



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Ausarbeitung

Moritz Philip Recke

Aufbauprojekt Bericht:
Knowledge Discovery in Open Government
Data am Beispiel Luftqualität

Moritz Philip Recke

Aufbauprojekt Bericht:
Knowledge Discovery in Open Government
Data am Beispiel Luftqualität in Hamburg

Ausarbeitung eingereicht im Rahmen des
Grundprojekts im Wintersemester 2015/2016

im Studiengang Master of Arts - Next Media
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Professor Dr. Kai von Luck

Abgegeben am 25.01.2016

Autor

Moritz Philip Recke

Thema der Ausarbeitung

Projektbericht zum Aufbauprojekt im 2. Semester des Studiengangs Master of Arts - Next Media mit dem Thema "Knowledge Discovery in Open Government Data am Beispiel Luftqualität in Hamburg".

Stichworte

Air Quality Index, Analyse, Arduino, AQI, City, Data, Data Driven Storytelling, Feinstaub, GovData, Hamburg, Human Sensing, Illustration, Lilypad, Luft, Luftqualität, Luftqualitätsindex LQI, KDD, Knowledge Discovery in Databases, Narration, Open Data, Open Government Data, Prototyp, Prototyping, Public Science, Sensoren, Shanghai, Smart City, Stadt, Stadtnarration, Storytelling, Transparenzportal, Umweltdaten, Urban Data, Visualisierung, Wetterdaten

Abstract

Im Rahmen des Aufbauprojektes wurden unterschiedliche Themengebiete entlang anwendungsbezogener Methoden bearbeitet, Konzepte und Lösungsansätze entwickelt und präsentiert. Gemeinsam mit Jessica Broscheit und Hannes Sieg wurde das Thema "Urban Storytelling" aus dem Grundprojekt weiterverfolgt und der erarbeitete Case Air aufgegriffen. Im Verlauf des Semesters hat sich die Gruppe auf Luftqualitätsdaten für Hamburg und einen internationalen Vergleich mit Shanghai fokussiert. Wesentliche Arbeitsschritte waren Knowledge Discovery in Databases in Open Government Data, die Aufbereitung der Daten für den internationalen Vergleich sowie die Entwicklung einer Narration im urbanen Kontext. Hierzu wurden Illustrationen und Animationen entworfen sowie ein Prototyp für eine Atemmaske zur Visualisierung der Daten konzipiert und mittels Lilypad Arduino konstruiert. Die Arbeitsergebnisse wurden der Studiengruppe mit einer multimedialen Präsentation präsentiert. Der Schwerpunkt meiner Aufgaben im Projekt lag auf Open Government Data, Knowledge Discovery in Databases sowie die Konzeption von Ansätzen für die Datennarration und die Projektpräsentation. Die Ausarbeitung beschreibt primär zugrundeliegende Prinzipien von Open Government Data und Knowledge Discovery in Databases sowie die Urban Storytelling Konzepte zur Inszenierung der erarbeiteten Erkenntnisse aus den Daten, nicht die konkreten Arbeitsergebnisse.

Inhaltsverzeichnis

1 Einleitung & Themenvorstellung	5
2 Open Government Data	6
2.1 Datenquellen für Open Government Data in Deutschland	7
2.2 Standards und Datenformate	8
3 Knowledge Discovery in Databases - KDD	9
4 Beispiel Luftqualität	11
4.1 Luftqualitätsindex (LQI)	11
4.2 Air Quality Index (AQI)	11
4.3 Vergleichbarkeit	12
5 Urban Storytelling Konzepte zur Luftqualität.....	13
5 Fazit & Ausblick	16
6 Literaturverzeichnis	17
7 Abbildungsverzeichnis	20
8 Anhang	21

1 Einleitung & Themenvorstellung

Für das Aufbauprojekt hat sich die Projektgruppe (gemeinsam mit Jessica Broscheit und Hannes Sieg) dazu entschieden, Ansätze des letzten Semesters im Grundprojekt weiterzuverfolgen. So wurde der Case Air unter dem Ansatz des "Urban Storytelling" weiterentwickelt. Dies ergab sich vor allem im Zusammenhang der parallel zum Projekt stattfindenden Klimakonferenz in Paris (United Nations Framework Convention on Climate Change, 21st Conference of the Parties - COP 21) im Dezember 2015.

Der Schwerpunkt dieser Ausarbeitung liegt auf Open Government Data, Knowledge Discovery in Databases sowie der Konzeption von Ansätzen für die Datennarration. In der Ausarbeitung von Jessica Broscheit (2016) wird der Design Thinking Prozess im Rahmen des Projektes und die konzeptionelle Ausgestaltung des entwickelten Prototypen genauer vorgestellt. Hannes Sieg (2016) erläutert Details zur Datenbeschaffung sowie -aufbereitung und bewertet die Open Government Data Ansätze am Beispiel des Transparenzportals Hamburg.

Anhand von Analysen öffentlich zugänglicher Daten zur Luftqualität in Deutschland konnte der Standort Hamburg international eingeordnet und u.a. mit der Partnerstadt Shanghai verglichen werden. Aufbauend auf diesen Daten wurde mit verschiedenen Möglichkeiten der Visualisierung und emotionaler Anreicherung experimentiert. Hierbei entstanden klassische Illustrationen und Animationen. Zur besseren Erfahrbarkeit der Daten in der realen Welt wurde darüber hinaus mit Lilypad Arduino ein Prototyp für eine Atemmaske entwickelt, welcher die Daten mittels Farb-LEDs visualisiert.

So wurde das abstrakte Konstrukt der Luftqualität in den urbanen Alltag überführt, um Passanten für das Thema sensibilisieren zu können. Durch die Verwendung von z.B. Feinstaub Sensoren in der Maske wäre auch eine dezentrale Datenlieferung an Global City Netzwerke durch Benutzer der Maske im Sinne von Public Science Ansätzen denkbar, was neue Möglichkeiten der Partizipation eröffnet.

Die Ausarbeitung beschreibt vordergründig den Prozess der Knowledge Discovery in den öffentlich zugänglichen Daten zur Luftqualität in Deutschland (Open Government Data) und dabei aufgetretene Probleme. Die Illustrationen und Prototypenentwicklung werden in Ansätzen vorgestellt sowie zugrundeliegende Überlegungen erläutert.

