

# **Entwicklung eines Trame-basierten Tools zur verständlichen Darstellung und Analyse klinischer Daten**

## **Bachelorarbeit**

Richard Meier zu Ummeln  
# 400311

30. April 2025

Gutachter: Prof. Dr. Benjamin Blankertz  
Prof. Dr. Marc Alexa



Technische Universität Berlin  
Fakultät IV – Elektrotechnik und Informatik  
Institut für Softwaretechnik und Theoretische Informatik  
Fachgebiet Neurotechnologie

# Kurzfassung

Große Datenbanken können unübersichtlich und komplex sein. Klinische Daten stellen dabei keine Ausnahmen dar. 2016 hat eine Forschungsgruppe der University of Waterloo für dieses Problem eine Lösung in Form eines web-basierten Visualisierungstools für die MIMIC-II-Datenbank entwickelt. MIMIC ist eine öffentliche Datenbank von Medizindaten und erhielt 2016 eine Version III. Diese Arbeit befasst sich mit der Entwicklung eines Prototypen für die MIMIC-III-Datenbank und versucht, diese Daten verständlich darzustellen, um sie besser analysieren zu können. Dabei werden unterschiedliche Literaturquellen herangezogen. Im Grundsatz wird das Paper des Vorgängertools benutzt, um die Lösung für MIMIC-II nachzubauen und mit sinnvollen Funktionen zu erweitern. Zusätzlich werden Richtlinien für gute wissenschaftliche Darstellungen berücksichtigt, um die Verständlichkeit zu erhöhen. Es entstand dadurch eine Lösung für MIMIC-III, die die grundlegende Funktionalität des Tools für MIMIC-II implementiert und erweitert. Diese kann von Forschenden verwendet werden, um sich einfacher in die klinischen Daten dieser Datenbank einzuarbeiten.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>MIMIC-III und MIMIC-Extract</b>	<b>2</b>
2.1	Visualisierungstool für MIMIC-II . . . . .	2
2.2	MIMIC-III . . . . .	3
2.3	MIMIC-Extract . . . . .	3
<b>3</b>	<b>Methoden</b>	<b>5</b>
3.1	Programmierung . . . . .	5
3.1.1	Preprocessing . . . . .	5
3.2	Architektur . . . . .	7
3.3	Qualitative Analyse . . . . .	8
3.3.1	Regeln zu guten Diagrammen . . . . .	8
3.3.2	Benchmark . . . . .	9
<b>4</b>	<b>Ergebnisse</b>	<b>10</b>
4.1	Benutzung . . . . .	10
4.2	Tool . . . . .	10
4.2.1	Navigationsleiste . . . . .	10
4.2.2	Seitenleiste . . . . .	10
4.2.3	Einführung . . . . .	11
4.2.4	Erforschen . . . . .	11
4.2.5	Vergleichen . . . . .	13
4.2.6	Analysieren . . . . .	15
4.3	Benchmark . . . . .	15
<b>5</b>	<b>Diskussion</b>	<b>17</b>
5.1	Performance . . . . .	17
5.2	Übersichtlichkeit . . . . .	17
5.3	Verständliche Diagramme . . . . .	18
5.4	Beitrag . . . . .	19
5.5	Limitationen . . . . .	20
5.6	Zukünftige Arbeiten . . . . .	20
<b>6</b>	<b>Fazit</b>	<b>21</b>
<b>7</b>	<b>KI-Tools</b>	<b>22</b>
7.1	Perplexity von Perplexity AI, Inc. . . . .	22
7.2	Code Completion von JetBrains s.r.o. . . . .	22

*Inhaltsverzeichnis*

7.3 Writefull von ThingLab Ltd . . . . . 22

# Abbildungsverzeichnis

2.1	CSV-Datei in einem Textbearbeitungsprogramm . . . . .	3
3.1	Design des Programms HDFView in Windows . . . . .	6
3.2	Funktionsweise des Tools . . . . .	8
4.1	Erste Seite des Tools . . . . .	11
4.2	Zweite Seite des Tools . . . . .	12
4.3	Dritte Seite des Tools . . . . .	13
4.4	Zoom-Funktion . . . . .	14
4.5	Hover-Funktion . . . . .	14
4.6	Vierte Seite des Tools . . . . .	16
4.7	Beispiel Boxplot . . . . .	16

# Tabellenverzeichnis

3.1	Tabellen der MIMIC-Extract-Datenbank und ihr Inhalt . . . . .	6
3.2	Zehn häufigste Diagnosen . . . . .	7
4.1	Tabelle im Explore-Tab . . . . .	12

# 1 Einleitung

Big Data ermöglicht es im Gesundheitswesen, Zusammenhänge zwischen Erkrankungen, Behandlungen und Ergebnissen aufzudecken und wirksame Maßnahmen zu entwickeln. [1] Big Data beschreibt dabei besonders große und komplexe Datensätze, die mit herkömmlichen Datenverarbeitungstools nicht verwaltet oder analysiert werden können. [2] Um die Daten auch verarbeiten zu können, braucht es allerdings gute Werkzeuge. Gleichzeitig ist die Einarbeitung in Datenbanken sehr zeitaufwändig. Ein Beispiel für solche großen und komplexen Datensätze ist die MIMIC-III (Medical Information Mart for Intensive Care III) Datenbank. [3] Auf der Springer Nature Website hat das MIMIC-III Paper über 300.000 Aufrufe und über 4.500 Zitationen laut dem Web of Science. [4] Diese Metriken zeigen die große Relevanz der kostenlos verfügbaren Datenbank.

Seit der Erstellung des Visualisierungstools für MIMIC-II [5] sind neun Jahre vergangen. Es erschienen zwei neue MIMIC-Versionen und weitere technische Innovationen. Darunter ein Framework der Firma *Kitware*, nämlich *Trame* [6] zum Visualisieren und wissenschaftlichen Analysieren von Daten. Diese Arbeit wird auf dem Paper basierend das Tool neu bauen und erweitern. In dieser Bachelorarbeit wird das *Trame*-Framework [6] genutzt, um auf dem Paper von Lee, Ribey und Wallace [5] aufzubauen und ein neues Visualisierungstool für MIMIC-III zu erstellen, sowie das Konzept des alten Tools im Hinblick auf die Analysefunktionalität zu erweitern. Als Fortsetzung dieser Entwicklung möchte ich mit meiner Arbeit dazu beitragen, das vorhergehende Tool weiterzuentwickeln und beabsichtige, dass mein Tool dadurch zukünftig wiederum von anderen wissenschaftlichen Beiträgen aufgegriffen und dadurch weiterentwickelt werden kann. Ziel ist die Entwicklung eines *Trame*-basierten Tools zur verständlichen Darstellung und Analyse klinischer Daten. Die Verständlichkeit soll dadurch erreicht werden, dass das resultierende Tool übersichtlich und einfach bedienbar sein soll. Da das alte Tool nicht mehr verfügbar ist, wird es von Grund auf mithilfe eines Frameworks neu gebaut. Zusätzlich wird versucht, die alte Funktionalität zu erweitern. *Trame* ist ein Python-Framework, das speziell zur Erstellung von wissenschaftlichen Analyse- und Modellierungstools entwickelt wurde. [6] Die vielen Beispiele der Firma hinter dem Tool [7] zeigen, dass es sich gut eignet, um verschiedenste Daten visuell darzustellen. „Python [...] ist hervorgegangen als ein leistungsstarkes Werkzeug für Datenanalyse wegen seiner umfangreichen Bibliotheken und der nutzerfreundlichen Beschaffenheit.“<sup>1</sup> Da *Trame* in Python als zusätzliches Package hinzugefügt und somit mit Python entwickelt wird, bietet dies eine gute Voraussetzung für das Zusammenspiel beider Komponenten und spricht für den Einsatz von *Trame* zur Entwicklung eines verständlichen Datenanalyse-Tools.

---

<sup>1</sup>Eigene Übersetzung, im Original: „Python, as a versatile programming language, has emerged as a powerful tool for data analytics due to its extensive libraries and user-friendly nature.“ [8]

## 2 MIMIC-III und MIMIC-Extract

In diesem Kapitel werden die zugrundeliegende Datenbank und die wissenschaftlichen Grundlagen der Arbeit vorgestellt.

MIMIC-III ist eine Sammlung von mehreren CSV-Dateien (Comma-Separated Values). [9, S. 7] CSV-Dateien sind grundsätzlich Textdateien, bestehend aus Strings, Trennzeichen und Zeilen. Textbearbeitungsprogramme stellen diese allerdings sehr schlecht lesbar dar (siehe Abbildung 2.1), vor allem, da es hierbei keine Raster gibt.

Eine Lösung stellen hierfür Tabellenverarbeitungsprogramme wie Excel dar. Diese einzeln durchzuschauen, kann allerdings einen hohen Aufwand bedeuten. Daten daraus zu extrahieren und sie zu analysieren, einen noch höheren. Hierbei kann die Pipeline von MIMIC-Extract helfen. [10] Die hier ausgegebenen Dateien bestehen lediglich aus einer HDF5-Datei (Hierarchical Data Format). Dies gilt allerdings auch nur für die Standardvariante, die man sich auf der „Google Cloud Platform“ holen kann. Sollte man eine eigene Version erstellen wollen, um einzelne Parameter anzupassen, muss das Programm heruntergeladen und aufwendig aufgesetzt werden. Nativ unterstützen Tabellenverarbeitungsprogramme wie Excel allerdings keine HDF5-Dateien. [11] Dies erschwert es also, mit diesen Programmen die MIMIC-Extract-Datenbank zu öffnen.

Auf andere Weise könnten diese mithilfe von Python-Skripten umständlich ohne Benutzeroberfläche verarbeitet werden. Alternativen mit Benutzeroberflächen gibt es derzeit noch nicht viele. Ein Beispiel ist dafür allerdings das Programm *HDFView*. [12]

Ein weiteres Problem, das sich ergibt, sind schlecht verarbeitete Werte. In der MIMIC-Extract-Datenbank befinden sich unter den 10 häufigsten Diagnosen bei der ersten Aufnahme z. B.

