



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Bachelorarbeit

David Stutz

Visualisierung und Auswertung offener
Mobilitätsdaten

David Stutz

Visualisierung und Auswertung offener
Mobilitätsdaten

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung

im Studiengang Wirtschaftsinformatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Olaf Zukunft
Zweitgutachter: Prof. Dr. Ulrike Steffens

Abgegeben am 23.08.2018

David Stutz

Thema der Arbeit

Visualisierung und Auswertung offener Mobilitätsdaten

Stichworte

Visualisierung, Mobilität, Big Data, Open Data

Kurzzusammenfassung

Diese Bachelorarbeit befasst sich mit dem Potenzial offener Daten, welches von zahlreichen Verkehrsverbänden und Verkehrsunternehmen wie auch Kommunen noch verkannt bzw. nicht genutzt wird. Die Verfügbarkeit von Daten ist für viele neue Mobilitätsangebote und Geschäftsmodelle Grundvoraussetzung. Ziel dieser Arbeit ist es die Thematik Open Data von Mobilitätsdaten näher zu betrachten und anhand einer Visualisierung und Auswertung deren Nutzen zu verdeutlichen.

David Stutz

Title of the paper

Visualization and analysis of open mobility data

Keywords

Visualization, Mobility, Big Data, Open Data

Abstract

This bachelor thesis deals with the potential of open data, which is still not recognized or used by numerous transport associations, transport companies and municipalities. The availability of data is a basic requirement for many new mobility offers and business models. The aim of this work is to take a closer look at the topic of open data of mobility data and to illustrate its benefits in form of a visualization and evaluation.

Inhaltsverzeichnis

1	Einleitung	6
2	Grundlagen	7
2.1	Mobilität	8
2.1.1	Intermodales Verkehrsmanagement	8
2.1.2	Modal Split	9
2.1.3	Mobilitätsdaten und Hackathons.....	10
2.1.4	Projekt HEAT	12
2.2	Big Data	13
2.2.1	Big Data im Anwendungsbereich Mobilität	13
2.2.2	Ist-Zustand in der Mobilität	14
2.2.3	Mögliche Datenquellen in der Mobilität.....	15
2.3	Visualisierung von Big Data	18
2.3.1	Definition Visualisierung	18
2.3.2	Visual Analytics für Bewegungsdaten	20
2.3.3	Werkzeuge zur Visualisierung	22
2.4	Open Data	26
2.4.1	Mobilitätsdaten und Open Data	29
3	Szenarien	34
3.1	Analyse und Konzeption.....	34
3.1.1	Weitere Vorgehensweise	34
3.1.2	Anforderung an Datenquellen und Daten	36
3.1.3	Eingesetzte Tools und Werkzeuge	38

3.1.4	Aufbau der Architektur	39
3.1.5	Visualisierung und Auswertung.....	41
3.2	Realisierung.....	42
3.2.1	Szenario 1 – Beeinflussung des Modal Split.....	42
3.2.2	Szenario 2 – HEAT Teststrecken	50
3.2.3	Szenario 3 – Evaluierung neuer Switchh Standorte	52
3.2.4	Szenario 4 – Verbesserung der Orientierungsmöglichkeiten	57
3.2.5	Nicht betrachtete Fragestellungen	60
4	Bewertung.....	61
4.1	Interpretation der Szenarien.....	61
4.2	Szenarioübergreifende Bewertung	66
5	Fazit	67
5.1	Ausblick	68

1 Einleitung

Mobilität ist ein Megatrend unserer Zeit (bpb, 2015) und Big Data gehört zu den großen Treibern der digitalen Transformation (Fallmann, 2015). Es handelt sich somit um große, gesellschaftlich relevante Themen, die ich auch aus persönlichem Interesse, unter anderem als Werkstudent bei einem Unternehmen des öffentlichen Personennahverkehrs und in meinem zurückliegenden Studium mit der Schwerpunktlegung auf Big Data, seit längerem verfolge.

Interessant und greifbarer werden die Themen in Bezug auf die Hansestadt Hamburg.

Die Mobilität spielt in einer Metropole wie Hamburg eine große Rolle. Im November 2017 ist Hamburg als Sieger um die Ausrichtung des Weltkongresses im Jahr 2021 zum Thema Intelligente Verkehrssysteme und Services hervorgegangen. Gemeinsam mit dem Bundesministerium für Verkehr und digitale Infrastruktur wird Hamburg Gastgeber des weltweit größten Kongresses in diesem Themenfeld. Hamburg wird in den nächsten Jahren Deutschlands Modellstadt für urbane Mobilitäts- und Logistiklösungen (Hamburg.de, 2017). Auch das Thema Open Data ist in der Freien- und Hansestadt Hamburg hochaktuell. Das 2012 beschlossene Hamburger Transparenzgesetz besagt, dass vorhandene Informationen unter Wahrung des Schutzes personenbezogener Daten unmittelbar der Allgemeinheit zugänglich zu machen und zu verbreiten sind. Dieses Gesetz soll die demokratische Meinungs- und Willensbildung fördern und eine Kontrolle des staatlichen Handelns ermöglichen (HmbtG, 2012) und hat zu einer Öffnung vieler erhobener Daten gesorgt. Mobilität, Big Data und Open Data: Drei Themengebiete, die zur Motivation dieser Thesis maßgeblich beigetragen.

Diese Bachelorarbeit befasst sich mit diesen drei Entwicklungen und wird die Eignung von visualisierten offenen Mobilitätsdaten durch verschiedene fiktive Szenarien näher betrachten. Die Ergebnisse werden daraufhin zusammengefasst und ausgewertet. Abschließend erfolgt eine ausführliche Bewertung mit Fazit. Durch dieses Vorgehen soll geprüft werden, ob sich öffentlich zugängliche Daten als nützlich erweisen und sich die Stakeholder im Mobilitätssektor um eine Öffnung ihrer Daten für die Allgemeinheit bemühen sollten.

Ziel dieser Arbeit ist es also, die Thematik Open Data von Mobilitätsdaten näher zu beleuchten und anhand einer Visualisierung und Auswertung deren Nutzen zu verdeutlichen.

2 Grundlagen

Zu Beginn werden die für diese Bachelorarbeit benötigten Grundlagen zusammengefasst. Dieses Kapitel ist deshalb in die Abschnitte Mobilität, Big Data, Visualisierung von Big Data und Open Data gegliedert.

Unter *Mobilität* wird zuallererst eine Eingrenzung des weit gefassten Begriffs vorgenommen und vorgestellt, mit welcher Art Mobilität sich diese Arbeit befasst.

Anschließend wird die Idee *Open Data* vorgestellt und in den Kontext zu Mobilitätsdaten gesetzt. Zusätzlich wird in diesem Kapitel unter anderem der Begriff Big Data näher betrachtet, verschiedene Open Data Initiativen in der Mobilitätsbranche vorgestellt, und auf Hackathons und deren möglichen Nutzen näher eingegangen.

Im Abschnitt *Visualisierung* wird der Begriff definiert und vorgestellt, wie Bewegungsdaten angemessen dargestellt werden könnten. Es werden abschließend Werkzeuge zur Visualisierung von Bewegungsdaten recherchiert.

2.1 Mobilität

Spätestens seit dem Beginn des 19. Jahrhunderts sind Gesellschaft und Wirtschaft durch eine stetig steigende Mobilität und Beschleunigung geprägt.

Die Basis des Lebens und der Ökonomie bildet Mobilität und kaum etwas kennzeichnet das Leben in der globalisierten Gesellschaft so sehr. Der Mobilität kommt ein unverzichtbarer Ermöglichungscharakter zu und bedeutet Beweglichkeit, Veränderung und Wandlungsfähigkeit. Für den Einzelnen, aber auch für die Gesellschaft. Was einerseits Risiken und Unsicherheit bedeuten kann, bietet andererseits neue Chancen, größere Optionenvielfalt und die Möglichkeit, Neues zu entdecken und zu erfahren. (Vgl. Megatrend Mobilität, 2018)

Mobilität ist somit Grundvoraussetzung unseres Lebens und Wirtschaftens geworden und ist ein Megatrend unserer Zeit. (Vgl. bpb, 2015) In der globalen Wirtschaft hängt von ihr ab, ob wir in Zukunft konkurrenzfähig bleiben. Von unserer privaten Mobilität wird bestimmt, ob wir künftig die guten Jobs bekommen und unsere Lebensqualität verbessern können.

„Heute stehen wir am Beginn eines multimobilen Zeitalters. Damit steigt auch die Suche nach Möglichkeiten, um Mobilitätsanforderungen und -wünsche ökonomisch, bequem und nachhaltig umzusetzen. Die Konsequenz ist, dass immer mehr Bereiche von Wirtschaft und Gesellschaft vom Megatrend Mobilität erfasst werden.“ (Megatrend Mobilität, 2018)

Diese Arbeit wird sich nicht mit dem Individualverkehr, also mit von Privatfahrzeugen abgewickelterm Verkehr, sondern mit den Mobilitätsangeboten des öffentlichen Personennahverkehrs der Hansestadt Hamburg befassen.

2.1.1 Intermodales Verkehrsmanagement

Intermodalität bezieht sich auf das Aneinanderreihen nicht gleichartiger Verkehrsmittel für eine bestimmte Wegestrecke, Modus ist in diesem Kontext das Verkehrsmittel, welches zur Verfügung gestellt wird (Vgl. Flügge, 2016).

„Die gängigsten Verkehrsmittel bzw. Verkehrsmodi für intermodale Mobilität sind Auto, Bus, Bahn, E-Bike, Fahrrad, Flugzeug, Motorrad, S-Bahn, U-Bahn und Zug. In ihrer Ausprägung können sie von Privatpersonen, privatwirtschaftlichen oder öffentlichen Anbietern auf dem Markt betrieben werden. Oft findet sich auch der Begriff der Verkehrsträger im Sinne des Straßenverkehrs, Schienenverkehrs, Seeverkehrs oder Flugzeugverkehrs wieder.“ (Flügge, 2016)

Ein über die Stadtgrenzen Hamburgs hinweg bekanntes Konzept zur Bündelung verschiedener Mobilitätsanbieter ist Switchh:

An 16 über die Stadt verteilten Switchh Punkten stehen Parkplätze und die Fahrzeuge der Partner car2go, DriveNow und cambio bereit. An fast allen Switchh Punkten befindet sich eine StadtRAD-Station und eine S- oder U-Bahn-Haltestelle in unmittelbarer Nähe. Die

Fahrzeuge von car2go und DriveNow können innerhalb der Geschäftsgebiete kostenlos auf allen öffentlichen Parkplätzen in Hamburg abgestellt und angemietet werden. Die Switchh Punkte werden stetig mehr (Vgl. Switchh, 2018). Switchh und den entsprechenden Stationen wird im späteren Szenario 2.3.2 größere Aufmerksamkeit gewidmet.

2.1.2 Modal Split

Der Modal Split (auch Modal Share genannt) ist eine Kenngröße zur Aufteilung der Verkehrsnachfrage auf verschiedene Verkehrsmittel. (Vgl. Kirchhoff, 2002)

Er stellt die prozentuale Verteilung des Verkehrsaufkommens oder der Verkehrsleistung differenziert nach den Verkehrsmitteln dar. Der Modal Split dient unter anderem zur Verkehrsplanung (Vgl. Randelhoff, 2018).

Modal Split in Hamburg

Zu Fuß	Fahrrad	ÖPNV	Kfz
28%	12%	18%	42%

Tabelle 1: Modal Split in Hamburg (Vgl. Infas 2008)

Das rot-grüne Regierungsbündnis Hamburgs hat im Koalitionsvertrag 2015 festgelegt, dass der Anteil des Fahrrades am Verkehr bis in die 2020er-Jahre auf 25 Prozent steigen soll (Koalitionsvertrag, 2015).

Ebenfalls angestrebt wird von der Hamburgischen Bürgerschaft ein weiterer Anstieg der zu Fuß zurückgelegten Wege. (Vgl. Bürgerschaft FHH, 2017). Des Weiteren soll der ÖPNV als Rückgrat der Mobilität weiter gestärkt werden.

„Die Kapazitäten im ÖPNV können und sollen auf verschiedene Weise erhöht werden. Beispielhaft seien hier Netzergänzungen, höhere Taktdichten oder größere Fahrzeuge genannt. Der Ausbau des Netzes für den schienengebundenen ÖPNV soll für einen Sprung bei Kapazität und Qualität sorgen und schnellbahnwürdige Räume anbinden.

Die Fahrgastzahlensteigerungen dürfen nicht zu unzumutbaren Beeinträchtigungen des Fahrkomforts führen (z.B. durch vollere Busse oder Züge). Die Herausforderung im ÖPNV besteht auch darin, in allen Bereichen technische Neuerungen sinnvoll umzusetzen und die Nutzung erneuerbarer Energien auszubauen. **Information und Service der Fahrgäste sind wichtig. Gut informierte Kunden, die sich auf einen guten Service vor Ort verlassen können, sind in der Regel zufriedene Kunden. Zielführend können hierbei Angebote zur Routensuche, Hinweise zu Anschlussverbindungen, Orientierungsmöglichkeiten an Verknüpfungsanlagen und aktuelle Störungsmeldungen sein.**“ (Bürgerschaft FHH, 2017)

Im späteren Szenario in Kapitel 3 wird analysiert, inwieweit die Visualisierung und Auswertung von Verkehrsdaten auf die erwähnten Punkte wie z.B. Information und Service oder Orientierungsmöglichkeiten hilfreich sein könnte und eventuell sogar auf die Ziele der

Hamburger Bürgerschaft hinwirken könnte, den Modal Split hin zu einer ökologischeren Ausrichtung zu beeinflussen.

2.1.3 Mobilitätsdaten und Hackathons

Ein Hackathon ist eine Veranstaltung, bei der Computerprogrammierer und andere an der Softwareentwicklung beteiligte Personen, darunter Grafikdesigner, Interfacedesigner, Projektmanager und andere, oft auch Themenexperten, intensiv an Softwareprojekten mitarbeiten. (Vgl. Leckart, 2012)

„Der Name setzt sich zusammen aus Hacken und Marathon: Das sind Wochenendveranstaltungen, bei denen interessierte Programmierer in weniger als 48 Stunden versuchen, mit den Daten Aufgaben zu lösen. Eine Kraftanstrengung sind diese Events auf jeden Fall. Die Entwickler motiviert hierbei, im Wettbewerb – alleine oder im Team – als Erste etwas mit neuen, „unverbrauchten“ Daten zu schaffen. Dabei entstehen Prototypen für innovative, lebensnahe Anwendungen.“ (digitalspirit, 2016)

Auf der ganzen Welt sind Regierungen und andere Institutionen aller Ebenen bestrebt ihre Bürgerinnen und Bürger durch Hackathons mit in die Nutzung offener Daten einzubeziehen. Durch die Befreiung von Daten schaffen Regierungen Chancen für die Entwicklung neuer Anwendungen für mobile Endgeräte, oft mit dem Ziel, die Dienste zu verbessern. Durch die Ausrichtung dieser Veranstaltungen bemühen sich die Institutionen, ihr Engagement für transparentere und offenere Wege zu signalisieren und fördern gleichzeitig die Entwicklung neuer Apps. Trotz ihrer Popularität ist der „zivile“ Hackathon ist ein relativ neues Phänomen, das noch nicht ausreichend bezüglich Nutzen und Auswirkungen erforscht wurde. Zivile Hackathons erschließen den aktuellen Zeitgeist von sozialer Innovation und Unternehmertum indem sie zivilgesellschaftlich denkende Hacker und Programmierer mit Regierungen verbinden. (Vgl. Johnson, 2014: 349–57)

Der Trend zu Open Data und daraus resultierend die Zunahme von Hackathons zeichnet sich klar ab und einige Institutionen in Deutschland bieten bereits regelmäßige Hackathons an, um die Spezialisten auf die zur Verfügung gestellten Daten loszulassen. Beispiele für Mobilitäts-Hackathons in Deutschland wurden hier recherchiert und aufgelistet.

DB Data Train Challenge (2015) – Auf dem ersten von der Deutschen Bahn initiierten Hackathon zeigte sich, wie wertvoll die Daten der Deutschen Bahn sind. Jedoch standen die genutzten Daten nach dem Hackathon nicht mehr zur Verfügung und dies sorgte bei vielen Teilnehmern zunächst für Frustration. Die geschaffenen Ergebnisse waren nach Beendigung des Hackathons nicht mehr nutzbar. Kenntnis daraus war, dass für die DB ein solches Format ohne dauerhaft offene Daten nicht funktioniert. Aus dieser Erfahrung heraus wurde im Konzern beschlossen, die Daten sukzessive zu öffnen und ein konzernweites Portal für offene Daten zu schaffen. (Vgl. digitalspirit, 2018)

„Um dies zu erreichen, hat sich ein Open-Data-Team gebildet, in dem überzeugte Mitarbeiter aus verschiedenen Geschäftsfeldern zusammenarbeiten. Derzeit sind überwiegend Mitglieder der Konzerninitiative Infrastruktur 4.0 sowie Mitarbeiter der DB System vertreten. In einem gemeinsamen Auftakt dieses Open-Data-Teams entstand das jetzige Datenportal data.deutschebahn.com (digitalspirit, 2018)

Ein Hackathon hat somit dafür gesorgt, dass die Deutsche Bahn ein Open Data Portal ins Leben gerufen hat und für einen Paradigmenwechsel im Umgang mit Mobilitätsdaten gesorgt hat.

Design the Smart Mobility – ITS Hackathon Hamburg - Im Zuge der DB Open Data Initiative werden freizugängliche Datensätze unter freier Lizenz veröffentlicht. Bei den stattfindenden Hackathons soll gezeigt werden, was mit diesen Daten alles möglich ist. Dazu werden Experten gesucht, deren Herz für Daten, Codierung und ungewöhnliche Ideen schlägt. Die Daten sollen die Teilnehmer befähigen, eigene Anwendungen zu kreieren. Während dieser Events besteht die Möglichkeit, sich mit internen und externen Experten zu vernetzen und sich für eine zukünftige Zusammenarbeit zu empfehlen. (Vgl. dbmindbox, 2018)

BMVI Data Run – „24 Stunden lang lädt das BMVI dazu ein, innovative Mobilitätslösungen auf Basis von Daten aus dem Geschäftsbereich des BMVI in Kombination mit kommunalen Daten voranzutreiben. Mitmachen können sowohl aktiv codende Teilnehmer als auch interessierte Zuschauer, die sich für Mobilität 4.0 und Daten interessieren. Neben der Möglichkeit zur Entwicklung kreativer Lösungen erhalten die Teilnehmenden die Möglichkeit zum Networking.“ (Deutscher Mobilitätspreis, 2018)

Auch das Bundesministerium für Verkehr und Infrastruktur bemüht sich um größtmögliche Transparenz und führt mit dem BMVI Data Run regelmäßig zivile Hackathons durch. Der BMVI Data Run fand 2018 bereits zum dritten Mal statt.

Mobilityhackathon – Der *Mobilityhackathon* ist ein gemeinsames Event der Hamburger Hochbahn AG und der eos.uptrade GmbH. Dort besteht die Möglichkeit für Teilnehmer, im Team innovative Ideen auszuhacken und diese in Codes umzusetzen. Zur Verfügung gestellt werden APIs und Hardware mit der Software entwickelt werden soll. Bei der API handelt es sich unter anderem um die Geofox Schnittstelle, die annähernd alle Verkehrsdaten des HVV anbietet.

2.1.4 Projekt HEAT

Um einmal einen Blick auf die Zukunft der Mobilität in Hamburg zu richten, wird in diesem Abschnitt ein spannendes Projekt der Hamburger Hochbahn vorgestellt. Im Rahmen des ITS-Weltkongresses, welcher im Jahr 2021 in Hamburg stattfinden wird (BWVI, 2017) hat dieses Projekt einen hohen Stellenwert und erweckt auch bei den Bürgern großes Interesse.

„Es ist ein in Deutschland bislang einmaliges Forschungs- und Entwicklungsprojekt: Noch in diesem Jahr will die HOCHBAHN mit elektrisch autonom fahrenden Kleinbussen - dann noch mit einem Fahrzeugbegleiter, der jederzeit eingreifen kann - im Testbetrieb über die Straßen der HafenCity rollen. "Hamburg Electric Autonomous Transportation", kurz HEAT, heißt das Projekt der HOCHBAHN mit namhaften Partnern, u.a. aus der deutschen Industrie. Als Vorreiter der Mobilität für Hamburg will die HOCHBAHN deshalb jedoch von Beginn an aktiv an der Entwicklung teilhaben.“ (Hochbahn, 2018)

Relevant wird dieses Projekt für diese Arbeit in der späteren Realisierung. Eventuell kann durch eine Fragestellung HEAT unterstützt werden.

2.2 Big Data

Um Open Data thematisieren zu können, muss zuerst der allumgängliche Begriff „Big Data“ betrachtet werden. Für diesen Begriff existiert keine eindeutige Definition, jedoch relativ unumstritten ist die Definition der Eigenschaften von Big Data durch Gartner (Vgl. Gartner, 2016). Das darin verwendete 3-V- Modell stellt die Herausforderungen des Datenwachstums als dreidimensional dar. Die drei Dimensionen beziehen sich auf ein ansteigendes Volumen (engl. volume) der Daten, auf eine ansteigende Geschwindigkeit (engl. velocity), mit denen Daten erzeugt und verarbeitet werden und auf eine steigende Vielfalt (engl. variety) der erzeugten Daten. Im folgenden Abschnitt werden diese drei Dimensionen in den Kontext der Mobilität gesetzt.