2 Open Government Data

Um Open Government Data zu betrachten, ist es zunächst sinnvoll, sich dem Begriff Open Data zu nähern. Offene Daten sind dadurch charakterisiert, dass sie durch Jeden ohne technische und rechtliche Einschränkungen genutzt, weiterverarbeitet und verbreitet werden können. Hierzu müssen entsprechende Datensätze vollständig, maschinenlesbar und einfach modifizierbar zur Verfügung gestellt werden (vgl. Open Knowledge Foundation Deutschland 2016, okfn.de).

Inhaltlich gibt es hierbei also keine Grenzen. Es kann sich um jede Art von Daten handeln, von Geodaten über Statistiken, Transport-, Verkehrs-, und Finanzdaten bis hin zu Lehrmaterialien und wissenschaftlichen Forschungsergebnissen. Es können hierbei nicht nur Datenbestände öffentlicher Einrichtungen, sondern auch solche privatwirtschaftlicher Unternehmen, Hochschulen, Medienunternehmen oder Non-Profit-Einrichtungen gemeint sein (vgl. von Lucke 2011, S. 5).

Open Government Data bezeichnet in diesem Zusammenhang Informationen des Staates und seiner Verwaltung, die entsprechend der Open Data Ansätze zur Verfügung gestellt werden (vgl. Bundesministerium des Innern 2012, S. 25 ff.). Die rechtliche Grundlage für die Bereitstellung dieser öffentlichen Daten ist in Deutschland das Informationsfreiheitsgesetz (IFG - Gesetz zur Regelung des Zugangs zu Informationen des Bundes) und ermöglicht es den Bürgern, Einsicht in Informationen der öffentlichen Verwaltung zu bekommen.

In vielen Ländern gibt es ähnliche Rechtsgrundlagen, so dass der Zugang zu Informationen dieser Art bereits in vielen Staaten gewährleistet ist (vgl. Right2INFO.org 2016, right2info.org). Eine proaktive und kostenlose Bereitstellung solcher Informationen ist hierbei nicht zwingend erforderlich und wird in vielen Fällen ergänzend geregelt, so wie in Hamburg durch das Hamburgische Transparenzgesetz (vgl. Hamburger Justizbehörde 2015, hamburg.de).

Ohne näher auf Potentiale für mehr Transparenz und volkswirtschaftlichen Nutzen einzugehen, ist klar, dass es eine große Menge an Daten gibt, die für private, wirtschaftliche, wissenschaftliche oder öffentliche Zwecke genutzt werden können. Die jeweiligen Zustände von Open Government Data Projekten sind national und international sehr unterschiedlich. So sind im Ausland entsprechende Projekte oft deutlich weiter fortgeschritten und werden in einem höheren Tempo weiter entwickelt (vgl. Barnickel und Kleemann 2012, S. 136 ff.).

2.1 Datenquellen für Open Government Data in Deutschland

In Deutschland ist die Landschaft der Datenquellen für Open Government Data sehr vielfältig. Aufgrund der föderalen Struktur der Bundesrepublik Deutschland sind die zugrundeliegenden technischen Infrastrukturen sowie die rechtlichen Grundlagen sehr heterogen. Die Gliederung in Bund, Länder und Kommunen führt aufgrund der zugrundeliegenden Ressorthoheit und damit verbundener Abgrenzung u.a. zu getrennten und nicht einheitlichen Datenbanken, IT-Systemen, Prozessen und Rechtsgrundlagen (vgl. Dietrich 2011, pbp.de).

Jede Verwaltungsebene in nahezu allen größeren Ballungszentren bietet Zugang zu Informationen im weitesten Sinne von Open Government Data. So gibt es beispielsweise für Hamburg Luftdaten auf Länderebene unter luft.hamburg.de sowie auf Bundesebene unter umweltbundesamt.de. Hierbei werden die Daten aber nicht zwingend auf die gleiche Weise oder in gleicher Detailtiefe zur Verfügung gestellt. Auch gibt es keinen Verweis zu regionalen oder nationalen Datenquellen.

Das Luftmessnetz Hamburg bietet beispielsweise eine deutlich größere Menge an Messkomponenten sowie Möglichkeiten zur detaillierten Einsichtnahme in Messperioden und Zeiträume. Eine Suche ist allerdings nur über die manuelle Auswahl der Messstationen möglich. Das Umweltbundesamt hingegen bietet weniger Messkomponenten und eingeschränkte Möglichkeiten der Einsichtnahme, allerdings lässt sich Deutschlandweit nach Messwerten zu Schadstoffen suchen (vgl. Abb. 1). Beide Portale bieten csv-Dateien als download an, die Formatierung der Daten ist jedoch auch völlig unterschiedlich.

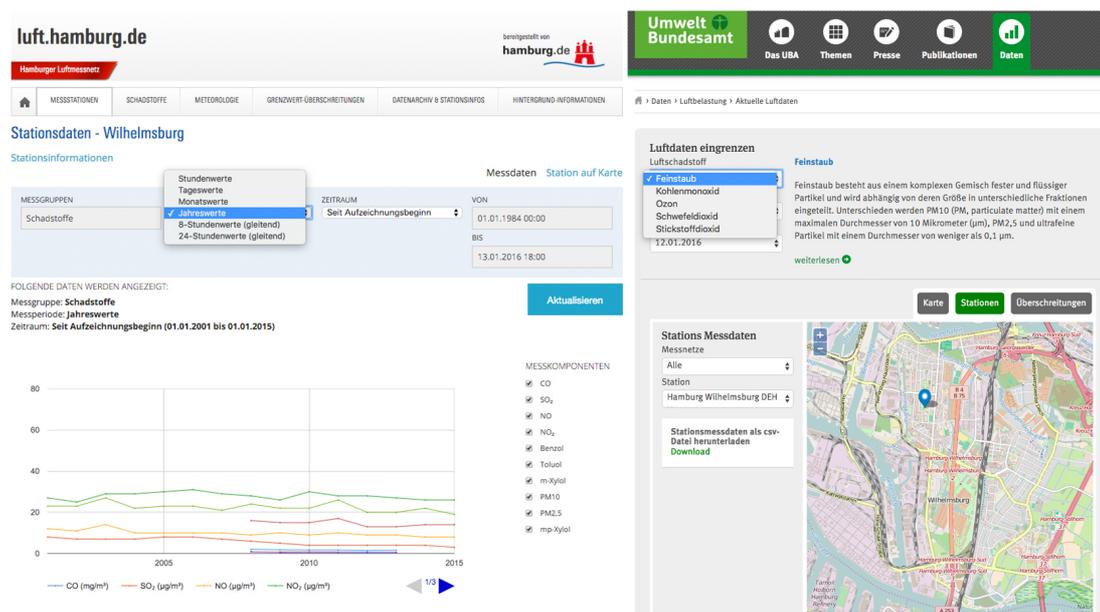


Abb. 1: Luftdaten Luftmessnetz Hamburg und Umweltbundesamt.

