassen pro Gemeinde. Einige Gemeinden haben keine amtlichen Strassen oder eine andere häufigste generische Bezeichnung (other: hellgrau).

Die relative Häufigkeit der generischen Strassenbezeichnungen auf Kantonsebene, als gestapeltes Säulendiagramm dargestellt (Abbildung 17), liefert weitere Erkenntnisse. So hat der Kanton Appenzell Ausserrhoden mit 80% einen sehr hohen Anteil von Strassen ohne generische Strassenbezeichnung. Die nächsthöchsten Anteile liegen etwas über 40% in den Kantonen Glarus und Uri. Den niedrigsten Anteil an Strassennamen ohne generische Bezeichnung hat der Kanton Waadt mit rund 5%. Beim

Betrachten der Säulendiagramme fallen in der Westschweiz jeweils Kantonspärchen mit ähnlichen Profilen auf. Die Pärchen sind Genf-Lausanne, Neuchâtel-Jura sowie Fribourg und Wallis. Wie, wegen der in der Schweiz einzigartigen Sprachverteilung, zu erwarten, besitzen die Kantone Graubünden und Tessin eine einzigartige Verteilung.

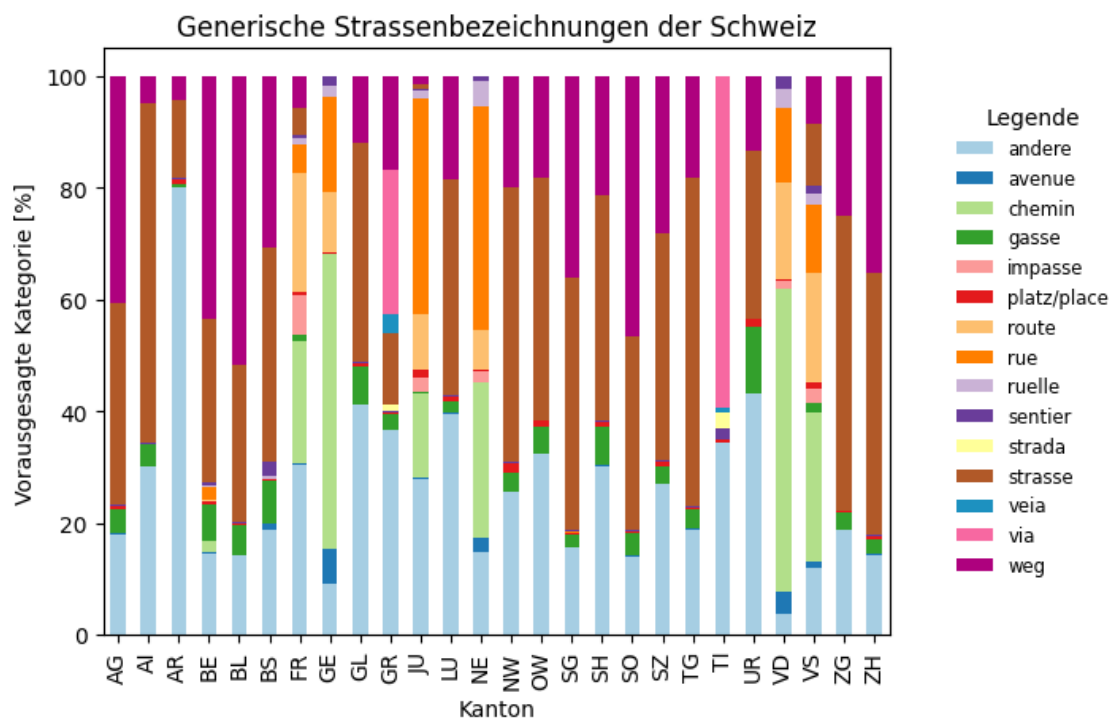


Abbildung 17 Relative Häufigkeit der generischen Strassenbezeichnung pro Kanton

Dieses gestapelte Säulendiagramm zeigt die prozentuale Häufigkeit generischen Strassenbezeichnungen der amtlichen Strassen der Schweiz

Ein Blick auf die Verteilung der generischen Strassenbezeichnungen in den beiden grössten Schweizer Städten Genf und Zürich (Abbildung 18; Abbildung 19) liefert weitere Einblicke in deren räumliche Verteilung. Die *quais* liegen in beiden Städten ausschliesslich an einem Gewässer, ansonsten ist die Verteilung unterschiedlich. In Genf



Abbildung 18 Visualisierung der generischen Strassenbezeichnung der Stadt Genf

erkennt man, für die Schweiz einzigartig, die langgezogenen, geraden *boulevards* im Stadtzentrum. Ebenfalls im Stadtzentrum befinden sich die kurzen *passagen*. Radial etwas von Stadtzentrum entfernt folgen die *avenues* und noch etwas weiter entfernt die *routes*. In Zürich befinden sich in der Altstadt mehrheitlich *gassen* und darum herum ein Geflecht aus *strassen* und *wegen*.

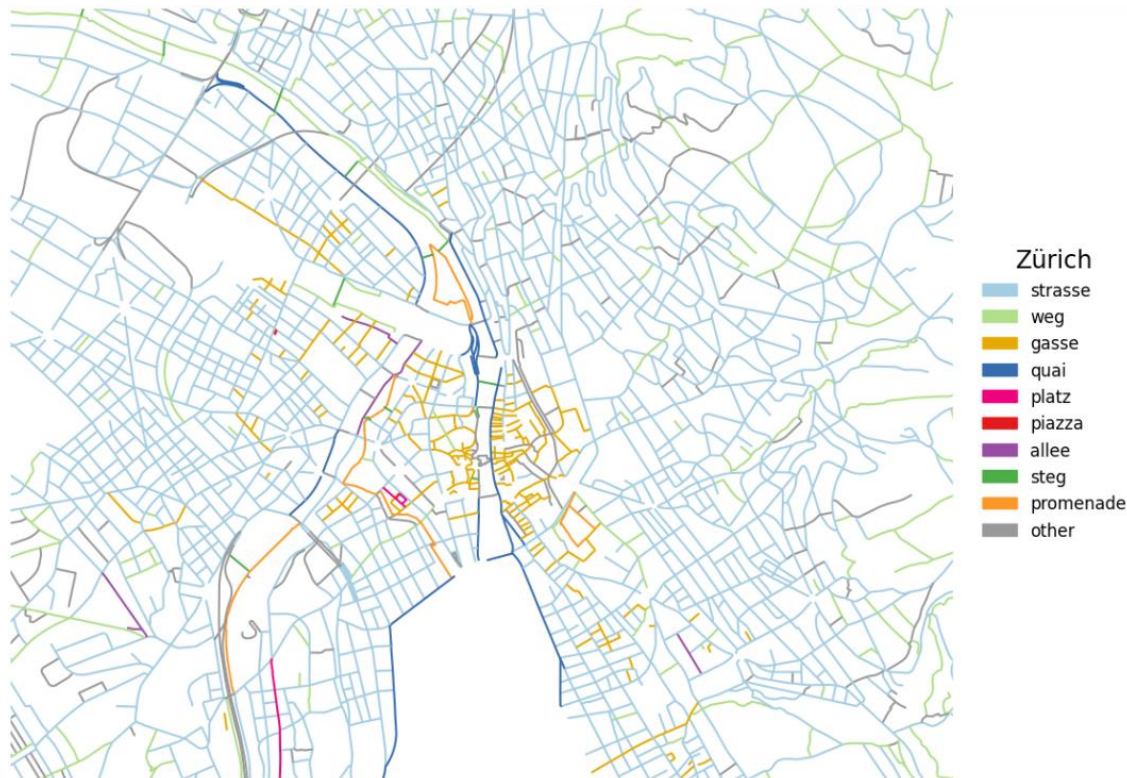


Abbildung 19 Visualisierung der generischen Strassenbezeichnung der Stadt Zürich

3.2.2 Spezifische Benennung generell

Die Resultate der spezifischen Strassenbenennungen beziehen sich ausschliesslich auf die rund 161'000 Strassen der deutsch- und französischsprachigen Schweiz. Die mehr rund 11'000 Strassen der italienisch- und romanischsprachigen Gemeinden sind nicht berücksichtigt. Mehr als 60% der 161'000 Strassen konnten einer der neun spezifischen Benennungskategorien, Natur, Siedlung, Ortschaft, Person, Lebewesen, Berg, Gewässer, Religion und Beruf, zugeordnet werden. Knapp 40% konnten entsprechend keiner dieser Kategorien zugeordnet werden. Abbildung 21 zeigt die quantitative Verteilung der Kategorien. Am häufigsten sind die Strassennamen mit knapp 30'000 nach Naturnamen benannt, was fast 18% aller Namen sind. Dahinter folgen die nach Siedlungsnamen benannten Strassennamen mit rund 10%. Am wenigsten Strassen sind mit rund 1% nach Berufen und mit knapp 2% nach religiösen Namen benannt.

Am längsten sind mit durchschnittlich rund 750m die nach Ortschaften und Bergen benannten Strassen, gefolgt von den nach Gewässern benannten Strassen mit mehr als 650m. Am kürzesten sind die nach Lebewesen benannten Strassen mit durchschnittlich rund 330m. Die durchschnittliche Strassenlänge aller anderen Kategorien liegt zwischen grob bei 500m. Wie bei den generischen Benennungen, sind auch bei der Darstellung der häufigsten spezifischen Benennung pro Gemeinde, die Sprachregionen ersichtlich (Abbildung 200). So dominieren in der Romandie Gemeinden, welche ihre Strassennamen am häufigsten nach Ortschaften oder Personen benennen. In der

Spezifische Strassenbezeichnungen nach Anzahl

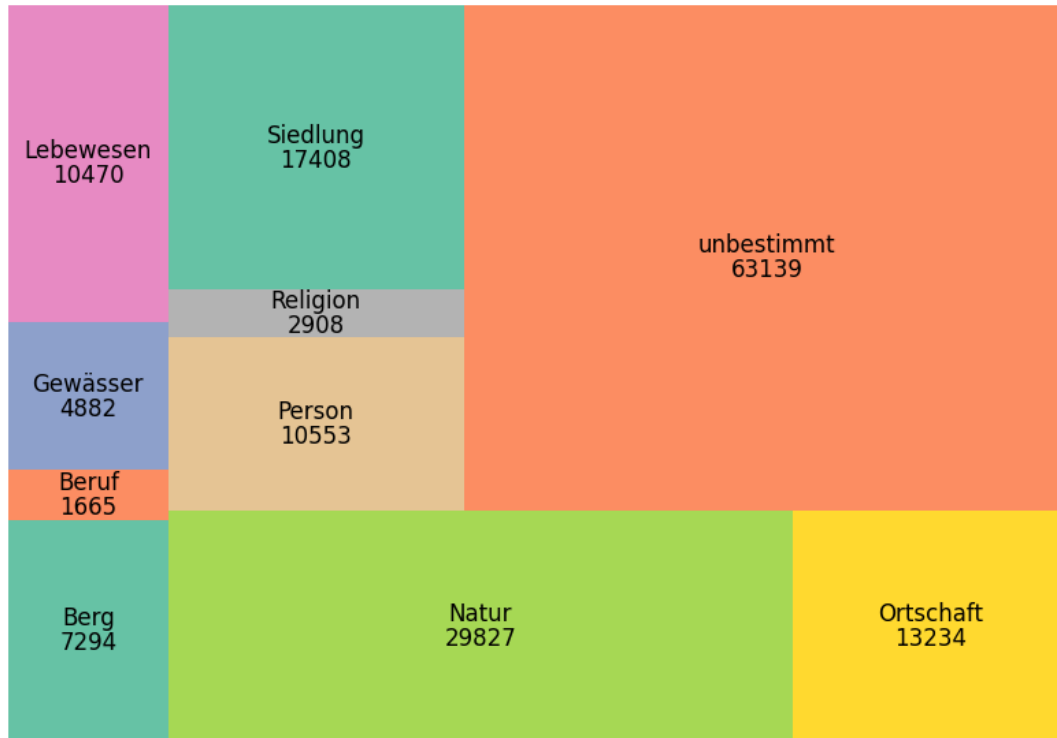


Abbildung 21 Häufigkeit der spezifischen Strassenbezeichnungen

In diesem Kacheldiagramm (Tree Map) sehen sie die Grösse der Kategorien, nach welchen die amtlichen Strassen der Schweiz benannt sind. Die italienisch- und romanischsprachigen Strassen sind nicht enthalten. Ohne letztere sind es insgesamt 161'380 Strassen