2.2.1 Big Data im Anwendungsbereich Mobilität

Diese Arbeit befasst sich mit Daten, die im Mobilitätssektor generiert werden. Betrachtet man diese im Kontext des 3-V- Modells, wird schnell klar, dass hier von Big Data gesprochen werden kann bzw. muss. Das Volumen, die Geschwindigkeit und auch die Vielfalt der Daten wächst rasant, nicht zuletzt durch die steigenden Nutzerzahlen im ÖPNV (Vgl. VDV, 2018), aber auch durch die technologische Revolution im Mobilitätswesen, welche durch die fortschreitende Digitalisierung ausgelöst wurde. (Vgl. Fasel, 2016: 337)

So sind in den letzten Jahren zahlreiche Apps entwickelt worden, mit denen Fahrplandaten abgerufen werden können und die dadurch ebenfalls Unmengen an auswertbaren Daten produzieren. Auch neue Mobilitätsangebote wie Bike- und Carsharing drängen auf den Markt und erzeugen umfangreiche Nutzerdaten.

Das Vorgehen um mit diesen immensen Datenmengen angemessen zu agieren, lässt sich sehr gut anhand des Big Data 5-Phasen Lifecycle Modells darstellen (s. Abbildung 1).

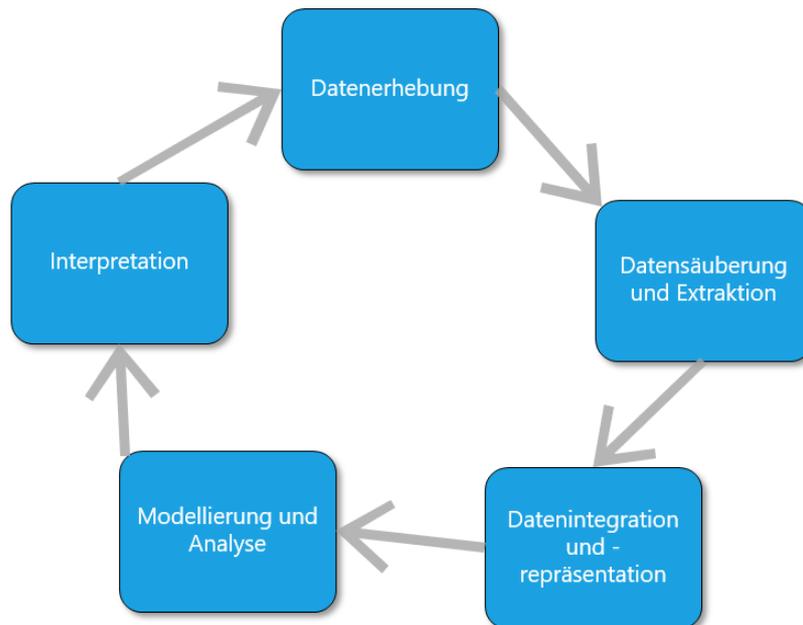


Abbildung 1: Big Data Lifecycle (Vgl. Jagadish, 2014: 86-94)

Auf diesen Big Data Lifecycle wird im 3. Kapitel näher eingegangen.

2.2.2 Ist-Zustand in der Mobilität

„Heutzutage werden Daten in der Mobilität oft nur für einen bestimmten Zweck von einem bestimmten Stakeholder gesammelt. Busbetreiber sammeln beispielsweise die Positionsdaten ihrer Busse für die Anzeige von Verspätungen, Mitarbeiter der Stadtplanung beauftragen Gutachter zur Maßnahmenbewertung, das Verkehrsmanagement verbaut Induktionsschleifen und Kameras für die adaptive Regelung von Lichtsignalanlagen. Ein Big Data-Experte würde argumentieren: Warum werden diese Daten nicht integriert und stehen allen zur Verfügung? Es könnten sich daraus völlig neue Sichtweisen mit einem Mehrwert für alle ergeben.“ (Fasel, Meier, 2016: 239)

Dieses Zitat war unter anderem der Anlass, sich näher mit der Thematik Big Data und Mobilitätsdaten auseinander zu setzen und deckt sich mit der bisher durchgeführten Recherche. Der Wert der Mobilitätsdaten rückt mehr und mehr in den Vordergrund bzw. in das Bewusstsein. Aus hiesiger Sicht scheint es viele vereinzelnde Datentöpfe verschiedenster Institutionen zu geben. Daraus resultiert die Frage ob eine Aggregation dieser Daten nicht große Chancen bieten würde. Dieser Ansicht wird nun weiter nachgegangen, zu diesem Zweck werden die Integration, Verfügbarmachung und Auswertung gesammelter Daten im weiteren Verlauf dieser Arbeit eine übergeordnete Rolle spielen.

Die mobilere Lebensweise, steigende Bevölkerungszahlen und die ansteigende Urbanisierung erfordern einen weit reichenden Wandel. Big Data in der Mobilität kann ein wesentlicher Baustein der Neuorganisation vorhandener Mobilitätssysteme sein. (Vgl. Fasel, Meier, 2016: 236)

2.2.3 Mögliche Datenquellen in der Mobilität

Die Auflistung in Tabelle 2 gibt einen Überblick, welche Daten und Datenquellen für Big Data in der Mobilität von Interesse sein könnten. Es ist nicht immer von vorneherein ein Nutzen offensichtlich, aber der Gedanke „Was wäre, wenn diese Daten zusammengeführt werden?“ ist sehr interessant (Fasel, Meier, 2016: 239).

Tab. 12.2 Liste von möglichen Datenquellen und in der FCD Modellregion Salzburg (fett) genutzte Datenquellen

Quelle	Daten
Infrastrukturbetreiber	Verkehrsgraphen (Rad, Straße, Schiene usw.); Echtzeitmessungen (Fahrbahntemperatur; Nebel, Witterung usw.); aktuelle Regelungsstrategie; Status der Wechselverkehrskennzeichen; Parkplatzstellplätze; Erhaltungszustand oder sicherheitskritische Schäden; Trassierungsparameter; Umstiegspunkte; Einrichtungen und öffentliche Gebäude; Daten zu Maut oder Gebühren; aktueller Fahrbahnzustand (Schnee, Streuung usw.)
Diansteanbieter	Öffentlicher Verkehr: Positionsdaten (FCD) , Fahrpläne, Echtzeitinformationen, Streckenführung, Haltestellen, Betriebszeiten, Preis- und Ticketinformation; Taxi: Taxistandorte, Ein- und Ausstiegspunkte; Parkplatz- und Parkhausauslastungen und Kapazitäten; Serviceeinrichtungen; Standorte und Auslastungen (Car Sharing, Fahrradausleih); Daten zu Ambulanzdiensten; Verkehrsmeldungen
Verkehrsmittel	Positionsdaten (FCD) ; Daten aus PKWs (CAN-Bus, OBD II); Daten aus LKWs oder Bussen (FMS); Daten aus Smartphone-Apps (FCD, Sensordaten, Kommentare); eigene Hardwareplattformen (Arduino usw.)
Kommerzielle Flotten	Positionsdaten (FCD) ; Dispositionsdaten von Logistikunternehmen oder Flottenbetreibern; Flottencharakteristika und -zusammensetzung
Reisende	Routenwahl; Start- und Zielpunkte; Optimierungsstrategie für Routenwahl
Behörden	Meldungen für Baustellen und Sperren (geplant und tatsächlich); Unfallmeldungen; Unfallstatistiken; Verordnungen (Einbahnen, Parkraumnutzung usw.) ; Quell-Ziel-Matrizen; Zensusdaten;
Kalenderdaten	Ferien; Feiertage; Events, Veranstaltungen und Großereignisse; erwarteter Reiseverkehr
Infrastrukturumfeld	Wettervorhersagen; Echtzeit und historisch: meteorologische Daten Immissionsdaten, Umweltschutzdaten, Lärm
Kommunikationsunternehmen	Floating Phone Data (Mobilfunkzellen); Verbindungsaufbau (WLAN, Bluetooth)

Tabelle 2: Liste von möglichen Datenquellen (Fasel, Meier, 2016: 240)

Es ergeben sich aus Tabelle 2 bei Kombination dieser unterschiedlichen Quellen verschiedene Fragestellungen, die bei Analyse dieser einen großen Mehrwert bieten könnten, z.B. Positionsdaten und Großereignisse oder der aktuelle Fahrbahnzustand und Wetterdaten. Im weiteren Verlauf wird Hamburg betrachtet und in diesem Zusammenhang möglichst viele verfügbare Datenquellen der Region gesammelt um diese sinnvoll in Relation zu setzen. Dies geschieht in Kapitel 3.

2.3 Visualisierung von Big Data

Dieser Abschnitt befasst sich mit der Thematik Visualisierung von großen digitalen Datenmengen, im Speziellen von Bewegungsdaten. Es erfolgt eine Begriffsdefinition und anschließend die Auseinandersetzung mit verschiedenen Visual Analytics Methoden um die Fragen zu beantworten, ob sie sich für Bewegungsdaten eignen und wenn ja welche sich speziell für diese Anforderung eignet. Im Anschluss erfolgt eine Analyse der aktuell am Markt befindlichen Produkte zur visuellen Auswertung von großen Datenmengen.

2.3.1 Definition Visualisierung

Der Begriff Visualisierung bedeutet unter anderem etwas optisch so betonen und herausstellen, dass es Aufmerksamkeit erregt oder auch Ideen in ein Bild umsetzen. (Vgl. Duden, 2015)

Diese Definition ist jedoch sehr grob gefasst. In Kombination mit großen Datensätzen ist hier Visualisierung von Big Data die treffendere Begrifflichkeit für diese Arbeit und im Kontext von Analyse und Auswertung, ist hier die Bezeichnung Visual Analytics zu nennen.

Visual Analytics ist ein interdisziplinärer Ansatz, der die Vorteile aus unterschiedlichen Forschungsgebieten verbindet. Mit der Visual-Analytics-Methode wird versucht Erkenntnisse aus großen und komplexen Datensätzen zu gewinnen. Dieses Vorgehen kombiniert die Stärken der automatischen Datenanalyse mit den Fähigkeiten des Menschen, schnell Muster oder Trends visuell zu erfassen. Durch geeignete Interaktionsmechanismen können Daten visuell exploriert und Erkenntnisse gewonnen werden. (Vgl. Keim, 2004: 36-42)

Die Analyse von Bewegungsdaten ist ein geeignetes Ziel zur Synergie von vielfältigen Technologien, einschließlich Visualisierung, Berechnungen, Datenbankabfragen, Datentransformationen und andere computergestützte Operationen. (Vgl. Andrienko, 2008: 375-378)

Als Workflow bei der Durchführung von Visual Analytics hat sich das nachfolgende Mantra (s. Abbildung 2) etabliert.

*“Analyse First -
Show the Important -
Zoom, Filter and Analyse Further -
Details on Demand”*

Abbildung 2: Visual Analytics Mantra (Keim, Mansmann, 2008)

Zuerst analysieren - das Wichtige anzeigen - zoomen, filtern und weiter analysieren, Details bei Bedarf. Das bedeutet, es wird zunächst analysiert und dann angezeigt. Danach wird eine kleine eingegrenzte Teilmenge durch Anwendung von Filter und Zoom-Operationen gebildet. Schließlich wird diese Untermenge einer sorgfältigeren Analyse unterzogen. Während dieses interaktiven Analyseprozesses werden die Einsichten gewonnen.

Im weiteren Verlauf wird nun näher auf Visual Analytics im Kontext mit Bewegungsdaten eingegangen. Auch die Szenarien in Kapitel 3 werden grundlegend auf der Visual-Analytics-Methodik basieren.

2.3.2 Visual Analytics für Bewegungsdaten

In diesem Abschnitt wird geprüft welche Kombination von visuellen und computergestützten Techniken die Analyse von massiven Bewegungsdaten unterstützen können und wie diese Techniken interagieren sollten. Abbildung 3 zeigt den Scope von Visual Analytics. Zwischen Information Analytics und Scientific Analytics ist Geospatial Analytics aufgeführt.

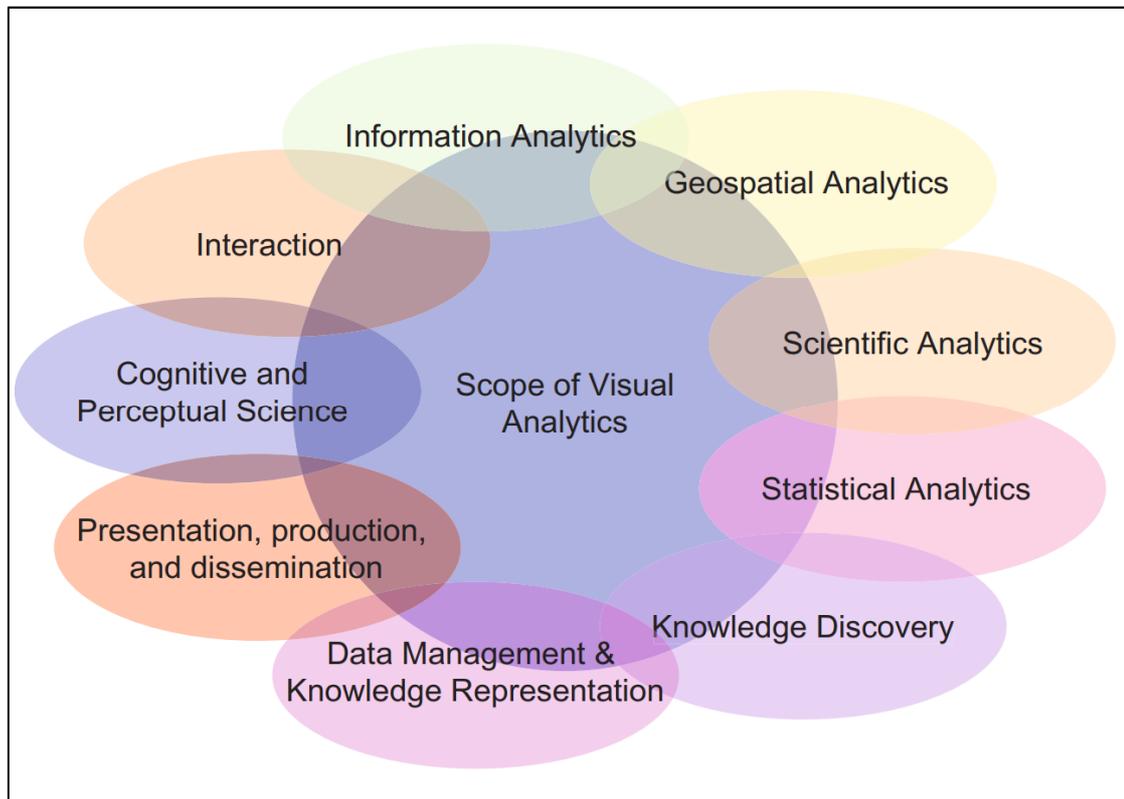


Abbildung 3: Scope von Visual Analytics (Hutchison, 2008: 79)

Geospatial Analytics bzw. Geodatenanalyse ist die Erfassung, Darstellung und Manipulation von Bildern, GPS, Satellitenfotografie und historischen Daten, die explizit in Form von geographischen Koordinaten oder implizit in Form einer Straßenadresse, Postleitzahl oder Waldbestandskennung beschrieben werden, wie sie auf geographische Modelle angewendet werden.

Die Geodatenanalyse entstand in den 1960er Jahren in Kanada für die Katalogisierung der natürlichen Ressourcen unter Verwendung der ersten geografischen Informationssysteme (GIS). Geographische Informationssysteme werden verwendet, um alle Arten von Phänomenen, die die Erde, ihre Systeme und Bewohner betreffen, vorherzusagen, zu verwalten und zu erfahren. Geospatial Analyst filtert relevante aus irrelevanten Daten heraus und wendet sie an, um die in der scheinbaren Unordnung von geografisch sortierten Daten verborgene Ordnung zu konzeptualisieren und zu visualisieren. Auf diese Weise können sie genaue Trendanalysen, -modelle und -vorhersagen erstellen. Analysten müssen jedoch wachsam bleiben, um räumliche Irrtümer, Verzerrungen oder Missverständnisse und kausale Zusammenhänge zu vermeiden: Die Geodatenanalyse wird manchmal als ebenso intuitiv wie die Wissenschaft angesehen. (Vgl. WhatIs, 2014)

Eine einfache Kombination aus Visualisierung und rechnerischer Analyse reicht jedoch nicht aus. Die Herausforderung besteht darin, analytische Werkzeuge und Umgebungen zu entwickeln, in denen die Leistungsfähigkeit von Berechnungsmethoden mit dem menschlichen Hintergrundwissen synergetisch kombiniert wird.

Wissen, flexibles Denken, Phantasie und Einsicht sind die Hauptziele des multidisziplinären Forschungsfeldes Visual Analytics. Sie ist definiert als die Wissenschaft des analytischen Denkens, die durch interaktive visuelle Schnittstellen erleichtert wird. (Vgl. Fisher, 2005: 69-104)

Dieser Herausforderung, der sinnhaften, nützlichen visuellen Analyse von Daten stellt sich diese Arbeit im weiteren Verlauf.

2.3.3 Werkzeuge zur Visualisierung

Modernste Echtzeitinformationen und fortschrittliche Analyselösungen sind eine teure Geldverschwendung, wenn man die Ergebnisse nicht klar und effektiv an die Menschen kommunizieren kann, deren Aufgabe es ist mit diesen zu arbeiten. Immer mehr Unternehmen verstehen, wie wichtig es ist, Datenanalyse in allen Bereichen der Belegschaft zur Verfügung zu stellen. Somit sind die Anforderungen an heutige Analysetools hoch. Die nachfolgende Abbildung 4 soll einen groben Eindruck über die Vielzahl der aktuell auf dem Markt befindlichen Big Data Tools vermitteln.

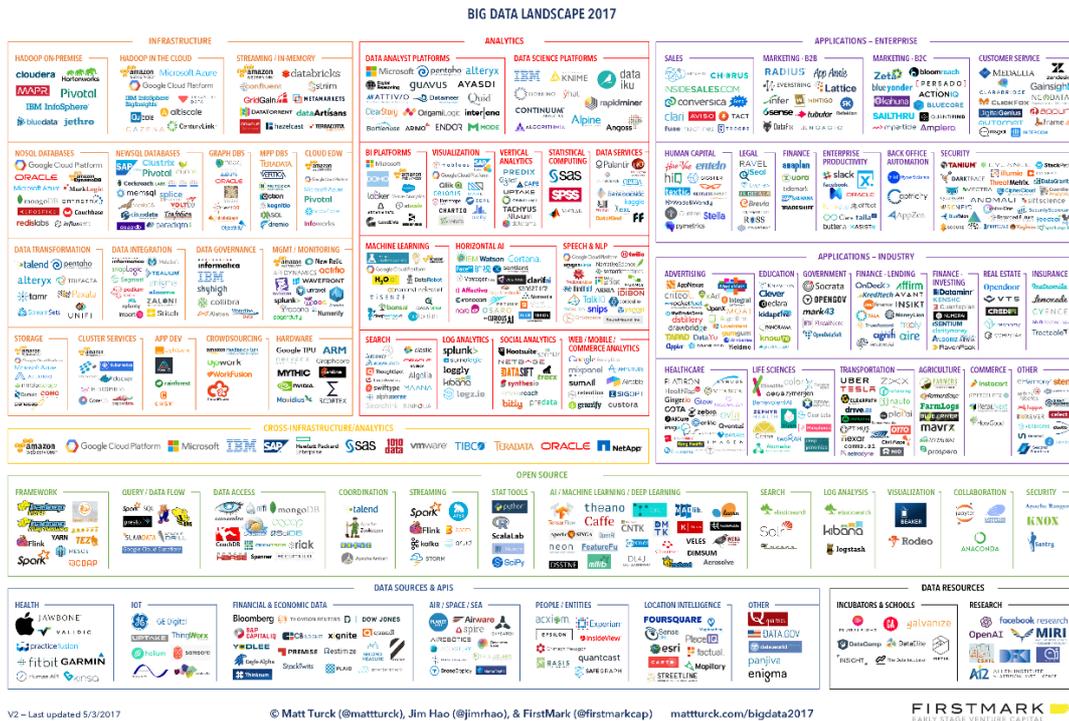


Abbildung 4: Big Data Landscape (Turck, 2017)

Der Ausschnitt des Big Data Landscape in Abbildung 5 zeigt die Rubrik *Analytics* im Detail. Es lässt sich also feststellen, dass die Anzahl der Big Data Produkte und Tools nicht unerheblich ist und sich daraus schlussfolgern lässt, dass der Markt dafür sehr groß und hart umkämpft zu sein scheint. Unter *Visualization* sind die Anbieter tableau, SAP, Google, QlikQ, celeonis, MAPD, Periscope, ZEPL, CHARTIO, ZOOMDATA und plotly aufgeführt.