- „CORONARY ARTERY DISEASE“,
- „CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT /SDA“ und
- „CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT/SDA“.

Da hier dieselben Diagnosen teils unterschiedlich formatiert werden, ist es schwierig, Datenbank-Operationen anzuwenden oder die Datensätze zu analysieren. Die Datenbank müsste daher noch weiter angepasst werden, was in einem Tabellenverarbeitungsprogramm schwierig umsetzbar ist. MIMIC-Extract ist außerdem bereits 5 Jahre alt und wurde seit 3 Jahren nicht mehr aktualisiert. Eine fehlende „requirements.txt“ macht es zusätzlich schwieriger, die alte Umgebung wieder aufzusetzen.

### 2.1 Visualisierungstool für MIMIC-II

Das grundsätzliche Konzept des neuen Tools basiert auf einer Arbeit von Lee, Ribey und Wallace [5]. In ihrer Arbeit haben sie ein Visualisierungstool erschaffen, das es ermöglicht, die MIMIC-II-Datenbank zu erforschen und einzelne Kohorten zu vergleichen. Dieses Tool ist mittlerweile nicht mehr öffentlich einsehbar, und Anfragen darauf blieben bis zum Ende meiner Arbeit unbeantwortet. Die Funktionalität wurde allerdings in ihrem Paper weitgehend dargestellt.

## 2 MIMIC-III und MIMIC-Extract

```
row_id,subject_id,gender,dob,dod,dod_hosp,dod_ssn,expire_flag
9467,10006,F,2094-03-05 00:00:00,2165-08-12 00:00:00,2165-08-12 00:00:00,2165-08-12 00:00:00,1
9472,10011,F,2090-06-05 00:00:00,2126-08-28 00:00:00,2126-08-28 00:00:00,,1
9474,10013,F,2038-09-03 00:00:00,2125-10-07 00:00:00,2125-10-07 00:00:00,2125-10-07 00:00:00,1
9478,10017,F,2075-09-21 00:00:00,2152-09-12 00:00:00,,2152-09-12 00:00:00,1
9479,10019,M,2114-06-20 00:00:00,2163-05-15 00:00:00,2163-05-15 00:00:00,2163-05-15 00:00:00,1
9486,10026,F,1895-05-17 00:00:00,2195-11-24 00:00:00,,2195-11-24 00:00:00,1
9487,10027,F,2108-01-15 00:00:00,2190-09-14 00:00:00,,2190-09-14 00:00:00,1
9489,10029,M,2061-04-10 00:00:00,2140-09-21 00:00:00,,2140-09-21 00:00:00,1
9491,10032,M,2050-03-29 00:00:00,2138-05-21 00:00:00,2138-05-21 00:00:00,2138-05-21 00:00:00,1
9492,10033,F,2051-04-21 00:00:00,2133-09-09 00:00:00,,2133-09-09 00:00:00,1
9494,10035,M,2053-04-13 00:00:00,2133-03-30 00:00:00,,2133-03-30 00:00:00,1
9495,10036,F,1885-03-24 00:00:00,2185-03-26 00:00:00,2185-03-26 00:00:00,2185-03-26 00:00:00,1
9497,10038,F,2056-01-27 00:00:00,2147-03-17 00:00:00,2147-03-17 00:00:00,2147-03-17 00:00:00,1
9499,10040,F,2061-10-23 00:00:00,2150-09-05 00:00:00,2150-09-05 00:00:00,2150-09-05 00:00:00,1
9501,10042,M,2076-05-06 00:00:00,2150-12-03 00:00:00,,2150-12-03 00:00:00,1
9502,10043,M,2109-04-07 00:00:00,2191-02-07 00:00:00,,2191-02-07 00:00:00,1
9503,10044,F,2071-02-11 00:00:00,2152-10-20 00:00:00,2152-10-20 00:00:00,2152-10-20 00:00:00,1
9504,10045,F,2061-03-25 00:00:00,2129-12-01 00:00:00,2129-12-01 00:00:00,2129-12-01 00:00:00,1
9505,10046,F,2141-03-15 00:00:00,2195-03-13 00:00:00,,2195-03-13 00:00:00,1
9514,10056,F,2046-02-27 00:00:00,2129-09-16 00:00:00,,2129-09-16 00:00:00,1
9517,10059,M,2081-01-03 00:00:00,2150-08-29 00:00:00,2150-08-29 00:00:00,2150-08-29 00:00:00,1
9519,10061,F,2031-05-19 00:00:00,2108-04-03 00:00:00,,2108-04-03 00:00:00,1
9522,10064,M,2058-04-23 00:00:00,2127-03-19 00:00:00,2127-03-19 00:00:00,,1
9523,10065,F,2111-07-18 00:00:00,2193-11-06 00:00:00,2193-11-06 00:00:00,2193-11-06 00:00:00,1
9525,10067,M,2101-06-10 00:00:00,2130-10-06 00:00:00,2130-10-06 00:00:00,,1
9527,10069,F,2146-10-23 00:00:00,2188-02-27 00:00:00,2188-02-27 00:00:00,2188-02-27 00:00:00,1
9531,10074,M,2081-12-26 00:00:00,2167-04-24 00:00:00,,2167-04-24 00:00:00,1
9533,10076,M,2038-05-10 00:00:00,2107-03-30 00:00:00,2107-03-30 00:00:00,2107-03-30 00:00:00,1
9539,10083,F,2110-03-25 00:00:00,2192-12-12 00:00:00,,2192-12-12 00:00:00,1
9544,10088,M,2029-07-09 00:00:00,2107-07-20 00:00:00,,2107-07-20 00:00:00,1
9545,10089,M,2046-04-18 00:00:00,2132-08-08 00:00:00,2132-08-08 00:00:00,2132-08-08 00:00:00,1
9546,10090,M,2096-02-27 00:00:00,2126-04-06 00:00:00,,2126-04-06 00:00:00,1
9549,10093,M,2053-09-08 00:00:00,2141-01-26 00:00:00,2141-01-26 00:00:00,2141-01-26 00:00:00,1
```

Abbildung 2.1: CSV-Datei in einem Textbearbeitungsprogramm

## 2.2 MIMIC-III

Die Arbeit verwendet als Datenbank MIMIC-III. „MIMIC-III ist eine große, kostenlos-verfügbare Datenbank bestehend aus anonymisierten gesundheitsbetreffenden Daten verbunden mit über vierzigtausend Patienten, die sich zwischen 2001 und 2012 in Intensivstationen des Beth Isreal Deaconess Medical Centers befanden.“<sup>1</sup> Die Datenbank enthält Daten von 61.532 Intensivaufenthalten und 46.520 eindeutigen Patient:innen. Als Akademiker:in kann man Kurse mit insgesamt elf Lektionen zu den Themen Ethik und Datenschutz ablegen, um Zugriff auf die Daten zu erhalten. Es gibt bereits vier Versionen dieser Datenbank. In dieser Arbeit soll es ausschließlich um die dritte Version gehen. In dem Github-Projekt des Papers befinden sich Code-Beispiele in Form von Jupyter-Notebooks zum Erstellen von Graphen. [13] Um diese Beispiele abwandeln zu können und eigene Graphen selbst erstellen zu können, wird allerdings das Verständnis für SQL vorausgesetzt, was eine große Hürde für viele Menschen oder auch insbesondere Gesundheitsforschende darstellen kann.

## 2.3 MIMIC-Extract

Lee, Ribey und Wallace [5] haben sich auch bereits der Aufbereitung von MIMIC-II gewidmet und ein Tool zur Extraktion, der Vorverarbeitung und der Repräsentation entworfen. Für MIMIC-III braucht es ebenfalls eine Lösung zur Vorverarbeitung. Um diese vernachlässigen zu können, setzt die Verarbeitungskette an der vorherigen Forschung von Wang, McDermott, Chauhan u. a. [10] an. In ihrem Paper beschreiben sie eine Pipeline, um Daten aus MIMIC-III zu extrahieren und die Datenbank vorzuverarbeiten. Dies beinhaltet die Funktionalität ihres MIMIC-Extract Programms. Das Programm ist unter dem open-source Projekt auf Github erhältlich. [14] Eine vorgefertigte Datenbank der vorverarbeiteten Daten mit Standardwerten kann ohne weiteren Aufwand kostenlos heruntergeladen werden.

<sup>1</sup>Eigene Übersetzung, im Original: „MIMIC-III (‘Medical Information Mart for Intensive Care’) is a large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital.“ [9, S. 1]

## *2 MIMIC-III und MIMIC-Extract*

Dadurch kann die vorhergehende Aufbereitung der klinischen Daten vernachlässigt und mithilfe von MIMIC-Extract kompensiert werden. Im weiteren Verlauf werde ich daher bei der Datenbank, wenn nicht anders beschrieben, immer von der MIMIC-Extract sprechen.[10]

## 3 Methoden

Zur Problemlösung werden verschiedene Methoden eingesetzt. Ich nutze die genannten Papers zu MIMIC-III, um einen Überblick über die Datenbank zu bekommen. Ich nutze außerdem das Paper zum Visualisierungstool von MIMIC-II und das Paper zu Trame, um das Wissen auf mein eigenes Tool zu übertragen. Ähnlich zu dem Tool zu MIMIC-II wird hier eine Web-App gebaut, die mit dem Browser aufrufbar ist. Zur Umsetzung des Programms verwende ich die Websites, Dokumentationen und Git-Repositories der Packages und des Frameworks.

Zum Durchsuchen der Datenbanken verwende ich Excel, HDFView und Jupyter-Notebooks. Die Non-Profit-Organisation *The HDF Group* bietet hier HDFView als eine Lösung an. [12] Das Programm hat laut Aussagen der Organisation über 3.500 Downloads im Monat. Es ist allerdings erst nach Registrierung erhältlich. Die Benutzeroberfläche von HDFView ist veraltet (siehe Abbildung 3.1) und das Programm stürzt bei besonders großen Tabellen ab, weshalb auf Jupyter-Notebooks [15] benutzt wurden. Diese haben es ebenfalls ermöglicht, den Code für das Tool auszuprobieren und bestimmte Werte auszulesen.