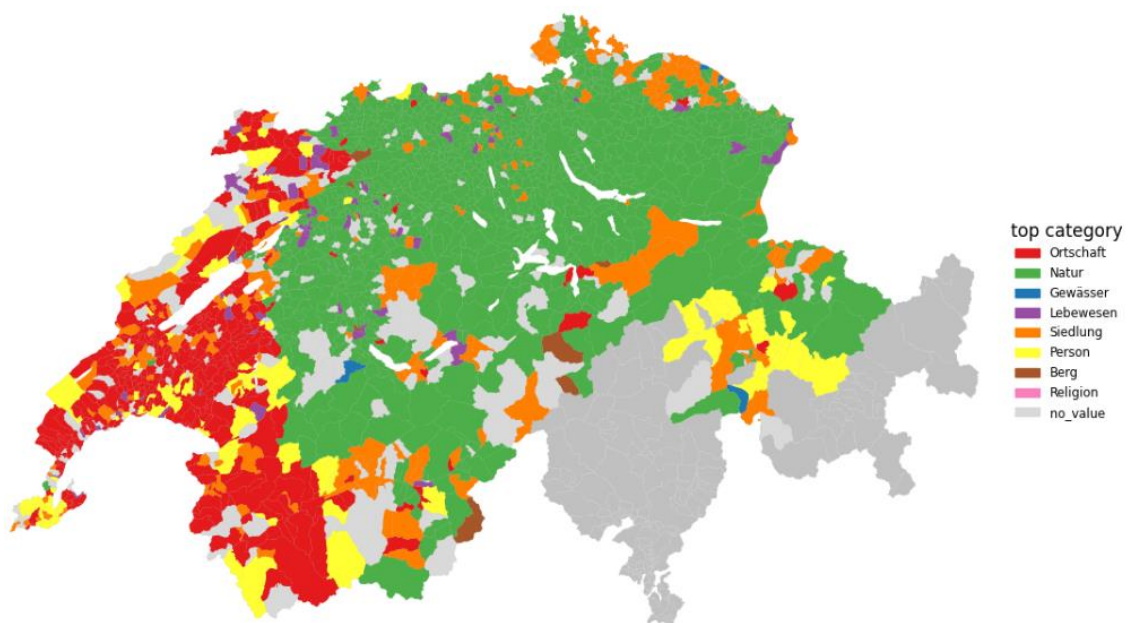


Abbildung 20 Top spezifische Strassenbezeichnung pro Gemeinde

Diese Karte zeigt die Kategorie der amtlichen Strassen, nach welcher die meisten Strassen einer Gemeinde benannt sind (z.B. sind in der Stadt Genf am meisten Strassen nach Personen benannt). Einige Gemeinden haben keine amtlichen Strassen oder keine konnte kategorisiert werden (no_value: hellgrau). Italienisch- und romanischsprachige Gemeinden wurden nicht berücksichtigt (grau eingefärbt).

Deutschschweiz dominieren klar die Naturnamen. Im Gegensatz zu der räumlichen Verteilung der generischen Benennungen, sind zusätzlich auf den ersten Blick auch Muster erkennbar, die die Sprachregionen übergreifen. So sind die Gemeinden, die ihre Strassennamen am häufigsten nach Siedlungsnamen oder Lebewesen benennen, relativ gleichmässig und etwas gesprenkelt über die Sprachregionen, verteilt. In den Kantonen Graubünden und Wallis sind die Kategorien vermischter als in der übrigen Schweiz. Dies mag u.a. an ihrer Mehrsprachigkeit und besonders häufigen Verwendung von Mundartnamen liegen, welche häufig nicht oder falsch kategorisiert werden. Beim Betrachten der Abbildung 20 sind zusätzlich folgende drei Sachverhalte zu berücksichtigen. Erstens ist bei vielen Gemeinden der Anteil an Strassennamen, welche keiner der 9 Kategorien zugeordnet werden kann, grösser als der Anteil der grössten Kategorie. D.h. die Aussagekraft dieser Abbildung, in Gebieten mit einem hohen Anteil an nicht kategorisierten Strassennamen (Abbildung 23) ist gering. Zweitens sind alle Gemeinden, bei welchen zwei oder mehrere Kategorien am häufigsten vorkommen, grau dargestellt. Ebenfalls grau sind Gemeinden, welche keine amtlichen Strassen besitzen oder keine kategorisiert werden konnten. Drittens, wie generell bei den Resultaten der spezifischen Benennungen wurden italienisch- und romanischsprachige Gemeinden nicht berücksichtigt und sind deshalb grau dargestellt.

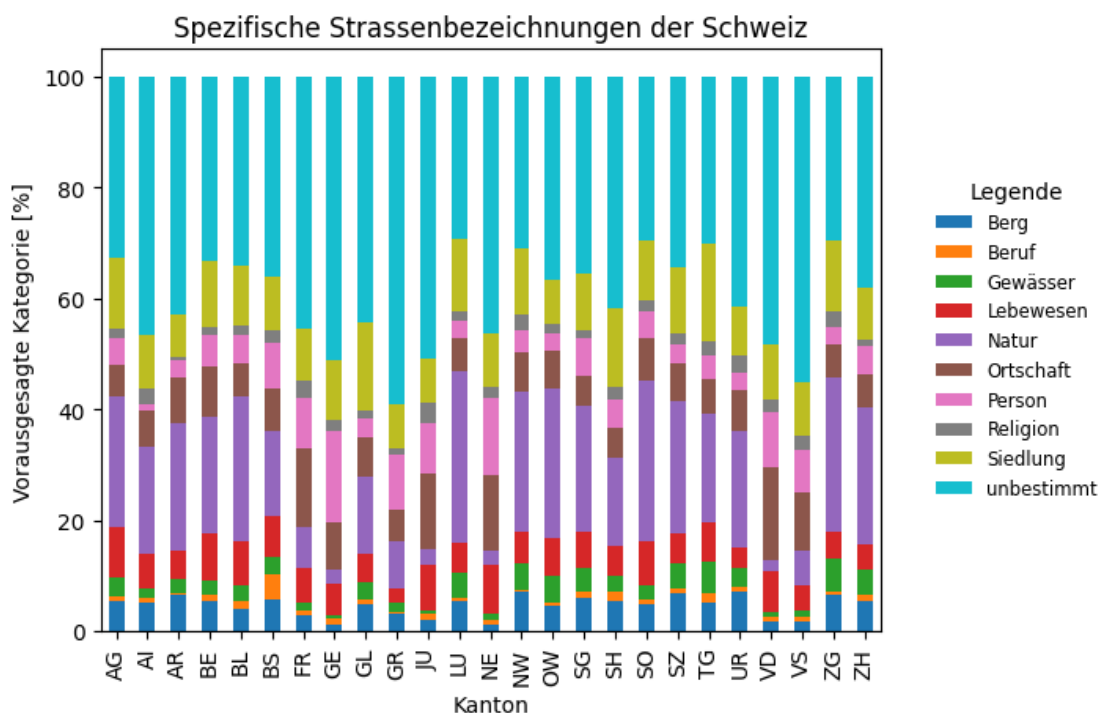


Abbildung 22 Relative Häufigkeit der Namenskategorie pro Kanton

Dieses gestapelte Säulendiagramm zeigt die prozentuale Häufigkeit der Kategorien der spezifischen Strassenbezeichnungen der amtlichen Strassen der Schweiz. Italienisch- und romanischsprachige Strassen wurden nicht analysiert, weshalb der Kanton Tessin nicht vorkommt und bei Kanton Graubünden nur die deutschsprachigen Strassen betrachtet wurden.

Die Darstellung der relativen Häufigkeit der Namenskategorien pro Kanton (Abbildung 22) bestätigt diverse bereits getätigte Aussagen. So z.B. die recht homogene Verteilung der Kategorien Siedlung und Lebewesen über alle Kantone oder der, im Verhältnis zu den französischsprachigen Kantonen, hohe Anteil der Kategorie Natur in den Deutschschweizer Kantonen. Der Kanton Basel-Stadt benennt seine Strassen mit grossem Abstand am häufigsten nach Berufen. Da Basel-Stadt der am dichtesten besiedelte, städtischste Schweizer Kanton ist, könnte dies darauf hindeuten, dass die Benennung der Strassen nach Berufen ein eher städtisches Phänomen ist. Diese

These wird in dieser Arbeit allerdings nicht genauer analysiert. Die Kantone mit dem höchsten Anteil an Strassen, wo die Kategorie nicht bestimmt werden konnte, sind Graubünden und Wallis mit einem Anteil von fast 60%. Wie bereits erwähnt liegt der Grund u.a. bei der häufigen Verwendung von Mundartnamen wie *Zum undru Hüs*, *Zum Chrizji* oder *Uf der Halda* und, insbesondere in Graubünden, die verbreitete Verwendung von italienisch- und romanischsprachigen Namen in heute mehrheitlich deutschsprachigen Gemeinden. Diese Aussage wird aufgrund einer manuellen Analyse der Strassennamen einiger Gemeinden mit besonders hohem Anteil an nicht kategorisierten Strassennamen getätigt (Abbildung 23). Neben dem auffällig hohen Anteil an nicht klassifizierten Namen der Gemeinden im Wallis und Graubünden sind in Abbildung 23 auch relativ viele Gemeinden in der Region des Kantons Jura und die eher peripheren Gemeinden in der Region Genf mit einem hohen Anteil ersichtlich. In dieser Abbildung wurden Gemeinden mit insgesamt weniger als 10 amtlichen Strassen grau eingefärbt, um eine Verzerrung durch sehr kleine Samples zu verhindern. Dies ist auch in den nachfolgenden Choroplethenkarten zu einzelnen Kategorien der Fall. Besonders viel Strassen konnten in der Region Thurgau sowie in vielen Gemeinden zwischen Bern und Zug kategorisiert werden.

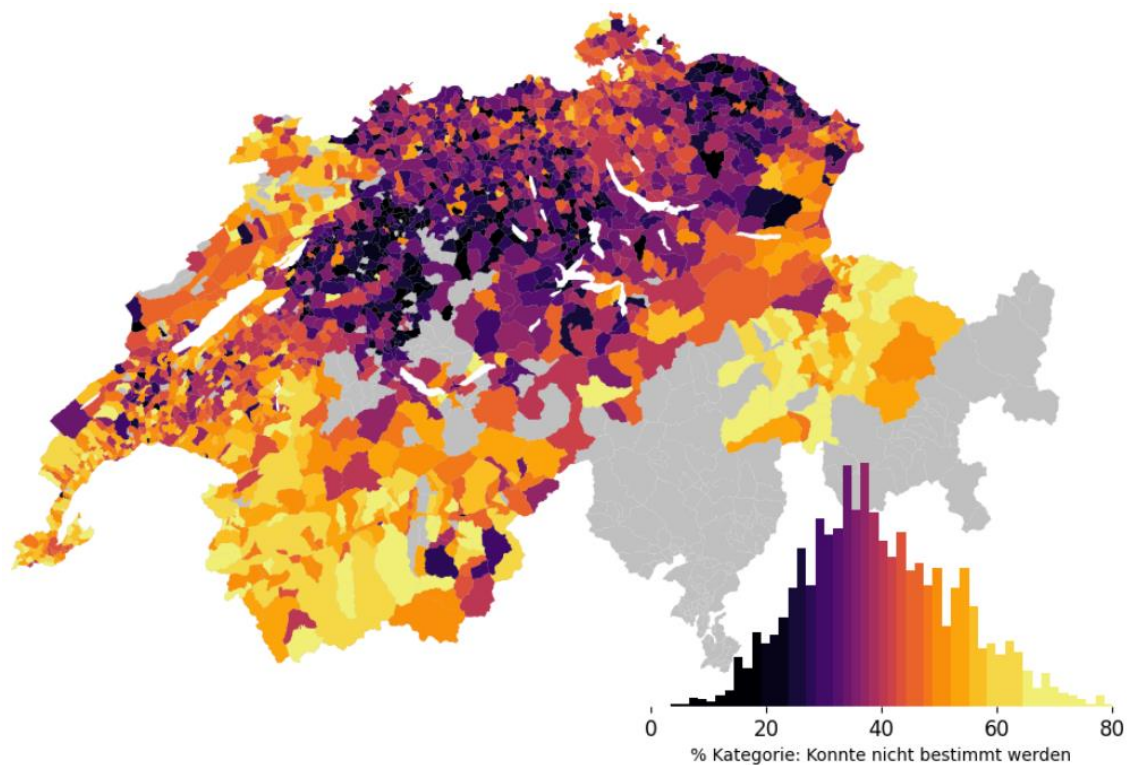


Abbildung 23 Anteil nicht klassifizierte Strassennamen pro Gemeinde

Diese Choroplethenkarte zeigt den Anteil (in Prozent) der amtlichen Strassen, welche nicht kategorisiert werden konnten pro Schweizer Gemeinde. Die Daten werden nach Quantilen in 20 Klassen dargestellt, wobei Gemeinden mit weniger als 10 amtlichen Strassen sowie italienisch- und romanischsprachige Gemeinden nicht berücksichtigt wurden (grau eingefärbt).