Abbildung 5: Big Data Landscape (Turck, 2017) - Ausschnitt mit Markierung

Für die spätere Umsetzung konnten deshalb aufgrund des begrenzten Zeitrahmens nicht alle Analysetools mit Spezialisierung auf Visualisierung in Betracht gezogen werden. Die in Tabelle 3 aufgeführten Werkzeuge standen daher zur engeren Wahl, da sie führend auf dem Self-Service-BI-Markt sind. (Vgl. SearchEnterpriseSoftware.de, 2017)

Microsoft Power BI	Tableau	QlikView
<p>PowerBI ist eine All-in-One BI- und Analyseplattform, die als Service oder als Desktop-Client angeboten wird, aber besonders für ihre Visualisierungsfähigkeiten geschätzt wird. Die Visualisierungen werden direkt aus den Berichten erstellt und können mit den Benutzern im gesamten Unternehmen geteilt werden. Neben einer Vielzahl von eingebauten Visualisierungs-Stilen werden ständig neue durch die Community erstellt oder mit den Developer Tools selbst von Grund auf neu erstellt und mit anderen Benutzern geteilt werden. Es beinhaltet auch eine Oberfläche, die es erlaubt, Visualisierungen unterschiedlicher Komplexität aus einfachen Suchbegriffen zu erstellen. Es wird konsequent als eines der am einfachsten zu bedienenden Werkzeuge für die visuelle Datenexploration eingestuft.</p>	<p>Tableau wird oft als der Goldstandard der Datenvisualisierungstools angesehen und erfreut sich daher eines breiten Einsatzes mit berichteten 57.000 aktiven Benutzerkonten. Ein großer Teil seiner Attraktivität beruht auf seiner Flexibilität und obwohl es vielleicht nicht so anfängerfreundlich, existiert ein riesiges Support-Netzwerk in Form einer globalen Gemeinschaft von Anwendern, die über viele Branchen verteilt sind. Durch seine Leistungsfähigkeit eignet es sich besonders gut für Big Data-Operationen mit schnellen, sich ständig ändernden Datensätzen und ist daher für die einfache Anbindung an eine Vielzahl von Industriestandard-Datenbanken wie MySQL, Amazon AWS, Hadoop, SAP und Teradata konzipiert. Es stehen drei Basis-Distributionen zur Verfügung - Desktop, Server und Cloud-basiert. Neu hinzugekommen sind die Hyper In-Memory-Funktionen, die die Analyse großer Datenmengen erheblich beschleunigt.</p>	<p>QlikView ist eine weitere sehr beliebte Option für die Erstellung und gemeinsame Nutzung von Visualisierungen, die auf den von einem Unternehmen verwendeten Daten basieren. Qlik hat in den letzten Jahren daran gearbeitet, das Produkt zugänglicher und benutzerfreundlicher zu machen, unabhängig von den technischen Möglichkeiten bei der Arbeit mit Daten. Dies bedeutet jedoch nicht, dass es auf Leistung oder Funktionen verzichtet, da aufschlussreiche Visualisierungen innerhalb von Minuten möglich sind, die dank der geräteunabhängigen Infrastruktur sofort mit jedem geteilt werden können. Häufig wird es zusammen mit der QlikSense-Plattform des Anbieters verwendet, um End-to-End-Analysen und Berichte zu erstellen. Darüber hinaus verfügt es über erweiterte Sicherheitsfunktionen, die es ermöglichen, verschiedene Zugriffsstufen auf die Daten je nach den Bedürfnissen der einzelnen Benutzer einzustellen.</p>

Tabelle 3: Comparing Data Visualization Software (Vgl. Marr, 2018)

In puncto Funktionalität und Design hat Tableau eine Vorreiterrolle eingenommen. Doch in jüngster Zeit hat sich der Wettbewerb verschärft, insbesondere durch Microsoft Power BI, das deutlich weniger kostet, aber auch weniger Funktionalität zur Verfügung stellt. Die Software von Qlik zeichnet sich wiederum durch die Fähigkeit aus, interaktive Grafiken mit komplexen Daten zu erstellen. (Vgl. SearchEnterpriseSoftware.de, 2017)

Für die Umsetzung des Szenarios werden bei der Entscheidung der Toolauswahl die hier gewonnenen Erkenntnisse zum Tragen kommen und in Kapitel 3.1.3 näher erläutert.

2.4 Open Data

In diesem Abschnitt erfolgt die Auseinandersetzung mit dem Begriff *Open Data* und deren Vor- und Nachteile. Des Weiteren werden verschiedene Open Data Initiativen der Mobilitätsbranche vorgestellt und abschließend näher auf den Trend zu Hackathons eingegangen.

Als Open Data werden Daten bezeichnet, die von jedermann ohne jegliche Einschränkungen genutzt, weiterverbreitet und weiterverwendet werden dürfen. Die Forderung nach Open Data beruht auf der Annahme, dass frei nutzbare Daten zu mehr Transparenz und Zusammenarbeit führen. (Vgl. Lucke, 2010: 2).

In Bezug auf Mobilitätsdaten ist eine Rechtfertigung für offene Daten, dass öffentliche Gelder die Generierung der Daten erst ermöglicht haben, also müssen sie auch öffentlich zugänglich sein. Die Freie Hansestadt Hamburg war 2016 beispielsweise zu 85,5% Eigentümer an dem Hamburger Verkehrsverbund. Die HGV Hamburger Gesellschaft für Vermögens- und Beteiligungsmanagement mbH gehört vollständig der Freien Hansestadt Hamburg und ist 100%ger Eigentümer der Hamburger Hochbahn AG. (Vgl. Hamburger Beteiligungsbericht, 2016). Mit dem hamburgischen Transparenzgesetz von 2012 wird versucht, dieser Forderung der Öffentlichkeit Genüge zu tun. Die durch Open Data ermöglichte Zusammenarbeit erweitert die verwendbaren Daten und den Nutzen für die Allgemeinheit und den Autor. Berühmtestes Beispiel dafür ist die Wikipedia.

Im Verlauf werden nun weitere Pro Argumente aufgeführt, Gegenstimmen zusammengefasst und beide Meinungen gegenübergestellt.

Pro Argumente Open Data

Neben der Stärkung der Demokratie durch Bürgerbeteiligung und anderem sozialen Nutzen, bergen offene Daten auch wirtschaftlichen Nutzen. Im konservativen Fall haben offene Daten in Deutschland ein Potential von 12,1 Mrd. Euro, optimistisch betrachtet sind sogar 131,1 Mrd. Euro möglich. Dies würde jedoch bedeuten, Open Data als Kernkomponente einer nationalen Strategie für das 21. Jahrhundert zu positionieren und sein Potential für große gesellschaftliche Herausforderungen zu identifizieren und zu mobilisieren. (Vgl. Dapp, Marcus, 2016: 9-10).

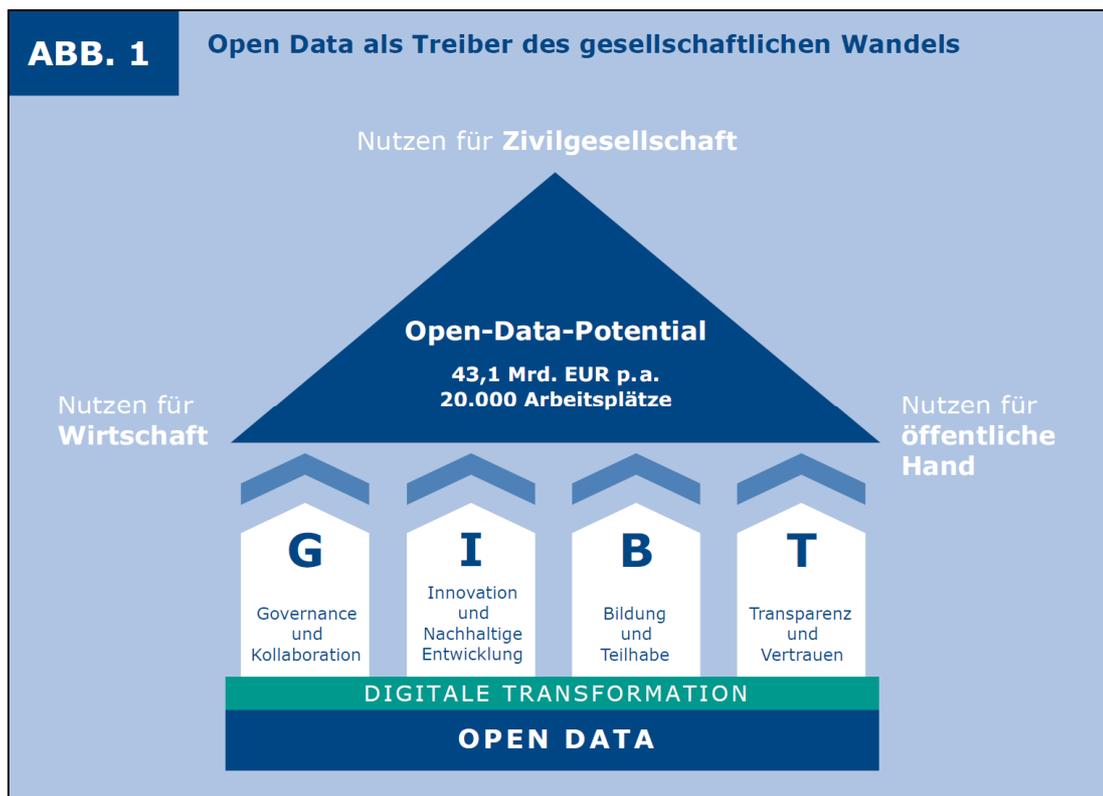


Abbildung 6: Open Data als Treiber (Dapp, M., 2016: 11)

Abbildung 6 stellt das Potential von Open Data für Deutschland eindrucksvoll dar. Gut erkennbar: die vier Potentialträger dieser Entwicklung. Sowohl die öffentliche Hand, als auch die Wirtschaft würde profitieren und dadurch einen Nutzen für die Zivilgesellschaft generieren.

Ein weiterer Vorteil ist, dass in den Behörden Geld gespart wird, da weniger bis keine Anfragen mehr an die Behörden gestellt werden. Für Unternehmen ist es einfacher und besser Geld zu erwirtschaften, weil die für bestimmte Projekte benötigten Daten besser bzw.

überhaupt verfügbar sind. Der Bürger kann sich besser über seine Umgebung informieren, sowie Zusammenhänge, Vorhaben und Zahlen und Fakten besser verstehen. (Vgl. Opendata Zürich, 2014: 13 - 17).

Contra Argumente Open Data

Jedoch gibt es auch kritische Stimmen, die sich gegen die Öffnung der Daten aussprechen. Hier kommen das Urheberrecht, aber auch datenschutzrechtliche Bedenken zum Tragen. „Viele argumentieren gegen offene und freizugängliche Informationen mit dem derzeit sowieso hochumstrittenen Urheberrecht. Denn einige Erkenntnisse sollen nicht zum Gemeingut, sondern bezahlt werden. Studien oder Marktforschungsergebnisse zu Themen des öffentlichen Sektors könnten unter diesen Bereich fallen. Auch richtet sich viel Kritik an den Umstand, dass durch Steuergelder erhobene freizugängliche Daten in direkter Konkurrenz zu kommerziellen Angeboten treten können – dahingehend hat sich die Causa Tagesschau-App in der Vergangenheit besonders hervorgetan. Hier hat u.a. der Axel-Springer-Verlag den Umstand kritisiert, dass Nachrichten umsonst zur Verfügung gestellt werden. Der Vorwurf der Wettbewerbsverzerrung kam zum Tragen.

Neben zusätzlichen datenschutzrechtlichen Bedenken wie der Frage der Anonymisierung, kommt es auch immer wieder zu dem Hinweis, dass je nachdem ob Daten frei oder kommerziell zugänglich gemacht werden, sich eine mögliche Haftung für eingebrachte Falschmeldungen erübrigen kann.“ (Weck, Andreas, 2012)

Gegenüberstellung

Pro Open Data	Contra Open Data
mehr Transparenz und Zusammenarbeit	Datenschutzrechtliche Bedenken
Stärkung der Demokratie durch Bürgerbeteiligung	Durch Steuergelder erhobene Daten in Konkurrenz zu kommerziellen Angeboten
hoher wirtschaftlicher Nutzen	Anonymisierung
Nutzen für Zivilgesellschaft	
Einsparung von Behördengeldern durch Wegfall von Anfragen	

Tabelle 4: Gegenüberstellung Open Data

2.4.1 Mobilitätsdaten und Open Data

„Weil sie was davon haben. Und wir haben auch was davon – weil wir besseren Kundenservice anbieten können“ (Schwaninger IT-TRANS, 2016)

Dieses Zitat von Günter Schwaninger, Projektmanager der Deutschen Bahn beschreibt den unbestreitbaren Vorteil, den die Öffnung der Daten den Mobilitätsunternehmen bieten kann. Besserer Kundenservice ist gleichbedeutend mit einer Attraktivitätssteigerung des ÖPNV und somit steigenden Nutzerzahlen. Dieses Potential von offenen Daten hat die Deutsche Bahn früh erkannt, und ist mit ihrem Datenportal Vorreiter der Open Data Bewegung.

Jedoch ist die Deutsche Bahn nicht die einzige Institution, die ein hohes Potential in offen zugänglichen Mobilitätsdaten erkannt hat. Im Folgenden Abschnitt wurden verschiedene Open Data Initiativen recherchiert, analysiert und gegenübergestellt. Betrachtet wurde dabei Deutschland und im Speziellen in Hamburg.

Mobilitätsdatenmarktplatz MDM

In Deutschland hat das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) – im Hinblick auf die steigende Nachfrage nach behördlichen Verkehrsdaten – der Bundesanstalt für Straßen (BASt) die Aufgabe übertragen, dynamische Informationen der Straßenbauverwaltungen bereitzustellen. Zur Wahrnehmung dieser Aufgabe wurde der MDM eingerichtet (vgl. BASt, 2018).

„Der Mobilitäts Daten Marktplatz (MDM) ist ein zentrales Online-Portal, das Verkehrsdaten bereitstellt. Durch den vereinfachten Datenaustausch mit Dritten sowie den Zugang für private Dienstleistungsanbieter eröffnen sich neue Möglichkeiten im Bereich des Verkehrsmanagements und Serviceangebote.“ (BMVI MDM, 2018)

mCloud

„In der mCLOUD sind Daten aus den Themenbereichen Straßenverkehr, Bahnverkehr, Luft- und Raumfahrt, Klima und Wetter sowie Gewässer und Wasserstraßen zu finden. Die Daten sind über Schlagworte recherchierbar und direkt zum Download oder zur dynamischen Einbindung in eigene Anwendungen verlinkt. Die Daten selbst liegen nicht in der mCLOUD. Stattdessen verweist die mCLOUD auf Datenschnittstellen und Download-Links der bereitstellenden Behörden und Unternehmen.“ (BMVI mCLOUD, 2018)

Die mCLOUD stellt Daten bereit, damit auf deren Grundlage neue digitale Mobilitätslösungen entwickelt werden können. Durch die Förderinitiative mFUND wird

dazu eine Vielzahl von Projekten gefördert, die die zu Verfügung gestellten Daten nutzen und daraus Ideen entwickeln. (Vgl. BMVI mCloud, 2018)

Open Data Portal der Deutschen Bahn

Die Deutsche Bahn hat 2015 erstmals eine Schnittstelle zu ihren Fahrplandaten bereitgestellt. Diese standen vorher nur ausgewählten Partnern wie Google zur Verfügung. Interessierte Entwickler sollen damit eigene Anwendungen programmieren können und auch Feedback geben. Über eine XML/JSON basierte REST-Schnittstelle werden die Dienste angeboten.

Seit der Öffnung werden nun schrittweise freigegebene Daten auf dem Datenportal, zum Beispiel zum Streckennetz, zu Bahnhöfen oder zur Mobilfunkabdeckung entlang der Bahnstrecken zur Verfügung gestellt. Dabei ist ihr auch der Austausch mit Entwicklern wichtig, auch durch Wettbewerbe und Hackathons. (Vgl. Heise Online, 2016)

Bei der Nutzung der Daten müssen die Marken- und Urheberrechte der Deutschen Bahn beachtet werden. Somit ist z.B. die Verwendung des DB-Logos ausgeschlossen.

Die Bahn möchte mit dem Portal schnell und unkompliziert Daten zur Verfügung stellen und treibt die Weiterentwicklung weiter voran und berücksichtigt das Feedback der Nutzer. Es wird kontinuierlich versucht, weitere neue Daten zur Verfügung zu stellen und das betreuende Team bemüht sich, innerhalb des Konzerns neue Datenbereitsteller zu finden, die sich als Open Data eignen. (Vgl. FAQ Deutsche Bahn, 2018)

Geoportal der Metropolregion Hamburg

Das Geoportal ist im Zuge des Transparenzgesetzes der Freien und Hansestadt Hamburg entstanden, um der Forderung nach offenen Daten gerecht zu werden.

Es ist die zentrale Stelle für den Zugriff auf unterschiedlichste Geodatenbestände, wie Raumordnung und Regionalplanung, Schutzgebiete, Gewerbeflächen oder auch Standortinformationen der IHKs der Metropolregion Hamburg. Die Daten werden auf einer ATIKIS-Internetkarte (Amtliches Topographisch-Kartographisches Informationssystem), Luftbildern bis zum Maßstab 1 : 5000 und der Karte 1 : 5000 abgebildet. (Vgl. Geoportal der Metropolregion Hamburg – MetaVer, 2016)

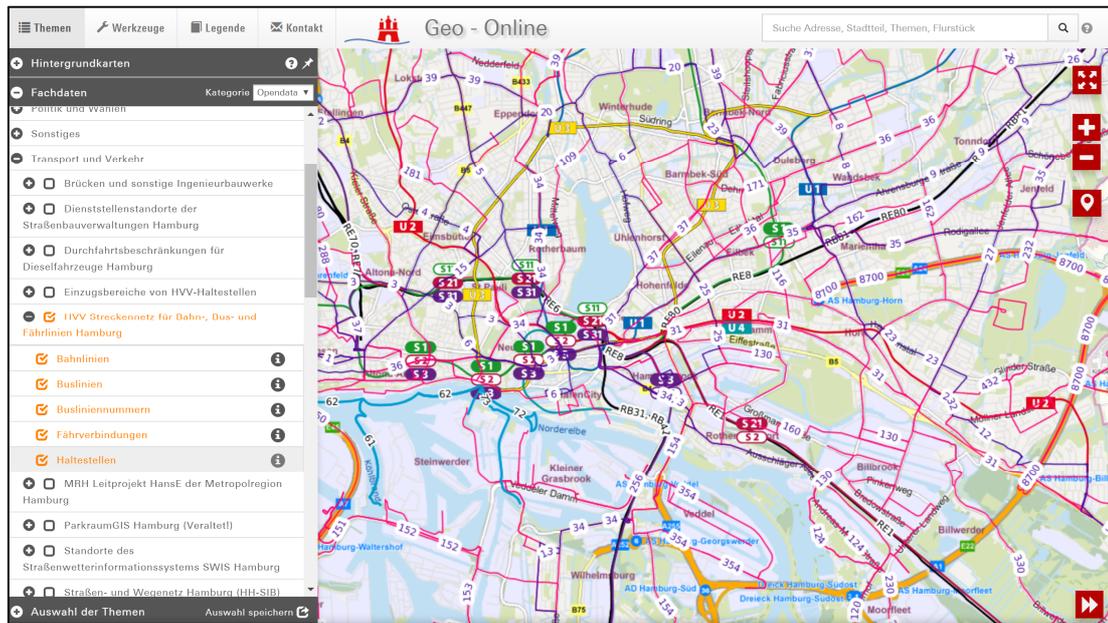


Abbildung 7: Geoportal - Verkehrsdaten (Geo-Online, 2018)

Abbildung 7 zeigt einen Ausschnitt des Geoportals. Ausgewählt wurden hier zur Veranschaulichung die Daten für Transport und Verkehr der Rubrik „Fachdaten“. Diese Daten und die Daten werden im späteren Verlauf näher betrachtet.

Die ausgewählten Daten können per Link im WMS Format abgerufen werden. WMS ist ein spezieller Webservice und stellt eine Schnittstelle zum Abrufen von Auszügen aus Landkarten dar. Des Weiteren ist es je nach Datensatz möglich XLS bzw. GML Daten herunter zu laden.

GEOFOX Verkehrsnetzplan

Das Unternehmen HBT GmbH betreibt für den Hamburger Verkehrsverbund das Fahrgastinformationssystem GEOFOX. Über die Webseite ist ein Verkehrsnetzplan mit zahlreichen Livedaten für jeden zugänglich abrufbar.

Verkehrsznetzplan

Geben Sie eine Haltestelle, eine Adresse oder einen besonderen Ort im HVV-Gebiet ein um sich das umliegende Verkehrsnetz anzeigen zu lassen. Oder geben Sie eine Liniennummer ein, um den Linienverlauf sehen zu können. Weitere Anzeigeoptionen finden Sie unterhalb der Karte.

Haltestelle, Adresse,
besonderen Ort oder Linie eingeben

In Karte anzeigen

→ Verfügbarkeit der Aufzüge im HVV (Live-Auskunft)



Abbildung 8: Verkehrsznetzplan des HVV Ausschnitt Hamburg Dammtor (Geofox.de, 2018)

Die verfügbaren Daten werden in Echtzeit angezeigt und auf einer Karte des OpenStreetMap-Projekts dargestellt (s. Abbildung 8). Das GEOFOX System verfügt nach eigener Auskunft über eine SOAP Schnittstelle, die jedoch nicht für die Allgemeinheit zugänglich ist.

Gegenüberstellung der recherchierten Quellen

Bezeichnung - Gibt die Bezeichnung des jeweiligen Portals an

Anbieter - Auflistung der Datenanbieter, die die Daten zur Verfügung stellen, bzw. das jeweilige Portal betreiben.

API - Unter API wurde aufgeführt ob eine Programmierschnittstelle zum Abrufen der Daten angeboten wird.