### 3.1 Programmierung

Als Entwicklungssoftware benutze ich die Entwicklungsumgebung Pycharm von der Firma JetBrains s.r.o. [16] Zur Erstellung des Back-ends und Front-ends wird Trame als Allzwecklösung verwendet. Die Diagramme werden mit Plotly gebaut, womit es möglich ist, hochinteraktive Graphen zu erstellen.

#### 3.1.1 Preprocessing

Die Datenbank, die von MIMIC-Extract ausgegeben wird, kann mit den Standard-Parametern auch von Google heruntergeladen werden. Diese HDF5-Datei wird dann noch einmal von Pandas [17], einer Bibliothek von Python zur Verarbeitung von Datenmengen, in der *preprocess.py* weiterverarbeitet und so an die Schnittstelle des Tools angepasst. Hier wird auf Einfachheit gesetzt. Das Programm soll es den Nutzenden vereinfachen, die Datenbank zu lesen und daher auch keine neuen Hürden schaffen. Aus diesem Grund ist es wichtig, dass die Nutzenden keine weiteren Tools herunterladen oder noch aufwändig ein Python-Skript konfigurieren müssen.

Die MIMIC-Extract Pipeline reduziert die 26 CSV-Dateien von MIMIC-III bereits auf fünf Tabellen, die in der Tabelle 3.1 aufgelistet sind. Mit dem Skript in der „preprocess.py“ wurde die „all\_hourly\_data.h5“ von MIMIC-Extract noch weiter vorverarbeitet für das Tool. Generell wurden hier nur die Tabellen „patients“ und „vital\_labs\_mean“ übertragen, da sie für die Analyse am relevantesten sind. Um die Performance zu erhöhen, wurde auch die MIMIC-Extract-Datenbank noch stark beschnitten. Es wurde einmal nach den Top-20 Diagnosen, die die Subjekte bei der Aufnahme im Krankenhaus bekamen, gefiltert. Einen Auszug der Top-10 Diagnosen wird in Abbildung 3.2 angegeben. Weitere

### 3 Methoden

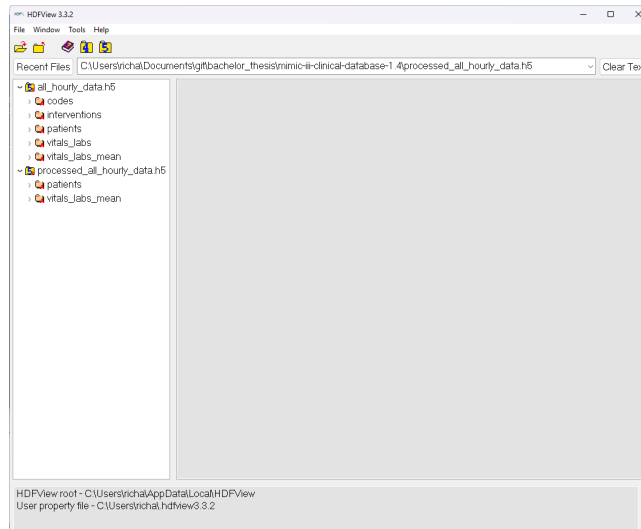


Abbildung 3.1: Design des Programms HDFView in Windows

Tabelle	Inhalt
patients	Patient:innen und ihre demographischen Werte oder statischen Ergebnisse
vitals_labs	Vital- und Laborwerte (Stundenmittelwert, Anzahl, Standardabweichung)
vitals_labs_mean	Vital- und Laborwerte (Stundenmittelwert)
interventions	Stündliche binäre Indikatoren für durchgeführte Interventionen
codes	ICD-9 Codes

Tabelle 3.1: Tabellen der MIMIC-Extract-Datenbank und ihr Inhalt

### 3 Methoden

Diagnose	Anzahl
PNEUMONIA	825
CORONARY ARTERY DISEASE	751
SEPSIS	612
CHEST PAIN	553
CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT /SDA	552
INTRACRANIAL HEMORRHAGE	542
GASTROINTESTINAL BLEED	490
CONGESTIVE HEART FAILURE	449
ALTERED MENTAL STATUS	433
CORONARY ARTERY DISEASE\CORONARY ARTERY BYPASS GRAFT/SDA	413

Tabelle 3.2: Tabelle mit den zehn Diagnosen, die in der Patienten-Spalte der MIMIC-Extract-Datei am häufigsten vorkommen zusammen mit ihrer Anzahl.

Diagnosen können allerdings mit dem Bearbeiten der Liste in der „preprocess.py“ wieder hinzugefügt werden.

In den Tabellen der Datenbank gibt es die Schlüsselwerte `subject_id` (Subjekt-ID), `hadm_id` (Krankenhaus-Aufenthalt-ID) und `icustay_id` (ID für den Aufenthalt auf der Intensivstation). Die Identifikationsnummern sind alle einzigartig und wurden nicht mehrfach vergeben. Daher wurden bei der Tabelle *patients* die Spalten „`hadm_id`“ und „`icustay_id`“ entfernt. Es wurden noch weitere Spalten entfernt. Die übrig gebliebenen Spalten in der *patients* Tabelle können in Tabelle 4.1 eingesehen werden.

Das Alter von Personen über 89 Jahren wurde in MIMIC-III generell auf über 300 Jahre verschoben, um das wahre Alter zu verschleiern und somit den gesetzlichen Regularien zu entsprechen. [9] Dies erschwert es allerdings auch, das Alter in Diagrammen darzustellen. Daher wird im Preprocessing das Alter auf 89 verschoben.

In der Tabelle „`vital_labs_mean`“ wurde eine Spalte entfernt, in der die angewandte Funktion stand, also ausschließlich der Wert „`mean`“. Da der Durchschnittswert nur auf die Stunden berechnet wurde, bildet das Preprocess-Skript aus den Stunden nochmal einen Durchschnittswert, um genau einen Wert für jeden Laborwerttyp pro Patient:in zu haben. Zum Schluss werden auch hier noch einmal die zusätzlichen Schlüsselattribute `hadm_id` und `icustay_id` entfernt.

Das Skript verpackt diese zwei neuen Tabellen dann in eine weitere HDF5-Datei für das Hauptprogramm.

## 3.2 Architektur

Die generelle Funktionsweise wird grob in der Abbildung 3.2 umrissen. Das Hauptprogramm nimmt die zuletzt erstellte HDF5-Datei als Datenbank auf. Diese wird von Pandas [17] eingelesen, zur besseren Darstellung ein wenig angepasst und als DataFrame im Arbeitsspeicher abgelegt. Vuetify [18] ist ein Framework für Webkomponenten, das Trame über eine Schnittstelle zur Verfügung gestellt wird. Das DataFrame wird in eine sequentiell einlesbare Liste umgewandelt, damit es als eine HTML-Tabelle dargestellt werden kann. Löst der:die Nutzer:in über den Browser ein Event aus, werden die notwendigen Daten über die Schnittstellen von Trame angefragt. Werden dabei auch State-Variablen verändert, so werden zusätzlich Methoden ausgelöst, die auf den Wert dieser Variablen hören. Da-

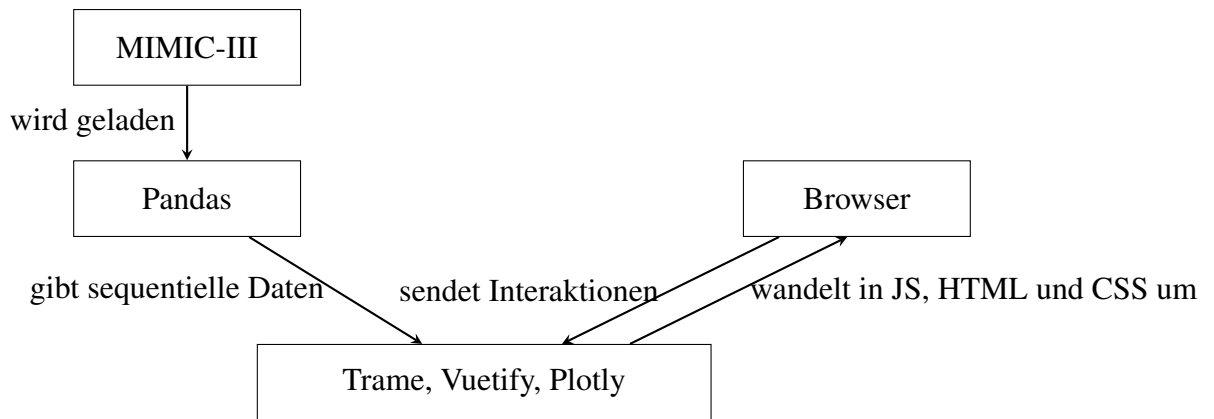


Abbildung 3.2: Grobe Darstellung der Funktionsweise des Tools (links die Serverseite, rechts die Clientseite), angelehnt an Trame-Einführungskurs [19, S. 3]

bei können serverseitig wiederum neue Pandas-Operationen auf dem DataFrame ausgeführt werden, mit denen dann durch Plotly z. B. neue Grafiken erzeugt und zurück an die Ausgabe im Browser geleitet werden. Das Design der Benutzeroberfläche wird hauptsächlich durch Standardklassen des Vuetify-Frameworks formatiert.

Kitware, die Firma hinter Trame [6], beschreibt ihr Framework als eine Möglichkeit, interaktive Applikationen mit wissenschaftlichen Visualisierungen zu erschaffen. [7] Der Name steht für das gleichnamige französische Wort und bedeutet „der Kern, der alles miteinander zusammenbindet“. Das Framework entstand im Jahr 2021. [19] Beim Programmieren mit dem Framework wird eine Kombination aus vielen Sprachen und Dialekten benutzt. Darunter fallen Python, HTML, JavaScript, CSS und Vuetify. Dank dieses Werkzeugs kann die Entwicklung mit einer fertigen Client-Server-Umgebung starten und dieser Schritt übersprungen werden.