Dieses Beispiel ist kein Einzelfall und zeigt wesentliche Probleme bei der Nutzung von Open Government Data auf. So gibt es zwar große Mengen an Daten, die jedoch über diverse Datenquellen und Datenformate verteilt sind und teilweise gar nicht oder nur eingeschränkt weiterverarbeitet werden können.

Der derzeitige Zustand erschwert den Zugriff auf Verwaltungsdaten und deren Verwendung also enorm und erfordert nicht nur eine Vereinheitlichung der Datenbestände sondern vor allem auch eine Vereinheitlichung von technischen Standards, Datenformaten sowie entsprechende Dokumentation für einen unkomplizierten Austausch und Nutzung von Open Government Data.

Hierzu gibt es bereits ein vom Bundesministerium des Innern initiiertes Projekt für ein GovData Portal, welches derzeit als Prototyp betrieben wird und Daten aller Verwaltungsebenen bündeln und zukünftig zentral zugänglich machen soll (vgl. Bundesministerium des Innern 2015, bmi.bund.de).

2.2 Standards und Datenformate

In Hinsicht auf eine Vereinheitlichung der Standards, Datenformate und Lizenzen ist in Deutschland noch viel zu tun. Die Daten werden derzeit oft in nicht standardisierten Formaten wie Tabellen, Fließtexten oder PDF-Dateien bereitgestellt. Diese Formate sind weitestgehend ungeeignet für maschinelle Verarbeitung und Einbindung in web-basierte Anwendungen. Es werden unterschiedliche Vokabulare und Klassifikationen verwendet, darüber hinaus ist die rechtliche Grundlage für Nutzungslizenzen uneinheitlich. Kämen international anerkannte Standard-Lizenzen zum Einsatz, so könnte in dieser Hinsicht Kompatibilität hergestellt werden (vgl. Dietrich 2011, pbp.de).

In technischer Hinsicht wäre der Einsatz einer Standard-Softwarelösung ebenfalls sinnvoll. Hier ist das Open Source Projekt Comprehensive Knowledge Archive Network (CKAN) die weltweit verbreitetste Lösung (vgl. Open Knowledge Foundation 2016, ckan.org). Alternativ ist als kommerzielles Produkt Socrata Open Data Portal zu nennen, welches vor allem in den USA weite Verbreitung findet und Städte wie New York City aber auch ganze Bundesstaaten wie Washington oder Illinois als Kunden vorweisen kann (vgl. Socrata 2015, socrata.com).

Hierbei handelt es sich um Systeme, welche Datensammlungen verwalten und dabei ähnlich wie klassische Content Management System funktionieren. Die Daten können von den Nutzern direkt durchsucht, betrachtet, analysiert und über standardisierte technische Schnittstellen auch maschinell abgerufen werden. So

ermöglicht eine RESTful API alle Funktionen des Kernsystems anzusprechen und Daten als JSON-Objekte weiterzuverarbeiten (vgl. Ikonomou 2015, S. 28 ff.).

Auch in Deutschland kommt CKAN bereits zum Einsatz. So setzt beispielsweise das Transparenzportal Hamburg auf das Open Source System und bietet die Standard-API an (vgl. Transparenzportal Hamburg 2016, transparenz.hamburg.de). Auch Berlin und das GovData Datenportal für Deutschland setzen auf CKAN. GovData versucht dabei auch eine einheitliche Metadatenstruktur für die Daten aller Verwaltungsebenen zu etablieren (vgl. GovData 2016, govdata.de).

Somit ist ein wesentlicher Grundstein für eine technische Vereinheitlichung gelegt. Würden nun die Datenlieferanten in Bund, Ländern und Kommunen statt Fließtexten und PDF-Dateien dokumentierte, maschinenlesbare Datensätze liefern, könnte das Potential von Open Government Data umfassender genutzt werden. Dies bleibt auf absehbare Zeit sicher die größte Herausforderung, auf dessen organisatorische, kulturelle und politische Hintergründe an dieser Stelle jedoch nicht weiter eingegangen werden soll.

3 Knowledge Discovery in Databases - KDD

Stehen große Mengen an Daten zur Verfügung, so bietet es sich an, in diesen nach Wissen/Erkenntnissen, also Mustern von Auffälligkeiten, zu suchen. Dies ist ein komplexer und anspruchsvoller Vorgang, der durch den Knowledge Discovery in Databases (KDD) Prozess beschrieben wird. Geprägt wurde der Begriff bereits 1889 im Rahmen der IJCAI-89 Konferenz und stellt diesen als ein interdisziplinäres Forschungsfeld zur Wissensgenerierung vor (vgl. Piatetsky-Shapiro 1990).

Hierbei kommen Felder wie Machine Learning, Statistik, Mustererkennung, Artificial Intelligence, Visualisierung und High Performance Computing zum Einsatz. Im Knowledge Discovery in Databases (KDD) Prozess ist es das Ziel, gültiges, bisher unbekanntes und potentiell nützliches Wissen aus Datenbanken zu extrahieren (vgl. Fayyad, Piatetsky-Shapiro, Smyth 1996 I, S. 1 ff.). Oft wird der Begriff Data-Mining verwendet, welcher allerdings nur einen Teilschritt des Knowledge Discovery in Databases (KDD) Prozesses darstellt. Vielmehr durchläuft der Prozess typischer Weise zunächst eine ganze Reihe von iterativen Schritten mit vorbereitenden Untersuchungen und Transformationen der zu untersuchenden Daten (vgl. Abb. 2). Dies lässt die Komplexität der Vorgangs erkennen.