In den folgenden sechs Kapiteln werden die räumlichen Verteilungen einiger Kategorien in Detail analysiert und besprochen.

3.2.3 Ortschaften

In der Schweiz sind insgesamt mehr als 13'000 Strassen nach Ortschaften benannt, dies sind fast 8% aller Strassen. Davon sind, mit weit über 100 Benennungen, mit Abstand am meisten nach den beiden Städten Bern und Zürich benannt, allein die *Bernstrasse* gibt es 81-mal. Dabei profitieren die beiden Städte auch von Strassennamen, welche beide Städtenamen beinhalten wie der *Bern-Zürichstrasse* oder der *Zürich-Bernstrasse*. Hinter Zürich und Bern folgen viele Ortschaften mit bis mehr als 50 Benennungen wie z.B. Lausanne. Einige Strassen sind nach bekannten Städten anderer Länder benannt. So sind 7 Strassen nach der französischen Hauptstadt Paris und 4 Strassen nach dem US-amerikanischen Metropole New York benannt. Auch nach den Städten Moskau und Berlin sowie dem Land Kongo ist je eine Strasse benannt. Die *Neu-Yorkstrasse* in Rüti ist gemäss Tagesanzeiger⁴⁷ nach dem Bauernhof am Strassenende benannt. Dort wollte ein junger Bauer in den 1860er-Jahren nach Übersee fahren und vom, den Legenden zufolge, unermesslichen Reichtum Amerikas profitieren. Doch sein Vater war von diesen Plänen nicht überzeugt, spendierte seinem Sohn den Bauernhof und bewog ihn zu bleiben. Als zusätzliche Linderung des Fernwehs benannte er seinen Hof *Neu York*.

Strassen benannt nach den Ortschaften Lausanne:blau, Neuchâtel:orange, Fribourg:grün, Bern:rot, Zürich:violett

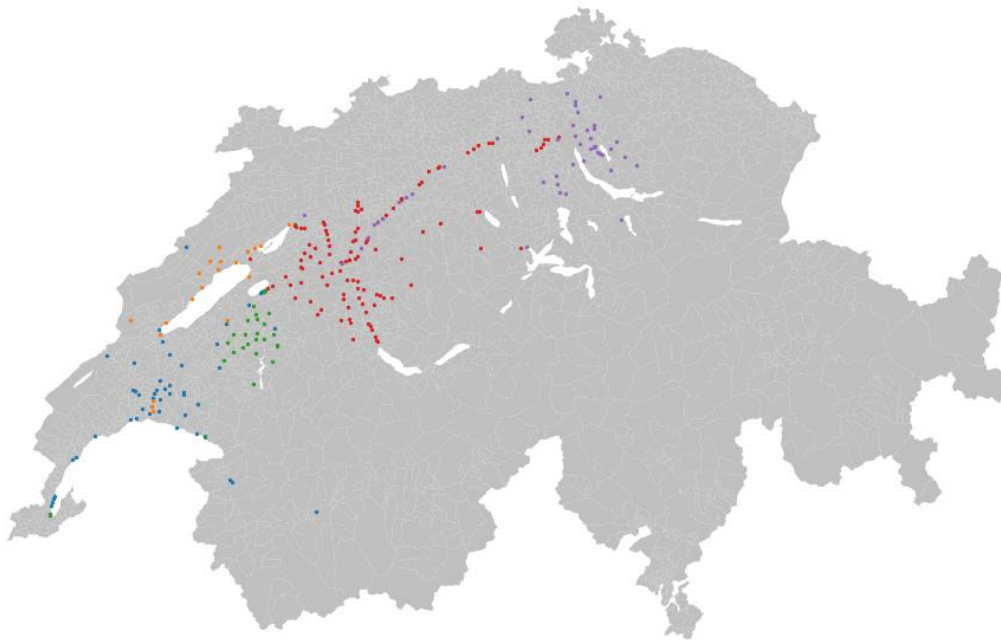


Abbildung 24 Verteilung der Strassen bekannt nach bekannten Ortschaften

Abbildung 24 zeigt die räumliche Verteilung der Strassen, welche nach den Städten Bern, Zürich, Lausanne, Neuchâtel und Fribourg benannt sind. Die Strassen aller fünf Städte bilden ein klar ersichtliches Cluster rund um die jeweilige Stadt. In der Stadt Bern lassen sich zudem Einfallsstrassen erkennen. Besonders markant zeigt sich eine Verbindung der Städte Zürich und Bern.

Beim Betrachten der Verteilung des relativen Anteils, der nach Ortschaften benannten Strassen, auf Gemeindeebene (Abbildung 25) lässt sich die Romandie erkennen. In den Gemeinden der Romandie, ist dieser Anteil gehäuft über 15%. Im Gegensatz

⁴⁷ https://www.tagesanzeiger.ch/ich-war-noch-niemals-in-neu-york-rueti-zh-670181166126?idp=OneLog&new_user=yes. Abgerufen am 28. Oktober 2022

dazu ist dies bei Deutschschweizer Gemeinden selten der Fall. In vielen kleineren Gemeinden des Kantons Waadt liegt der Anteil bei über 30%. Mit über 53% liegt der Anteil in der Waadtländer Gemeinde Bettens am höchsten. Von Dorfzentrum Bettens aus führen radial mehrere Strassen in die Nachbargemeinden Daillens, Oulens, Echallens und Boussens und sind nach diesen benannt. Dieses Muster lässt sich in vielen in dieser Region liegenden Gemeinden finden, wobei die Strassen häufig auch nach Ortschaften innerhalb einer Gemeinde oder einer ehemaligen Gemeinde benannt sind. Auch in einigen Deutschschweizer Gemeinden im Oberraargau, wie z.B. in Bleienbach BE lässt sich diese Muster auffinden.

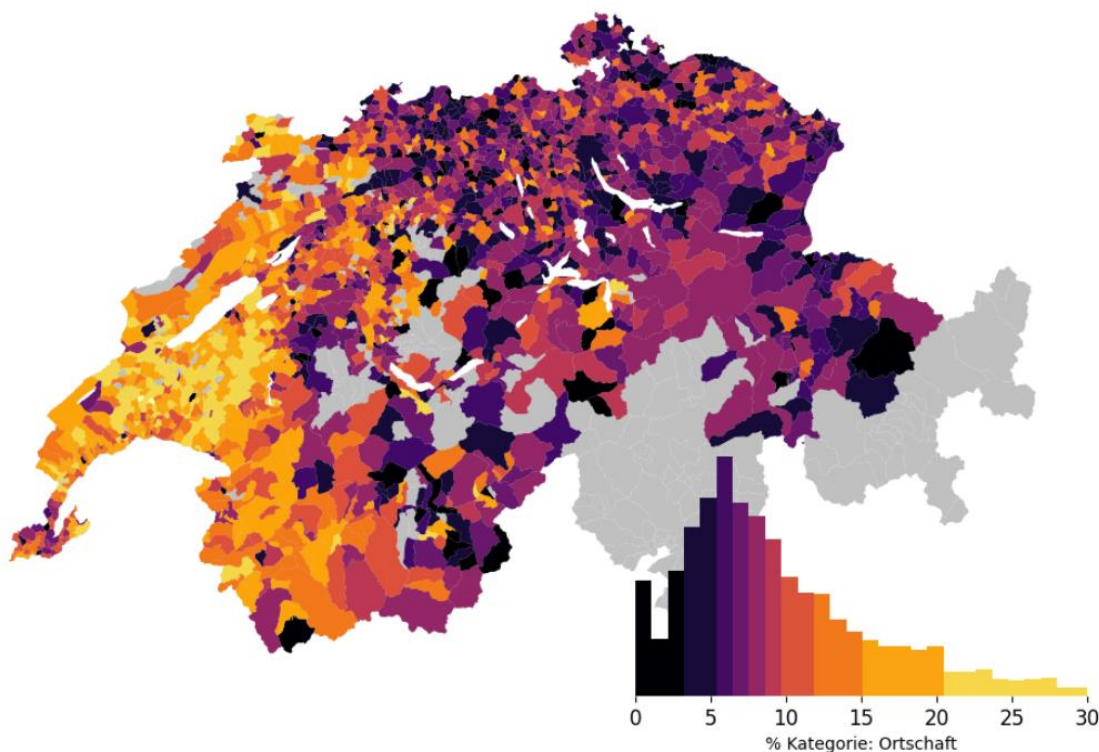


Abbildung 25 Anteil der nach Ortschaften benannten Strassen pro Gemeinde

Diese Choroplethenkarte zeigt den Anteil (in Prozent) der amtlichen Strassen, welche nach der Kategorie «Ortschaft» benannt sind pro Schweizer Gemeinde. Die Daten werden nach Quantilen in 10 Klassen dargestellt, wobei Gemeinden mit weniger als 10 amtlichen Strassen sowie italienisch- und romanischsprachige Gemeinden nicht berücksichtigt wurden (grau eingefärbt).

3.2.4 Personen

Nach einigen wenigen Persönlichkeiten sind in der Schweiz mehr als 40 Strassen benannt. Dazu zählen der legendäre Schweizer Freiheitskämpfer Wilhelm Tell, der Oberbefehlshaber der Schweizer Armee während des Zweiten Weltkriegs Henri Guisan und der Begründer des abendländischen Mönchtums Martin von Tours bekannt als Sankt Martin. Dahinter gibt es eine grössere Gruppe von Persönlichkeiten, nach welchen mehr als 10 Strassen benannt sind. Dazu zählen unter anderem der Wandermönch Gallus, der Dichter Gottfried Keller, der Politiker Stefano Franscini, der Reformator Huldrych Zwingli und der Philosoph Jean-Jacques Rousseau. Johanna Spyri, die Heidi-Erschafferin, folgt mit 8 nach ihr benannten Strassen als erste Frau.

Die Visualisierungen des Anteils an nach Personen benannten Strassen pro Gemeinde (Abbildung 26) und der Variation gegenüber den jeweiligen Nachbargemeinden (Abbildung 27) zeigen drei gut ausgeprägte schweizweite Muster. Erstens liegt

der Anteil in Städten verhältnismässig hoch. So grenzen sich auf der Choroplethenkarte (Abbildung 26) u.a. die Städte Zürich, Bern, St. Gallen oder Neuchâtel farblich gegenüber den umliegenden Gemeinden ab. Andere wie Winterthur weisen einen statistisch signifikant höheren Anteil als die Nachbargemeinden auf und sind deshalb auf der LISA-Cluster Visualisierung orange (Abbildung 27). Zweitens ist der Anteil in der Romandie höher als in der Deutschschweiz. Die grössten Städte der Romandie Genf, Lausanne, Fribourg und La Chaux-de-Fonds haben allesamt einen Anteil von mehr als 20%. Drittens haben viele Gemeinden im Oberwallis und in Graubünden ebenfalls einen hohen Anteil an, nach Personen benannten Strassen. Diese letzte Feststellung entspricht nicht der Realität, sondern ist auf Mängel der Zuordnung zurückzuführen. Zum einen sind viele Strassennamen in diesen Gemeinden in Mundart geschrieben. Zum anderen existieren auch in den Deutschschweizer Bündner Gemeinden viele romanisch- und italienischsprachigen Namen, respektive es handelt sich um gemischt-sprachige Gemeinden. Das Modell ist nicht auf diese beiden Sachverhalte trainiert und findet fälschlicherweise in diesen Strassennamen Personennamen auf Deutsch oder Französisch.