## 3.3 Qualitative Analyse

Die Ergebnisse werden unter anderem anhand der in diesem Kapitel genannten Kriterien bewertet. Zur qualitativen Analyse wird ein Paper mit Regeln zu guten wissenschaftlichen Diagrammen verwendet und für die Analyse des Programms ein Performance- und Accessibility-Benchmark.

### 3.3.1 Regeln zu guten Diagrammen

Rougier, Droettboom und Bourne [20] benennen in ihrem Werk zehn Regeln, nach denen das Design von Abbildungen verbessert werden kann und wie häufige Fehler vermieden werden können. Eine Zusammenfassung dieser zehn Regeln lautet:

1. Zielgruppe: Enthält die Grafik alle wichtigen Informationen für ein breiteres Publikum?
2. Botschaft: Hat die Abbildung eine Botschaft? Ist sie schnell wahrnehmbar? Ist sie einschlägig?
3. Abbildung an Medium anpassen: Größe entsprechend des Bildschirms und ist die Interaktivität mit der Abbildung ausgeschöpft (Zoom, Drag)? Wenn der Leser mehr Zeit hat (als in einer Präsentation), kann man mehr Details und Erklärungen einbauen.

### 3 Methoden

4. Bildunterschriften sind nicht optional: Wie wird die Grafik gelesen? Was sind Präzisierungen, die nicht visuell wiedergegeben werden können?
5. Nicht blind den Standardwerten vertrauen.
6. Farbeffizienz: Der Standard sollte schwarz sein. Mit Farben können Akzente gesetzt werden.
7. Objektivität präsentieren: Kolleg:innen fragen, was sie von der Grafik denken. Verhältnisse proportional richtig darstellen. Den einfachsten Typen verwenden. Labels, Ticks, Titel und den vollen Wertebereich wiedergeben.
8. Dekorationen, die dem Lesenden nichts sagen, sollten entfernt werden.
9. Botschaft und Lesbarkeit sind wichtiger als optionale Schönheitsmerkmale.
10. Das richtige Tool verwenden.
11. Sind die Plots barrierefrei?

Die elfte Regel ist meine eigene Ergänzung. Um die Grafik für alle Menschen verständlich machen zu können, sollten auch Leute einbezogen werden, die nicht alle oder auch gar keine Farben sehen können.

#### 3.3.2 Benchmark

Mithilfe von Trame werden vor allem Websites gebaut. Für Websites gibt es verschiedene Benchmark-Systeme und andere Messmethoden, die die Performance messen können. Ein weit verbreitetes Beispiel mit 1.000.000 Nutzer:innen ist „Lighthouse“. [21] Dies soll das Nutzererlebnis der Software als ersten Eindruck bewerten.

## 4 Ergebnisse

Im Folgenden werden das Tool und die vereinzelt Unterfunktionen als Ergebnis der Arbeit beschrieben.

### 4.1 Benutzung

Für die Installation wird die Freischaltung für die MIMIC-III-Datenbank benötigt. Laut Johnson, Pollard, Shen u. a. [9] kann dies eine Woche dauern. Des Weiteren wird Python benötigt. Die genauen Schritte befinden sich unter der Installationsanleitung in der „README.md“-Datei. Zum Starten des Programms müssen die benötigten Packages in der „requirements.txt“ installiert, die MIMIC-Extract-Datei ins Verzeichnis gelegt, die „preprocess.py“- und die „main.py“-Datei ausgeführt werden. Bei jedem weiteren Start muss dann nur noch die „main.py“ ausgeführt werden.

### 4.2 Tool

Im Folgenden wird das Hauptprogramm beschrieben. Das Tool bietet eine dauerhaft sichtbare Seitenleiste und vier verschiedene Modi, auf die im Folgenden genauer eingegangen wird.

#### 4.2.1 Navigationsleiste

Oben auf der Seite befindet sich eine dauerhaft präsente Navigationsleiste, die das Einklappen der Seitenleiste ermöglicht, wodurch der restliche Seiteninhalt vergrößert wird (siehe Abbildung 4.2). Daneben befindet sich der Titel der Seite.

Am rechten Ende der Navigationsleiste befinden sich die vier Knöpfe „Intro“, „Explore“, „Compare“ und „Analyse“ zum Wechseln zwischen den Modi.

#### 4.2.2 Seitenleiste

Auf der Seitenleiste können Bedingungen gesetzt werden, die die Subjekte aus der ersten Kohorte erfüllen sollen. Darunter fallen die erste Intensivstation der Person bei der Aufnahme im Krankenhaus, das Alter und das Geschlecht der Person. Zu den möglichen Optionen der ersten Intensivstation gehören CCU (Coronary Care Unit), CSRU (Cardiac Surgery Recover Unit), MICU (Medical Intensive Care Unit), SICU (Surgical Intensive Care Unit) und TSICU (Trauma Surgical Intensive Care Unit). Mit dem Schieberegler oder mit den Eingabefeldern kann das gewünschte Alter der Kohorte eingegrenzt werden. Als letzte Option kann zwischen den Geschlechtern männlich und weiblich gewählt werden.

## 4 Ergebnisse

MIMIC-III Data Visualization Tool

INTRO EXPLORE COMPARE ANALYZE

### Cohort 1 Characteristics

First ICU  
CCU CSRU MICU SICU TSICU

Age  
0 90

Gender  
FEMALE MALE OTHER

APPLY SELECTION

### Introduction

This tool aims to make the MIMIC-III database more understandable for humans.

#### General Information

MIMIC (Medical Information Mart for Intensive Care) is a large, freely-available database comprising deidentified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center. MIMIC-III contains data from 2001-2012. The data was collected from Metavision and CareVue bedside monitors." (Source: <https://mimic.mit.edu/docs/about/>)

The MIMIC-III database contains data of 61,532 ICU stays and 46,520 unique patients. This webtool uses the MIMIC-Extract version of that database with default parameters. To narrow it further down you have to use the preprocess script to make it more lightweight and focus on the most common diagnoses (with the best statistical basis). (Tool: <https://github.com/MLforHealth/MIMIC-Extract>)

Difference between the three types of ID: Each patient has a subject ID. A patient can have multiple admissions/visits at one hospital. A visit at the hospital can have multiple visits at an ICU. Dates of birth for patients aged over 89 were shifted to obscure their true age and comply with HIPAA regulations: these patients appear in the database with ages of over 300 years." (Source: <https://physionet.org/content/mimiciii/1.4/>)

#### Tool Explanation

##### Instructions

1. Start at the explore tab, select a cohort and see the effects.
2. Take a look at the graphs or download the plots as images if you made a discovery.
3. Come back to the introduction if you need further information.

##### Drawer

The drawer is the bar on your left. It combines all the functionality to select a cohort. On the compare tab you get a another drawer on the right to select a second cohort.

##### First ICU

The first unit that took care of the patient.

Term	Description
CCU	Coronary Care Unit
CSRU	Cardiac Surgery Recover Unit
MICU	Medical Intensive Care Unit
SICU	Surgical Intensive Care Unit
TSICU	Trauma Surgical Intensive Care Unit

##### Age

With the slider you can select a range of ages. You can also use the number input fields. The first box includes all subjects from that age and the second box excludes all from that point.

##### Gender

The gender of the patient.  
The gender column only contains values of "M" (Male) and "F" (Female) and therefore the "Other"-option has been disabled.

##### Explore

On this page you can search through the database and see the effects your selection and the resulting cohort.  
Here you can get a first impression of the subjects data. Click the selection box in the datatable for all subjects and optionally deselect single subjects.

##### Compare

On the next page "Compare" you can compare 2 cohorts. Select all the characteristics in the side panel you want your cohorts to have.

##### Analyze

On the last page you have more space to get deeper in the analysis of a specific cohort or the wholledatasets.

Powered by frame © 2025 Kware Inc.

Abbildung 4.1: Erste Seite des Tools zum Erklären des Tools und der Hintergründe.

Mit dem Knopf „Apply Selection“ kann die Auswahl finalisiert werden. Im Hintergrund werden beim Drücken dann Pandas-Methoden auf dem DataFrame angewandt, um einzelne Subjekte zu selektieren. Die selektierten Subjekte können auf dem Tab „Explore“ geprüft werden.

### 4.2.3 Einführung

Auf der ersten Seite „Intro“ (siehe Abbildung 4.1) stehen Erklärungen und weitere Informationen zu den angezeigten Daten. Für Fragen kann hier jederzeit hingesprungen werden. Ganz oben werden generelle Informationen zu dem Tool gegeben. Danach werden einzelne Funktionen erklärt und verschiedene Eigenheiten, die nicht intuitiv erscheinen.

### 4.2.4 Erforschen

Auf der zweiten Seite „Explore“ (siehe Abbildung 4.2) kann die Datenbank durchforstet werden. Dafür wird die Tabelle der Subjekte als HTML-Tabelle mit wenigen zusätzlichen Formatierungen dargestellt. Die einzelnen Spalten werden in Tabelle 4.1 erklärt. Es können hier auch einzelne Subjekte aus- bzw. ausgewählt werden. Eine Suchleiste ermöglicht es, die komplette Datenbank nach Werten der Personen zu durchsuchen – dabei werden alle Spalten in Betracht gezogen.

Der Tabellenkopf ermöglicht es, einzelne Spalten oder auch mehrere Spalten mithilfe von Priorisierung zu sortieren („multi-sort“). Ein Beispiel für eine mögliche Sortierung anhand mehrerer Eigenschaften wäre es, erst nach dem Geschlecht zu sortieren und dann innerhalb der Geschlechter nach dem Alter. Im Tabellenfuß können die Patient:innen pro Seite eingestellt werden. Zusätzlich gibt es Knöpfe, um auf die erste, vorherige, nächste und letzte Seite zu springen.