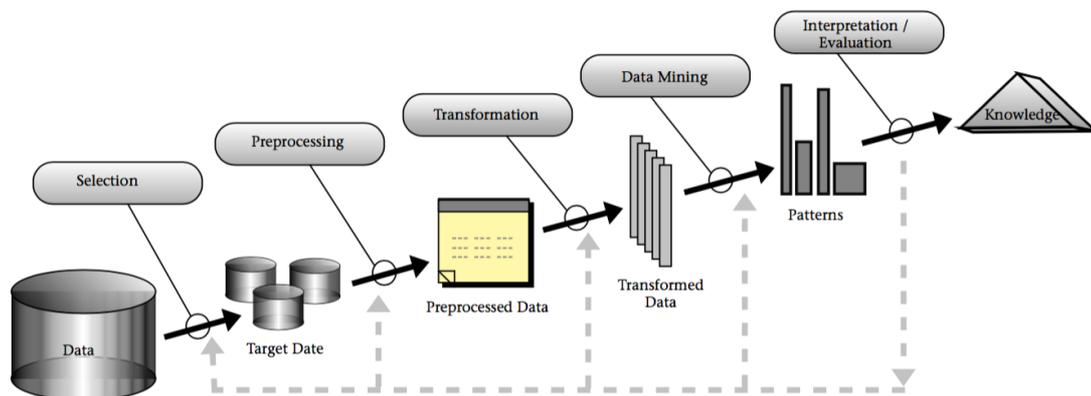


Abb. 2: An Overview of the Steps that Compose the KDD Prozess.

Die zu durchsuchenden Daten müssen zunächst im Zusammenhang mit der Fragestellung sinnvoll ausgewählt werden (Selection). Hieraus kann dann ein Subset von Daten oder Variablen ausgewählt werden, in welchem gesucht wird (Target Date). Durch eine Vorverarbeitung (Preprocessing) werden u.a. überflüssige Daten entfernt, fehlende Daten ergänzt bzw. der Umgang mit Fehlern definiert. Die Daten können dann beispielsweise durch eine Einschränkung der zu betrachtenden Variablen reduziert und durch die Auswahl einer sinnvollen Repräsentation projiziert werden (Preprocessed Data). Diese Daten werden im Anschluss für spezifische Data Mining Verfahren transformiert (Transformation), um dann mittels explorativer Analyse geeignete Modelle und Hypothesen auszuwählen (Transformed Data) und schließlich mit Data Mining Verfahren, wie z.B. Klassifizierung, Regression, Clustering etc., nach Mustern (Patterns) in den Daten zu suchen. Das Erkennen von Mustern kennzeichnet die Daten mit einer Struktur, welche dann bewertet und interpretiert (Interpretation/Evaluation) werden kann. Kommt man zu keinem sinnvollen Ergebnis muss der Prozess oder einzelne Prozessschritte erneut mit entsprechenden Anpassungen durchlaufen werden (vgl. Fayyad, Piatetsky-Shaprio, Smyth 1996 II, S. 41 ff.).

Bei der Bewertung der gefundenen Muster kann eine Visualisierung, z.B. mit Entscheidungsbäumen oder Verortungen, sinnvoll sein (vgl. Ester und Sander 2000, S. 6). Erst durch das erfolgreiche Durchlaufen dieses Prozesses können aus großen Datensätzen Erkenntnisse, also Wissen ermittelt werden. Dieses Wissen kann dann weiterverarbeitet oder zugänglich gemacht bzw. vermittelt werden, wobei z.B. statische und interaktive Visualisierungen der Daten und Muster für den Betrachter hilfreich sein können.

4 Beispiel Luftqualität

Im Rahmen des Aufbauprojektes hat sich die Projektgruppe mit Open Government Data zur Umwelt, genauer mit Luftqualitätsdaten beschäftigt und den KDD Prozess durchlaufen. Hierbei wurden vordergründig die Luftdaten der Stadt Hamburg sowie der Partnerstadt Shanghai in China für den Zeitraum eines Monats betrachtet. Hierbei ergab sich zunächst die Herausforderung, entsprechende Daten überhaupt zu finden, bzw. diese international miteinander vergleichen zu können.

So verwendet Hamburg weder den in der EU verwendeten Luftqualitätsindex CAQI (Common Air Quality Index), noch den international verbreiteten, von den USA entwickelten Air Quality Index AQI, der auch in Shanghai zum Einsatz kommt. Vielmehr wird ein auf Länderebene in Deutschland verbreiteter aber nicht von allen Bundesländern eingesetzter Indikator für Luftqualität verwendet.

4.1 Luftqualitätsindex (LQI)

In Hamburg kommt der Luftqualitätsindex LQI zum Einsatz, welcher als stundenaktueller Indikator für die allgemeine Luftqualität in der Stadt dient. Hierbei werden Schadstoffe in 6 Indexklassen mit den Stufen sehr gut (1), gut (2), befriedigend (3), ausreichend (4), schlecht (5) und sehr schlecht (6) eingeteilt. Der LQI betrachtet dabei die Schadstoffe Stickstoffdioxid (NO₂, 1-h-Mittelwert), Schwefeldioxid (SO₂, 1-h-Mittelwert), Ozon (O₃, 1-h-Mittelwert), Kohlenmonoxid (CO, 8-h-Mittelwert gleitend) und Feinstaub (PM₁₀, 24-h-Mittelwert gleitend) in µg/m³ und ordnet die gemessenen Schadstoffe einer Indexklasse zu.

Der jeweils schlechteste Index eines der gemessenen Schadstoffe wird als Indikator der Luftqualität für die jeweilige Messstation angegeben (vgl. Hamburger Luftmessnetz 2016, luft.hamburg.de).

4.2 Air Quality Index (AQI)

Der Air Quality Index AQI ist ein tagesaktueller Indikator für Luftqualität und unterscheidet die Indexklassen Good (0-50), Moderate (51-100), Unhealthy for Sensitive Groups (101-150), Unhealthy (151-200), Very Unhealthy (201-300) und Hazardous (301-500). Der AQI betrachtet ebenfalls die Schadstoffe Stickstoffdioxid (NO₂, 1-h-Mittelwert), Schwefeldioxid (SO₂, 1-h- bzw. 24-h-Mittelwert), Ozon (O₃, 1-h- bzw. 8-h-Mittelwert), Kohlenmonoxid (CO, 8-h-Mittelwert), allerdings in ppm (parts per million). Der Feinstaub (PM_{2,5}, 24-h-Mittelwert und PM₁₀, 24-h-Mittelwert) wird in µg/m³ gemessen.