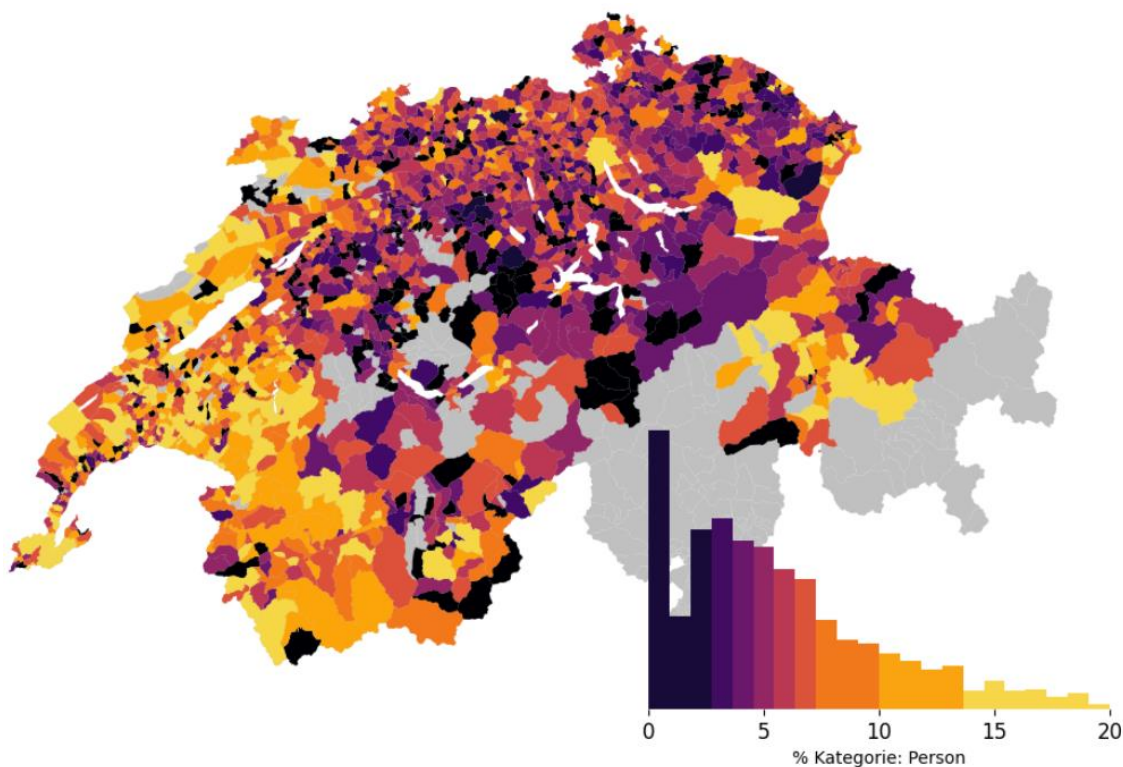


Abbildung 26 Anteil der nach Personen benannten Strassennamen pro Gemeinde

Diese Choroplethenkarte zeigt den Anteil (in Prozent) der amtlichen Strassen, welche nach Personen benannt sind pro Schweizer Gemeinde. Die Daten werden nach Quantilen in 10 Klassen dargestellt, wobei Gemeinden mit weniger als 10 amtlichen Strassen sowie italienisch- und romanischsprachige Gemeinden nicht berücksichtigt wurden (grau eingefärbt).

Von den schweizweit mehr als 10'000 nach Personen benannten Strassennamen konnte bei 3'261 das Geschlecht der Person, nach der die jeweilige Strasse benannt ist, erhoben werden. Davon sind 733 (23%) weiblich, 2527 (77%) männlich und eine ist eine Transfrau. In allen grösseren Schweizer Städten sind zurzeit deutlich mehr Strassen nach Männern als nach Frauen benannt. In der Stadt Zürich sind von den insgesamt 207 nach einer Person benannten Strassen (bei welchen das Geschlecht bekannt ist) 43 (20%) nach einer Frau und 164 (80%) nach einem Mann benannt. Die räumliche Verteilung der Strassen der Stadt Zürich nach Geschlecht scheint zufällig zu sein, allenfalls gibt es verhältnismässig viele lange nach Frau benannten Strassen (Abbildung 28). Die erste Frau, nach der in Zürich eine Strasse benannt wurde, heisst Eleonore Cramer-Mylius. Die Benennung geschah gemäss NZZ 1882 als Würdigung der Eleonorenstiftung⁴⁸. Mehrere Städte setzen sich aktuell dafür ein, diese starke Untervertretung der Frauen in der Namensgebung der Strassen auszugleichen. So wählte der Berner Gemeinderat bei der Strassenbenennung nach Personen auf dem Inselareal im Jahr 2020 ausschliesslich Frauennamen⁴⁹.

LISA Cluster Karte - Strassen benannt nach der Kategorie "Person" (Globaler Moran's I: 0.32, p-Wert: 0.001)

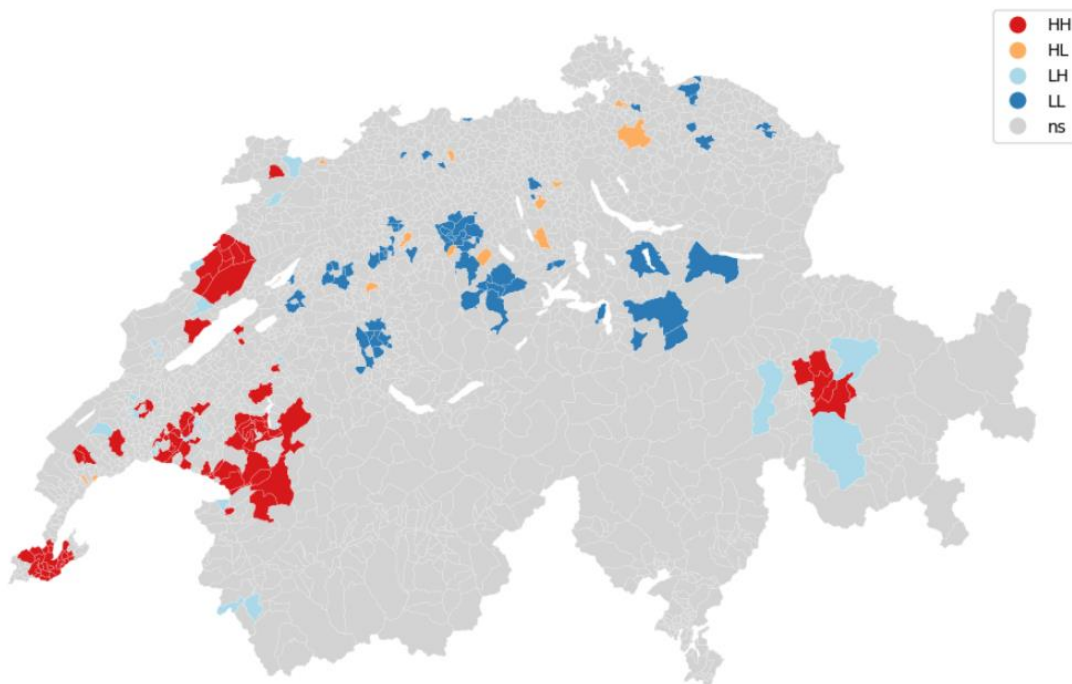


Abbildung 27 LISA-Cluster der nach Personen benannten Strassennamen pro Gemeinde

Die Karte zeigt die Variation des Anteils der amtlichen Strassen, welche nach Personen benannt sind (LISA-Cluster mit $p\text{-Wert} = 0.01$ und räumliche Gewichtung nach Queens Methode). Rot gefärbte Gemeinden sind Gemeinden mit einem hohen Anteil, die mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind (HH, z.B. die Region Genf). Blau Gemeinden haben einen niedrigen Anteil und sind von Gemeinden mit niedrigem Anteil umgeben (LL). Die orangen und hellblauen Gemeinden weisen einen grossen Unterschied im Anteil gegenüber den Nachbargemeinden auf (HL, z.B. Winterthur, und LH).

⁴⁸ <https://www.bluewin.ch/de/news/schweiz/gleichstellung-bei-strassennamen-werden-frauen-massiv-uebergangen-424612.html>. Abgerufen am 10. Oktober 2022

⁴⁹ https://www.bern.ch/mediencenter/medienmitteilungen/aktuell_ptk/namen-fuer-neue-strassen-und-wege-auf-dem-inselareal. Abgerufen am 10. Oktober 2022



Abbildung 28 Nach Personen benannte Strassen in Zürich nach Geschlecht

3.2.5 Lebewesen

Von den über 10'000 nach Lebewesen benannten Strassen sind mit über 1'000 am meisten nach *Rebe/Vigne* benannt. Dahinter folgen *Rose* und *Linde/Tilleul* mit über 700 Benennungen. Bei den Tieren wurden am meisten Strassen nach *Geiss/Chèvre*, *Wolf/Loup*, *Bär/Ours* und *Fuchs/Renard* benannt, allesamt über 200-mal. Weniger häufig, aber dennoch vorhanden, sind die Benennungen *Knoblauch* und *Elefant*.

Muster in der räumlichen Verteilung einer einzelnen Art von Lebewesen sind, zumindest für Laien, nur schwer erkennbar und die Verteilung eher zufällig. Eine Ausnahme bildet der Term *Vige*, welcher schön die Rebberge der Romandie erkennen lässt. Auf Kantonsebene liegt der Anteil, der nach Lebewesen benannten Strassen, praktisch in jedem Kanton um die 7, 8 Prozent. Auf Gemeindeebene hingegen variiert dieser Anteil zwischen 0 und über 50 Prozent (Gemeinden mit weniger als 10 Strassennamen sowie die italienisch- und romanischsprachigen Gemeinden sind nicht inbegriffen). In der Gemeinde Hagneck sind z.B. mehr als die Hälfte aller 23 Strassen nach Lebewesen benannt. In Hagneck findet man u.a. folgende Strassen: *Eichenstrasse*, *Akazienweg*, *Kirschbaumweg*, *Oleanderweg*, *Birkenweg*, *Pappelweg*, *Erlenweg*, *Lindenweg*, *Ulmengeweg*, *Ahornweg*, *Fliederweg*, *Arvenweg*, *Erlenweg*. Ein Blick auf die LISA-Cluster Karte (Abbildung 30) und die Karte des Anteils pro Gemeinde (Abbildung 29) zeigt denn auch diese grosse Variabilität auf Gemeindeebene.

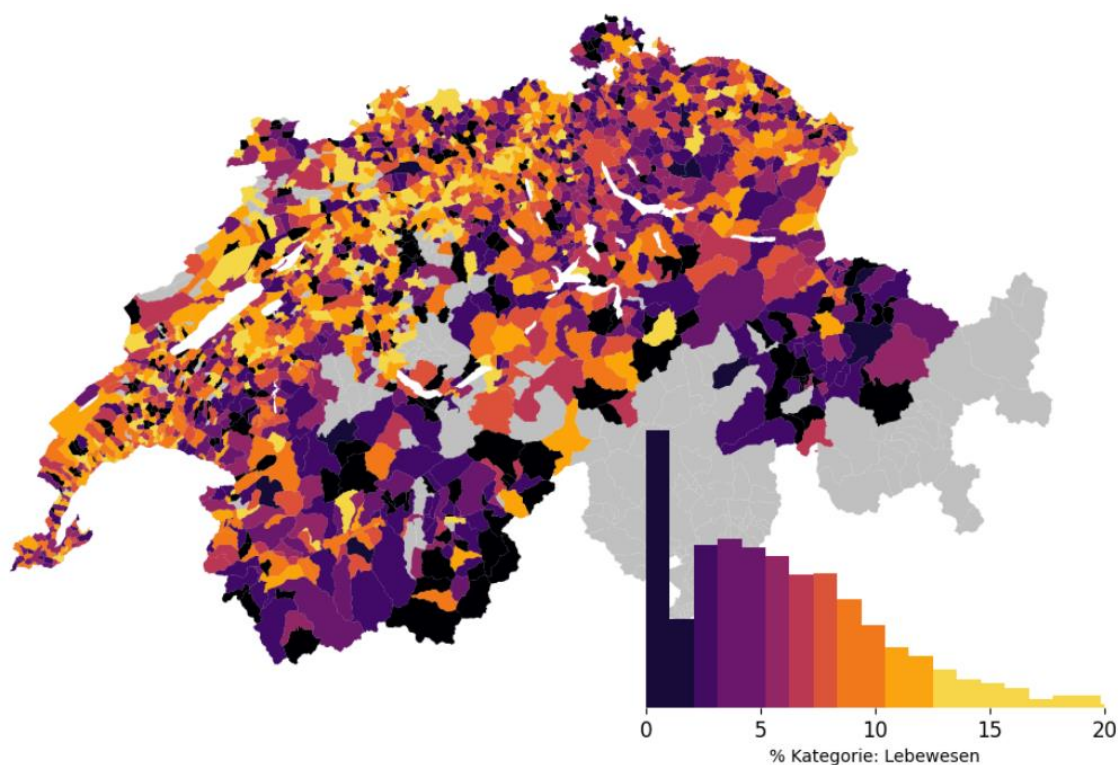


Abbildung 29 Anteil Strassennamen benannt nach Lebewesen pro Gemeinde

Diese Choroplethenkarte zeigt den Anteil (in Prozent) der amtlichen Strassen, welche nach der Kategorie «Lebewesen» benannt sind pro Schweizer Gemeinde. Die Daten werden nach Quantilen in 10 Klassen dargestellt, wobei Gemeinden mit weniger als 10 amtlichen Strassen sowie italienisch- und romanischsprachige Gemeinden nicht berücksichtigt wurden (grau eingefärbt).