## 4 Ergebnisse

Spalte	Wert
subject_id	Identifikationsnummer als Zahl
gender	Geschlecht als M (männlich) oder F (weiblich)
ethnicity	Eine oder mehrere ethnische Gruppen der Patient:in
age	Alter in Jahren als Bruchzahlen
insurance	Versicherung
diagnosis_at_admission	Erste Diagnose zur Aufnahme im Krankenhaus
los_icu	Länge des Intensivaufenthalts (Anzahl der Tage als Bruchzahl)
admission_type	Aufnahmetyp
first_careunit	Erste Intensivstation
hospital_expire_flag	Patient gestorben (expired) oder nicht gestorben (not expired)

Tabelle 4.1: Spalten der Tabelle im Explore-Tab als Key in der MIMIC-Extract Datei mit den jeweiligen möglichen Werten

MIMIC-III Data Visualization Tool

INTRO EXPLORE COMPARE ANALYZE

**Cohort 1 Characteristics**

First ICU: CCU CSRU MICU SICU TSICU

Age: 5 — 80

Gender: FEMALE MALE OTHER

APPLY SELECTION

**Exploration**

In your selected cohort are 515 unique patients.

Subject ID	Gender	Ethnicity	Age	Insurance	Diagnosis at admission	Length of ICU stay (days)	Admission type	First ICU (all except NWARD)	Patient died
<input type="checkbox"/> 46	M	WHITE	74	Medicare	GASTROINTESTINAL BLEED	0.77	EMERGENCY	MICU	Not expired
<input type="checkbox"/> 68	F	BLACK/AFRICAN AMERICAN	41.8	Medicare	PNEUMONIA	3.54	EMERGENCY	MICU	Not expired
<input type="checkbox"/> 94	M	ASIAN	74.43	Medicare	SEPSIS	1.1	EMERGENCY	CCU	Not expired
<input checked="" type="checkbox"/> 103	F	UNKNOWN/NOT SPECIFIED	60.33	Private	SUBARACHNOID HEMORRHAGE	3.02	EMERGENCY	TSICU	Not expired
<input type="checkbox"/> 112	M	WHITE	89	Medicare	GASTROINTESTINAL BLEED	0.84	EMERGENCY	MICU	Not expired
<input type="checkbox"/> 133	M	WHITE	77.33	Medicare	CORONARY ARTERY DISEASE	5.99	EMERGENCY	CSRU	Not expired
<input type="checkbox"/> 134	M	UNKNOWN/NOT SPECIFIED	80.83	Medicare	CORONARY ARTERY DISEASE	1.83	ELECTIVE	CSRU	Not expired
<input type="checkbox"/> 137	M	WHITE	74.72	Private	CORONARY ARTERY DISEASE	5.13	EMERGENCY	CSRU	Not expired
<input type="checkbox"/> 140	M	WHITE	53.52	Private	UPPER GI BLEED	2.15	EMERGENCY	MICU	Not expired
<input type="checkbox"/> 144	M	WHITE	54.27	Medicaid	SUBDURAL HEMATOMA	5.84	EMERGENCY	TSICU	Not expired

Patients per page: 10 11-20 of 9032 |< < > >|

Powered by itame © 2025 Kitware Inc.

Abbildung 4.2: Zweite Seite des Tools zum Veranschaulichen der Subjekte.

## 4 Ergebnisse

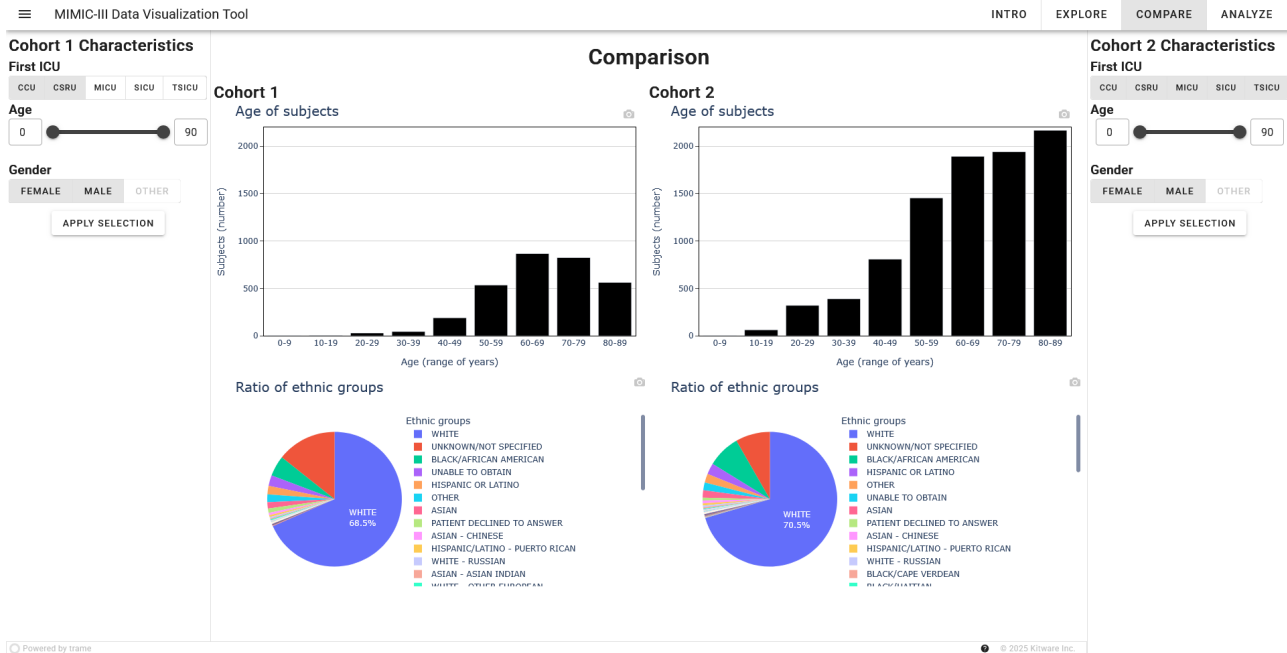


Abbildung 4.3: Dritte Seite des Tools zum Vergleichen von Kohorten.

### 4.2.5 Vergleichen

Auf der dritten Seite „Compare“ (siehe Abbildung 4.3) können zwei Kohorten miteinander verglichen werden. Dazu wird auf der rechten Seite eine neue Seitenleiste eingeführt, in der Bedingungen für die zweite Kohorte gesetzt werden können und die gleich aufgebaut ist wie die linke Seitenleiste. Der Hauptbereich wird hier grundsätzlich in zwei Teile aufgliedert. Auf der linken Seite befinden sich die Einstellungen und Darstellungen zur Kohorte 1 und auf der rechten Seite zur Kohorte 2, der Vergleichskohorte.

Vorgefertigt gibt es hier zwei Diagramme. Das erste Diagramm ist ein Balkendiagramm für das Alter mit einem Balken für jeden 10er-Schritt. Die y-Skalen der Diagramme werden hier vereinheitlicht. Beim Halten des Mauszeigers über einen Balken wird die genaue Anzahl der Subjekte angezeigt. Weiterhin ist es möglich, mit der Maus einen Bereich zu markieren, der dann vergrößert wird (siehe Abbildung 4.4).

Das zweite Diagramm ist ein Tortendiagramm, bei dem das Verhältnis der ethnischen Gruppen von den jeweiligen Kohorten mithilfe verschiedener Farben angegeben wird. Ist eine Sektion zu klein, um die Ethnie oder den prozentualen Anteil anzuzeigen, kann mit dem Mauszeiger drübergehalten werden, um sie mit einem kleinen Fenster wieder sichtbar zu machen (siehe Abbildung 4.5). Dadurch kann auch die zugehörige genaue Anzahl sichtbar gemacht werden.

Die Diagramme können hier und auf der nächsten Seite auch direkt mithilfe der Plotly-Funktion als Bilder exportiert werden. Standardmäßig ist für den Export aufgrund der Beliebtheit das Dateiformat PNG (Portable Network Graphics) eingestellt. Andere verfügbare Formate hierfür sind JPEG, WebP, SVG, PDF und EPS, wodurch nur der Parameter im Code verändert werden muss. [22]

## 4 Ergebnisse

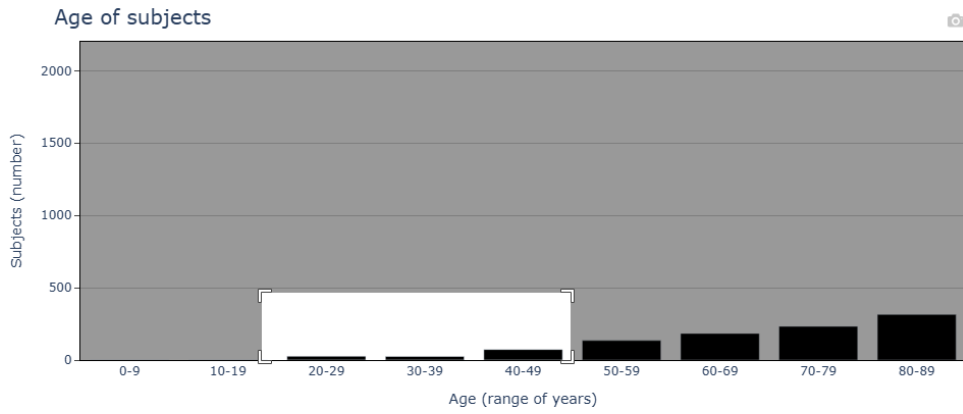


Abbildung 4.4: Beispiel der Zoom-Funktion von Plotly-Diagrammen

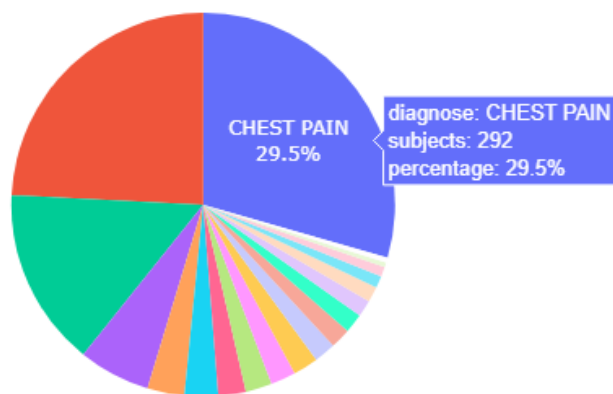


Abbildung 4.5: Beispiel der Hover-Funktion von Plotly-Diagrammen

### 4.2.6 Analysieren

Auf der vierten Seite „Analyze“ (siehe Abbildung 4.6) kann nach dem Auswählen einer Kohorte ein genauerer Blick auf diese geworfen werden. Es werden verschiedene vorgefertigte Diagramme bereitgestellt, die auch jeweils als Bilder heruntergeladen werden können. Ein Beispiel hierfür ist die Abbildung 4.7. Standardmäßig sind auf dieser Seite vier Diagramme sichtbar, wobei es von zwei Diagrammen jeweils drei weitere Versionen gibt.