Die gemessenen Schadstoffe werden ebenfalls einer Indexklasse zugeordnet und auch hier bestimmt der schlechteste gemessene Schadstoffwert den Index der Luftqualität für die jeweilige Messstation (vgl. AirNow 2015, airnow.gov).

4.3 Vergleichbarkeit

Stellt man die Indizes LQI und AQI gegenüber, sind schon auf den ersten Blick einige Unterschiede festzustellen (vgl. Abb. 3). So zielt der AQI schon in seinen Bezeichnungen viel deutlicher auf das Gesundheitsrisiko ab und erlaubt eine deutlich feinere Darstellung. Beim LQI sind starke, relative Veränderungen kaum erkennbar und nur an harten Bruchkanten festzustellen.

LQI		AQI	
1	sehr gut	0 - 50	good
2	gut	51 - 100	moderate
3	befriedigend	101 - 150	unhealthy for sensitive groups
4	ausreichend	151 - 200	unhealthy
5	schlecht	201 - 300	very unhealthy
6	sehr schlecht	301 - 500	hazardous

Abb. 3: LQI und AQI Indizes

Um also eine Vergleichbarkeit der Standorte Hamburg und Shanghai herzustellen, sollte ein einheitlicher Index verwendet werden. Aufgrund der höheren Verbreitung erscheint der AQI hier sinnvoll. Um aus dem LQI den AQI zu bestimmen, müssen aufgrund der unterschiedlichen Berechnungsmethoden die einzelnen Messwerte der Schadstoffe zur Bestimmung des LQI zur Verfügung stehen.

Da für Hamburg alle Schadstoffwerte in $\mu\text{g}/\text{m}^3$ angegeben werden, ist hierzu eine Überführung in ppm (parts per million) mittels der Mol-Massen der Schadstoffe notwendig. Durch eine einmalige Konfiguration der Transformation des LQI in den AQI sind die Indizes im Format des AQI vergleichbar. Eine Transformation des AQI in den LQI wäre ebenfalls möglich, dies setzt aber einen Zugriff auf einzelne Schadstoffwerte in Open Government Data Quellen internationaler Standorte voraus. Dies ist im Rahmen des Projektes nicht weiter betrachtet worden.

In den deutschen Open Government Data Portalen luft.hamburg.de sowie umweltbundesamt.de liegen die notwendigen Einzelwerte zur Bestimmung des Hamburger LQI und somit zur Berechnung des AQI vor. Allerdings treten bei der

Zusammenstellung der Daten weitere Probleme auf. So liefern beispielsweise nicht alle Messstationen Daten zu allen notwendigen Schadstoffen und die Datensätze des Luftmessnetzes Hamburg enthalten Lücken und Fehler (vgl. Anhang 1). Die Daten des Umweltbundesamtes sind nach Schadstoffen und nicht Messstationen sortiert und enthalten die Daten in einer anderen Formatierung (vgl. Anhang 2).

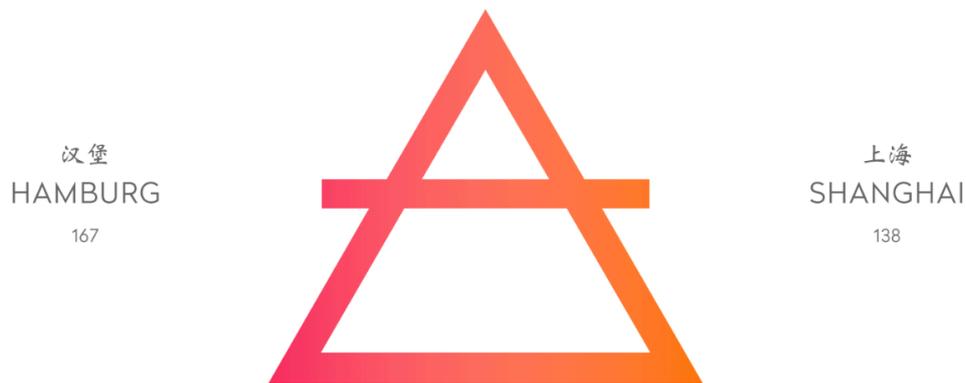
Für die Berechnung im Projekt wurden die Daten des Umweltbundesamtes verwendet. Die zur Verfügung gestellten csv-Dateien haben sich als besser verwertbar und vollständiger herausgestellt als die des Luftmessnetzes Hamburg. Darüber hinaus wurden Fehler mit entsprechenden Maßnahmen bereinigt und unvollständige Datensätze durch Daten der nächstgelegenen Messstationen ergänzt.

Durch die vorgenommenen Berechnungen konnten die Luftqualitätsdaten der Stadt Hamburg mit denen der Partnerstadt Shanghai auf Basis des AQI verglichen werden, welche beispielsweise auf dem Portal aqicn.org einzusehen sind.

5 Urban Storytelling Konzepte zur Luftqualität

In der medialen Berichterstattung wird oft auf die schlechte Luftqualität und hohe Schadstoffbelastung in chinesischen Städte wie Beijing oder Shanghai hingewiesen. Insbesondere im Vorfeld der Klimakonferenz in Paris im Dezember 2015 stand Beijing im internationalen Fokus (vgl. Fritz 2015, washingtonpost.com). Allerdings gibt es auch Berichte, nach denen die Luftqualität in Deutschland deutlich schlechter ist, als vermutet. So steigen beispielsweise auch in Hamburg die Schadstoffwerte weiter an (vgl. Meyer-Wallmann 2015, abendblatt.de).

Um sich diesen urbanen Mythen etwas systematischer zu nähern, wurde ein datengestützter Vergleich der Stadt Hamburg mit der chinesischen Partnerstadt Shanghai angestrebt. Als Datenquelle für Shanghai wurde aqicn.org ausgewählt, wo der Luftqualitätsindex AQI zum Einsatz kommt. Durch die Bestimmung des AQI für Hamburg anhand der gemessenen Schadstoffe repräsentativer Messstationen konnte der Vergleich hergestellt werden. In ausgewählten Testzeiträumen ist hierdurch erkennbar, dass die Luftqualität in Hamburg keinesfalls besser ist als in Shanghai. Insgesamt liegen die Städte im Durchschnitt des Betrachtungszeitraums in etwa gleich auf (vgl. Anhang 3). Es lässt sich also festhalten, dass es in Hamburg schlechter um die Luftqualität steht, als man annehmen würde.