Auf überregionaler Ebene lässt sich hingegen ein deutliches Muster erkennen. Der Korridor von der Region Aarau/Basel bis nach Lausanne beinhaltet deutlich mehr Gemeinden mit hohem Anteil an nach Lebewesen benannten Strassen als die anderen Regionen. Er beinhaltet auch fast alle Cluster an Gemeinden mit einem hohen Anteil, welche mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind. Im Gegensatz zu einigen anderen Benennungskategorien (u.a. Ortschaft oder Natur) verläuft dieses grossräumliche Muster über die deutsch-französische Sprachgrenze.

LISA Cluster Karte - Strassen benannt nach der Kategorie "Lebewesen" (Globaler Moran's I: 0.18, p-Wert: 0.001)

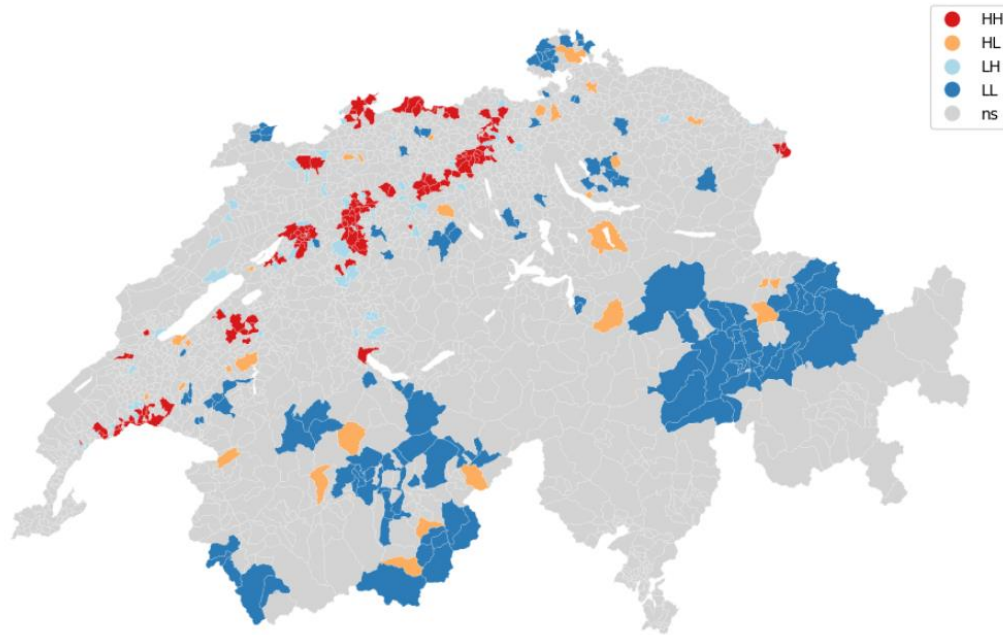


Abbildung 30 LISA-Cluster der nach Lebewesen benannten Strassennamen pro Gemeinde

Die Karte zeigt die Variation des Anteils der amtlichen Strassen, welche nach der Kategorie «Religion» benannt sind (LISA-Cluster mit p-Wert = 0.05 und räumliche Gewichtung nach Queens Methode). Rot gefärbte Gemeinden sind Gemeinden mit einem hohen Anteil, die mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind (HH, z.B. Region Biel). Blau Gemeinden haben einen niedrigen Anteil und sind von Gemeinden mit niedrigem Anteil umgeben (LL, z.B. Region Mittelbünden). Die orangen und hellblauen Gemeinden weisen einen grossen Unterschied im Anteil gegenüber den Nachbargemeinden auf (HL und LH, z.B. Neuchâtel).

3.2.6 Berge

Von den knapp 7'300 nach Bergen benannten Strassen sind die meisten nach den beiden grossen, die Schweiz durchlaufenden, Gebirgen Jura und Alpen benannt. Gezählt nach Namen einzelner Gipfel kommt die Rigi, mit 140 nach ihr benannten Strassen, am meisten vor. Auf die Rigi folgen der Säntis mit 124 und der Pilatus mit 89 Benennungen. Von der Berner Oberländer Bergen ist der Eiger mit 52 Benennungen die Nummer eins. Von den Jura-Bergen der Chasseral mit 29 Benennungen und in der Romandie der Mont-Blanc mit 21 Benennungen. Abbildung 31 zeigt, dass die nach spezifischen Gipfeln benannten Strassen, wenig überraschend, grossmehrheitlich im Sichtbereich der entsprechenden Gipfel liegen. Nach dem weltweit wohl bekanntesten Schweizer Gipfel, dem Matterhorn, ist keine einzige Strasse benannt. Dies ist doch einigermaßen überraschend, trotz der gegenüber den oben erwähnten Gipfeln, geringen Sichtbarkeit des Gipfels in bewohnten Gebieten.

Strassen benannt nach den Bergen Rigi:gelb, Säntis:blau, Eiger:grün, Chasseral:rot, Mont-Blanc:violett

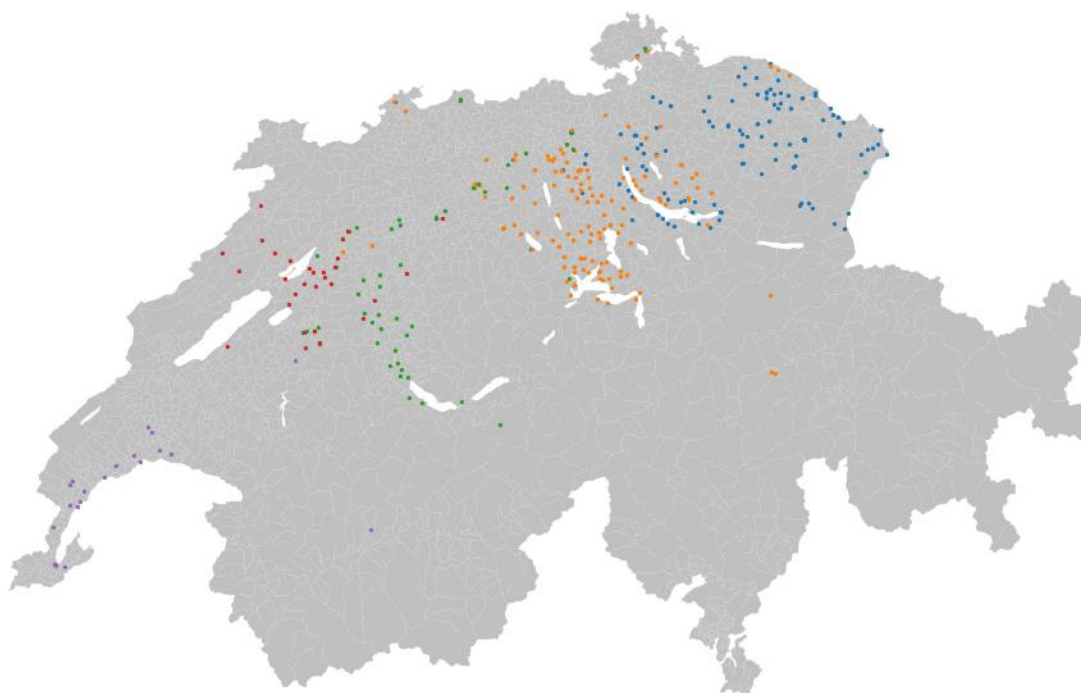


Abbildung 31 Verteilung der Strassen benannt nach den bekanntesten Bergen

Anzahl Strassen pro Berg: Rigi 140, Säntis 124, Pilatus 89, Chasseral 29, Mont-Blanc 21

Relativ betrachtet, wurden in der Region Zürcher Oberland besonders viele Strassen nach Bergen benannt (Abbildung 32). Eine kurze Internetrecherche zeigt denn auch, dass die Region sowohl historisch (viele Drumlins) als auch aktuell (viele Berggasthäuser) stark mit ihren Bergen verwurzelt ist⁵⁰. In der Region liegen mit Bachtel, Sch-nebelhorn und Hörnli auch mehrere Gipfel mit weiter Rund-sicht und die Alpen sind ebenfalls in Sichtweite.

⁵⁰ <https://www.zuerioberland-tourismus.ch/de/zuercher-oberland>. Abgerufen am 10. Oktober 2022

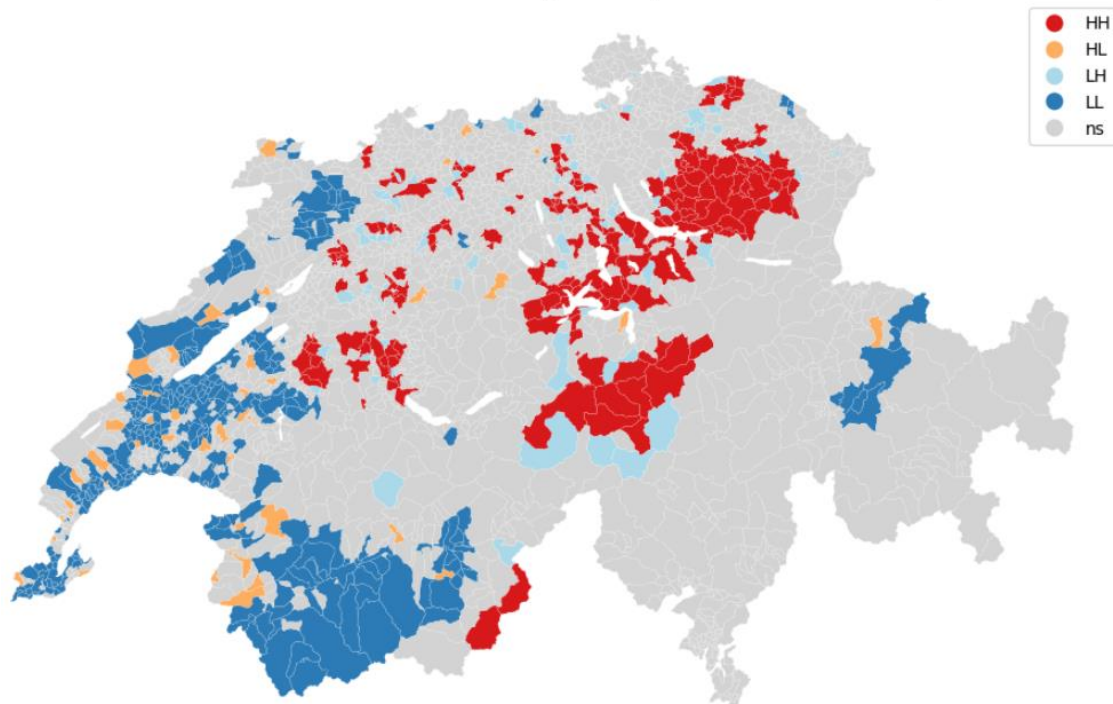


Abbildung 32 LISA-Cluster der nach Bergen benannten Strassennamen pro Gemeinde

Die Karte zeigt die Variation des Anteils der amtlichen Strassen, welche nach Bergen benannt sind (LISA-Cluster mit $p\text{-Wert} = 0.05$ und räumliche Gewichtung nach Queens Methode). Rot gefärbte Gemeinden sind Gemeinden mit einem hohen Anteil, die mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind (HH, z.B. Cluster Zürcher Oberland). Blau Gemeinden haben einen niedrigen Anteil und sind von Gemeinden mit niedrigem Anteil umgeben (LL, z.B. Region Genf). Die orangen und hellblauen Gemeinden weisen einen grossen Unterschied im Anteil gegenüber den Nachbargemeinden auf (HL und LH).

3.2.7 Gewässer

Von den fast 5'000 nach Gewässern benannten Strassen sind die meisten nach generischen Gewässernamen benannt wie z.B. *Faubourg du lac* oder *Am Bächli*. Davon kommt allein *see/lac* 1036-mal vor. Dies könnte u.a. daran liegen, dass die grössten Schweizer Seen, Ausnahme *Lac Léman*, ambige Namen tragen und oft nach Ortschaften benannt sind. Anders verhält sich dieser Sachverhalt insbesondere bei den grossen Flüssen. Am häufigsten Strassen (insgesamt 221) sind nach dem Rhein benannt, gefolgt von der Aare mit 102 Benennungen. Ein Blick auf die räumliche Verteilung (Abbildung 33) lässt auf den ersten Blick die Flussverläufe der grössten Schweizer Flüsse erkennen.

Strassen benannt nach den Gewässern Rhein:blau, Aare:gelb, Reuss:grün, Rhône/Rhone:rot, Limmat:violett

Anzahl Strassen pro Fluss:
 Rhein 221
 Aare 102
 Reuss 65
 Rhône/Rhone 34
 Limmat 30.