Das erste Diagramm ist ein Streudiagramm, welches die Länge des ICU-Aufenthalts in Relation zu dem Alter darstellt. Jedes einzelne Subjekt wird hier als Punkt repräsentiert. Als dritte Dimension wird hier zusätzlich angegeben, ob die Person gestorben ist. Das Diagramm bietet die Funktion an, einen bestimmten Bereich zu vergrößern. Beim Halten des Mauszeigers über einzelne Punkte werden die drei Größen des Diagramms noch einmal angezeigt. Zusätzlich wird in dem Diagramm eine LOWESS-Trendlinie (Locally Weighted Scatterplot Smoothing) eingezeichnet, um eine mögliche Abhängigkeit der Variablen beider Achsen zu prüfen und zwei Gruppen der dritten Dimension miteinander vergleichen zu können.

In dem Diagramm oben rechts wird als Tortendiagramm das Verhältnis angegeben, welche ersten Diagnosen den Patient:innen gegeben wurden.

Im unteren Bereich der Seite befinden sich zwei Boxplot-Diagramme. Das linke Diagramm stellt das Geschlecht und das rechte Diagramm das Alter in Abhängigkeit zu einer bestimmten, selbst wählbaren Größe. Die selbst wählbare Variable für die y-Achse umfasst den Hämoglobin-Wert, den Magnesium-Wert, den Wert des Blutdrucks und den Wert der Sauerstoffsättigung. Alle Werte sind Durchschnittswerte von den gemessenen Werten über den Aufenthalt hinweg.

Im Diagramm für die Geschlechter werden für jedes angewählte Geschlecht Boxplots erstellt. Diese stellen von oben nach unten obere Ausreißer, das Maximum, das obere Quantil, den Median, das untere Quantil, das Minimum und untere Ausreißer dar. Beim Halten des Mauszeigers über einen Boxplot werden die genauen Zahlen zu den Maßen angezeigt.

Das Diagramm auf der rechten Seite bietet die gleiche Funktionalität. Das Alter ist hier nach gerundeten Jahreszahlen gruppiert.

## 4.3 Benchmark

Die fertige Website wurde dem Lighthouse-Benchmark von Google Chrome unterzogen. Der Test wurde im standardmäßig ausgewählten „Navigation“-Modus für Desktop-Geräte und mit den Kategorien Performance und Accessibility ausgeführt. Das Tool erzielte in diesen Kategorien eine Wertung von 38 für Performance und 73 für Accessibility von jeweils 100 Punkten. In den jeweiligen Kategorien steht 0 bis 49 für „Schlecht“, 50 bis 89 für „Verbesserungswürdig“ und 90 bis 100 für „Gut“. Als Begründung für den schlechten Performance-Wert nennt der Lighthouse Report eine lange Zeit (6,7 Sekunden) bis der größte Text oder das größte Bild „gemalt“ wird, eine lange Dauer, bei der die Eingabe des Nutzenden blockiert war (0,9 Sekunden) und eine langsame Geschwindigkeit, wie schnell beim Seitenaufbau Inhalte angezeigt werden (4,2 Sekunden). Die Zeit bis zum Anzeigen des ersten Texts oder Bilds (0,5 Sekunden) und die Bewegung von visuellen Elementen (0) wurde hingegen als gut bewertet. [23] Bei der Accessibility wurde unter anderem bemängelt, dass die Buttons keinen zugänglichen Namen haben und die Formularelemente keine zugehörigen Beschriftungen.

## 4 Ergebnisse

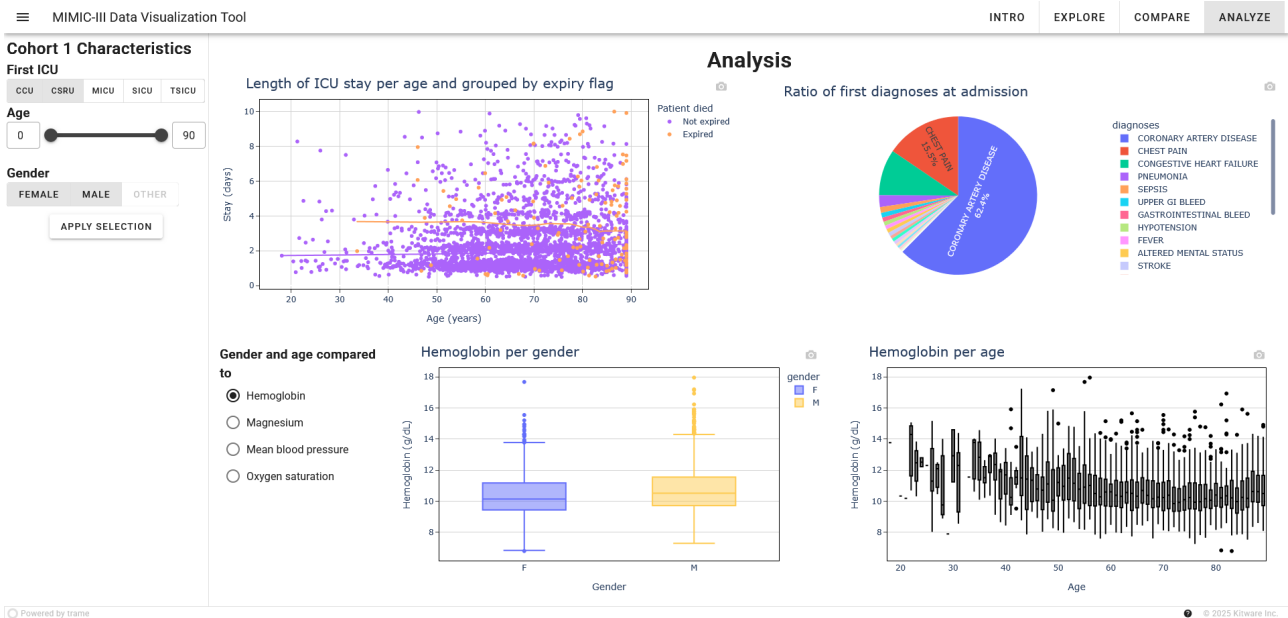


Abbildung 4.6: Vierte Seite des Tools zum genaueren Analysieren einer spezifischen Kohorte.

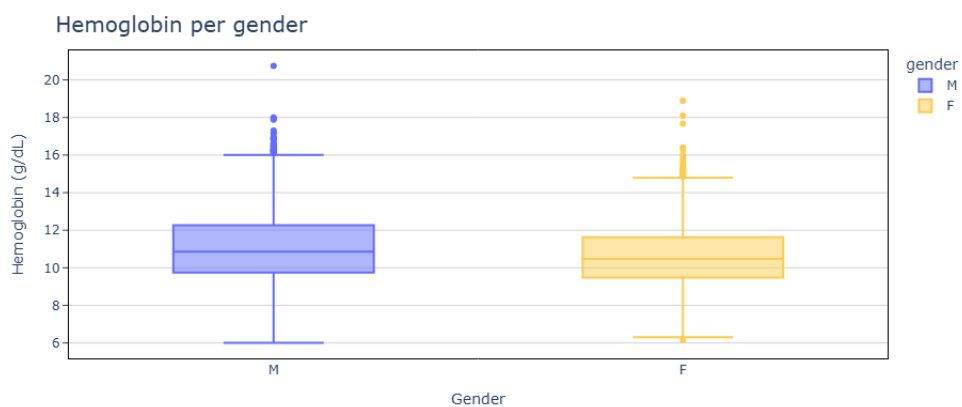


Abbildung 4.7: Boxplot mit dem durchschnittlich gemessenen Hämoglobin-Wert einer Person verglichen mit dem Geschlecht

## 5 Diskussion

Für all die zuvor genannten Probleme bietet das hier entwickelte Trame-basierte Tool eine Lösung. Das Tool ist open-source und kann daher einfach angepasst werden und steht kostenlos zur Verfügung. Es benötigt keine richtige Installation. Es muss nur das Projekt geclont [24] und die Packages mit einem Befehl über die „requirements.txt“ heruntergeladen werden. Das Main-Skript besteht auch grundsätzlich nur aus einer Datei, bei der einfach neue Darstellungen hinzugefügt werden können oder auch eine andere Datenbank benutzt werden kann, an die das Skript dann noch angepasst werden muss. Dies senkt einerseits die Einstiegskosten und andererseits auch die Geschwindigkeit, mit der das Tool eingesetzt werden kann.

Sobald die Installation abgeschlossen ist, ist es auch einfach, weitere Diagramme hinzuzufügen. Dies erfordert lediglich ein wenig Erfahrung mit Plotly.

Sollte das Aufsetzen der Python-Umgebung für manche Menschen noch zu kompliziert sein, kann es auch durch andere als Docker aufgesetzt und dann an diese Personen verteilt werden. [25] Da hier dann nur noch das Aufrufen einer URL notwendig ist, gibt es somit keine nennenswerte Einstiegskosten mehr. Theoretisch ist auch das Aufrufen der Website mit einem Handy oder Tablet möglich, allerdings ist die Webapp noch nicht an diese Geräte angepasst und bedarf noch weiterer Entwicklung.