AQI – OCT. 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Abb. 4: AQI Illustration: Hamburg vs. Shanghai

Für eine Inszenierung dieser Erkenntnis wurden eine animierte Deutschlandkarte und animierte Illustrationen zur Gegenüberstellung von Hamburg und Shanghai erstellt. Letztere setzen bewusst auf eine alternative Darstellung zur klassischen Verortung von Luftqualitätsdaten auf einer Landkarte und verwendet das Elementsymbol für Luft der Vier-Element-Lehre (vgl. Abb. 4). Dennoch stellen klassische Visualisierungen von Daten keinen emotionalen Bezug her. So wurde mit Möglichkeiten experimentiert, die Daten in eine reale, anfassbare und erlebbare Narration zu überführen, um das Thema emotional anzureichern.

Es wurde ein Konzept für eine Atemmaske entwickelt, welche die AQI Daten aufnehmen und mittels einer LED in Signalfarben visuell darstellen kann. Denkbar wäre hierbei, die Maske mit Konnektivität z.B. zu einer Smartphone App auszustatten, um so in Echtzeit auf lokale Daten zugreifen zu können. In der App und auf dem zugehörigen Webportal könnten zusätzlich Karten mit globalen Luftwerten einsehbar sein. Abstrakte Strukturen zur Einordnung von erhobenen Daten werden so in den realen urbanen Raum überführt und ändern sich je nach Standort des Nutzers in Echtzeit. So handelt der Benutzer einer Atemmaske nicht nur bewusst im besten Sinne seiner Gesundheit, auch werden Passanten durch die Signalfarbe ebenfalls auf die aktuelle Luftqualität aufmerksam gemacht, in die Narration einbezogen und ggf. für das Thema sensibilisiert oder sogar aktiviert.

Mit Lilypad Arduino wurde in der Projektgruppe ein erster Prototyp für eine Atemmaske konstruiert, welche mit den Daten des betrachteten Testzeitraumes ausgestattet wurde und entsprechende Lichtsignale abgeben kann (vgl. Abb. 5).



Abb. 5: Prototyp Atemmaske

Für die farbliche Darstellung wurde der AQI auf ein 3-Farb-System reduziert, der zwischen sehr guter und guter in blau, befriedigender in orange und schlechter Luftqualität in rot unterscheidet.

Eine weitere Iterationsstufe der Maske könnte die Installation von eigenen Schadstoffsensoren sein. So könnte eine mit Sensoren ausgestattete Datenmaske z.B. Kohlenmonoxid oder Natriumdioxid selber messen und die Werte in Echtzeit an ein bestehendes oder zu konzipierendes globales Netzwerk für Luftqualitätsdaten übertragen. Die Größe solcher Sensoren lässt erkennen, dass diese in absehbarer Zeit ggf. sogar direkt in mobile Geräte eingebaut oder an jeder anderen denkbaren Stelle installiert werden könnten (vgl. Pewatron 2016, pewatron.com).

Dies würde ein feinmaschiges Netz von Messstationen mit deutlich höherer Datendichte ermöglichen, so dass erstmalig überhaupt Betrachtungen der Luftqualität in der Fläche möglich wären. Auch ließe sich die Dauer aus Schadstoff-Aussetzung einzelner Benutzer betrachten. Durch eine solche Dezentralisierung der Datenbeschaffung, also Crowdsourcing, könnten im Sinne von City Governance neue Möglichkeiten der Partizipation entstehen (vgl. Motto, You, Sacco, Ma 2014, S. 439 ff.). Über die Bereitstellung von offen APIs könnte es Benutzern auch ermöglicht werden, eigene Lösungen rund um dieses Thema zu entwickeln, welche die Community ergänzen können. Dieses Konzept wird z.B. auch durch das Smart Citizen Projekt propagiert, welches neben einem Urban Sensor Kit auch eine Community Platform für Forschung und Entwicklung anbietet (vgl. Smart Citizen 2016, smartcitizen.me).

5 Fazit & Ausblick

Wie aufgezeigt, hatte die Projektgruppe im Verlauf des Semesters mit vielen Hürden im Knowledge Discovery in Databases (KDD) Prozess in Open Government Data in Deutschland zu kämpfen. Zusammengefasst kann das Potential von Open Government Data in Deutschland im Bereich der Luftqualität bisher nicht ausgeschöpft werden. Auch ist ein internationaler Vergleich vor dem Hintergrund diverser internationaler Standards für die Bestimmung der Luftqualität schwierig. Alle bestehenden Datenquellen und Lösungen sind individuelle Einzellösungen, die weder für historische Daten noch in Echtzeit brauchbare Anwendungen liefern. Es ist weiterhin zu vermuten, dass dieser Zustand auch bei anderen Inhalten/Daten der öffentlichen Verwaltung in Deutschland derzeit ähnlich ist.

Vor dem Hintergrund des entwickelten Prototypen stellt sich neben aufgezeigten rechtlichen, organisatorischen und technischen Hürden für Open Government Data Portale die Frage, wie dezentral erhobene Echtzeitdaten zur Luftqualität im Sinne eines Global City Netzwerkes angebunden werden können. So ist u.a. zu klären, wie mit der sicherlich uneinheitlicheren Datenqualität aus crowdsourced Quellen umzugehen ist. In einem nächsten Schritt wäre zu prüfen, inwiefern diese dann intelligent genutzt werden könnten, um daraus Erkenntnisse oder ggf. sogar planerische Maßgaben für die Zukunft abzuleiten.

Interessant können hierbei auch Public Science Ansätze sein, bei denen der Bürger selbst am wissenschaftlichen Prozess beteiligt wird, indem er durch seine Partizipation wissenschaftliche Fragestellung selbst mit entwickelt. So können physische Sensoren zwar bestimmen, was passiert, nicht jedoch warum oder wie Ereignisse oder Muster entstehen und sich entwickeln. Smart City Ansätze könnten dabei durch Human Sensing Methodiken ergänzt werden, indem z.B. Social Media Updates ausgewertet und in crowdsourced Erhebungen einbezogen werden (vgl. Doran, Gokhale, Dagnino 2013, S. 1323 ff.).