Dateiformate - Die Daten können in unterschiedlichen Dateiformaten vorliegen. Erwähnt werden nur die Dateiformate der öffentlich zugänglichen Daten.

Qualität und Umfang der Daten - Auf Grundlage der Recherche wurde eine Einschätzung der Datenqualität und des Umfangs vorgenommen.

Bezeichnung	Anbieter	API	Dateiformate	Qualität der Daten	Datenumfang
Mobilitätsdatenmarktplatz MDM	BMVI	keine	DATEX II	nicht ermittelbar	nicht ermittelbar
MCloud	BMVI	Keine	Rechercheplattform	-	-
Open Data Portal DB	Deutsche Bahn	REST	CSV, JSON	Export veraltet	partiell
Geoportal Hamburg	FHH	WMS	GML, XLS	Sehr gut	umfangreich
GEOFOX Verkehrsnetzplan	HBT	SOAP	JSON	nicht ermittelbar	sehr umfangreich

Tabelle 5: Gegenüberstellung der recherchierten Quellen

3 Szenarien

Um nun die Daten visualisieren und auswerten zu können werden in dem folgenden Kapitel unter Einbeziehung der Erkenntnisse aus den vorherigen Abschnitten 2.1, 2.2 und 2.4 öffentlich zugängliche Daten nach striktem systematischen Vorgehen erhoben, gesäubert, integriert, modelliert und interpretiert.

Dieses Kapitel gliedert sich in die zwei Abschnitte *Analyse und Konzeption* und *Realisierung*.

3.1 Analyse und Konzeption

Im folgenden Abschnitt werden die nun zu betrachtenden Begebenheiten erfasst und systematisch zergliedert. Anschließend werden die Rahmenbedingungen festgelegt sowie die zu erreichenden Ziele formuliert und die weiteren Arbeitsschritte definiert.

3.1.1 Weitere Vorgehensweise

Im Abschnitt 3.2 folgt nun die praktische Umsetzung. Daher ist es sinnvoll verschiedene Möglichkeiten zur weiteren Vorgehensweise zu betrachten, um am Ende zu einem guten Ergebnis zu gelangen.

Die in vorheriger Ausbildung und auch im Studium vermittelten Konzepte zum Ablauf in Projekten wären zum Beispiel das Wasserfall-Modell als phasenorientiertes Vorgehensmodell oder auch eine agile Vorgehensweise nach SCRUM.

Um dem Thema Big Data gerecht zu werden, mit dem sich diese Arbeit größtenteils befasst, und da es bei näherer Betrachtung am sinnvollsten erscheint, wird der weitere Ablauf der praktischen Umsetzung nach dem in Abschnitt 2.2.1 eingeführten Big Data LifeCycle durchgeführt. Die typische Bearbeitung eines Szenarios stellt sich demnach wie folgt dar:

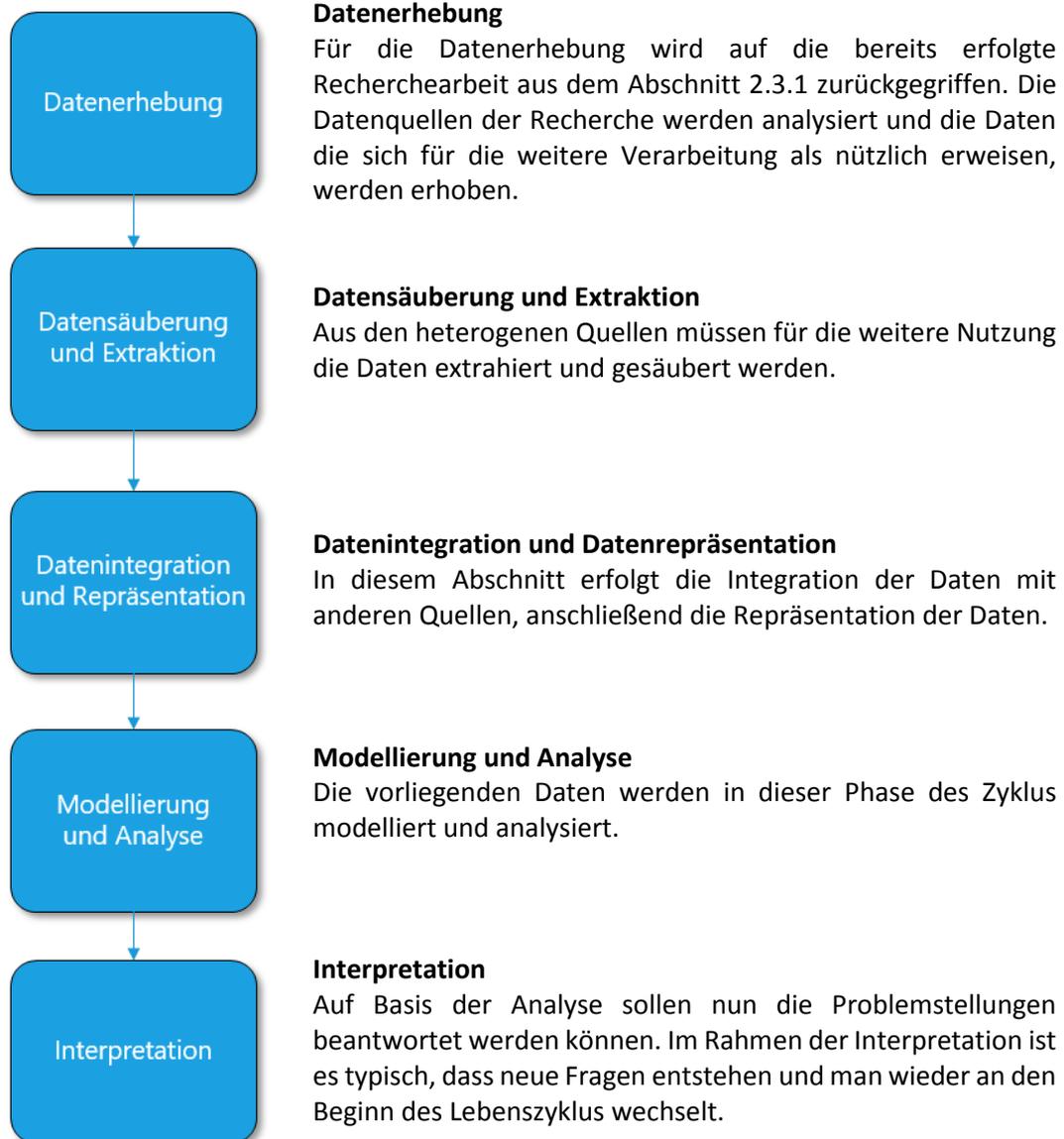


Abbildung 9: Szenarienablauf

3.1.2 Anforderung an Datenquellen und Daten

Wie erwähnt wurden in Abschnitt 2.2.1 eine intensive Recherche vorgenommen und unterschiedliche Datenquellen analysiert. Um noch einmal zu betonen: für den weiteren Verlauf werden ausschließlich Mobilitätsdaten der Stadt Hamburg betrachtet.

Ausgehend von dieser Recherche wurden vier wesentliche Datenquellen ermittelt, die für den weiteren Verlauf in Frage kommen (Abbildung 10).

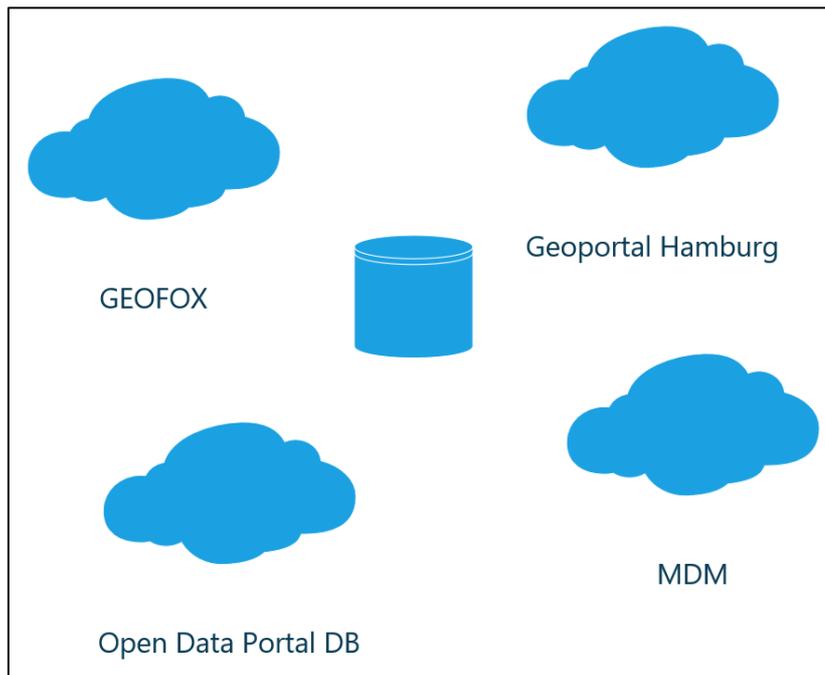


Abbildung 10: Ermittelte Datenquellen

MDM – Zu den Daten des Mobilitätsdaten Marktplatzes besteht kein direkter Zugang. Einen Zugang erhalten lediglich Personen mit institutionellem und/oder unternehmerischem Hintergrund. Da hier auch deutschlandweite Daten zusammengefasst werden, und nur die Mobilitätsdaten Hamburgs betrachtet werden, ist die Datenmenge zu umfangreich. Nichtsdestotrotz verfolgt der MDM ein interessantes Konzept und wird im Ausblick noch einmal näher betrachtet.

Open Data Portal DB – Über das Open Data Portal werden verschiedene Datensätze im XLS Format bezogen. Zudem wird die Nutzung der API in Betracht gezogen.

Geoportal Hamburg – Über das Geoportal können verschiedene Daten bezogen werden, jedoch sind die vorliegenden Daten sehr allgemein bzw. müssen interessante Daten erst durch eine spezielle Anfrage bei der Behörde erhoben werden. Zudem erzeugen das Parsen der GML Dateien, die über die WMS-API abrufbaren Daten zurückgegeben werden, einen erheblichen Mehraufwand.

GEOFOX – Weder ältere Datensätze noch Echtzeitdaten stehen der Öffentlichkeit zur Verfügung.

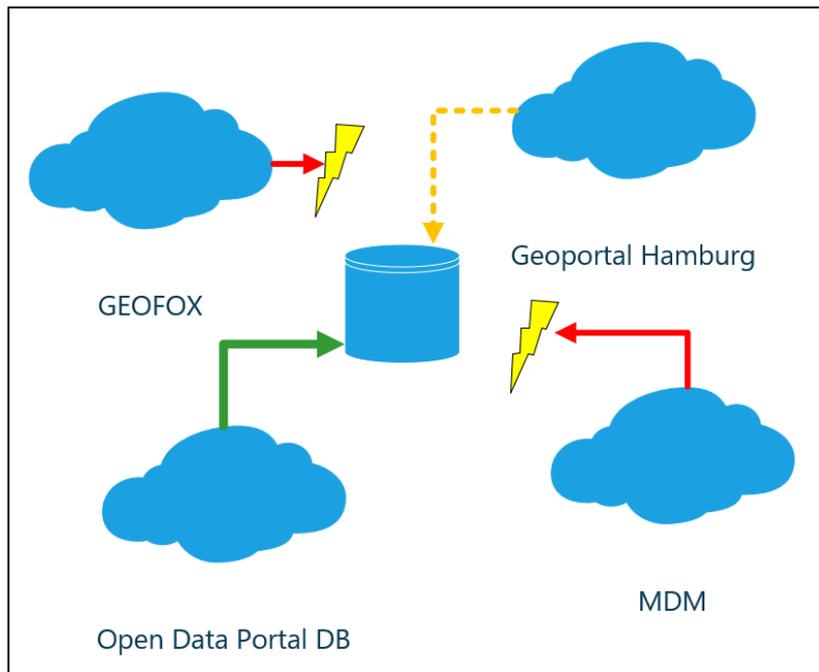


Abbildung 11: Datenquellen nach Analyse

Als Konsequenz aus der Analyse (Abbildung 11) und zur weiteren Konzeption wird in Betracht gezogen, bei Bedarf Mock-Up Daten zu erzeugen um eine adäquate Aussagekraft treffen zu können. Dies könnte unterschiedliche Daten des Geoportals betreffen oder auch Geofox Daten, bzw. Daten von Swichh. Des Weiteren werden die Ideen und Fragestellungen stark an den Daten, die erhoben werden können, ausgerichtet.

3.1.3 Eingesetzte Tools und Werkzeuge

Für die Umsetzung der Szenarien im Folgenden Abschnitt 3.2 wurde sich auf Grundlage der Recherche in Abschnitt 2.3.3 für Tableau entschieden. Die sehr breite Nutzerbasis und die Vielzahl an Datenbankschnittstellen waren die ausschlaggebenden Punkte für diese Entscheidung.

Zum Abruf von vorhanden APIs wird Python verwendet

Zu diesem Zweck wurde eine für Studenten kostenfreie Lizenz beschafft. Auf einem Windows 10 System wurde anschließend Tableau in der Version 10.4.0 installiert. Ebenfalls wurde Python 3.6 in der 32-Bit Version auf dem System installiert.

3.1.4 Aufbau der Architektur

Für die Verarbeitung der Mobilitätsdaten erfolgte ein Architekturentwurf mit den in Abbildung 12 aufgeführten drei Komponenten. Diese Aufteilung wurde bewusst einfach gehalten und erfüllt die Anforderungen, um aussagekräftige Visualisierungen aus Bestandsdaten zu realisieren. Die Plattform stellt die Rohdaten zur Verfügung, für die Datenverwaltung wird üblicherweise auf relationale Datenbanksysteme oder noSQL verwendet, in der Anwendungsschicht sind Big Data Tools, wie in Kapitel 2.3.3. vorgestellt, zu finden.



Abbildung 12: Architekturentwurf

Planung, Design und Aufbau eines Data-Warehouse mit ETL oder ELT Werkzeugen ist ein sehr spannendes Themengebiet im Zusammenhang mit Big Data und der Verarbeitung und Speicherung von Massendaten. Ebenfalls die Entwicklung komplexer Architekturen wie Lambda- bzw. Kappa und der Umgang mit verteilten Systemen. Jedoch ist dies nicht Bestandteil dieser Bachelorthesis. Wie eine Architektur in der Mobilität aufgebaut werden könnte, wird im Ausblick näher betrachtet.

Augenmerk wird weiterhin auf Visualisierung von großen Datenmengen gelegt. Die Ausprägungen der drei Komponenten sind in Abbildung 13 ersichtlich.

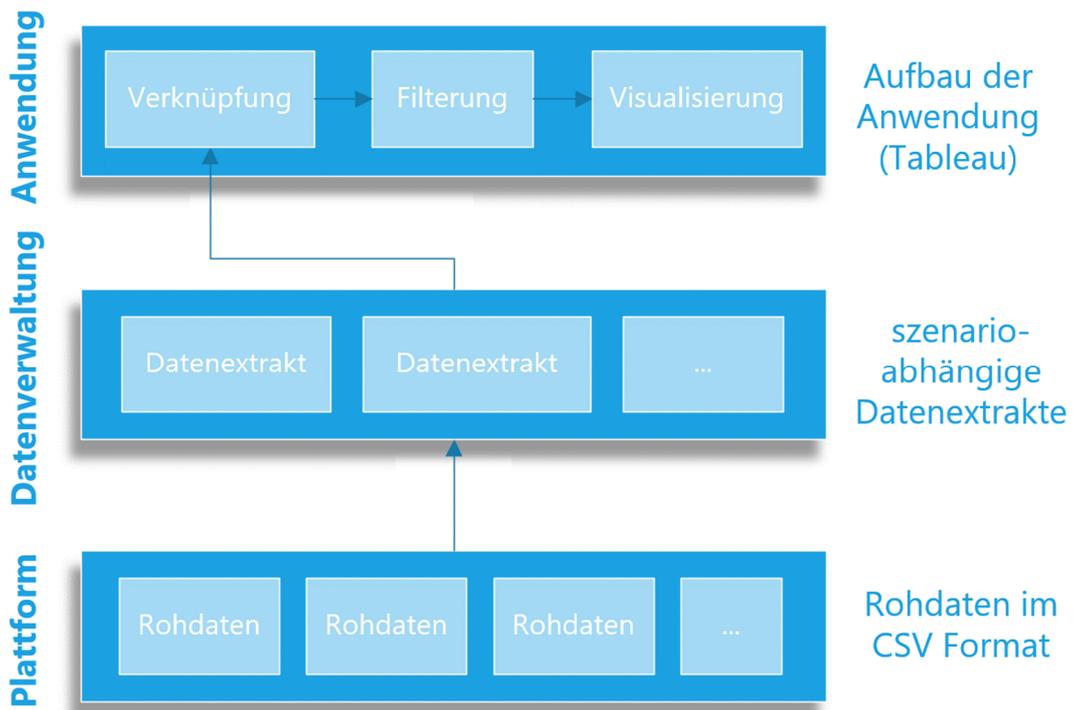


Abbildung 13: Schichten der Architektur im Detail

Die Plattform für dieses Projekt bildet ein handelsüblicher Client-PC. Auf diesem werden die verschiedenen Rohdaten gespeichert.

Tableau bietet den Vorteil, auch massive CSV Dateien in angemessener Zeit verarbeiten zu können und daraus Datenextrakte zu generieren. Aus den Rohdaten werden dafür Teilmengen gebildet, die in Form eines Datenextraktes gespeichert werden. Die Datenverwaltung stellen somit später die zu erzeugenden Datenextrakte sicher. Es wird sich um Bestandsdaten handeln, die im Laufe der Szenarien keiner Änderung oder Anpassung bedürfen. Auf eine Konzeption und Umsetzung eines Datenbanksystems zur Datenverwaltung wird daher verzichtet.

Die Anwendungskomponente stellt das Front-End dar. Dort werden die Daten verknüpft, kombiniert, repräsentiert und visualisiert.

3.1.5 Visualisierung und Auswertung

Die besten Berechnungstechniken für die Datenverarbeitung und -analyse sind wertlos, wenn nicht geeignete Methoden in Abhängigkeit von den Daten gewählt werden. Die Steuerung der Arbeit der Methoden, die Interpretation der erzielten Ergebnisse, das Verstehen, was als nächstes zu tun ist, obliegt der Verantwortung des Nutzers und erfordert auch ein gewisses Maß an Kreativität. Um eine effektive Einschätzung und Schlussfolgerungen zu ermöglichen, müssen die relevanten Informationen in angemessener Weise präsentiert werden.

Gearbeitet wird unter anderem mit statistischen, jedoch hauptsächlich mit raumbezogenen, geospatialen Daten und der Darstellung auf Karten, die bereits in Tableau integriert sind (OpenStreetMap). Durchgeführt werden einfache entscheidungsstützende Analysen.

3.2 Realisierung

In der Realisierungsphase wird nun das in der Analyse und Konzeption festgelegte Vorgehen auf verschiedene Fragestellungen angewendet. Üblich ist, dass nach der Interpretation der Daten im Big Data Kontext neue Fragen aufgeworfen werden und der Lebenszyklus von neuem beginnt. Diesem Umstand wird durch verschiedene Durchläufe Rechenschaft geleistet.

3.2.1 Szenario 1 – Beeinflussung des Modal Split

Für das erste Szenario wird Bezug nehmend auf Abschnitt 2.1.2 (Forderung der Bürgerschaft zur Nutzung von Fahrrädern) folgende Fragestellung formuliert:

Können durch Visualisierung von Mobilitätsdaten Erkenntnisse gewonnen werden, welche StadtRad-Stationen in Hamburg zur Verbesserung des Modal Splits hin zur Nutzung von mehr Fahrrädern unterstützt werden sollten?