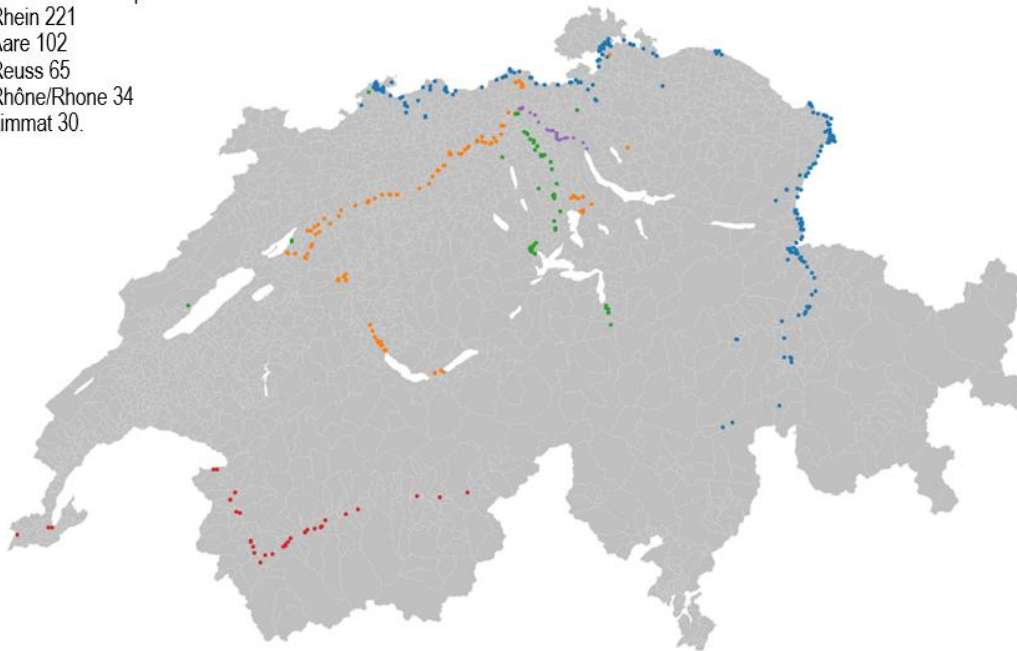


Abbildung 33 Verteilung der Strassen benannt nach den bekanntesten Flüssen

Auf Gemeindeebene zeigt sich, dass viele Gemeinden aus der Deutschschweiz, welche an einem See liegen, verhältnismässig viele Strassennamen nach Gewässern benannt besitzen (Abbildung 34). Dabei stechen insbesondere zwei Cluster heraus. Diejenigen der Gemeinden, welchem am Boden- oder am Zürichsee liegen. In Münsterlingen TG sind z.B. 14 von 72 Strassen nach Gewässern benannt. Es gibt in dieser Gemeinde u.a. folgende Strassennamen: *Seeblickstrasse, Seebachstrasse, Seeweg, Oberer Seeweg, Seedorfstrasse, Im Seegarte, Seestrasse, Seewiesenstrasse, Seegasse, Bachstrasse, Seewiesenweg*.

LISA Cluster Karte - Strassen benannt nach der Kategorie "Gewässer" (Globaler Moran's I: 0.3, p-Wert: 0.001)

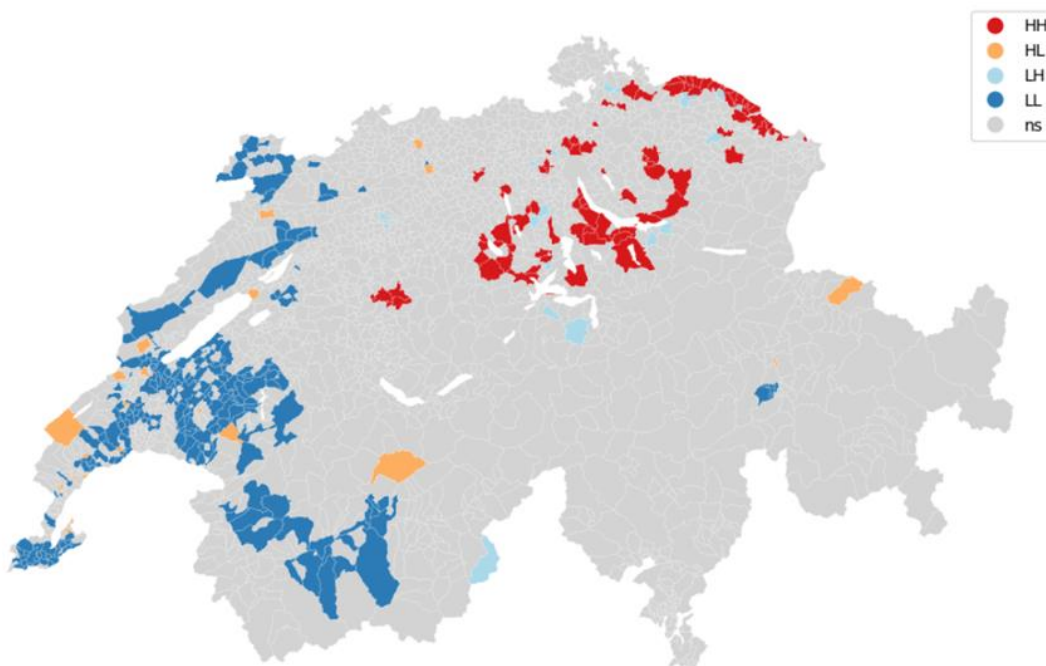


Abbildung 34 LISA-Cluster der nach Gewässern benannten Strassennamen pro Gemeinde

Abbildung 34 zeigt die Variation des Anteils der amtlichen Strassen, welche nach Gewässern benannt sind (LISA-Cluster mit p-Wert = 0.02 und räumliche Gewichtung nach Queens Methode). Rot gefärbte Gemeinden sind Gemeinden mit einem hohen Anteil, die mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind (HH, z.B. Gemeinden am Bodensee). Blau Gemeinden haben einen niedrigen Anteil und sind von Gemeinden mit niedrigem Anteil umgeben (LL, z.B. Region Unterwallis).

3.2.8 Berufe

Wegen der insgesamt geringen Anzahl von etwas über 1'600 an nach Berufen benannten Strassennamen sind aussagekräftige Muster nur spärlich vorhanden. Die am häufigsten vorhandenen Berufe sind Hirte, Senn, Schmied, Weber, Jäger, Müller, Fischer und Käser mit zwischen 50 und 150 Benennungen. Einzelne räumliche Muster sind dennoch auffindbar wie Abbildung 35 anhand der Verteilung der nach den Berufen Fischer und Käser benannten Strassen eindrücklich aufzeigt. Die nach dem Beruf Fischer benannten Strassen liegen häufig an einem grösseren Gewässer. Im Gegensatz dazu liegen die nach dem Beruf Käser benannten Strassen eher in ländlichen Gebieten abseits der grössten urbanen Zentren.

Strassen benannt nach den Berufen Fischer/Pêcheur:blau, Käser/Laiterie:orange



Abbildung 35 Verteilung der Strassen benannt nach Berufen

3.3 Evaluation

Wir bewerten die Klassifizierungsleistung unseres Modells der Kategorisierung der spezifischen Strassenbenennungen anhand einer Konfusionsmatrix (Abbildung 36) durch den Vergleich mit unserem Verifikationsdatensatz (Kapitel 2.3.4). Jede Zeile der Matrix steht für die Strassennamen des Verifikationsdatensatzes oder anders genannt *Ground Truth*, also der tatsächlichen Kategorie. Jede Spalte steht für Zuteilung der Strassennamen an eine Kategorie durch unser Modell vorausgesagt. So wurden z.B. von den insgesamt 54 Instanzen der Kategorie Berg in der *Ground Truth* 36 auch korrekt klassifiziert. 17 wurden fälschlicherweise der Kategorie Person zugeordnet, 4 wurden fälschlicherweise keiner Kategorie zugeordnet und 6 fälschlicherweise auf weitere Kategorien verteilt. Bei allen Klassen, einschliesslich der Klasse der nicht kategorisierbaren Strassennamen, hier *Andere* genannt, werden mindestens die Hälfte aller Strassennamen richtig vorausgesagt. Dieser Sachverhalt ist graphisch, durch die gut erkennbare Diagonale von links oben nach rechts unten, gut ersichtlich. Ebenfalls gut ersichtlich sind die hohen Anzahlen an nach Personen benannten Strassen im Verifikationsdatensatz und den damit verbundenen hohen Anzahlen an falsch kategorisierten nach Personen benannten Strassen. So wurden z.B. die nach Personen benannten Strassen *Ochsenbeinstrasse*, *Lilienthalstrasse* und *Angelikaweg* fälschlicherweise Lebewesen klassifiziert und die *Hallwilstrasse* und der *Rougemontweg* fälschlicherweise als Ortschaft. In der Konfusionsmatrix ebenfalls gut ersichtlich sind mit der Spalte *Andere* der, durch unser Modell nicht kategorisierten Strassennamen, wie z.B. die *Kurfisstenstrasse* oder das *Chemin du Muveran*, welche nicht als nach einem Berg benannte Strasse erkannt wurden. Weitere Beispiele an nicht kategorisierten Strassennamen sind die *Färberstrasse* oder die *Strehlgasse*, welche nach einem Beruf benannt wären.

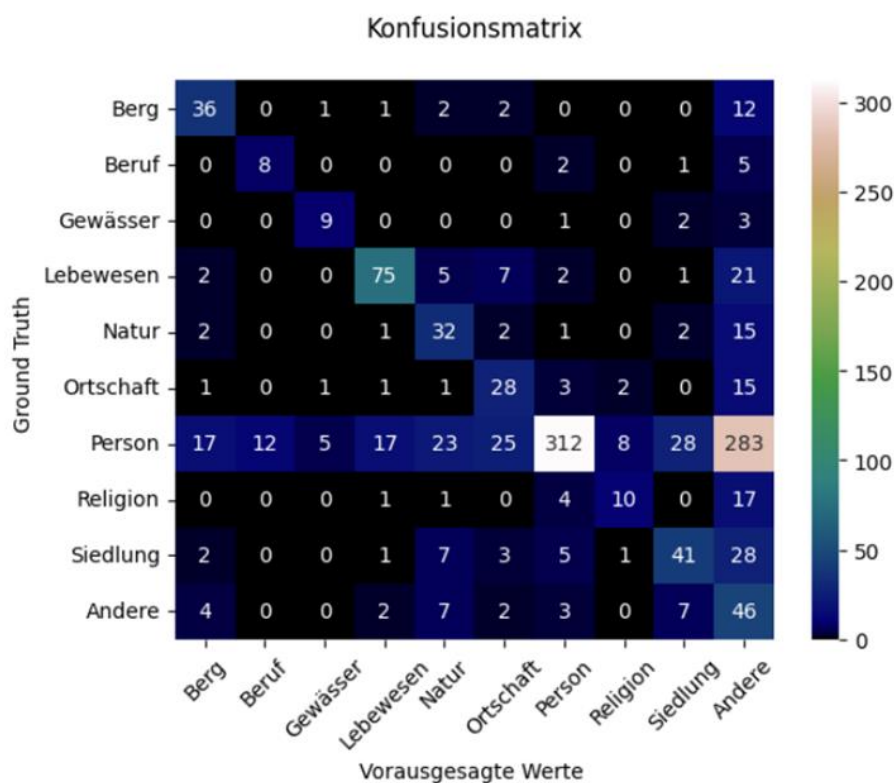


Abbildung 36 Konfusionsmatrix zur Evaluation des Modells der Kategorisierung der spezifischen Strassenbenennung

Gleiches gilt für die nach Lebewesen benannten Strassen *Chemin des Lys*, *Rue des Cygnes*, *Mirabellenweg*, die nach Ortschaften benannten Strassen *Avenue d'Evian*, *Avneu de Rhodanie*, *Rue de Madagascar*, die nach Siedlungsnamen benannten Stras-

sen *Gassstrasse*, *Lagerstrasse*, *Spielweg* sowie die nach Personen benannten Strassen *Pestalozzistrasse*, *Gotthelfstrasse*, *Ampèresteg*. Diese Beispiele zeigen zum einen, dass unser Modell noch einiges an Optimierungspotential besitzt, z.B. durch Kompletieren der Kataloge, aber auch, dass viele doppeldeutige Strassennamen ohne zusätzliche Informationen nicht ohne weiteres kategorisiert werden können. Wie z.B. die *Avenue de Milan*, welche wir als nach einem Lebewesen benannt vorausgesagt haben, aber nach der italienischen Stadt benannt ist.