Letztlich ist festzustellen, dass solche Ansätze bisher noch in einem sehr frühen Stadium sind und vor allem im wissenschaftlichen Kontext untersucht werden. Eine zeitnahe Möglichkeit zur Nutzung von crowdsourced Umweltdaten in Kombination mit Open Government Data in Deutschland ist nicht absehbar. Doch auch im internationalen Umfeld ist dies bisher nicht erkennbar. Zwar gibt es diverse Web-Projekte für crowdsourced Umweltdaten und öffentliche Datenquellen staatlicher Verwaltungen, eine sinnvolle Verknüpfung dieser Daten oder gar eine Integration von Human Sensing ist bisher jedoch nicht erfolgt.

6 Literaturverzeichnis

AirNow 2015

AirNow: Air Quality Index (AQI) Basics. Online: <http://airnow.gov/index.cfm?action=aqibasics.aqi> Abruf: 15.01.2016.

Barnickel und Kleemann 2012

Barnickel, Niels; Klessmann, Jens: Open Data - Am Beispiel von Informationen des öffentlichen Sektors. Herb, Ulrich (Hrsg.): Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft. universaar. Saarbrücken.

Bundesministerium des Innern 2012

Bundesministerium des Innern: Open Government Data Deutschland. Online: http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment_kurzfassung.pdf Abruf: 15.01.2016.

Bundesministerium des Innern 2015

Bundesministerium des Innern: Verwaltungsdaten für Jedermann. Online: http://www.bmi.bund.de/DE/Nachrichten/Dossiers/OpenData/opendata_node.html Abruf: 15.01.2016.

Dietrich 2011

Dietrich, Daniel: Offene Daten in Deutschland. Online: <http://www.bpb.de/gesellschaft/medien/opendata/64061/offene-daten-in-deutschland> Abruf: 15.01.2016.

Doran, Gokhale, Dagnino 2013

Doran, Derek; Gokhale, Swapna; Dagnino, Aldo: Human Sensing for Smart Cities. 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ASONAM'13, August 25-29, 2013, Niagara, Ontario, Canada.

Ester und Sander 2000

Ester, Martin; Sander, Jörg: Knowledge Discovery in Databases: Techniken und Anwendungen. Springer-Verlag. Berlin Heidelberg New York.

Fayyad, Piatetsky-Shapiro, Smyth 1996 I

Fayyad, Usama; Piatetsky, Gregory; Smyth, Padhraic: From Data Mining to Knowledge Discovery: An Overview. In Advances in Knowledge Discovery and Data Mining. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. and Uthurusamy, R (Hrsg.), 1-30. AAAI Press. Menlo-Park.

Fayyad, Piatetksy-Shapiro, Smyth 1996 II

Fayyad, Piatetksy-Shapiro, Smyth. Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic: From Data Mining to Knowledge Discovery in Databases. In AI Magazine Volume 17 Number 3 S. 37-54. AAAI Press. Menlo-Park.

Fritz 2015

Fritz, Angela: Beijing chokes on off-the-charts air pollution as thick smog settles over northern China. Online: <https://www.washingtonpost.com/news/capital-weather-gang/wp/2015/11/30/beijing-chokes-on-off-the-charts-air-pollution-as-thick-smog-settles-over-northern-china/> Abruf: 16.01.2016.

GovData 2016

GovData: Metadatenstruktur für Daten in Deutschland. Online: <https://www.govdata.de/metadaten-schema> Abruf: 15.01.2016.

Hamburger Justizbehörde 2015

Hamburger Justizbehörde: Hamburgisches Transparenzgesetz. Online: <http://www.hamburg.de/transparenzgesetz/> Abruf: 15.01.2016.

Hamburger Luftmessnetz 2016

Hamburger Luftmessnetz: Der Hamburg Luftqualitätsindex LQI. Online: <http://luft.hamburg.de/wir-ueber-uns/4242598/hamburger-luftqualitaetsindex/> Abruf: 15.01.2016.

Ikonomou 2015

Ikonomou, Vasilis: Open Data basierte digitale narrative Strukturen. Online: <http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/bachelor/ikonomou.pdf> Abruf: 15.01.2016.

Meyer-Wellmann 2015

Meyer-Wellmann, Jens: Hamburgs Luft wird wieder schlechter. Online: <http://www.abendblatt.de/hamburg/article205488231/Hamburgs-Luft-wird-wieder-schlechter.html> Abruf: 16.01.2016.

Motta, You, Sacco, Ma 2014

Motta, Gianmario; You, Linlin; Sacco, Daniele; Ma, Tianyi: CITY FEED: a Crowdsourcing System for City Governance. 2014 IEEE 8th International Symposium on Service Oriented System Engineering. IEEE.

Open Knowledge Foundation Deutschland 2016

Open Knowledge Foundation Deutschland 2016: Offene Daten. Online: <http://okfn.de/themen/offene-daten/> Abruf: 15.01.2016.

Open Knowledge Foundation 2016

Open Knowledge Foundation: CKAN instance around the world. Online: <http://ckan.org/instances/> Abruf: 15.01.2016.

von Lucke 2011

von Lucke, Jörn: Innovationsschub durch Open Data, Datenportale und Umsetzungswettbewerbe. Schauer, Reinbert; Tohm, Norbert und Hilgers, Dennis (Hrsg.): Innovative Verwaltungen - Innovationsmanagement als Instrument von Verwaltungsreformen. Trauner Verlag. Linz.

Pewatron 2016

Pewatron: Gas Sensors: MiCS-4514 CO/NO2 MOS gas sensor Online: <http://www.pewatron.com/en/products-solutions/physical-sensors/gas-sensors/info/mics-4514-cono2-mos-gas-sensor/> Abruf: 16.01.2016.

Piatetsky-Shapiro 1990

Piatetsky-Shapiro, Gregory: Knowledge Discovery in Real Databases: A Report on the AJCAI-89 Workshop. AI Magazine 11(5): 68-70. AAAI Press. Menlo-Park.

Right2INFO.org 2016

Right2INFO.org 2016: The Right To Information: Good Law And Practice. Online: <http://www.right2info.org/about#about-us> Abruf: 15.01.2016.

Smart Citizen 2016

Smart Citizen: Smart Citizen Platform. Online: <https://smartcitizen.me/> Abruf: 16.01.2016.

Socrata 2015

Socrata: Customer Stories. Online: <https://www.socrata.com/customer-stories/> Abruf: 15.01.2016.

Transparentportal Hamburg 2016

Transparentportal Hamburg: Hinweise zur API. Online: <http://transparenz.hamburg.de/hinweise-zur-api/> Online: 15.01.2016.