Idee: Übersicht von Radstationen mit vielen Buchungen (frequentierte Orte) oder auch Stationen mit den längsten Ausleihzeiten

Benötigte Daten: StadtRad

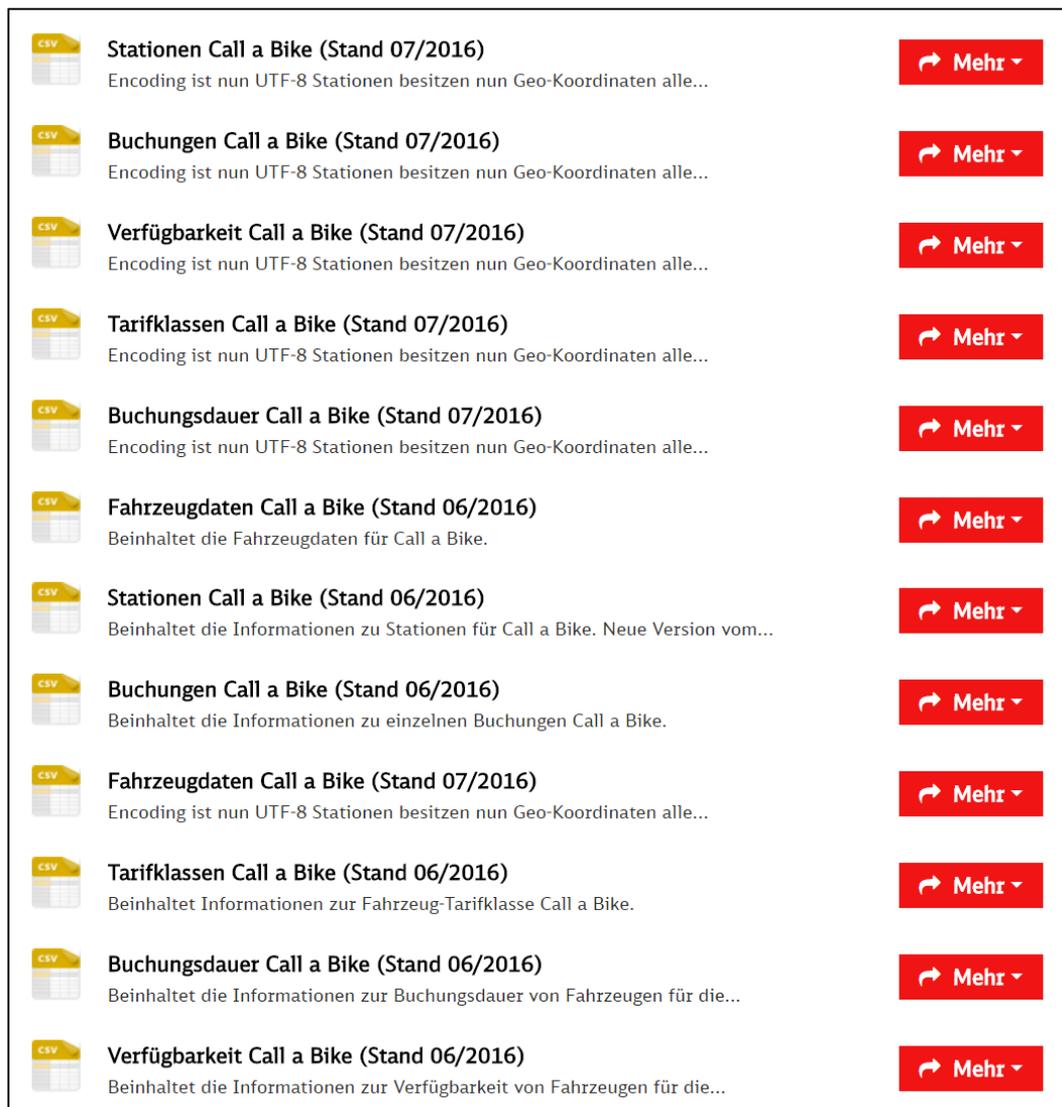
1. Datenerhebung

Für dieses Szenario wurden StadtRad-Daten und deren Buchungen benötigt.



Abbildung 14: Screenshot DB Open Data Portal – Call a Bike

Ausgewählt und heruntergeladen wurden zu diesem Zweck die Datensätze *Stationen Call a Bike (Stand 06/2016)* und *Buchungen Call a Bike (Stand 06/2016)* (s. Abbildung 15). Beide Datensätze liegen im CSV Format vor.



	Stationen Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Buchungen Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Verfügbarkeit Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Tarifklassen Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Buchungsdauer Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Fahrzeugdaten Call a Bike (Stand 06/2016) Beinhaltet die Fahrzeugdaten für Call a Bike.	↶ Mehr ▾
	Stationen Call a Bike (Stand 06/2016) Beinhaltet die Informationen zu Stationen für Call a Bike. Neue Version vom...	↶ Mehr ▾
	Buchungen Call a Bike (Stand 06/2016) Beinhaltet die Informationen zu einzelnen Buchungen Call a Bike.	↶ Mehr ▾
	Fahrzeugdaten Call a Bike (Stand 07/2016) Encoding ist nun UTF-8 Stationen besitzen nun Geo-Koordinaten alle...	↶ Mehr ▾
	Tarifklassen Call a Bike (Stand 06/2016) Beinhaltet Informationen zur Fahrzeug-Tarifklasse Call a Bike.	↶ Mehr ▾
	Buchungsdauer Call a Bike (Stand 06/2016) Beinhaltet die Informationen zur Buchungsdauer von Fahrzeugen für die...	↶ Mehr ▾
	Verfügbarkeit Call a Bike (Stand 06/2016) Beinhaltet die Informationen zur Verfügbarkeit von Fahrzeugen für die...	↶ Mehr ▾

Abbildung 15: Screenshot DB Open Data Portal

Für die weitere Analyse werden die Stationsdaten der Call a Bike Standorte benötigt, da über den Namen der Station auf die Geokoordinaten referenziert wird.

2. Datensäuberung und Extraktion

Die Rohdaten wurden gesichtet und geprüft und anschließend in Tableau geöffnet. Der Stationsdatensatz wurde aufgrund der geringen Größe direkt aufbereitet (Abbildung 16 u. 17)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
RENTAL_ZONE	RENTAL_ZONE_NAME	TYPE	CITY	COUNTRY	POI_AIRPORT	POI_LONG_D	POI_SUBURB	POI_UNDEERG	CLASSIFICATH	RENTAL_ZONE_CODE	COMPANY	COMPANY_G		
402826	Station	CAB-Aachen	stationbased	Aachen	DE	0	0	0	0	Carsharing	CAB-Aachen	Call a Bike	Call a Bike	
402828	Station	Fahndungsst:	stationbased	Aachen	DE	0	0	0	0	Carsharing	CAB-Aachen	Call a Bike	Call a Bike	
194529	Standort	normal	Andermatt	CH	0	0	0	0	0	Call a Bike	Standort	111120	Call a Bike	
402579	Station	CAB-Anderm:	stationbased	Andermatt	CH	0	0	0	0	Carsharing	CAB-Anderm:	Call a Bike	Call a Bike	
402581	Station	Fahndungsst:	stationbased	Andermatt	CH	0	0	0	0	Carsharing	CAB-Anderm:	Call a Bike	Call a Bike	
128493	Standort	normal	Aschaffenburg	DE	0	0	0	0	0	Call a Bike	Aschaffenburg	6370	Call a Bike	
401831	Station	CAB-Aschaffe	stationbased	Aschaffenburg	DE	0	0	0	0	Carsharing	CAB-Aschaffe	Call a Bike	Call a Bike	
145092	Standort	bsc	Augsburg	DE	0	0	0	0	0	Call a Bike	Augsburg Hbl	8600	Call a Bike	
402054	Station	CAB-Augsbur	stationbased	Augsburg	DE	0	0	0	0	Carsharing	CAB-Augsbur	Call a Bike	Call a Bike	
402056	Station	Fahndungsst:	stationbased	Augsburg	DE	0	0	0	0	Carsharing	CAB-Augsbur	Call a Bike	Call a Bike	
402057	Station	CAB-Bad Her:	stationbased	Bad Hersfeld	DE	0	0	0	0	Carsharing	CAB-Bad Her:	Call a Bike	Call a Bike	
402059	Station	Fahndungsst:	stationbased	Bad Hersfeld	DE	0	0	0	0	Carsharing	CAB-Bad Her:	Call a Bike	Call a Bike	
402027	Station	CAB-Bad Hon	stationbased	Bad Honau	DE	0	0	0	0	Carsharing	CAB-Bad Hon	Call a Bike	Call a Bike	
402029	Station	Fahndungsst:	stationbased	Bad Honau	DE	0	0	0	0	Carsharing	CAB-Bad Hon	Call a Bike	Call a Bike	
242611	Standort	normal	Baden-Baden	DE	0	0	0	0	0	Call a Bike	Hauptbahnhc	7650010	Call a Bike	
242612	Standort	normal	Baden-Baden	DE	0	0	0	0	0	Call a Bike	Kino	7650060	Call a Bike	
243011	Standort	normal	Baden-Baden	DE	0	0	0	0	0	Call a Bike	Wohnmobiln	7650020	Call a Bike	

Abbildung 16: Stationsdaten - unbearbeitet

Die überflüssigen Tabellenspalten und alle Stationen, die sich nicht in Hamburg befinden, wurden entfernt. Abbildung 17 zeigt die aufbereitete Datenstruktur, mit der weitergearbeitet wurde.

1	SWITCHH_ID;Switchh-KOORD;SWITCHH_RENTAL_ZONE
2	999901;Hannoversche Straße 85;switchh Punkt Harburg
3	999902;Johann-Meyer-Straße 56;switchh Punkt Bergedorf
4	999903;Bei der Reitbahn 3;switchh Punkt Bei der Reitbahn
5	999904;Paul-Neveermann-Platz 1;switchh Punkt Altona
6	999905;Nernstweg 32/34;switchh Punkt Nernstweg
7	999906;Sillemstraße 65;switchh Punkt Sillemstraße
8	999907;Heußweg 66;switchh Punkt Heußweg
9	999908;Schlump;switchh Punkt Schlump
10	999909;Lattenkamp 78;switchh Punkt Lattenkamp
11	999910;Goernestraße 21;switchh Punkt Kellinghusenstraße
12	999911;Rothenbaumchaussee 76-78;switchh Punkt Hallerstraße
13	999912;Theodor-Heuss-Platz;switchh Punkt Dammtor
14	999913;Rödingsmarkt 10;switchh Punkt Rödingsmarkt
15	999914;Ernst-Merck-Straße;switchh Punkt Hauptbahnhof Nord
16	999915;Beim Strohhause 38;switchh Punkt Berliner Tor
17	999916;Wagnerstraße 5;switchh Punkt Hamburger Straße
18	999917;Saarlandstraße 39;switchh Punkt Saarlandstraße
19	999918;Wiesendamm 6;switchh Punkt Barmbek
20	999919;Schlossstrasse;switchh Punkt Wandsbek Markt

Abbildung 17: Stationsdaten – aufbereitet

Über *Verbindungen* wurden die beiden Datensätze nun eingebunden. Die für die Analyse überflüssigen Spalten der Buchungsdaten wurden entfernt, um eine Teilmenge für das

anschließend angefertigte Extrakt zu erzeugen und so die Menge der Ausgangsdaten erheblich zu verringern.

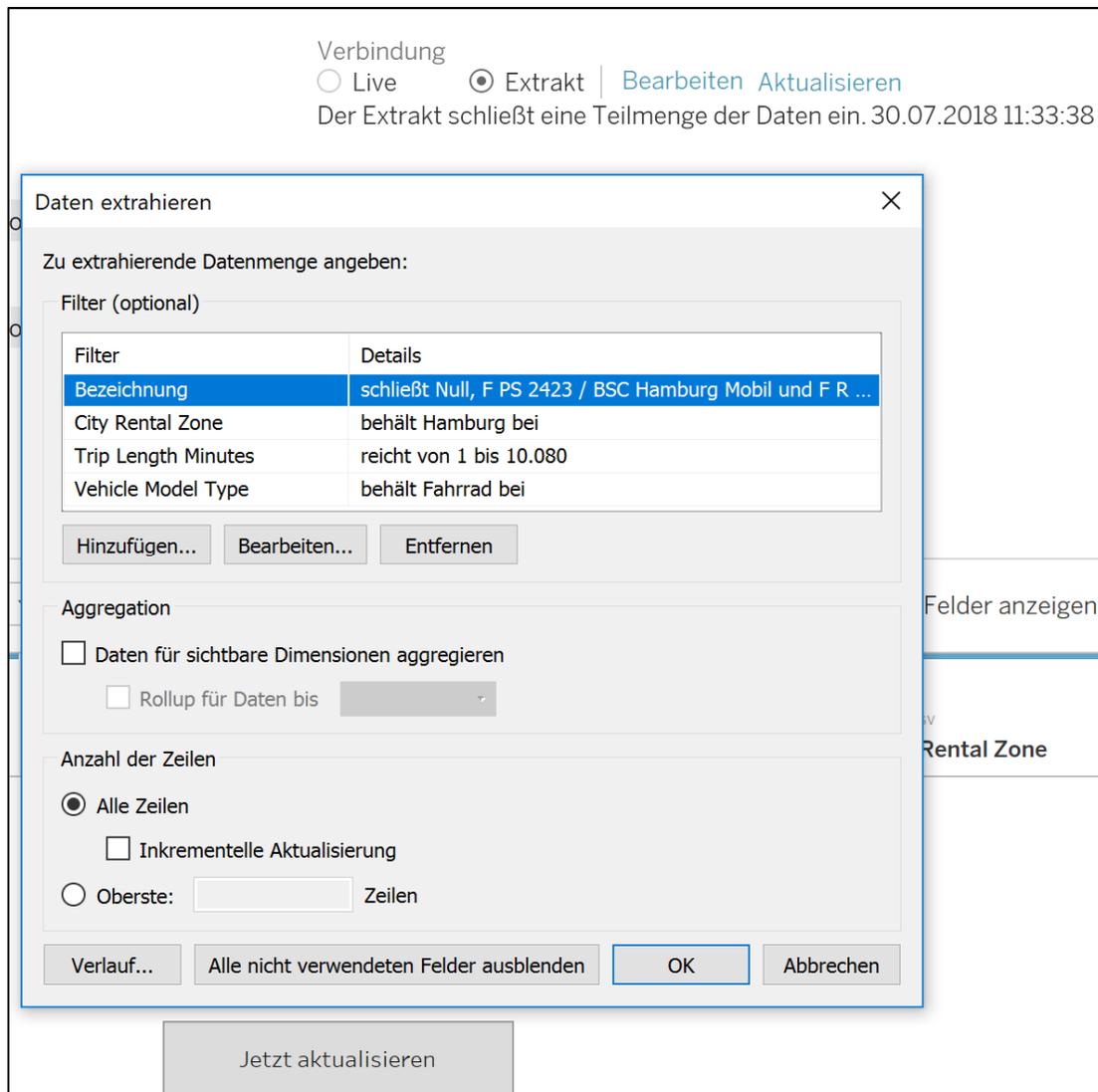


Abbildung 18: Datenextrakt in Tableau

Um die Daten sinnhaft zu extrahieren wurden mehrere Filter hinzugefügt (s. Abbildung 18).

- Über die Bezeichnung wurden Stationen ausgeschlossen, die fehlerhaft sind bzw. klar als Wartungsstationen erkennbar waren.
- Als City Rental Zone wurden die Daten nun auf Hamburg beschränkt
- Die Trip Length Minutes geben an, wie lange ein Fahrrad ausgeliehen wurde. Um fehlerhafte Messwerte auszuschließen, wurde nur Leihen unter 10080 Minuten (7 Tage) in das Extrakt aufgenommen.

- Der Vehicle Model Type wurde auf Fahrräder eingeschränkt.
Anschließend wurde das Extrakt erzeugt (Dateititel Extrakt_Szenarien1-2) und auf dem System gespeichert.

3. Datenintegration und Datenrepräsentation

Die Buchungsdaten und die Standortdaten müssen nun sinnhaft mittels JOIN kombiniert werden. Dies ermöglicht erst die spätere Visualisierung auf der Karte. Dafür bietet Tableau verschiedene JOIN Varianten an, die in Tabelle 6 aufgeführt sind.

Join Typ	Resultat	Symbol
Innen	Das Ergebnis ist eine Tabelle, die nur die Werte enthält, die in beiden Tabellen übereinstimmen.	
Links	Alle Werte aus der linken Tabelle und die entsprechenden Übereinstimmungen aus der rechten Tabelle werden als Ergebnis beim Left Join geliefert. Wenn ein Wert in der linken Tabelle keine entsprechende Übereinstimmung in der rechten Tabelle hat, bleibt ein Nullwert.	
Rechts	Nullwert, wenn die rechte Tabelle keine Übereinstimmung in der linken Tabelle hat.	
Vollständige äußere Verknüpfung	Auch <i>Full Outer Join</i> genannt. Enthält alle Werte aus beiden Tabellen. Wenn ein Wert aus einer der beiden Tabellen nicht mit der anderen Tabelle übereinstimmt, wird ein Nullwert hinzugefügt.	

Tabelle 6: JOIN Typen

Der Buchungsdatensatz enthält die Standortnamen, die im nächsten Schritt mit den Namen in der Stationstabelle verknüpft wurden (s. Abbildung 19)

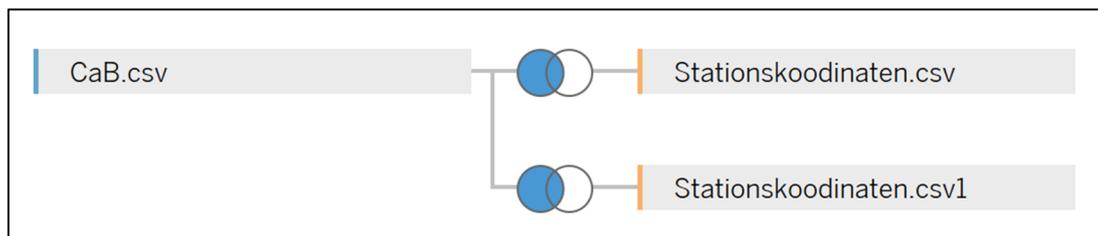


Abbildung 19: Verknüpfung der Datensätze

Für die Kombination der Datensätze wurde Bezug nehmend auf Tabelle 6 ein Linker JOIN gewählt, da die Buchungsdaten die Bezugsgrößen sind und so sichergestellt ist, dass jede Buchung einem Standort zugewiesen wird. Verknüpft wurden die Abholstationen und die Rückgabestationen. Dazu wurden die Stationsdaten dupliziert und entsprechend kombiniert.

4. Modellierung und Analyse

Im ersten Schritt werden die vorhandenen Daten nun so modelliert, dass die Stationen mit den längsten Ausleihzeiten sichtbar gemacht werden. Ausgangspunkt sind wie bei allen weiteren Szenarien die Geodaten der Stationen. Diese wurden als Datentyp der entsprechenden geografischen Rolle (Longitude = Längengrad, Latitude = Breitengrad) zugewiesen. Anschließend auf die Spalten und Zeilen verteilt (s. Abbildung 20)



Abbildung 20: Zuordnung von Longitude und Latitude

Im nächsten Schritt wurde die Vehicle Hal Id mit der Kennzahl *Anzahl* dem Markierungsbereich hinzugefügt. Dies bewirkt, dass alle Fahrräder einzeln betrachtet werden. Durch Hinzufügen der Start Rental Zone wird nun die Menge aller Fahrräder, die an der Abholstation entliehen wurden, aufgeführt (s. Abbildung 21).

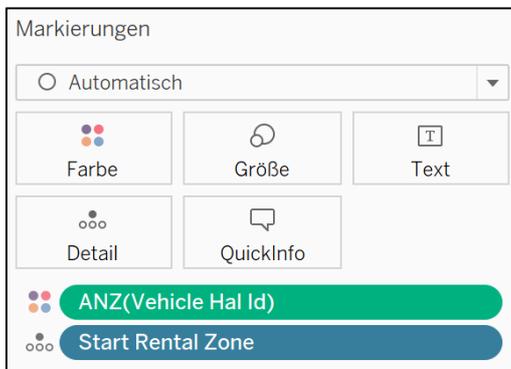


Abbildung 21: Markierungsbereich

Für eine gute visuelle Unterscheidbarkeit wurde eine farbliche Kennzeichnung gewählt. Wie der Abbildung 22 entnommen werden kann, ist das Maximum der Farbskala 121091 Fahrräder. Umso mehr Fahrräder also an einer Station geliehen wurden, desto dunkler ist die Färbung. Eine typische Heat-Map wurde erzeugt.

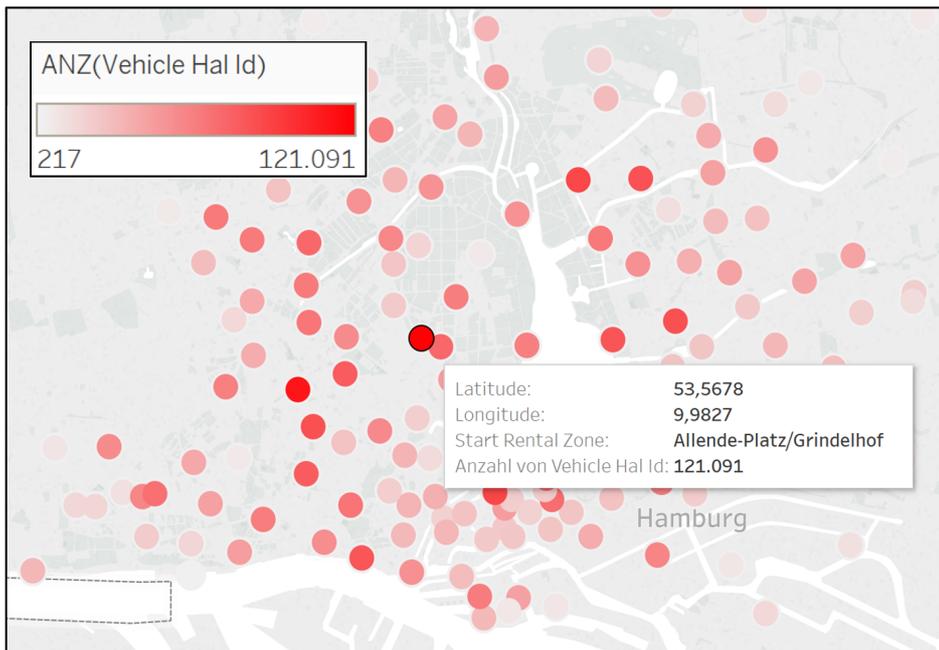


Abbildung 22: Visualisierung Szenario 1 – Frequentierte Orte

Die Station Allende-Platz/Grindelhof ist demnach die Station mit den meisten entliehenen Fahrrädern Hamburgs und entspricht auch der maximalen Färbung auf der Karte (121.091 ausgeliehene Fahrräder).

Im nächsten Schritt wurden die längsten Ausleihzeiten betrachtet. Dazu wurde das Datenblatt dupliziert und alle Markierungen entfernt, bis auf die Geokoordinaten und die Start Rental Zone.