	precision	recall	f1-score	support
<i>Berg</i>	0.56	0.67	0.61	54
<i>Beruf</i>	0.4	0.5	0.44	16
<i>Gewässer</i>	0.56	0.6	0.58	15
<i>Lebewesen</i>	0.76	0.66	0.71	113
<i>Natur</i>	0.41	0.58	0.48	55
<i>Ortschaft</i>	0.41	0.54	0.46	52
<i>Person</i>	0.94	0.43	0.59	730
<i>Religion</i>	0.48	0.3	0.37	33
<i>Siedlung</i>	0.5	0.47	0.48	88
<i>Andere</i>	0.1	0.65	0.18	71
<i>accuracy</i>			0.49	1227
<i>macro avg</i>	0.51	0.54	0.49	1227
<i>weighted avg</i>	0.75	0.49	0.55	1227

Tabelle 11 Kennzahlen zur Performance des Modells zur Vorhersage des spezifischen Strassenkategorie

In Tabelle 11 sind einige Kennzahlen zur Performance des Modells zur Vorhersage der spezifischen Strassenkategorie ersichtlich. Die Genauigkeit, auf Englisch *precision*, beschreibt den Anteil an korrekt vorausgesagten Strassennamen einer Kategorie. So ist z.B. die Genauigkeit der Klasse *Person* mit 0.94 sehr hoch, was bedeutet, dass die von uns als *Person* klassierten Strassen, effektiv Strassen mit hoher Wahrscheinlichkeit sind. Die Genauigkeit ist, wegen der stark unterschiedlichen Anzahl Strassennamen pro Klasse des Verifikationsdatensatzes, allerdings nur bedingt aussagekräftig. Die Sensitivität auch *Trefferquote* genannt, auf Englisch *recall*, gibt die Wahrscheinlichkeit an, mit der nach einer Kategorie benannte Strassen auch korrekt klassifiziert werden. Dieser Wert ist bei der Klasse *Person* mit 0.43 viel tiefer als die Genauigkeit. Dies bedeutet, dass viele nach Personen benannten Strassen nicht als solches erkannt werden. Beide Werte sind wichtig und hängen voneinander ab. Der F1-Wert kombiniert die Genauigkeit und die Sensitivität mittels des gewichteten harmonischen Mittels. Den höchsten F1-Wert besitzt die Klasse *Lebewesen*, welche unser Modell auffallend gut voraussagt. Im Gegensatz dazu sagt unser Modell Strassen, welche nach religiösen Namen oder Berufen benannt sind, relativ schlecht voraus. Der makrogemittelte F1-Wert wird aus dem Mittelwert aller F1-Werten berechnet. Beim gewichteten durchschnittlichen F1-Wert wird zusätzlich die Anzahl Werte pro Klasse berücksichtigt. Uns sind keine anderen Modelle bekannt, mit welchem wir diese Werte vergleichen könnten.

4 Diskussion

In den folgenden Kapiteln wird auf das allgemeine Modell und einige der verwendeten Pythonbibliotheken eingegangen. Die Diskussion zu den Resultaten und den daraus erstellten Karten und Diagrammen erfolgt aus Gründen der Übersichtlichkeit jeweils direkt im Kapitel 3.2.

4.1 Auswertung Strassennamen

Es hat sich gezeigt, dass die Auswertung der generischen Benennungen relativ einfach zu bewerkstelligen war. Bei den Resultaten wird, ausser in den Regionen mit einem hohen Anteil an *Mundart-Strassennamen*, von einer hohen Genauigkeit ausgegangen, welche jedoch nicht empirisch ausgewiesen wird.

Bei den spezifischen Benennungen müssen die Resultate, etwa die Angaben der erstellten Karten und Diagramme, kritisch auf deren Aussage betrachtet werden. Das gewählte Vorgehen definierte neun Kategorien, an welchen die Strassennamen zugeteilt wurden. Da die Zuweisung jeweils pro Kategorie erfolgte, konnten die einzelnen Strassennamen an mehrere Kategorien zugeteilt werden. Somit bestand eine Herausforderung bei der Auflösung von Mehrfachzuteilungen. In dieser Arbeit wurde dafür manuell eine Rangliste über die Kategorien erstellt mit welcher die definitive Zuerteilung erfolgte, was nicht optimal ist.

Mit einer Quote von 60% der Strassennamen, welche mindestens in einer Kategorie enthalten sind, wurde ein unerwartet hoher Wert erreicht. Dieser Wert ist jedoch mit Vorsicht zu geniessen. Die Evaluation der Zuteilung ergab, eine Genauigkeit von rund 50%, wie in der Konfusionsmatrix in Abbildung 36 ersichtlich ist. Diese gibt bei einzelnen Kategorien auch eine hohe Präzision vor. Es muss jedoch davon ausgegangen werden, dass die Evaluation nicht sehr aussagekräftig ist. Dies auf Grund des Prüfdatensatzes, der *Ground Truth*, welcher sehr einseitig auf städtischen Strassen mit Bezug zu Personen besteht. Dieses Problem konnte nicht reduziert werden. Sämtliche Versuche zur Erstellung weiterer Prüfdatensätze erbrachten keine befriedigenden Resultate.

Insgesamt muss diese Zuteilung der spezifischen Benennungen als eine explorative Arbeit betrachtet werden, da offensichtliche Probleme des Modells aus zeitlichen Gründen nicht verbessert werden konnten. Die Resultate ergaben dennoch interessante und auch überraschende Erkenntnisse. Diese können als These genutzt werden, um etwa weitere Auswertungen vorzunehmen oder mit Fachkräften historische Prozesse genauer zu untersuchen.

4.2 Pandas / Geopandas

Für das Analysieren und Bearbeiten der zu verarbeitenden Daten wurde ausschliesslich pandas respektive Geopandas verwendet. Die beiden Bibliotheken erlauben ein sehr schnelles Laden von vielen Daten. Die Bearbeitung der rund 220'000 Strassen aus dem Amtlichen Verzeichnis der Strassen war ohne Einschränkungen und Wartezeiten möglich. Auch das Laden von Daten aus OpenStreetMap mittels Overpass kann einfach bewerkstelligt werden.⁵¹

Als Vorteil bei der Analyse von Geodaten mit geopandas hat sich die interaktive Karte herausgestellt. Sie wird mittels `geopandas.GeoDataFrame.explore`⁵² aufgerufen und generiert eine leaflet basierte Karte⁵³. Ein Nachteil dieser Karte ist, dass sie in GitHub

⁵¹ Beispiel in Notebook du-Wikidata.ipynb ersichtlich

⁵² <https://geopandas.org/en/stable/docs/reference/api/geopandas.GeoDataFrame.explore.html>. Abgerufen am 21. November 2022

⁵³ Beispiel in du-StrassenVZ-BS.ipynb ersichtlich

nicht dargestellt wird. Für die Darstellung von Karten auf GitHub wurde deshalb entweder die Geopanda Plot Funktion oder matplotlib verwendet.

4.3 SPARQLWrapper

Um Objekte von Wikidata via dem Wikidata Query Service (WDQS) abzufragen, hat sich SPARQLWrapper bewährt. SPARQL ist die Abfragesprache für strukturierte Datenbanken.

Die grösste Herausforderung, bei der Abfrage über WDQS, stellte sich vor allem bei längeren Abfragen (>5min). Dabei ist es wichtig, den «Access best practices» des WDQS zu folgen. Vor allem zwei Punkten haben sich dabei herausgehoben:

1. Follow the User-Agent policy -- send a good User-Agent header
2. If you get a 429 Too Many Requests response, stop sending further requests for a while (see the Retry-After response header)⁵⁴

Daraus ergab sich für den SPARQLWrapper folgenden Header mit der User-Agent Angabe:

```
wdUrl = "https://query.wikidata.org/sparql"
user_agent = 'Streetnamequery/1.0 (https://github.com/CaptainInler/cassda-
zertifikatsarbeit)'
sparql = SPARQLWrapper(wdUrl, agent=user_agent)
```

Die effektive Abfrage des WDQS-Endpunktes:

```
for i in range(0,5):
    try:
        results = sparql.query()
        break
    except HTTPError as e:
        print(f"{i} Anfragen ausgeführt")
        if e.status == 429:
            print(f'Statuscode 429 aufgetreten: Anfrage geht in {e.headers.get("retry-
after")})sec weiter')
            sleep(int(e.headers.get("retry-after"))+2)
            continue
        else:
            raise
```

Der for-Loop stellt sicher, dass die Abfrage wiederholt wird, sollte sie durch einen 429 Status unterbrochen werden. Der sleep-Timer wurde bewusst um die zwei Sekunden Sicherheitsmarge erhöht.

Eine mögliche Alternative zum SPARQLWrapper besteht mit <https://pypi.org/project/Wikidata/>. Diese Bibliothek wird allerdings seit zwei Jahren nicht mehr aktiv gepflegt⁵⁵ und deshalb in dieser Arbeit nicht näher untersucht. Zudem sind die meisten Beispiele im Internet zu automatisierten Abfragen des WDQS mit dem SPARQLWrapper beschrieben.

4.4 SpaCy

Die diversen Funktionalitäten von SpaCy erwiesen sich in dieser Arbeit oft an der Grenze zwischen geeignet und ungeeignet. So haben wir uns z.B. knapp für die Verwendung von NER und gegen die Lemmatisierung entschieden. Ein aus unserer Sicht vielversprechender Ansatz zur Nutzung von SpaCy zur Kategorisierung von Strassenamen ist die individuelle Anpassung der Standard-NER-Klassen oder das Hinzufügen von eigenen Klassen z.B. durch Integration von Katalogen⁵⁶.

⁵⁴ https://www.wikidata.org/wiki/Wikidata:Data_access#Access_best_practices. Abgerufen am 20. November 2022

⁵⁵ <https://github.com/dahlia/wikidata/commits/main>. Abgerufen am 20. November 2022

⁵⁶ <https://www.machinelearningplus.com/nlp/training-custom-ner-model-in-spacy>. Abgerufen am 25.11.2022

4.5 Pandarallel

Normalerweise werden Daten in pandas sequenziell abgearbeitet. Bei den rund 172'000 Strassennamen dauert eine Analyse aller Datensätze durch spacy je nach Computerleistung eine lange Zeit (>30min). Mittels pandarallel werden die Abfragen parallelisiert durchgeführt⁵⁷. Dazu muss die eigentliche Abfrage in eine separate Funktion ausgegliedert werden.

```
def checkForEntity(num):
    doc = nlp(num) #spacy Natural language processing
    if doc.ents:
        for ent in doc.ents:
            if len(num) == ent.end_char: #for two word term. e.g.: General Guisan
                return ent.label_

streetnames["name_ENT"] = streetnames["name"].parallel_apply(checkForEntity)
```

Pandarallel hat sich als gutes Werkzeug herausgestellt. Die einzelnen Abfragen konnten einfach parallelisiert werden und boten eine merkbare Beschleunigung bei Manipulationen aller Strassennamen.

4.6 Erstellung von Graphiken und Statistiken

Die von uns getesteten Python-Bibliotheken, welche sich zum Erstellen von Statistiken und Visualisierungen am besten eignen sind matplotlib und seaborn. Beide sind sehr benutzerfreundlich, einzig die Optimierung oder das *Customizing* der Farben und Legenden verlief nicht immer geradlinig. Für einige Darstellungen mit speziellen Anforderungen haben wir andere oder zusätzliche Bibliotheken verwendet. U.a. datashader für die Darstellung eines grossen Geometriedatensatzes oder legendgram für die Darstellung einer Choroplethenkarte, welche mit einem Histogramm verknüpft ist. Für die Erstellung der Konfusionsmatrix und Berechnung der Metriken erwiesen sich seaborn und sklearn als optimal. Schliesslich erwies sich pySAL als optimale Wahl für die Berechnung von räumlichen Autokorrelationen.

⁵⁷ <https://pypi.org/project/pandarallel/>. Abgerufen am 21. November 2022

5 Fazit und Ausblick

In dieser Zertifikatsarbeit wurden Strassennamen anhand der Herkunft des Namens und deren geografischen Lage analysiert. Das Data Mining erfolgte anhand der Phasen des CRISP-DM Modells, welches die verschiedenen Arbeiten und auch die Datenablage insgesamt gut strukturierten. Wie im *life cycle* Modell von CRISP Modell vorgesehen, verlief diese Arbeit iterativ, was zu Wiederholungen der Phasen bei neuen Erkenntnissen führte.

Für die Durchführung der einzelnen Arbeitsschritte wurden JupyterLab Notebooks erstellt, welche auf GitHub einsehbar sind. Die JupyterLab Notebooks erlauben auf eine übersichtliche Art die einzelnen Arbeitsschritte abzuspeichern, da auch die resultierenden Grafiken dargestellt sind. Somit kann relativ einfach dem erstellten Code gefolgt werden, ohne ihn ausführen zu müssen. Das Arbeiten mit den Pythonbibliotheken stellte sich nach einer steilen Lernkurve als überaus produktiv und sinnvoll heraus. Vor allem Pandas / Geopandas bieten sich als gutes Werkzeug für die Arbeit mit den tabellarisch flach strukturierten Grundlagedaten dar. Dank den einfachen Möglichkeiten der Datenvisualisierungen waren auch schnell erste Resultate sichtbar, die zu weiteren Nachforschungen animierten.