### 5.1 Performance

Die Verarbeitung der Datenbank basiert hauptsächlich auf Pandas [17] und seiner DataFrame-Klasse. Dadurch ergibt sich auf der Serverseite eine schnelle Laufzeit. Insgesamt benötigt das Rendern der Website aber etwas Zeit. Dies schlug sich auch in dem Benchmark nieder. Die von Lighthouse berichteten Probleme beim Accessibility-Wert könnten in zukünftigen Versionen noch angepasst werden. Der Performance-Score wird derzeit noch stark von der Vuetify-Tabelle beeinflusst, die die Ladezeiten erheblich bremst. Auch dort gäbe es von Seiten von Trame Möglichkeiten, diese gegen eine performantere Tabelle auszutauschen. Vor allem der Performance-Wert ist grundsätzlich schlecht gewesen, allerdings wurde das Programm auch nur auf einem Testsystem getestet, weswegen hier noch weitere Tests nötig wären für eine allgemeingültige Aussage.

### 5.2 Übersichtlichkeit

Die Informationen zu der Datenbank sind stark verstreut in dem Paper, der Dokumentation und der Datenbank selbst. Wenn man MIMIC-Extract benutzt, dann gibt es dazu auch noch neue Informationen über die HDF-Datei, allerdings muss man für viele Informationen dann wieder auf das Original zurückgreifen. Durch die Erklärungen und das organisierte Darstellen in der Web-App können bereits viele Fragen vorab beantwortet werden, ohne dass sie auf verschiedenen Websites zusammengesucht werden müssen. Das Dashboard ermöglicht es, alle Daten gesammelt anzuzeigen. Beim Verwenden

von MIMIC-III über die CSV-Dateien ist dies umständlicher, da hier die richtigen Definitionen oder Erklärungen aus verteilten Websites oder Dokumenten herausgesucht werden müssen. [9][4]

Dank dem Vuetify Trame-Widget hat das Tool ein neuartiges Material-Design, das die Übersichtlichkeit verbessert und auch bei weiteren Erweiterungen garantiert. Die einzelnen Komponenten sind sehr reaktiv. Vuetify passt sich an den Zoom an und vergrößert die Komponenten mit. Das ermöglicht es Menschen, die nicht gut sehen können, die Seite einfach an sich anzupassen. Plotly unterstützt ebenso die Möglichkeit zu zoomen.

### 5.3 Verständliche Diagramme

Ein Hauptaspekt der Verständlichkeit des Tools sind die Diagramme. Diese sind in anderen Arbeiten oft schwierig zu lesen oder schlecht umgesetzt. Die Verwendung der Kriterien von Rougier, Droettboom und Bourne [20] wird anhand der Abbildung 4.7 veranschaulicht.

#### **Enthält die Grafik alle wichtigen Informationen für ein breiteres Publikum?**

Das Diagramm erfüllt die Grundvoraussetzungen eines gewöhnlichen Diagramms. Die Abbildung kann also einfach exportiert werden und für eine wissenschaftliche Arbeit verwendet werden, ohne Informationen hinzufügen zu müssen. Als Abbildung in einer Arbeit würde hier allerdings noch die Bildunterschrift fehlen. Diese ist sehr kontextabhängig und könnte ohne den Einsatz von KI nicht automatisch generiert werden.

#### **Hat die Abbildung eine Botschaft? Ist sie schnell wahrnehmbar? Ist sie einschlägig?**

Grundsätzlich kann nicht garantiert werden, dass die Abbildungen der Nutzenden eine Botschaft haben. Die Beispielgrafik zeigt allerdings, dass die Männer in der Kohorte im Durchschnitt einen höheren Hämoglobin-Wert hatten als die Frauen.

#### **Abbildung an Medium angepasst: Größe entsprechend des Bildschirms und ist die Interaktivität mit der Abbildung ausgeschöpft (Zoom, Drag)? Wenn der Leser mehr Zeit hat (als in einer Präsentation), kann man mehr Details und Erklärungen einbauen.**

Die Größe ist auf der Website an einen typischen Bildschirm angepasst, allerdings ist dies abhängig von den Einstellungen der Nutzenden. Die Plotly-Grafiken sind sehr interaktiv und haben auch eine Zoom- und Drag-Funktion. Da die Nutzenden für das Benutzen des Tools unbegrenzt Zeit haben, könnten auch noch mehr Informationen eingebaut werden. Allerdings widerspricht dies auch anderen minimalistischen Regeln. Drei Dimensionen sollten für ein Diagramm im Normalfall auch ausreichen.

#### **Bildunterschriften sind nicht optional: Wie wird die Grafik gelesen? Was sind Präzisierungen, die nicht visuell wiedergegeben werden können?**

Wie im ersten Punkt beschrieben, ist dieser Punkt für Grafiken auf der Website eigentlich nicht relevant.

### **Nicht blind den Standardwerten vertrauen.**

Die Plotly-Werte wurden sowohl mit einem generellen Konfigurationsobjekt als auch mit individuellen Parametern zu jedem Diagrammtyp verändert.

### **Farbeffizienz: Der Standard sollte schwarz sein. Mit Farben können Akzente gesetzt werden.**

In diesem Fall wurden die Farben bewusst als Akzente genutzt.

### **Objektivität präsentieren: Kolleg:innen fragen, was sie von der Grafik denken. Verhältnisse proportional richtig darstellen. Den einfachsten Typen verwenden. Labels, Ticks, Titel und den vollen Wertebereich wiedergeben.**

Die y-Achse hat die Werte auf derselben Höhe wie das Diagramm rechts daneben, um eine Verfälschung der Wahrnehmung zu vermeiden. Es wurden sowohl Beschriftungen, Ticks als auch ein Titel verwendet. Der volle Wertebereich wurde hier bewusst vermieden, da die Grafik ansonsten kleiner geworden wäre und damit die Lesbarkeit gelitten hätte.

### **Dekorationen, die dem Lesenden nichts sagen, sollten entfernt werden.**

Das Diagramm verfügt über keine Dekorationen, außer der Farben.

### **Botschaft und Lesbarkeit sind wichtiger als optionale Schönheitsmerkmale.**

Die Botschaft und Lesbarkeit standen hier im Vordergrund.

### **Das richtige Tool verwenden.**

Plotly ist ein besonders fortgeschrittenes und weit verbreitetes Tool und sollte damit in die Kategorie der richtigen Tools fallen. Ein Indikator dafür sind die hohen Interaktionszahlen auf dem GitHub-Projekt von Plotly. [26]

### **Sind die Plots barrierefrei?**

Die Farbtöne wurden bewusst gewählt, da sie sich gut eignen, um Menschen mit verschiedenen Farbblindheiten nicht zu benachteiligen. [27]

## 5.4 Beitrag

Das Tool adressiert einige Punkte der Arbeit von Lee, Ribey und Wallace [5]. In ihrer Arbeit haben die Autor:innen angemerkt, dass die zukünftige Forschung ein „Analyze“-Feature hinzufügen könnte, welches in Python implementiert werden sollte. Das jetzige Tool unterstützt noch nicht die erwähnten Tests oder die Entwicklung von Machine-Learning-Algorithmen, hat aber trotzdem einen ersten Grundrahmen für diese Funktionen geschaffen.

Weiterhin erwähnen sie die Möglichkeit einer Exportfunktion, um Abbildungen oder Rohdaten zu exportieren. Durch die Speicherfunktion von Plotly wurde dies zu einem Teil erfolgreich umgesetzt. Zu guter Letzt wurde erwähnt, dass das Tool auf die neu erschienene MIMIC-III-Version verbessert werden sollte. Auch dies wurde hiermit umgesetzt.

### 5.5 Limitationen

Wegen der zuvor genannten Probleme stellt die MIMIC-Extract-Datenbank keine optimale Lösung für das Programm dar und sollte ausgetauscht werden gegen eine Alternative. Die Schnittstelle zur HDF5-Datei ist allerdings sehr starr und würde ohne größere Anpassungen keinen Austausch der Datenbank vertragen.

Wie durch die Benchmarks sichtbar geworden ist, skaliert das Programm nicht sehr gut mit großen Tabellen. Die Ladezeiten werden dadurch zu lang, um benutzerfreundlich zu bleiben. Für den Prototypen wurde auch nur eine reduzierte Version der Datensätze benutzt. Allerdings gibt es auch hier bereits Lösungen von Trame, die sich noch in der Vorabversion befinden, um bei Tabellen nur die wirklich angezeigten Zeilen zu rendern. [28]

Die Graphen in dem Tool sind derzeit bis auf einige Ausnahmen noch wenig modular und können nur mit Python- und Plotly-Kenntnissen erweitert werden.

Die Datenbank muss mit einem mehrstündigen Online-Kurs freigeschaltet werden, für den man sich auch erst bewerben muss. Dies stellt einen erhöhten Aufwand dar und heißt, dass das Tool zumindest für einen Großteil der Bevölkerung keinen Nutzen darstellt.

### 5.6 Zukünftige Arbeiten

Eine mögliche Lösung wäre, eine eigene Verarbeitungspipeline für die MIMIC-Datenbank zu bauen und somit von MIMIC-Extract unabhängig zu machen. Ansonsten kann das Tool auch auf die MIMIC-IV-Version erweitert werden.

Offen bleiben die beschriebenen Funktionen aus der Arbeit von Lee, Ribey und Wallace [5], die „Analyse“-Seite mit der Funktion zu erweitern, einfache statistische Tests durchzuführen oder Machine-Learning-Algorithmen zu entwickeln.

Mit dem Framework ist es einfach, noch weitere Analysetools sowie weitere Graphen zu implementieren. Es können auch noch weitere Darstellungen eingebaut werden, wie zum Beispiel 3D-Ansichten des menschlichen Körpers, die verdeutlichen, auf welche Körperteile sich die jeweiligen Diagnosen beziehen. [7] Dies würde das Verständnis der Datenbank auch für Personen weiter vereinfachen, die ihre Profession nicht im Medizinbereich haben.

Zukünftige Arbeiten könnten sich darauf fokussieren, die Ansicht auf Mobilgeräten zu verbessern, um den Umgang noch weiter zu vereinfachen.

Meine Arbeit bezog sich im Großen auf die Implementierung einer Lösung. In einer Folgearbeit könnte die Validierung des Tools in seinem Kontext überprüft werden und inwiefern es den Anforderungen der Verständlichkeit und Analyse gerecht wird. Hier kann im Speziellen betrachtet werden, ob die Grafiken von Personen als verständlich bewertet werden.