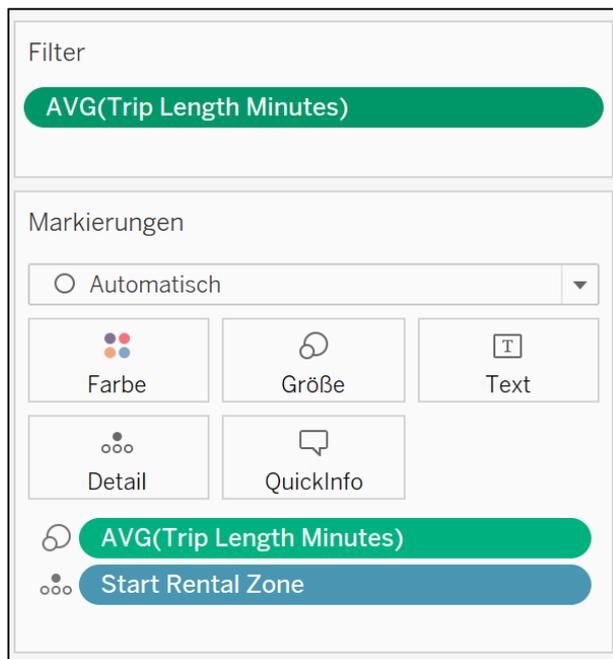


Abbildung 23: Filter und Markierungen - Länge der Leihzeiten

Wie in Abbildung 23 zu erkennen, wurde die Kennzahl Trip Length Minutes hinzugefügt und der Mittelwert gebildet. Als Filter wurde eine minimale Durchschnitts-Verleihzeit von 20 Minuten ausgewählt. Der maximale Durchschnittswert (pro Station) liegt bei 122,10 Minuten und wurde als obere Grenze beibehalten. Für diese Visualisierung wurde eine Darstellung der Durchschnitts-Leihminuten nach Größe ausgewählt.

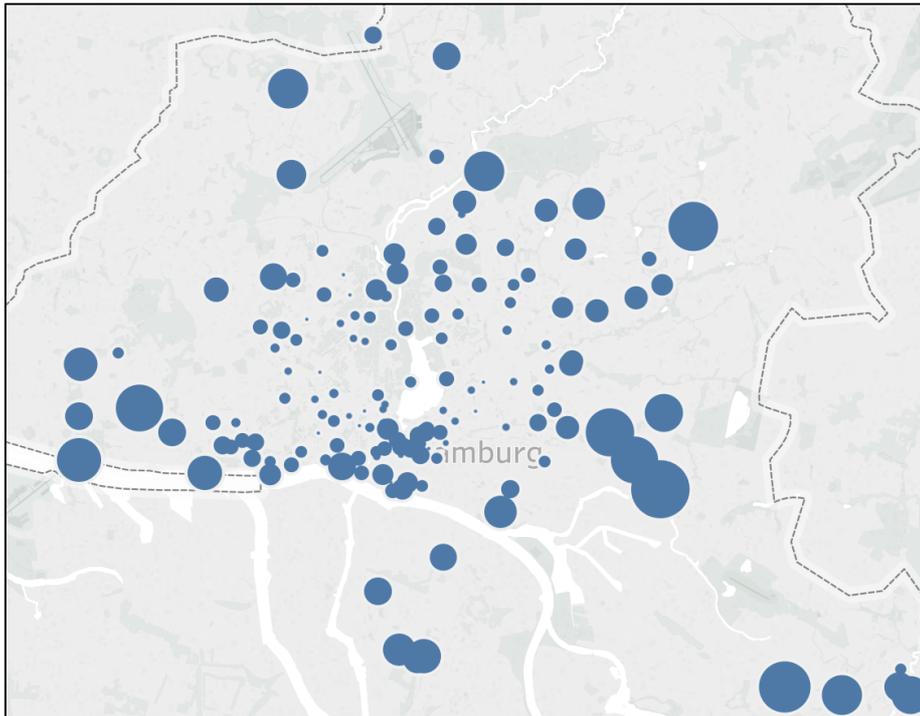


Abbildung 24: Visualisierung Szenario 1 - Leihminuten

Erkennbar ist auf den ersten Blick, dass die Durchschnittsleihminuten der entliehenen Räder an den Stationen außerhalb des Zentrums deutlich höher sind (s. Abbildung 24). Vor allem im Südosten sind sehr lange Ausleihzeiten erkennbar. Aus dieser Erkenntnis hat sich eine weitere Fragestellung ergeben, diese wurde in Szenario 3 analysiert.

3.2.2 Szenario 2 – HEAT Teststrecken

Die nachfolgende Frage wurde Bezug nehmend auf Abschnitt 2.1.4, das HEAT Projekt der Hamburger Hochbahn, formuliert.

Kann mit Hilfe der vorliegenden Daten für das Projekt HEAT ein Informationsgewinn generiert werden?

Idee: Ermittlung von sehr frequentierten Kurzleihstrecken von Stadträdern (~1 Kilometer) um diese als Teststrecken für autonome Fahrzeuge zu evaluieren

Benötigte Daten: StadtRad

4. Modellierung und Analyse

Die Schritte 1-4 des Big Data Zyklus konnten für dieses Szenario übersprungen werden, da die Daten aus dem Szenario 1 für die Fragestellung direkt genutzt werden. Es erfolgte lediglich eine neue Modellierung und Analyse.

Für die Visualisierung wurden Anzahl der Fahrräder und Start- und End Rental Zone hinzugefügt. Die weiter detaillierte Auswertung der Abbildung 25 erfolgt im Kapitel *Bewertung*.

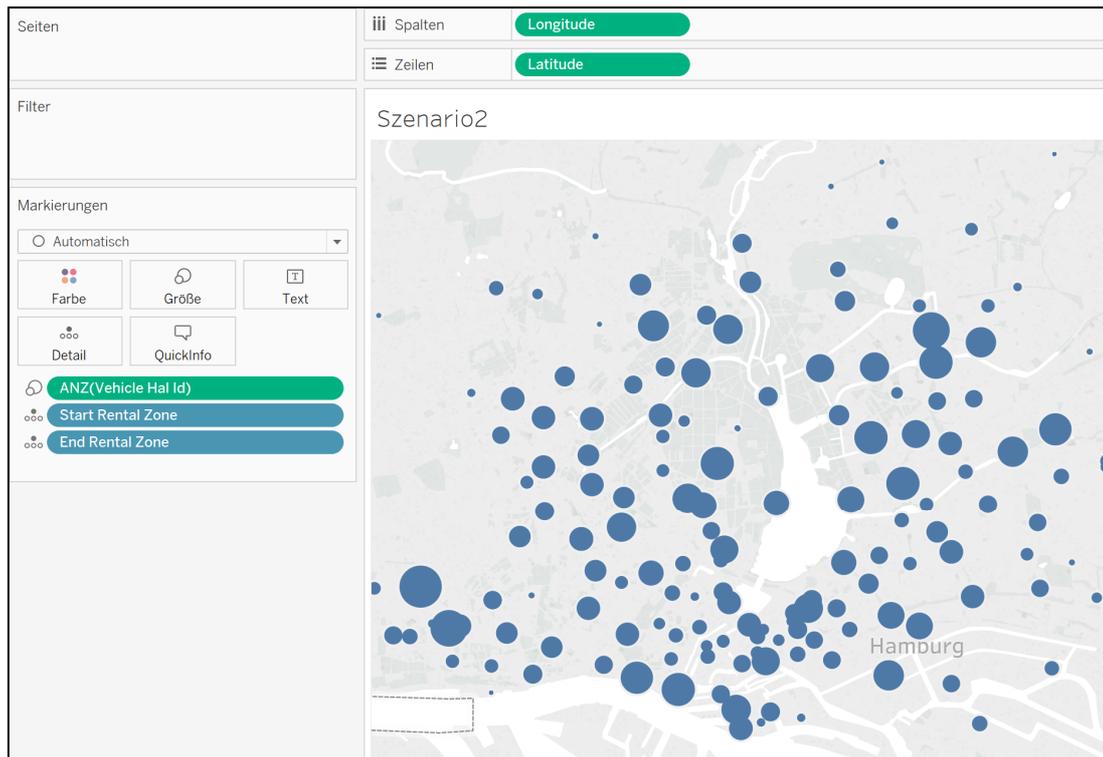


Abbildung 25: Frequentierte Abhol- und Rückgabestationen

3.2.3 Szenario 3 – Evaluierung neuer Switchh Standorte

Der dritte Durchlauf befasst sich mit der Evaluierung von neuen Switchh Stationen. Aus Szenario 1 ist hervorgegangen, dass vor allem im Südosten Hamburgs sehr viele Fahrräder verhältnismäßig lange ausgeliehen werden.

Welche neuen Gebiete könnten für weitere Switchh Stationen interessant sein?

Benötigte Daten: StadtRad, Switchh Stationen

Idee: StadtRad-Stationen mit den längsten Ausleihzeiten mit den Switchh Stationen verknüpfen, um Versorgungslücken sichtbar zu machen.

1. Datenerhebung

Für dieses Szenario wurde nun aufgrund fehlender Datenquellen mit Mock-Up Daten gearbeitet. Das bedeutet in diesem Fall, dass die Daten zur Veranschaulichung aufgrund von technisch nicht abrufbaren Quellen manuell gewonnen wurden. Switchh bietet keine Möglichkeit an diese Daten zu gelangen. Die Daten wurden daher händisch gesammelt und erfasst.

Dafür wurden die Switchh Stationen über die Website abgerufen (s. Abbildung 26) und anschließend über Google Maps die Geokoordinaten ermittelt.

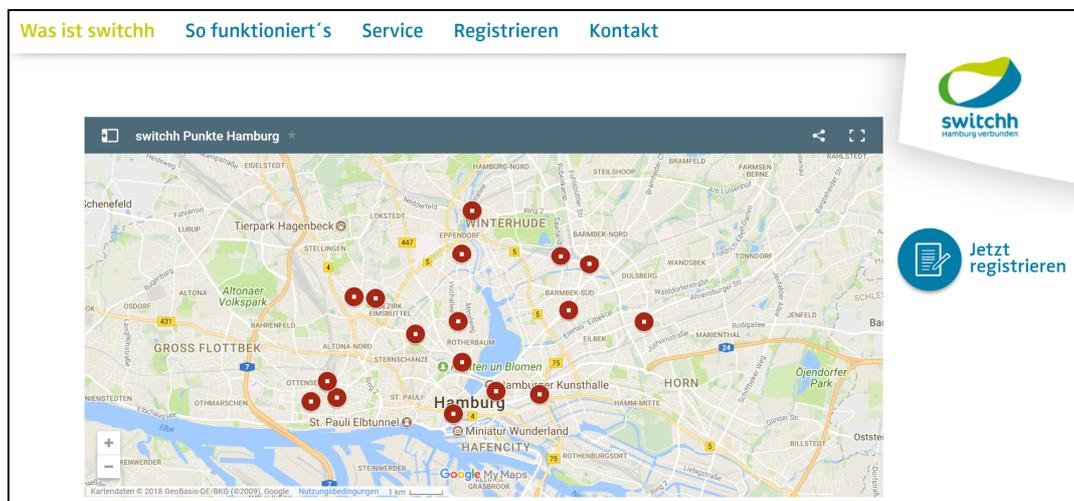


Abbildung 26: Switchh Stationen Hamburg (Switchh 2018)

Die Daten wurden, wie in Tabelle 7 ersichtlich, in die bereits vorhandene CSV Datei mit den Stationsdaten von StadtRad hinzugefügt.

switchh	Hannoversche Straße 85	9,9435796	53,5152578
switchh	Johann-Meyer-Straße 56	10,0445212	53,5323369
switchh	Bei der Reitbahn 3	9,9459435	53,5628809
switchh	Paul-Neermann-Platz 1	9,9350647	53,5521257
switchh	Nernstweg 32/34	9,9197438	53,5521244
switchh	Sillemstraße 65	9,9414913	53,57816
switchh	Heußweg 66	9,8672139	53,5520617
switchh	Schlump	9,9859308	53,5715917
switchh	Lattenkamp 78	9,9930413	53,60041
switchh	Goernestraße 21	9,9882123	53,589584
switchh	Rothenbaumchaussee 76-78	9,989319	53,578545
switchh	Theodor-Heuss-Platz	9,9789078	53,5768446
switchh	Rödingsmarkt 10	9,9941271	53,5613716
switchh	Ernst-Merck-Straße	10,002908	53,5538153
switchh	Beim Strohhouse 38	10,0217015	53,5530867
switchh	Wagnerstraße 5	10,0063806	53,5530854
switchh	Saarlandstraße 39	9,9538508	53,5530636
switchh	Wiesendamm 6	10,0170767	53,5698548
switchh	Schlossstrasse	10,0288465	53,5624105

Tabelle 7: Switchh Stationsdaten

Im nächsten Schritt wurde eine neue CSV Datei (Switchh.csv, s. Tabelle 8) mit einer selbst erzeugten ID, den Switchh Standorten und den entsprechenden Switchh Stationsnamen erzeugt, um diese später mit den StadtRad Daten zu vereinigen und den Stationsdaten zu verknüpfen.

SWITCHH_Switchh-KOORD	SWITCHH_RENTAL_ZONE
999901 Hannoversche Straße 85	switchh Punkt Harburg
999902 Johann-Meyer-Straße 56	switchh Punkt Bergedorf
999903 Bei der Reitbahn 3	switchh Punkt Bei der Reitbahn
999904 Paul-Nevermann-Platz 1	switchh Punkt Altona
999905 Nernstweg 32/34	switchh Punkt Nernstweg
999906 Sillemstraße 65	switchh Punkt Sillemstraße
999907 Heußweg 66	switchh Punkt Heußweg
999908 Schlump	switchh Punkt Schlump
999909 Lattenkamp 78	switchh Punkt Lattenkamp
999910 Goernestraße 21	switchh Punkt Kellinghusenstraße
999911 Rothenbaumchaussee 76-	switchh Punkt Hallerstraße
999912 Theodor-Heuss-Platz	switchh Punkt Dammtor
999913 Rödingsmarkt 10	switchh Punkt Rödingsmarkt
999914 Ernst-Merck-Straße	switchh Punkt Hauptbahnhof Nord
999915 Beim Strohhouse 38	switchh Punkt Berliner Tor
999916 Wagnerstraße 5	switchh Punkt Hamburger Straße
999917 Saarlandstraße 39	switchh Punkt Saarlandstraße
999918 Wiesendamm 6	switchh Punkt Barmbek
999919 Schlossstrasse	switchh Punkt Wandsbek Markt

Tabelle 8: Switchh ID und Zonen

Weitere Informationen zu den Switchh Stationen, wie Auslastung oder Anzahl der Leihvorgänge, standen bis dato nicht zur Verfügung.

2. Datensäuberung und Extraktion

In diesem Schritt wurden die bereits gesäuberten Daten aus Szenario 1 herangezogen. Die Mock-Up Daten wurden selbst erzeugt und mussten deshalb nicht bereinigt werden. Für dieses Szenario wurde ein neues Datenextrakt erzeugt. Die maximalen Leihminuten wurden wie im vorherigen Fall auf 10800 Minuten reduziert, um unbrauchbare und verfälschende Daten auszuschließen. Bei der Ausleihzone wurden Hamburg und NULL ausgewählt, da die Switchh-Daten die Spalte City Rental Zone nicht beinhalten und dadurch bei der Verknüpfung mit NULL aufgefüllt werden (s. Abbildung 27).

Daten extrahieren

Zu extrahierende Datenmenge angeben:

Filter (optional)

Filter	Details
CITY_RENTAL_ZONE	behält Null und Hamburg bei
TRIP_LENGTH_MINUTES	reicht von 1 bis 10.080 und behält Null-Werte bei

Hinzufügen... Bearbeiten... Entfernen

Aggregation

Daten für sichtbare Dimensionen aggregieren

Rollup für Daten bis []

Anzahl der Zeilen

Alle Zeilen

Inkrementelle Aktualisierung

Oberste: [] Zeilen

Verlauf... Alle nicht verwendeten Felder ausblenden OK Abbrechen

Abbildung 27: Datenextrakt für Szenario 3

3. Datenintegration und Datenrepräsentation

Die erzeugten Mock-Up Daten der Switchh Stationen und die StadtRad Daten wurden nun miteinander verbunden. Dazu wurden beide Datensätze über *Neue Vereinigung* im Bereich *Datenquelle* hinzugefügt. Über das Spaltenmenü wurden anschließend die Spalten mit gleichem Inhalt zusammengeführt (Abbildung 28).

#	Abc
Project	Project
VEHICLE_HAL_ID & SWITCHH_ID	START_RENTAL_ZONE & Switchh-KOORD

Abbildung 28: Zusammengeführte Spalten

Um die Stationsdaten beider verbundenen Datensätze zu matchen, wurde, wie bereits in den vorherigen Szenarien, die Start und End Rental Zone miteinander verknüpft. Die erfolgte Integration der drei unterschiedlichen Datenquellen ist in Abbildung 29 dargestellt.

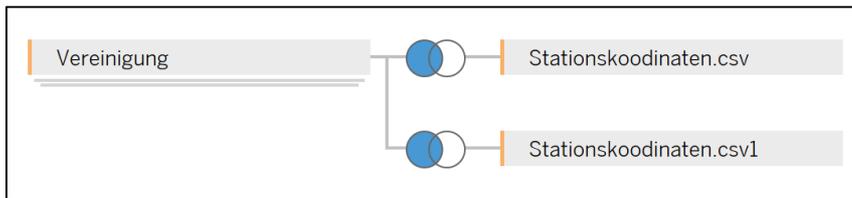


Abbildung 29: Intergration der 3 Datenquellen

4. Modellierung und Analyse

Wie in den Szenarien zuvor wurden die Geodaten auf die Zeilen und Spalten verteilt. Interessant war hier nun der Umgang mit den Switchh Daten.

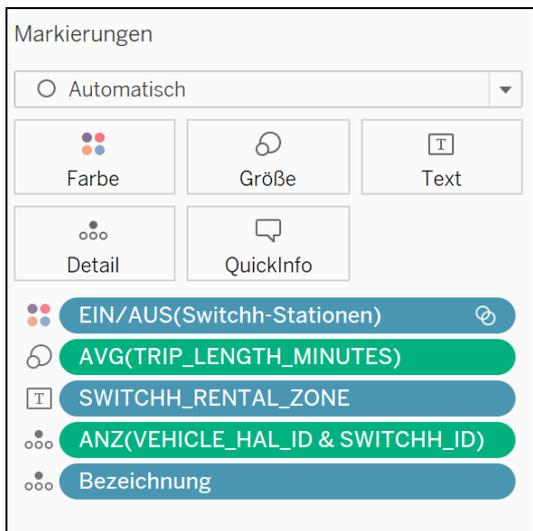


Abbildung 30: Modellierung Szenario 3

Hierfür wurde eine Gruppe *Switchh-Stationen* gebildet, um die 19 Switchh Stationen (Stand Juli 2018) auf der Karte zu kennzeichnen.

Zusätzlich wurden die durchschnittlichen Leihzeiten, der Name der entsprechenden StadtRad Station und die Anzahl der entliehenen Fahrräder auf die Karte gebracht. Das Feld *Bezeichnung* liefert den Namen der Station (s. Abbildung 30).



Abbildung 31: Visualisierung Szenario 3 - Switchh Stationen

In Kapitel 4 erfolgt die Auswertung der modellierten Daten. Abbildung 31 zeigt einen ersten Ausschnitt der Karte, dort wurden zur Veranschaulichung die Switchh Stationen ausgewählt.

3.2.4 Szenario 4 – Verbesserung der Orientierungsmöglichkeiten

Das letzte Szenario befasst sich nicht mit Visualisierungen im Allgemeinen. In sehr konzeptionellen Rahmen wird nun die nachfolgende Fragestellung bearbeitet. Im ersten Szenario wurde festgestellt, dass der Bereich Hamburg – Dammtor ein sehr frequentierter Bereich ist. Zudem befinden sich in unmittelbarer Nähe zwei Radstationen. In Kapitel 2.1.2 wurde auf die Strategie der Bürgerschaft eingegangen, die Kapazitäten im ÖPNV durch Verbesserung von Information, Service und Orientierungsmöglichkeiten der Kunden erhöhen zu wollen.

Ist es möglich, einen besseren Service und verbesserte Orientierungsmöglichkeiten an U/S-Bahn oder auch Switchh Stationen zu ermöglichen?