Bei der Analyse der generischen Strassenbenennungen stellte sich heraus, dass diese Auswertung einfacher ist und schneller zu Resultaten führten. Dabei zeigten sich interessante geografische Muster.

Bei den spezifischen Strassenbenennungen wurden neun Kategorien gebildet. Anhand verschiedener Methoden wurden die Strassennamen diesen Kategorien zugeteilt. Diese Zuteilung stellte sich als der anspruchsvollste Teil dieser Arbeit heraus, welcher durchaus verbessert werden kann. Die durchgeführte Verifikation der aufbereiteten Daten zeigt ein für die erste Auswertung zufriedenstellendes Resultat.

Als Ausblick ist zu erwähnen, dass die erstellten Auswertungen, insbesondere der spezifischen Benennungen, gut mit Anpassungen an den Abfragen der Kataloge und Wikidata verbessert werden können. So könnte das Model zusätzlich mit Faktoren der räumlichen Nähe zu Objekte ergänzt werden. Bei Personen ist etwa eine Berücksichtigung der Person mit einem Bezug zum entsprechenden Ort, wie über den Geburts- oder Todesort sowie der Wirkungsstätte denkbar. Zusätzliche wäre auch denkbar weitere Datensätze, etwa *swissnames*⁵⁸, in die Auswertung mit einzubeziehen.

Der Datensatz könnte auch in Form einer explorativen Applikation zur Verfügung gestellt werden, ähnlich wie z.B. die Applikation *StreetToPerson* (Gurtovoy, et al., 2022).

Im Bereich der Auswertungen sind keine Grenzen gesetzt. So ist ein Vergleich der Strassen benannt nach dem alten Rom mit den römischen Handelswegen denkbar. Ebenfalls aufschlussreich könnte eine Auswertung mittels zusätzlichem Zeitparameter sein, etwa in welcher Epoche die Strassen nach welcher Kategorie benannt wurden. Hilfreich dabei sind die beiden OGD-Datensätze von Basel-Stadt und der Stadt Zürich.

Auch ein Blick über die Landesgrenze könnte spannende Erkenntnisse bieten. Etwa die These, ob in den Nachbarländern tatsächlich viel mehr Strassen nach Personen benannt sind.

⁵⁸ <https://www.swisstopo.admin.ch/de/geodata/landscape/names3d.html>. Abgerufen am 23. November 2022

6 Glossar

6.1 Odonym

Gelegentlich auch *Hodonym*. Von griechisch ὁδός, *hodós*, *Strasse* und ὄνομα, *ónoma*, *Name*.

Wird in dieser Arbeit als Synonym für Strassenname verwendet. Zum Beispiel ist das Wort *Riggerbachstrasse*, welches die entsprechende Strasse in Olten beschreibt, ein Odonym.

6.2 Etymologie

Von griechisch ἔτυμος *étymos* (*wahr, wahrhaft, echt, wirklich*) und λόγος *lógos* (*-kunde, -lehre, -wissenschaft*). Lehre vom wahren Sinn der Wörter.⁵⁹

⁵⁹ <https://www.dwds.de/wb/Etymologie>. Abgerufen am 21. November 2022

7 Quellenverzeichnis

- Purves, Ross S. 2022.** *Part 2: Nuts and bolts: Information retrieval basics*. Folien zur Vorlesung "Geographic Information Retrieval" (unveröffentlicht), Olten : Fachhochschule Nordwestschweiz, 2022.
- Bancilhon, Melanie, et al. 2021.** *Streetonomics: Quantifying Culture Using Street Names*. 2021. <https://doi.org/10.1371/journal.pone.0252869>.
- Brunner, Tobias J. und Pruves, Ross S. 2008.** Spatial Autocorrelation and Toponym Ambiguity. *GIR '08: Proceedings of the 5th Workshop on Geographic Information Retrieval*. 2008.
- Bühler, Michael und Gmür, Daniel. 2022.** Projektskizze. *Spatial Data Mining mit Python: Möglichkeiten und Tools*. Olten : Fachhochschule Nordwestschweiz, 2022. S. 2. unveröffentlicht.
- Bundesamt für Landestopografie swisstopo. 2021.** «Spezifikation Amtliches Verzeichnis der Strassen / Amtliches Verzeichnis der Gebäudeadressen». [Online] 17. März 2021. [Zitat vom: 06. November 2022.] <https://www.swisstopo.admin.ch/de/geodata/amtliche-verzeichnisse/strassenverzeichnis.html>.
- Bundesamt für Statistik, Sektion Mobilität.** <https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaet-verkehr/erhebungen/stl.html>. [Online] [Zitat vom: 11. 11 2022.]
- GeoNV.** Verordnung über die geografischen Namen vom 01. Juli 2017. SR 510.625.
- Gurtovoy, Daria und Gottschalk, Simon. 2022.** *Linking Streets in OpenStreetMap to Persons in Wikidata*. New York, NY, USA : ACM, 2022. S. 4. <https://doi.org/10.1145/3487553.3524267>.
- IBM Corporation. 2021.** «CRISP-DM Help Overview». *SPSS Modeler*. [Online] 17. August 2021. [Zitat vom: 09. September 2022.] <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>.
- Kohler, Barbara; Schweizer Radio und Fernsehen. 2016.** SRF. *Das sind die verrücktesten Strassennamen der Schweiz*. [Online] 20. April 2016. [Zitat vom: 13. November 2022.] <https://www.srf.ch/radio-srf-1/radio-srf-1/das-sind-die-verruecktesten-strassennamen-der-schweiz>.
- KVAV-ZH.** Kantonale Verordnung über die amtliche Vermessung vom 27. Juni 2012.
- seaborn.kdeplot.** <https://seaborn.pydata.org/generated/seaborn.kdeplot.html>. [Online] [Zitat vom: 12. 11 2022.]
- Senn, Lea; Fehr, Reto; watson % FixxPunkt AG. 2022.** watson. *Das sind die speziellsten, witzigsten und häufigsten Strassennamen der Schweiz*. [Online] 24. Januar 2022. [Zitat vom: 13. November 2022.] https://www.watson.ch/wissen/schweiz/133417259-das-sind-die-witzigsten-strassennamen-der-schweiz#h4_1.
- SV-BE.** Strassenverordnung des Kanton Bern vom 01.01.2020.

8 Anhang

8.1 Übersicht Kapitel Notebook und Autor

Kapitel	Notebook	Autor
Vorwort		dg
1 Einleitung		dg
2 Vorgehen		dg
2.1 Business understanding		dg
2.2 Data understanding		dg
2.2.1 Aufbau Odonym		dg
2.2.2 Amtliches Verzeichnis der Strassen	du-StrassenVZ.ipynb	dg
2.2.3 Wikidata	du-Wikidata.ipynb	dg
2.2.4 Strassennamen Kanton Basel-Stadt	du-StrassenVZ-BS.ipynb	dg
2.2.5 Strassennamenverzeichnis Stadt Zürich		dg
2.2.6 OpenStreetMap		mb
2.2.7 Kataloge		mb
2.3 Data preparation		mb
2.3.1 Data Cleaning	Extract_street_terms_from_street_names.ipynb	mb
2.3.2 Extraktion generische Strassenbenennung	Extract_street_terms_from_street_names.ipynb	mb
2.3.3 Extraktion spezifische Strassenbenennung	Extract_street_terms_from_street_names.ipynb spacy_windows.ipynb	mb
2.3.4 Verifikationsdatensatz	verification.ipynb verification_merge.ipynb	mb
2.3.5 Weitere Verifikationsansätze	dp-StrassenBS-Menschen.ipynb dp-StrassenBS-Wikidata.ipynb	dg
2.4 Modelling		mb

2.4.1 Wahl und Definition der Kategorien		mb
2.4.2 Wikidata	mapping wiki person.ipynb	dg
2.4.3 Zuordnen anhand von Katalogen Zuordnen	mapping%20dictionary.ipynb	mb
2.4.4 Named-Entity Recognition	spacy_windows.ipynb	mb
2.4.5 Selbstkartierung Selbst	mapping%20dictionary.ipynb Merge_mappings_to_specific_categories.ipynb	mb
3 Resultate		mb/dg
3.1 Ablage GitHub		dg
3.2 Verteilung und Statistiken		mb
3.2.1 Generische Benennung	street_designations_statistics.ipynb street_designations_communes.ipynb street_designations_line_geometries.ipynb	mb
3.2.2 Spezifische Benennung bis und mit 0 Abbildung 34 zeigt die Variation des Anteils der amtlichen Strassen, welche nach Gewässern benannt sind (LISA-Cluster mit p-Wert = 0.02 und räumliche Gewichtung nach Queens Methode). Rot gefärbte Gemeinden sind Gemeinden mit einem hohen Anteil, die mit Gemeinden mit ebenfalls hohen Anteilen umgeben sind (HH, z.B. Gemeinden am Bodensee). Blau Gemeinden haben einen niedrigen Anteil und sind von Gemeinden mit niedrigem Anteil umgeben (LL, z.B. Region Unterwallis). Berufe	street_designations_statistics.ipynb street_designations_communes.ipynb Extract_street_terms_from_street_names.ipynb street_designations_line_geometries.ipynb	mb
3.3 Evaluation	verification.ipynb verification_merge.ipynb	mb
4 Diskussion		mb/dg
4.1 Auswertung Strassennamen		mb/dg
4.2 Pandas / Geopandas		dg
4.3 SPARQLWrapper		dg

4.4 SpaCy		mb
4.5 Pandarallel		dg
4.6 Erstellung von Graphiken und Statistiken		mb
5 Fazit und Ausblick		dg

8.2 Wikidataquery: Personen mit Bezug zur Schweiz

```
SELECT DISTINCT ?item ?itemLabel ?itemDescription
WHERE {
  {?item wdt:P31 wd:Q5.
   ?item wdt:P19/wdt:P131* wd:Q39.}
  UNION
  {?item wdt:P31 wd:Q5.
   ?item wdt:P20/wdt:P131* wd:Q39.}
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],de" }
```

8.3 Wikidataquery: Strassen mit Statement benannt nach

```
# Für die Darstellung als Karte das Resultat als «Map» anzeigen lassen.
SELECT DISTINCT * WHERE {
  ?item wdt:P31 wd:Q79007;
  wdt:P138 ?object;
  wdt:P131* wd:Q39;
  wdt:P625 ?geo .
}
```

8.4 Wikidataquery: Label von wikidataobjekten

```
# %s ist die Variable, welche durch einen effektiven Namen ersetzt werden muss
SELECT ?subject ?subjectLabel ?instanceLabel WHERE {
  ?subject rdfs:label "%s"@de;
  wdt:P31 ?instance.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de" . }
```

8.5 Wikidataquery: Familienname von Personen

```
# %s ist die Variable, welche durch einen effektiven Namen ersetzt werden muss
SELECT ?subject ?subjectLabel WHERE {
  ?subject wdt:P734 ?familyname;
  wdt:P31 wd:Q5.
  ?familyname rdfs:label "%s"@de;
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de" . }
```

8.6 Wikidataquery: Label von allen Objekten

```
# %s ist die Variable, welche durch einen effektiven Namen ersetzt werden muss
SELECT ?subject ?subjectLabel ?instanceLabel WHERE {
  ?subject rdfs:label "%s"@de;
  wdt:P31 ?instance.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de" . }
```

8.7 Wikidataquery: Abfrage über Suche

```
SELECT ?subject ?subjectLabel ?givennameLabel ?familynameLabel ?sexLabel ?birth ?death
?placebirth ?placedeath ?imageLabel
WHERE {
  hint:Query hint:optimizer "None".
  SERVICE wikibase:mwapi {
    bd:serviceParam wikibase:api "EntitySearch";
                        wikibase:endpoint "www.wikidata.org";
                        mwapi:search "%s";
                        mwapi:language "de".
    ?subject wikibase:apiOutputItem mwapi:item .
  }
  ?subject wdt:P31 wd:Q5.
  OPTIONAL {?subject wdt:P735 ?givenname;}
  OPTIONAL {?subject wdt:P734 ?familyname;}
  OPTIONAL {?subject wdt:P21 ?sex;}
  OPTIONAL {?subject wdt:P569 ?birth;}
  OPTIONAL {?subject wdt:P570 ?death;}
  OPTIONAL {?subject wdt:P19 ?placebirth;}
  OPTIONAL {?subject wdt:P20 ?placedeath;}
  OPTIONAL {?subject wdt:P18 ?image.}
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de" . }
}
LIMIT 1
```