7 Abbildungsverzeichnis

Abb. 1: Luftdaten Luftmessnetz Hamburg und Umweltbundesamt

Hamburger Luftmessnetz 2015: Stationsdaten Wilhelmsburg. Online: <http://luft.hamburg.de/clp/messstationen-aktuelle-messdaten/clp1/station/61WB?> Abruf: 15.01.2016.

Umwelt Bundesamt 2015: Aktuelle Luftdaten. Online: <http://www.umweltbundesamt.de/daten/luftbelastung/aktuelle-luftdaten> Abruf: 15.01.2016.

Abb. 2: An Overview of the Steps that Compose the KDD Process.

entnommen aus Fayyad, Piatetsky-Shapiro, Smyth 1996 I. Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic: From Data Mining to Knowledge Discovery in Databases. AI Magazine Volume 17 Number 3 37-54. AAAAI.

Abb. 3: LQI und AQI Indizes

Hamburger Luftmessnetz: Der Hamburger Luftqualitätsindex LQI. Online: <http://luft.hamburg.de/wir-ueber-uns/4242598/hamburger-luftqualitaetsindex/> Abruf: 15.01.2016

AirNow: Air Quality Index (AQI) Basics. Online: <http://airnow.gov/index.cfm?action=aqibasics.aqi> Abruf: 15.01.2016

Abb. 4: AQI Illustration: Hamburg vs. Shanghai

Darstellung von Jessica Broscheit, Datengrundlage vgl. Anhang 3

Abb. 5: Prototyp Atemmaske

Foto von Jessica Broscheit: Konstruktion einer Atemmaske mit Lilypad Arduino

8 Anhang

Anhang 1: Luftmessnetz Hamburg: Wilhelmsburg 24.12.15 - 22.01.16

Station	Wilhelmsburg	Wilhelmsburg	Wilhelmsburg	Wilhelmsburg	Wilhelmsburg	Wilhelmsburg
Komponente	SO ₂	NO	NO ₂	PM10	PM2	5
Einheit	µg/m ³					
Messzeit	Tageswerte	Tageswerte	Tageswerte	Tageswerte	Tageswerte	
22.01.2016						
21.01.2016	4	11	46	24	24	
20.01.2016	4	13	54	16	16	
19.01.2016	5	17	47	26	25	
18.01.2016	5	40	46	20	19	
17.01.2016	3	3	25	9	11	
16.01.2016	3	7	34	20	19	
15.01.2016	3	7	32	14	14	
14.01.2016	7	9	33		13	
13.01.2016	5	14	40		10	
12.01.2016	3	18	33	14	14	
11.01.2016	3	19	35	9	10	
10.01.2016	3	15	38	16	16	
09.01.2016	3	18	41	17	13	
08.01.2016	5	16	28	19	16	
07.01.2016	8	8	23	85	81	
06.01.2016	13	5	24	41	38	
05.01.2016	9	7	25	48	45	
04.01.2016	5	4	20	44	42	
03.01.2016	3	2	10	31	29	
02.01.2016	3	2	10	25	23	
01.01.2016	6	19	24	111	93	
31.12.2015	4	12	25	47	46	
30.12.2015	3	4	17	23	22	
29.12.2015	3	2	12	15	13	
28.12.2015	3	37	31	10	7	
27.12.2015	3	2	15	13	10	
26.12.2015	3	2	5	9	9	
25.12.2015	5	2	12	11	6	
24.12.2015	4	2	21	11	8	

Formartierungsfehler

lückenhafte Daten

Anhang 2: Umweltbundesamt: Wilhelmsburg SO2 24.12.15 - 22.01.16

# DEHH059 Hamburg Wilhelmsburg: Tagesmittel für Schwefeldioxid [$\mu\text{g}/\text{m}^3$]	
24.12.15	1:00;4
25.12.15	1:00;5
26.12.15	1:00;3
27.12.15	1:00;3
28.12.15	1:00;3
29.12.15	1:00;3
30.12.15	1:00;3
31.12.15	1:00;4
1.1.16;	1:00;6
2.1.16;	1:00;3
3.1.16;	1:00;3
5.1.16;	1:00;9
6.1.16;	1:00;13
7.1.16;	1:00;8
8.1.16;	1:00;5
12.1.16;	1:00;3
13.1.16;	1:00;5
14.1.16;	1:00;7
15.1.16;	1:00;3
16.1.16;	1:00;3
17.1.16;	1:00;3
18.1.16;	1:00;5
19.1.16;	1:00;5
20.1.16;	1:00;4
21.1.16;	1:00;4

unerklärte Daten,
vermutlich Uhrzeit
Daten sind aber Durchschnittswerte

Anhang 3: AQI Werte für Hamburg und Shanghai im Oktober 2015

Tag	Zeit	Hamburg LQI	Hamburg AQI	SHANGHAI AQI
01/10/15	7:00:00 PM	3	85	76
02/10/15	7:00:00 PM	3	102	114
03/10/15	7:00:00 PM	4	110	74
04/10/15	7:00:00 PM	5	181	34
05/10/15	7:00:00 PM	5	171	70
06/10/15	7:00:00 PM	4	159	78
07/10/15	7:00:00 PM	4	93	21
08/10/15	7:00:00 PM	4	100	107
09/10/15	7:00:00 PM	3	73	167
10/10/15	7:00:00 PM	3	59	30
11/10/15	7:00:00 PM	3	75	163
12/10/15	7:00:00 PM	3	91	157
13/10/15	7:00:00 PM	4	106	198
14/10/15	7:00:00 PM	3	104	149
15/10/15	7:00:00 PM	4	141	107
16/10/15	7:00:00 PM	4	100	87
17/10/15	7:00:00 PM	3	83	122
18/10/15	7:00:00 PM	3	98	114
19/10/15	7:00:00 PM	4	84	112
20/10/15	7:00:00 PM	3	100	99
21/10/15	7:00:00 PM	4	89	153
22/10/15	7:00:00 PM	3	102	117
23/10/15	7:00:00 PM	4	143	112
24/10/15	7:00:00 PM	3	87	107
25/10/15	7:00:00 PM	3	89	82
26/10/15	7:00:00 PM	4	145	156
27/10/15	7:00:00 PM	4	145	161
28/10/15	7:00:00 PM	5	167	138
29/10/15	7:00:00 PM	5	167	68
30/10/15	7:00:00 PM	5	167	57
31/10/15	7:00:00 PM	5	167	93