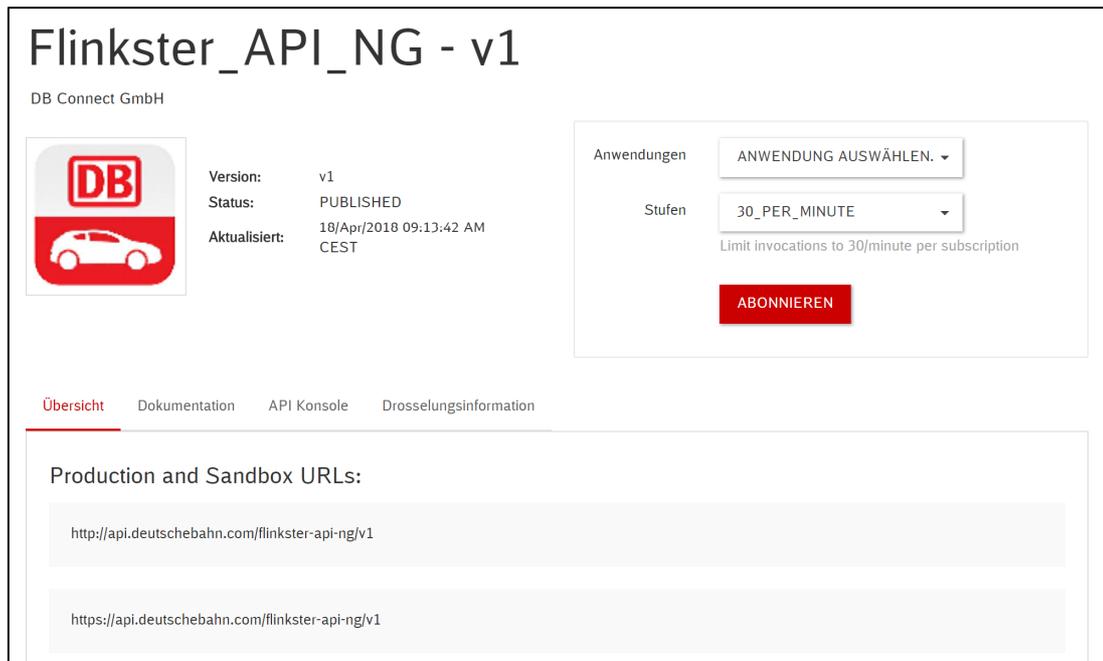
Idee: Monitore in Eingangshallen von U/S-Bahn Stationen, die eine aktuelle Auskunft über die Radstationen in der Nähe eines Bahnhofs liefern. So könnte auf das ÖPNV Angebot und die Verfügbarkeiten direkt beim Aus- oder Umstieg hingewiesen werden.

Benötigte Daten: Echtzeitdaten um Verfügbarkeiten von Standorten abzufragen. Die nächsten Schritte befassen sich konzeptionell mit der Gewinnung der Daten für ein solches Projekt.

1. Datenbeschaffung

Das Portal der Deutschen Bahn bietet eine REST Schnittstelle an, mit der Echtzeitdaten an verschiedenen Stationen abgerufen werden können (s. Abbildung 32).

Um eine Verbindung mit der Schnittstelle herzustellen wurden im Rahmen der Bachelorarbeit eine Registrierung durchgeführt und die API eingerichtet.



The screenshot shows the 'Flinkster_API_NG - v1' API portal. At the top left is the DB logo and a car icon. To the right, it displays metadata: Version: v1, Status: PUBLISHED, and Aktualisiert: 18/Apr/2018 09:13:42 AM CEST. On the right side, there are two dropdown menus: 'Anwendungen' with 'ANWENDUNG AUSWÄHLEN.' and 'Stufen' with '30_PER_MINUTE'. Below the 'Stufen' menu, it says 'Limit invocations to 30/minute per subscription'. A red 'ABONNIEREN' button is positioned below these options. At the bottom, there are navigation links: 'Übersicht', 'Dokumentation', 'API Konsole', and 'Drosselungsinformation'. Below the navigation, a section titled 'Production and Sandbox URLs:' contains two text boxes with the following URLs: 'http://api.deutschebahn.com/flinkster-api-ng/v1' and 'https://api.deutschebahn.com/flinkster-api-ng/v1'.

Abbildung 32: Flinkster Schnittstelle (API-Portal 2018)

Anschließend wurde mittels Python ein Script erstellt, mit welchem aktuelle Daten im Umkreis der Station Hamburg Dammtor abgerufen werden können (s. Abbildung 33)

```
#Dieses Script stellt eine Verbindung zum DB Open Data Portal her und
#ruft die Echtzeitdaten vom Standort Dammtor ab
#
# Author David Stutz WI6 im Rahmen der Bachelorarbeit

#Import des http Clients
import http.client

#Herstellen der Connection
conn = http.client.HTTPSConnection("api.deutschebahn.com")

#Zertifizierung mit der Schnittstelle
headers = {
    'authorization': "Bearer f2748c08e0ffcd06b44b2146a35dbdc8",
    'content-type': "application/x-www-form-urlencoded",
    'host': "api.deutschebahn.com",
    'x-amz-date': "20180409T182647Z",
    'cache-control': "no-cache",
    'postman-token': "dbd6c477-80b2-3944-73ee-a4e2cac20944"
}

#GET Request auf die API
conn.request("GET", "/flinkster-api-ng/v1/bookingproposals?lat=53.560833&lon=9.989444&radius=10000&providernetz=2&limit=50",
            headers=headers)
res = conn.getresponse()
data = res.read()

#Öffnen und Befüllen der JSON Datei
jsonfile = open('cab.json', 'w')
jsonfile.write(data.decode("utf-8"))
```

Abbildung 33: Python Script zum Abruf der REST Schnittstelle

Dem GET request wurden dazu alle wichtigen Informationen mitgegeben.

```
conn.request("GET", "/flinkster-api-ng/v1/bookingproposals?
lat=53.560833&lon=9.989444&radius=10000&providernetz=2&limit=50",
headers=headers)
```

Geokoordinaten: *lat=53.560833&lon=9.989444*

Radius der abgefragt werden soll: *radius=10000*

Abruf von Flinkster (1) oder StadtRaddaten(2): *providernetz=2*

Limit der abzurufenden Räder: *limit=50*

Die Daten werden anschließend in einem JSON Dokument gespeichert.

2. Datensäuberung und Modellierung

Aufgrund der begrenzten Zeit wurden die gewonnen Rohdaten nicht weiter aufbereitet bzw. verarbeitet. Welche Möglichkeiten die weitere Bearbeitung dieser Daten bieten könnte, wird in Kapitel 4.1 im Detail erläutert. Der Big Data Lebenszyklus wird somit nicht weiter durchlaufen und das Szenario beendet.

3.2.5 Nicht betrachtete Fragestellungen

Die nachfolgenden Fragestellungen wurden für diese Arbeit ebenfalls als Szenario in Betracht gezogen, jedoch aufgrund von fehlenden belastbaren Daten verworfen und nicht weiter konkretisiert.

- Zu welchen Zeiten (Großevents, Wetter, Uhrzeit) sind Bus und Bahnen besonders frequentiert? Wie beeinflussen unterschiedliche Begebenheiten den Modal Split?
- Beeinflusst eine Switchh Verleihstation die Nutzerzahlen der U/S-Bahn oder die Nutzerzahlen von StadtRad?

Dies sind aus hiesiger Sicht ebenfalls interessante Fragestellungen, die jedoch unbeantwortet bleiben werden. Es fehlen die Daten des HVV um hier sinnvolle Visualisierungen zu erzeugen.

4 Bewertung

In diesem Kapitel erfolgt die Bewertung der gesammelten Erkenntnisse. Die einzelnen Szenarien aus Kapitel 3 werden, wie in Abschnitt 3.1.1. beschrieben und nach dem vorgestellten Big Data Lifecycle üblich, nun interpretiert. Anschließend wird eine allgemeine Bewertung vorgenommen.

4.1 Interpretation der Szenarien

Szenario 1 - Beeinflussung des Modal Split

Im ersten Schritt wurde ermittelt, welche Stationen sehr frequentiert sind. Die fünf Stationen mit den meisten ausgeliehenen Fahrrädern sind in Tabelle 9 aufgeführt.

Latitude:	53,5678
Longitude:	9,9827
Start Rental Zone:	Allende-Platz/Grindelhof
Anzahl von Vehicle Hal Id:	121.091
Latitude:	53,5627
Longitude:	9,9610
Start Rental Zone:	Schulterblatt/Eifflerstraße
Anzahl von Vehicle Hal Id:	111.035
Latitude:	53,5522
Longitude:	9,9955
Start Rental Zone:	Jungfernstieg / Ballindamm
Anzahl von Vehicle Hal Id:	87.665
Latitude:	53,5838
Longitude:	10,0100
Start Rental Zone:	Goldbekplatz / Semperstraße
Anzahl von Vehicle Hal Id:	87.213
Latitude:	53,5696
Longitude:	10,0271
Start Rental Zone:	Mundsburg / Schürbeker Straße
Anzahl von Vehicle Hal Id:	82.623

Tabelle 9: Frequentierte Stationen Top 5

Diesen Stationen sollte eine größere Beachtung geschenkt werden, da es vorstellbar ist, dass es hier zu Engpässen zu Stoßzeiten kommen könnte. Diese Stationen sind interessant für einen Ausbau, bzw. die Umgebung lukrativ für die Einrichtung einer weiteren Station.

Im zweiten Teil dieses Szenarios wurde geprüft, von welchen Stationen Fahrräder am längsten ausgeliehen wurden. Auffällig war hier, dass vor allem Stadträder von Stationen der Randgebiete Hamburgs ungewöhnlich lange ausgeliehen werden. Es lässt vermuten, dass dort viele Fahrräder für einen längeren Ausflug genutzt werden. Einen weiteren Anhaltspunkt für diese Theorie könnte hier die Ermittlung des Wochentags liefern, an dem die Fahrräder ausgeliehen werden. Dies wäre mit den vorliegenden Datensätzen durchaus möglich, wurde aber aufgrund des knappen Zeitrahmens nicht weiterverfolgt.

Nördlich	
Latitude:	53,6408
Longitude:	9,9501
Start Rental Zone:	Nordalbinger Weg/Paul-Sorge-Straße
Durchschn. Trip Length Minutes:	55,55
Östlich	
Latitude:	53,6062
Longitude:	10,1202
Start Rental Zone:	Rahlstedter Weg/Berner Heerweg
Durchschn. Trip Length Minutes:	74,66
Südlich	
Latitude:	53,4997
Longitude:	9,9965
Start Rental Zone:	Barkassenanleger/Mengestraße
Durchschn. Trip Length Minutes:	43,79
Westlich	
Latitude:	53,5474
Longitude:	9,8620
Start Rental Zone:	Elbchaussee/Teufelsbrück
Durchschn. Trip Length Minutes:	63,29

Tabelle 10: Stationen mit den längsten Ausleihezeiten nach Himmelsrichtung

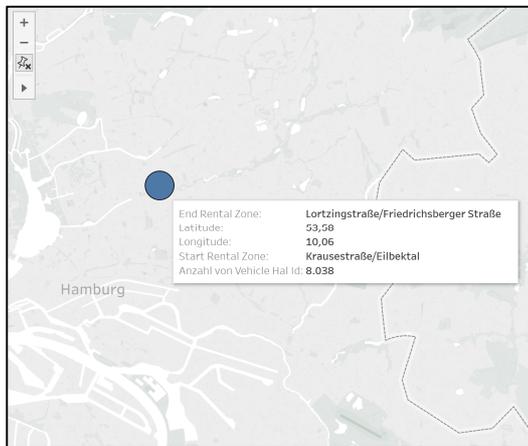
Diese Stationen (s. Tabelle 10) könnten beispielsweise mit Elektrofahrrädern oder Lastenrädern ausgerüstet werden, aber auch weitere Mobilitätsangebote könnten in Betracht kommen. Aus diesen Ergebnissen wurde Szenario 3 abgeleitet.

Zusammengefasst: mit den vorliegenden Daten wurden Informationen ermittelt, die bei der Veränderung des Modal Split hin zur Nutzung von mehr Fahrrädern durchaus nützlich sein können.

Szenario 2 – HEAT Teststrecken

Aus den modellierten Daten des zweiten Szenarios wurden nun exemplarisch zwei Strecken ermittelt, die nach rein subjektiver Einschätzung gute Bedingungen für eine Teststrecke des autonomen Busses des Projektes HEAT darstellen. Aus den vorliegenden Stadrad-Daten wurde die Erkenntnis gewonnen, dass beide Strecken sehr frequentiert sind. Des Weiteren sind beide Strecken relativ kurz (etwas über einen Kilometer), jedoch zu lang um diese Strecke zu Fuß zurückzulegen. Abbildung 33 zeigt die Station *Krausestraße* im Osten Hamburgs. Mit 8038 geliehenen Fahrrädern auf der Route Lortzingstraße – Krausestraße würde sich hier ein Einsatz eines autonom fahrenden Kleinbusses höchstwahrscheinlich anbieten. Mit Google Maps wurden die Streckeneigenschaften ermittelt. Die Strecke hat eine Gesamtlänge von 1,5km und würde mit einem Fahrzeug in 5 Minuten zurückgelegt werden.

Erste ermittelte Strecke



Analyse mit Google Maps

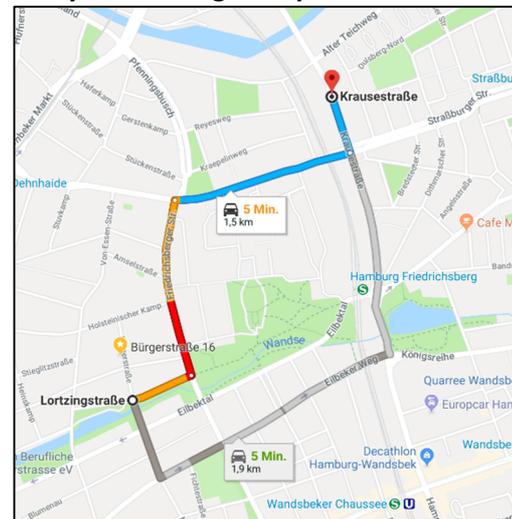
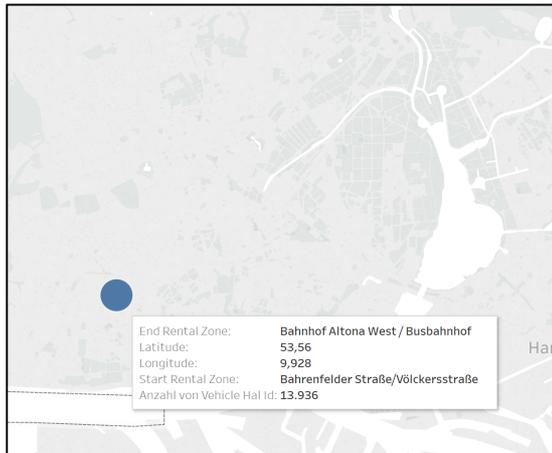


Abbildung 34: Lortzingstraße – Krausestraße (Rechts: Google Maps 2018)

Die Zweite Strecke ist mit 13936 ausgeliehenen Fahrrädern noch frequentierter als die erste Strecke. Sie befindet sich in Altona im Westen Hamburgs und umfasst 1,3km. Diese Strecke könnte ebenfalls interessant für autonom fahrende Fahrzeuge sein (s. Abbildung 35)

Zweite ermittelte Strecke



Analyse mit Google Maps

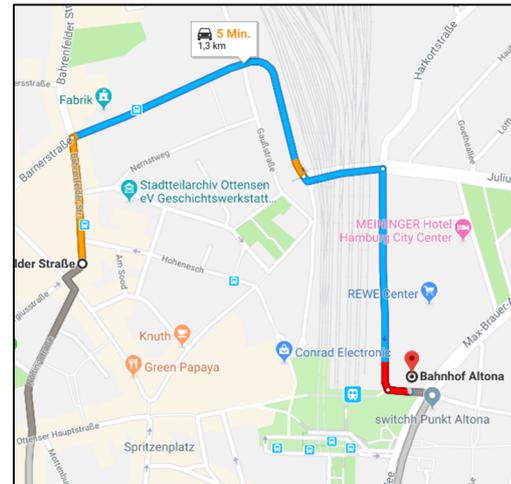


Abbildung 35: Bahrenfelder Straße – Bahnhof Altona West (Rechts: Google Maps 2018)

Zusammengefasst: aus den vorliegenden Daten wurden Informationen gewonnen, die zeigen, dass einzelne Stationen eine überdurchschnittlich hohe Frequenz zwischen Station zu Station aufweisen und diese Erkenntnis für das Projekt HEAT, aber auch für andere Themen, nützlich sein könnten. Es ist zum Beispiel durchaus auch eine Versorgung mit On-Demand Mobilitätsangeboten auf diesen Strecken vorstellbar.

Szenario 3 – Evaluierung neuer Switchh Standorte

Durch die Kombination der StadtRad Daten und der Mock-Up Daten der Switchh Stationen konnten ebenfalls interessante Erkenntnisse gewonnen werden. Im Südosten Hamburgs befinden sich vier StadtRad Stationen, von denen aus Fahrräder verhältnismäßig lange ausgeliehen werden. Die Station mit der stärksten Freqüentierung und den längsten Ausleihzeiten in diesem Bereich ist die Station Horner Rennbahn/Meurer Weg. Wie in Abbildung 36 kenntlich gemacht, befindet sich an diesem Standort bzw. auch in unmittelbarer Nähe keine Switchh Station mit weiteren Mobilitätsangeboten, über die beispielsweise ein Fahrzeug ausgeliehen werden könnte. Es wäre deshalb ratsam, die Gegend der Horner Rennbahn für einen Ausbau einer weiteren Switchh Station in Betracht zu ziehen. Durch die Zusammenführung können durchaus auch noch weitere Informationen gewonnen werden, jedoch zeigt dieses exemplarische Beispiel das Potential von kombinierten Mobilitätsdatensätzen und ist daher ausreichend.

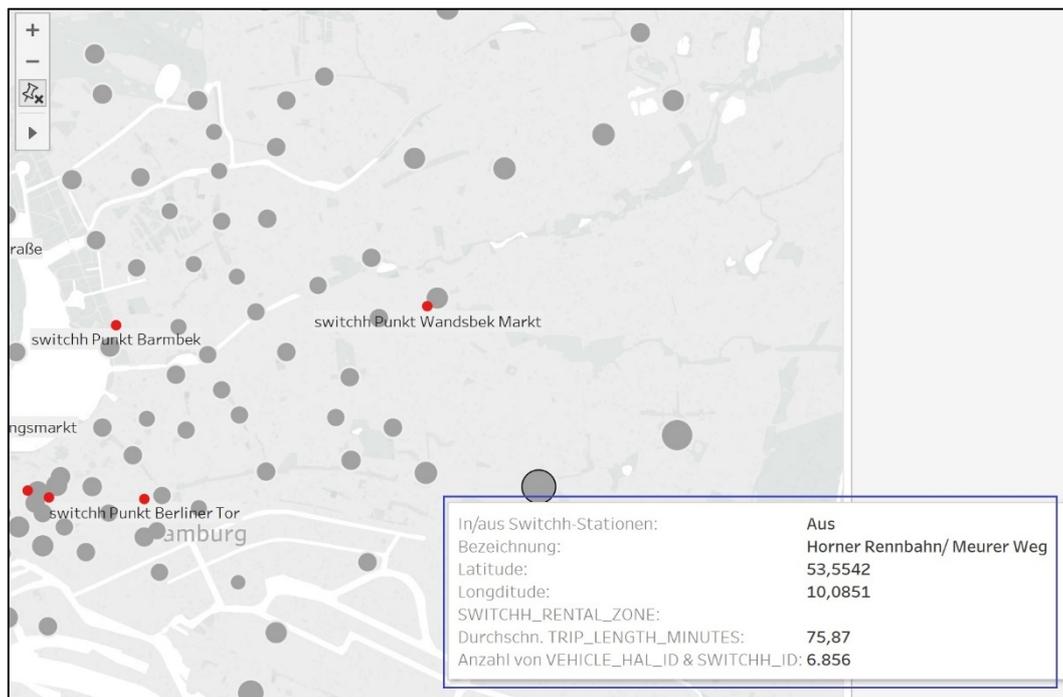


Abbildung 36: StadtRad Station Horner Rennbahn

Zusammengefasst: Die Verbindung der beiden Datentöpfe hat in diesem kleinen Rahmen schon einen erheblichen Mehrwert geliefert, obwohl es sich lediglich um manuell eingespielte Mock-Up Daten handelt. Mit weiteren Informationen zu den jeweiligen Switchh Stationen, zum Beispiel in dem Umfang der öffentlich zugänglichen StadtRad-Daten (Länge der Ausleiherzeit, frequentierte Strecken, gefragte Stationen) wären noch detailliertere und aufschlussreichere Auswertungen möglich.

Szenario 4 – Verbesserung der Orientierungsmöglichkeiten

Mit den zur Verfügung gestellten Daten und dem vorgestellten Vorgehen, also dem Abrufen von Echtzeitdaten, könnte in der Theorie mit wenig Aufwand ein grafisches Informationssystem entwickelt werden, welches den Nutzern des ÖPNV einen großen Mehrwert bieten würde. Bei Verlassen der S/U-Bahn Station könnten mit einer Infotafel direkt auf die in der Nähe befindlichen weiteren Mobilitätsangebote (in dem Fall StadtRad) mit den entsprechenden Verfügbarkeiten hingewiesen und der Reisende dazu animiert werden, beispielsweise direkt auf ein StadtRad umzusteigen.

Zusammengefasst: die Echtzeitdaten die über die REST Schnittstelle abgerufen werden können, ermöglichen interessante neue Blickwinkel und Ideen und befördert kreative Ansätze, die durchaus einen Mehrwert für den Kunden bieten könnten.

4.2 Szenarioübergreifende Bewertung

Fachliche Einschätzung

Die vier bearbeiteten Szenarien haben unterschiedliche Ziele verfolgt und gezeigt, welches Potential offene Mobilitätsdaten bieten können.