## 6 Fazit

Ziel der Arbeit war es, ein Trame-basiertes Tool zur verständlichen Darstellung und Analyse klinischer Daten zu entwickeln.

Im Rahmen dessen ist ein umfangreiches Tool entstanden, welches die Funktionen des Tools von Lee, Ribey und Wallace [5] sogar erweitert. Im Vergleich zum Vorgänger meines Tools wurde die neuere Version von MIMIC verwendet und die Möglichkeit, Grafiken zu exportieren, hinzugefügt. Die Funktionen auf der Datenbank umfassen die Selektion von Kohorten, die Möglichkeit des Vergleichs der Kohorten sowie das genauere Betrachten anhand einer Tabelle und das Erstellen von Diagrammen basierend auf den Daten.

Dieses Tool steht allen kostenfrei zur Verfügung. Es kann mit minimalem Aufwand eingesetzt werden. Daher ist es nun Personen, die an MIMIC-III forschen, möglich, noch einfacher in die Materie einzusteigen.

Ich erhoffe mir, dass zukünftige Arbeiten ebenfalls mein Tool aufgreifen und es für MIMIC-IV oder auch andere klinische Datenbanken weiterentwickeln.

## 7 KI-Tools

### 7.1 Perplexity von Perplexity AI, Inc.

Ich habe Perplexity [29] zur Quellenrecherche benutzt. Alle KI-Vorschläge wurden sorgfältig von mir geprüft. Ich habe die Vorschläge lediglich als Anhaltspunkte für meine Recherche benutzt und keine Inhalte der KI in meine Arbeit übertragen. Standardmäßig ist Perplexity auf das Modell „Best“ eingestellt und sagt, dass es damit das beste Modell für die Anfrage auswählen würde. Die verfügbaren Modelle sind *Sonar*, *Claude 3.7 Sonnet*, *GPT-4.1*, *Gemini 2.5 Pro*, *Grok 3 Beta*, *R1 1776*, *o4-mini* und *Claude 3.7 Sonnet Thinking*.

### 7.2 Code Completion von JetBrains s.r.o.

Ich habe JetBrains' Code Completion [30] (Modell: Python) als Teil der PyCharm-IDE (Version: 2024.1.7) zur Programmierunterstützung genutzt, aber damit keine vollständigen Funktionen erzeugt. Alle KI-Vorschläge wurden sorgfältig von mir geprüft und nur akzeptiert, wenn sie meinen Absichten entsprachen.

### 7.3 Writefull von ThingLab Ltd

Ich habe Writefull [31] (Modell: GPT) als Teil der Overleaf-Toolbar (Version: 2025.31.0) zur Korrektur der Orthografie meiner Bachelorarbeit benutzt. Alle KI-Vorschläge wurden sorgfältig von mir geprüft und nur akzeptiert, wenn sie meinen Absichten entsprachen.

## Literatur

- [1] X. Han. „Is more data always better?“ University of Groningen. (27. März 2024), Adresse: <https://www.rug.nl/news/2024/04/is-more-data-always-better> (besucht am 30.04.2025).
- [2] M. Chen. „Big Data, große Möglichkeiten: So extrahieren Sie den maximalen Wert,“ Oracle. (23. Sep. 2024), Adresse: <https://www.oracle.com/de/big-data/what-is-big-data/> (besucht am 30.04.2025).
- [3] D. Chrimes und C. Kim, „Comparison of MIMIC-III and MIMIC-IV for big data analytics of health informatics,“ in *2023 IEEE International Conference on Big Data (BigData)*, Dez. 2023, S. 6128–6130. DOI: [10.1109/BigData59044.2023.10386585](https://doi.org/10.1109/BigData59044.2023.10386585). Adresse: <https://ieeexplore.ieee.org/document/10386585> (besucht am 30.04.2025).
- [4] „Article metrics - MIMIC-III, a freely accessible critical care database | scientific data,“ Adresse: <https://www.nature.com/articles/sdata201635/metrics> (besucht am 30.04.2025).
- [5] J. Lee, E. Ribey und J. R. Wallace, „A web-based data visualization tool for the MIMIC-II database,“ *BMC Medical Informatics and Decision Making*, Jg. 16, Nr. 1, S. 15, 4. Feb. 2016, ISSN: 1472-6947. DOI: [10.1186/s12911-016-0256-9](https://doi.org/10.1186/s12911-016-0256-9).
- [6] S. Jourdain, P. O’Leary und W. Schroeder, „Trame: Platform Ubiquitous, Scalable Integration Framework for Visual Analytics,“ *IEEE Computer Graphics and Applications*, März 2025. DOI: [10.1109/MCG.2025.3540264](https://doi.org/10.1109/MCG.2025.3540264).
- [7] „Trame: Revolutionizing visual analytics with python.“ (), Adresse: <https://www.kitware.com/trame-revolutionizing-visual-analytics-with-python/> (besucht am 28.04.2025).
- [8] M. A. Kabir und M. R. Ahmed, *Python for data analytics: A systematic literature review of tools, techniques, and applications*, Rochester, NY, 13. Nov. 2024. DOI: [10.69593/ajsteme.v4i04.146](https://doi.org/10.69593/ajsteme.v4i04.146). (besucht am 30.04.2025).
- [9] A. E. W. Johnson, T. J. Pollard, L. Shen u. a., „MIMIC-III, a freely accessible critical care database,“ *Scientific Data*, Jg. 3, Nr. 1, S. 160035, 24. Mai 2016, Publisher: Nature Publishing Group, ISSN: 2052-4463. DOI: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35).
- [10] S. Wang, M. B. A. McDermott, G. Chauhan, M. C. Hughes, T. Naumann und M. Ghassemi, „MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III,“ in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2. Apr. 2020, S. 222–235. DOI: [10.1145/3368555.3384469](https://doi.org/10.1145/3368555.3384469).
- [11] Ariel. „Answer to Load HDF5 in Excel?“ Stack Overflow. (18. Apr. 2022), Adresse: <https://stackoverflow.com/a/71914097/22103145> (besucht am 29.04.2025).
- [12] „HDF view,“ The HDF Group. (), Adresse: <https://www.hdfgroup.org/download-hdf-view/> (besucht am 24.04.2025).

## Literatur

- [13] *MIT-LCP/mimic-iii-paper*, 27. Apr. 2025. Adresse: <https://github.com/MIT-LCP/mimic-iii-paper> (besucht am 28. 04. 2025).
- [14] *MLforHealth/MIMIC\_Extract*, 29. Apr. 2025. Adresse: [https://github.com/MLforHealth/MIMIC\\_Extract](https://github.com/MLforHealth/MIMIC_Extract) (besucht am 30. 04. 2025).
- [15] „Project jupyter,“ Jupyter. (), Adresse: <https://jupyter.org> (besucht am 28. 04. 2025).
- [16] „PyCharm: The only Python IDE you need,“ JetBrains. (), Adresse: <https://www.jetbrains.com/de-de/pycharm/> (besucht am 30. 04. 2025).
- [17] T. p. d. team, *pandas-dev/pandas: Pandas*, Version v2.2.3, 20. Sep. 2024. DOI: [10.5281/zenodo.13819579](https://doi.org/10.5281/zenodo.13819579).
- [18] „Vuetify — a vue component framework,“ Vuetify. (), Adresse: <https://vuetifyjs.com/en/> (besucht am 30. 04. 2025).
- [19] S. Jourdain, „Trame intro.“ Adresse: [https://github.com/Kitware/trame-tutorial/raw/master/course/introduction/trame\\_intro\\_ref.pdf](https://github.com/Kitware/trame-tutorial/raw/master/course/introduction/trame_intro_ref.pdf) (besucht am 29. 04. 2025).
- [20] N. P. Rougier, M. Droettboom und P. E. Bourne, „Ten simple rules for better figures,“ *PLOS Computational Biology*, Jg. 10, Nr. 9, 11. Sep. 2014, Publisher: Public Library of Science, ISSN: 1553-7358. DOI: [10.1371/journal.pcbi.1003833](https://doi.org/10.1371/journal.pcbi.1003833).
- [21] „Lighthouse - chrome web store.“ (23. Apr. 2024), Adresse: <https://chromewebstore.google.com/detail/lighthouse/blipmdconlkpinefehnjammfjpmbjk> (besucht am 29. 04. 2025).
- [22] „Static image export in Python.“ (), Adresse: <https://plotly.com/python/static-image-export/> (besucht am 29. 04. 2025).
- [23] „Lighthouse-Leistungsbewertung,“ Chrome for Developers. (), Adresse: <https://developer.chrome.com/docs/lighthouse/performance/performance-scoring?hl=de> (besucht am 29. 04. 2025).
- [24] M. zu Ummeln. „Richard meier zu ummeln / bachelorarbeit · GitLab,“ GitLab. (29. Apr. 2025), Adresse: <https://git.tu-berlin.de/rimzu/bachelorarbeit> (besucht am 30. 04. 2025).
- [25] „Docker | trame.“ (), Adresse: <https://kitware.github.io/trame/guide/deployment/docker.html> (besucht am 28. 04. 2025).
- [26] „Plotly,“ GitHub. (), Adresse: <https://github.com/plotly> (besucht am 30. 04. 2025).
- [27] Extensis. „So gestalten Sie Designs für Farbenblinde.“ (), Adresse: <https://www.extensis.com/de-de/blog/so-gestalten-sie-designs-f%C3%BCr-farbenblinde> (besucht am 29. 04. 2025).
- [28] *Kitware/trame-datagrid*, 15. Apr. 2025. Adresse: <https://github.com/Kitware/trame-datagrid> (besucht am 28. 04. 2025).
- [29] „Perplexity,“ Perplexity AI. (), Adresse: <https://www.perplexity.ai> (besucht am 28. 04. 2025).
- [30] „Code completion | PyCharm,“ PyCharm Help. (), Adresse: <https://www.jetbrains.com/help/pycharm/auto-completing-code.html> (besucht am 24. 04. 2025).

## *Literatur*

- [31] „Writefull for Overleaf — Writefull.“ (), Adresse: <https://www.writefull.com/writefull-for-overleaf> (besucht am 24. 04. 2025).