Obwohl für die ersten beiden Szenarien nur eine Datenquelle genutzt wurde, waren die Ergebnisse bereits überzeugend. Kombiniert mit anderen Datensätzen nehmen der Informationsgehalt und die Möglichkeiten, diese Daten sinnvoll einzusetzen, zu. Auch fernab der Visualisierung sind die offenen Mobilitätsdaten in vielen Anwendungsbereichen durchaus sinnvoll einsetzbar und bieten einen Mehrwert für andere anbieterübergreifende Projekte. Dies hat die Recherche für Abschnitt 2.1.3 ebenfalls gezeigt. Bei den Mobilitäts-Hackathons wurden zahlreiche interessante Projekte realisiert und vielversprechende Prototypen entwickelt. Dies zum Teil jedoch auch mit Daten, die nicht der Allgemeinheit zur Verfügung stehen, sondern lediglich im Rahmen des jeweiligen Hackathons.

Technische Bewertung

Für Massendaten unterschiedlicher Quellen eignet sich das Vorgehen wie in Kapitel 3 jedoch nur bedingt. Vor allem die verwendete Architektur würde den Ansprüchen nicht genügen. Grund dafür sind die erforderliche Robustheit, Skalierbarkeit und Erweiterbarkeit die eine moderne Big Data Architektur bieten sollte. Auf Plattform-Ebene ist zu erwähnen, dass die genutzte Hardware (Intel Core i5 2,5Ghz und 4 GB Arbeitsspeicher) bei der Erstellung der Extrakte und bei der Berechnung der Visualisierungen voll ausgelastet war.

Die Methodik des Bearbeitens der Daten in Zyklen hingegen hat sich als sehr sinnvoll erwiesen. Auch für geschlossene Daten eignet sich die Vorgehensweise, wie sie in Kapitel 3 durchgeführt wurde und weiter abstrahiert ist das zyklische Vorgehen auch für völlig andere Massendaten durchaus geeignet, da die Big Data Tools, in diesem Fall Tableau, sich sehr gut dafür eignen.

5 Fazit

Diese Bachelorarbeit hat sich mit der Frage beschäftigt, welches Potential offene Mobilitätsdaten den Datenhaltern und der Gesellschaft bereithält. Dazu wurden verschiedene Thematiken wie Open Data, Big Data und deren Visualisierung und Mobilität behandelt. Anschließend erfolgte in verschiedenen Szenarien die strukturierte Visualisierung und Auswertung der recherchierten offenen Mobilitätsdaten.

Die Ergebnisse zeigen, dass die Öffnung von Datenquellen in der Mobilität, aber auch allgemein einen deutlichen Informationsgewinn und Innovationsschub erwirken kann. Und das nicht nur durch Visualisierungen, sondern auch durch anderweitige Verwendung der Daten.

Die Initiative der Deutschen Bahn zur Öffnung der Unternehmensdaten ist bisher eher noch Ausnahme als die Regel, aber es bleibt festzuhalten, dass die zur Verfügung gestellten Daten eine Goldgrube für spannende Analysen aber auch Projekte darstellen, die ohne die Öffnung der Daten für die Allgemeinheit keinesfalls möglich wären. Dies untermauern ebenfalls die vielseitigen Projekte, die in den verschiedenen Mobility Hackathons realisiert wurden.

Festzuhalten bleibt jedoch, dass viele Anbieter die Freigabe ihrer Daten nicht in Betracht ziehen oder die Daten in einem Format vorliegen die die Weiterverarbeitung erschweren oder gar verhindern.

Big Data in der Mobilität ist weniger das Ziel als die logische Konsequenz der fortschreitenden Digitalisierung und es scheint, dass Digitalisierung und Datenintegration alle Stakeholder voranbringt. Und eine Öffnung weiterer Mobilitätsdaten würde den Unternehmen und Institutionen des Mobilitätssektors und der Gesellschaft einen noch größeren Vorteil verschaffen.

5.1 Ausblick

Die Bearbeitung dieser Arbeit hat gezeigt, dass eine übergeordnete moderne Big Data Architektur zur Zentralisierung aller Mobilitätsdaten sehr sinnvoll wäre. Vorstellbar wäre ein Aufbau in Form eines deutschlandweiten Data-Warehouse, in dem die neuesten Erkenntnisse im Umgang mit der Verarbeitung von Massendaten mit einfließen könnten. Hier wären die in Abschnitt 3.1.4 erwähnten Big Data-Architekturen Lambda und Kappa nochmals zu nennen, die eine Batchverarbeitung von ruhenden Datenquellen und auch eine Echtzeitverarbeitung ermöglichen könnten. Zusätzliche Komponenten, wie zum Beispiel für Machine Learning könnten ebenfalls integriert werden. Diese Architektur könnte dann offene Schnittstellen anbieten, oder auch die Möglichkeit Massendaten in gängigen Datenformaten zu beziehen. Das Konzept der Datenzentralisierung verfolgt bereits das BMVI mit dem Mobilitäts-Daten-Marktplatz.

Mobilitäts-Daten-Marktplatz und Datex II

Mit der Einrichtung des MDM haben sich die Bedingungen zur Weitergabe der Verkehrsinformationen wesentlich verbessert. Der Anschluss von Datenabnehmern erfolgt über standardisierte Datex II-Datenmodelle, die zur Festlegung der Dateninhalte gemeinsam mit den Städten entwickelt wurden. Für Datenbereitsteller und -abnehmer vereinfacht sich die Handhabung, weil nicht unterschiedliche Schnittstellen und Formate bedient werden müssen und die Anforderungen an die städtischen IT-Sicherheitsanforderungen besser erfüllt werden können. Von diesem vereinfachten Zugang wird erwartet, dass mit zunehmender Teilnehmerzahl am Marktplatz auch unter der Nutzung neuer Übertragungsmöglichkeiten Verkehrsteilnehmer künftig verbesserte Informationen zur Nutzung der kommunalen Verkehrsnetze erhalten könnten. Ein im Rahmen des MDM entwickelter Muster-Datenüberlassungsvertrag (oder in Zukunft evtl. Open Data-Lizenzen) reduzieren den Aufwand der rechtlichen Regelung zur Datennutzung. (Vgl. Stieler, 2015). In wie weit sich das Datex II Format etablieren wird und ob sich genügend Datenanbieter beteiligen ist jedoch nicht sicher und gilt es zu prüfen.

In wie weit dieses Portal des Bundesministeriums für Verkehr und Infrastruktur angenommen wird und ob es sich etabliert, wird die Zukunft zeigen und ist bis dato nicht sicher. Für die Metropolregion Hamburg sind mit dem Hamburger Verkehrsverbund rein subjektiv betrachtet jedoch die optimalen Voraussetzungen geschaffen, um in einem gemeinschaftlichen Projekt ein Data-Warehouse zu entwickeln um alle Mobilitätsdaten der Akteure zu zentralisieren und diese Datensätze, wenn auch nur partiell, der Allgemeinheit zur Verfügung zu stellen.

Quellenverzeichnis

Fallmann, Daniel. „The Big Six: Sechs Big Data-Megatrends 2016“. (2015). Mindbreeze. Zugriffen 27. Juli 2018. <https://www.mindbreeze.com/de/blog/the-big-six-sechs-big-data-megatrends-2016.html>.

Hamburg.de „ITS-Weltkongress 2021 findet in Hamburg statt“. (2017). Zugriffen 2. August 2018. <https://www.hamburg.de/pressearchiv-fhh/9806016/2017-11-01-bwvi-its-weltkongress-2021/>.

HmbtG „Das Hamburgische Transparenzgesetz“. hamburg.de. (2012). Zugriffen 27. Juli 2018. <http://transparenz.hamburg.de/das-hmbtg/>.

Hochbahn „Das Projekt HEAT – autonome Kleinbusse in der HafenCity“. (2018). Zugriffen 27. Juli 2018. https://www.hochbahn.de/hochbahn/hamburg/de/Home/Naechster_Halt/Ausbau_und_Projekte/projekt_heat.

BWVI „ITS-Weltkongress 2021 findet in Hamburg statt“. hamburg.de. (2017). Zugriffen 27. Juli 2018. <https://www.hamburg.de/pressearchiv-fhh/9806016/2017-11-01-bwvi-its-weltkongress-2021/>.

„Megatrend Mobilität“, 24. Mai 2018. (2018) Zugriffen 25.07.2018. <https://www.zukunftsinstitut.de/dossier/megatrend-mobilitaet/>.

bpb - Bundeszentrale für politische Bildung. „Megatrends? | bpb“. (2015). Zugriffen 25. Juli 2018. <http://www.bpb.de/shop/zeitschriften/apuz/209968/megatrends>.

switchh. „switchh Punkte Hamburg“. (2018). Zugriffen 25. Juli 2018. https://www.google.com/maps/d/viewer?mid=1Rt6CBV9QVnIEfd5MtSm8-hW6_6U&hl=de.

Flügge, Barbara, Hrsg. *Smart Mobility*. (2016). Wiesbaden: Springer Fachmedien Wiesbaden.

Infas. „Mobilität im Großraum Hamburg“. (2008). Regionalkonferenz Verkehr & Mobilität am 4. November 2011.

Kirchhoff, P. „Städtische Verkehrsplanung“. (2002). Teubner Verlag, Stuttgart.

Randelhoff, Martin. „[Kurz erklärt] Was ist der Modal Split und was sagt er aus? » Zukunft Mobilität“. (2018). *Zukunft Mobilität* (blog), 9. April 2018. <https://www.zukunft->

mobilitaet.net/167600/analyse/was-ist-der-modal-split-grenzen-verkehrsmittelwahl-einschraenkungen-wege-verkehrsleistung/.

Koalitionsvertrag über die Zusammenarbeit in der 21. Legislaturperiode der Hamburgischen Bürgerschaft. „Zusammen schaffen wir das moderne Hamburg“. (2015). Zugegriffen 26.07.2018. https://www.spd-fraktion-hamburg.de/uploads/tx_wfpresse/koalitionsvertrag_download.pdf

Bürgerschaft der FHH. „Mitteilung des Senats an die Bürgerschaft Mobilität in Hamburg – Ziele“. (2017). Zugegriffen 25.07.2018. <https://www.hamburg.de/contentblob/11221894/e14ddb50d45ee9778ed827bba297fdd/data/drucksache-21-7748-mobilitaet-in-hamburg-ziele.pdf>

Beyer, Marc. (2011). „Gartner Says Solving ‚Big Data‘ Challenge Involves More Than Just Managing Volumes of Data“. Zugegriffen 10. Mai 2018. <https://www.gartner.com/newsroom/id/1731916>.

VDV. (2018). „Statistik Personenverkehr - Verband Deutscher Verkehrsunternehmen - VDV“. Zugegriffen 10. Mai 2018. <https://www.vdv.de/statistik-personenverkehr.aspx>.

Fasel, Daniel, und Meier, Andreas Hrsg. Big Data: Grundlagen, Systeme und Nutzungspotenziale. (2016). Edition HMD. Wiesbaden Springer Vieweg.

H. V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, Cyrus Shahabi. (2014). Communications of the ACM, Vol. 57 No. 7, Pages 86-94.

Dapp, Marcus M., Dian Balta, Walter Palmetshofer, Helmut Krcmar, und Pencho Kuzev. (2016). Open Data. The Benefits: das volkswirtschaftliche Potential für Deutschland. Sankt Augustin Berlin: Konrad-Adenauer-Stiftung e.V.

Opendata Zürich. „Open data der stadt zürich was bringt es“. (2014). Präsentationen & Vorträge. <https://de.slideshare.net/Opendatazurich/open-data-der-stadt-zrich-was-bringt-es>.

FHH Beteiligungsbericht. (2016). Zugegriffen 2. Juni 2018. <http://beteiligungsbericht.fb.hamburg.de/Download.html>.

Weck, Andreas. OpenData – wer nicht wagt, der nicht gewinnt?! - Netzpiloten.de. (2012). *Netzpiloten Magazin* (blog), 1. September 2012. <https://www.netzpiloten.de/opendata-wer-nicht-wagt-der-nicht-gewinnt/>.

„IT-TRANS 2016: Digitalisierung im ÖPNV“. (2016). *Intelligente Welt* (blog), Zugegriffen 8. März 2016. <http://intelligente-welt.de/oepnv-im-rampenlicht-die-wichtigsten-erkenntnisse-der-it-trans-2016/>.

„BMVI - Mobilitäts Daten Marktplatz (MDM)“. (2018). Zugegriffen 3. Juni 2018. <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mobilitaets-daten-marktplatz.html>.

BAST, Bundesanstalt für Straßenwesen: MDM – Die Herausforderung: Ein geordneter Markt für Verkehrsdaten. (2018). Zugegriffen 03.06.2018 <http://www.mdm-portal.de/herausforderung.html>.

Stieler, Patric, Volker Kanngiesser, und Florian Hilti. „MobilitätsDatenMarktplatz – welche Chancen ergeben sich für Städte und Gemeinden?“. (2015). Germany: Wichmann Verlag. http://gispoint.de/gisopen-paper?id=127&tx_browser_pi1%5BshowUid%5D=1571.

„BMVI - mCloud“. (2018) Zugegriffen 3. Juni 2018. <https://www.bmvi.de/DE/Themen/Digitales/mFund/mCloud/mcloud.html>.

Jörn von Lucke, Christian Geiger „Open Government Data - Frei verfügbare Daten des öffentlichen Sektors“. (2010). zeppelin university. Zugegriffen 22.06.2018. <https://www.zu.de/institute/togi/assets/pdf/TICC-101203-OpenGovernmentData-V1.pdf>.

„Open Data: Deutsche Bahn gibt Schnittstelle für Fahrplandaten frei“. (2016). Heise Online. Zugegriffen 24. Juni 2018. <https://www.heise.de/newsticker/meldung/Open-Data-Deutsche-Bahn-gibt-Schnittstelle-fuer-Fahrplandaten-frei-3118919.html>.

„FAQ - Open-Data-Portal – Deutsche Bahn Datenportal“. (2018). Zugegriffen 24. Juni 2018. <http://data.deutschebahn.com/faq>.

„Geoportal der Metropolregion Hamburg - MetaVer“. (2016). Zugegriffen 23. Juli 2018. <http://www.metaver.de/trefferanzeige?docuuid=322BC056-BC3D-4ED1-A481-EE3068A17433>.

„Geo-Online“. (2018). Zugegriffen 23. Juli 2018. <https://geoportal-hamburg.de/geoportal/geo-online/index.html#>.

Leckart, Steven. „The Hackathon Is On: Pitching and Programming the Next Killer App“. (2012). *Wired*, 17. Februar 2012. https://www.wired.com/2012/02/ff_hackathons/all/1/.

digitalspirit. „Hackathons und Crowdsourcing sichern Innovation“. (2016). Zugegriffen 20. April 2018. <https://digitalspirit.dbsystel.de/hackathons-und-crowdsourcing-sichern-innovation/>.

Johnson, Peter, und Pamela Robinson. „Civic Hackathons: Innovation, Procurement, or Civic Engagement?: Civic Hackathon: Procurement or Civic Engagement?“. (2014). *Review of Policy Research* 31, Nr. 4.

„Deutscher Mobilitätspreis | 3rd BMVI Data-Run – Intelligente emissionsarme Mobilitätslösungen in Städten“. (2018). Zugegriffen 23. Juli 2018. <https://deutscher-mobilitaetspreis.de/journal/bmvi-data-run>.

dbmindbox „DB Open Data Hackathons“. (2018). Zugegriffen 23. Juli 2018. <https://dbmindbox.com/de/db-opendata-hackathons/>.

Dudenredaktion (Bibliographisches Institut), Hrsg. *Duden, das Fremdwörterbuch*. 11., vollständig überarbeitete und aktualisierte Auflage. (2015). Der Duden in zwölf Bänden, Band 5. Berlin: Dudenverlag, 2015.

Fisher, Brian. (2005). „Illuminating the Path: An R&D Agenda for Visual Analytics“.

Andrienko, G., N. Andrienko, I. Kopanakis, A. Ligtenberg, und S. Wrobel. „Visual Analytics Methods for Movement Data“. In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*, herausgegeben von Fosca Giannotti und Dino Pedreschi. (2008). Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

Keim, D.A., C. Panse, M. Sips, und S.C. North. (2004). „Visual Data Mining in Large Geospatial Point Sets“. *IEEE Computer Graphics and Applications* 24, Nr. 5, September 2004.

Keim, Daniel A., Florian Mansmann, Jörn Schneidewind, Jim Thomas, und Hartmut Ziegler. „Visual Analytics: Scope and Challenges“. (2008). In *Visual Data Mining*, 76–90. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2008.

Hutchison, David, Michael H Böhlen, Takeo Kanade, Josef Kittler, Jon M Kleinberg, Friedemann Mattern, Arturas Mazeika, u. a. *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. (2008). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2008.

WhatIs „What Is Geospatial Analysis? - Definition from WhatIs.Com“. (2014). Zugegriffen 24. Juli 2018.

SearchEnterpriseSoftware.de „BI-Anbietervergleich: Tableau vs. Power BI vs. Qlik Sense“. (2017). Zugegriffen 5. August 2018. <https://www.searchenterprisesoftware.de/lernprogramm/BI-Anbietervergleich-Tableau-vs-Power-BI-vs-Qlik-Sense>.

Marr, Bernard. “Comparing Data Visualization Software: Here Are The 7 Best Tools For 2018“. (2018). Zugegriffen 24. Juli 2018. <https://www.forbes.com/sites/bernardmarr/2018/06/20/comparing-data-visualization-software-here-are-the-7-best-tools-for-2018/#1de36a076d0b>.

Turck, Matt „Matt-Turck-FirstMark-2017-Big-Data-Landscape.png (4217×2875)“. (2017). Zugegriffen 26. Juli 2018. <http://mattturck.com/wp-content/uploads/2017/05/Matt-Turck-FirstMark-2017-Big-Data-Landscape.png>.

DB Open Data Portal „Call A Bike - Open-Data-Portal – Deutsche Bahn Datenportal“. (2017). Zugegriffen 28. Juli 2018. <http://data.deutschebahn.com/dataset/data-call-a-bike>.

API-Portal „API-Portal - Flinkster_API_NG“. (2018). Zugegriffen 28. Juli 2018. https://developer.deutschebahn.com/store/apis/info?name=Flinkster_API_NG&version=v1&provider=DBOpenData#.

Abbildungsverzeichnis

Abbildung 1: Big Data Lifecycle (Vgl. Jagadish, 2014: 86-94).....	14
Abbildung 2: Visual Analytics Mantra (Keim, Mansmann, 2008).....	19
Abbildung 3: Scope von Visual Analytics (Hutchison, 2008: 79).....	20
Abbildung 4: Big Data Landscape (Turck, 2017).....	22
Abbildung 5: Big Data Landscape (Turck, 2017) - Ausschnitt mit Markierung	23
Abbildung 6: Open Data als Treiber (Dapp, M., 2016: 11).....	27
Abbildung 7: Geoportal - Verkehrsdaten (Geo-Online, 2018)	31
Abbildung 8: Verkehrsnetzplan des HVV Ausschnitt Hamburg Dammtor (Geofox.de, 2018)32	
Abbildung 9: Szenarienablauf	35
Abbildung 10: Ermittelte Datenquellen	36
Abbildung 11: Datenquellen nach Analyse	37
Abbildung 12: Architekturentwurf.....	39
Abbildung 13: Schichten der Architektur im Detail	40
Abbildung 14: Screenshot DB Open Data Portal – Call a Bike.....	42
Abbildung 15: Screenshot DB Open Data Portal.....	43
Abbildung 16: Stationsdaten - unbearbeitet	44
Abbildung 17: Stationsdaten – aufbereitet.....	44
Abbildung 18: Datenextrakt in Tableau	45
Abbildung 19: Verknüpfung der Datensätze.....	46
Abbildung 20: Zuordnung von Longitude und Latitude	47
Abbildung 21: Markierungsbereich.....	47
Abbildung 22: Visualisierung Szenario 1 – Frequentierte Orte.....	48
Abbildung 23: Filter und Markierungen - Länge der Leihzeiten	49
Abbildung 24: Visualisierung Szenario 1 - Leihminuten.....	50
Abbildung 25: Frequentierte Abhol- und Rückgabestationen	51
Abbildung 26: Switchh Stationen Hamburg (Switchh 2018).....	52
Abbildung 27: Datenextrakt für Szenario 3.....	55
Abbildung 28: Zusammengeführte Spalten	55
Abbildung 29: Intergration der 3 Datenquellen.....	56
Abbildung 30: Modellierung Szenario 3.....	56
Abbildung 31: Visualisierung Szenario 3 - Switchh Stationen.....	57
Abbildung 32: Flinkster Schnittstelle (API-Portal 2018).....	58
Abbildung 33: Python Script zum Abruf der REST Schnittstelle.....	59
Abbildung 34: Lortzingstraße – Krausestraße (Rechts: Google Maps 2018)	63
Abbildung 35: Bahrenfelder Straße – Bahnhof Altona West (Rechts: Google Maps 2018)...	64
Abbildung 36: StadtRad Station Horner Rennbahn	65

Tabellenverzeichnis

Tabelle 1: Modal Split in Hamburg (Vgl. Infas 2008).....	9
Tabelle 2: Liste von möglichen Datenquellen (Fasel, Meier, 2016: 240).....	16
Tabelle 3: Comparing Data Visualization Software (Vgl. Marr, 2018)	24
Tabelle 4: Gegenüberstellung Open Data	28
Tabelle 5: Gegenüberstellung der recherchierten Quellen	33
Tabelle 6: JOIN Typen.....	46
Tabelle 7: Switchh Stationsdaten.....	53
Tabelle 8: Switchh ID und Zonen.....	54
Tabelle 9: Frequentierte Stationen Top 5	61
Tabelle 10: Stationen mit den längsten Ausleihezeiten nach Himmelsrichtung	62

Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, den _____