

Business Intelligence (WIB5/WNB5)

Projektdokumentation (2A-GR11) **Sommersemester 2020**

an der Fakultät Wirtschaftsinformatik
der Hochschule Furtwangen

vorgelegt von

Adrian Coman, Eric Sartorius

Dozentin:

Prof. Dr. Monika Frey-Luxemburger

Eingereicht am

7. August 2020

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS.....	III
1 AUFGABENSTELLUNG UND ZIEL DES PROJEKTES	4
2 BUSINESS CASE.....	6
2.1 Motivation und Ziele	6
2.2 Herausforderungen, Maßnahmen und Kennzahl(en).....	6
2.3 Datenquelle.....	7
3 SOLLKONZEPT	8
3.1 Beschreibung	8
3.2 Anwendungsarchitektur	8
3.3 Kennzahl(en)-Definition.....	9
3.4 Datenmodell.....	9
3.5 Prototyp	10
3.5.1 Datenbank	10
3.5.2 Excel-Tabellen.....	11
3.5.3 Visualisierung in Excel	12
4 BI ANWENDUNG	14
4.1 Dashboard und Datenanalyse.....	14
4.2 Datenbereitstellung und ETL-Prozess	16
5 RESÜMEE UND AUSBLICK	18
LITERATURVERZEICHNIS	XIX

Abbildungsverzeichnis

Abbildung 1 Projektplan	5
Abbildung 2 Anwendungsarchitektur.....	9
Abbildung 3 Kennzahldefinition	9
Abbildung 4 Flache Tabelle.....	9
Abbildung 5 Snowflake Schema	10
Abbildung 6 Datenbanktabellen (ERM).....	11
Abbildung 7 Excell-Tabelle.....	12
Abbildung 8 Prototyp des Dashboards in Excel (1).....	13
Abbildung 9 Prototyp des Dashboards in Excel (2).....	13
Abbildung 10 Dashboard im Überblick	15
Abbildung 11 Datenanalyse unter Einsatz des Dashboards	15
Abbildung 12 Datenbereitstellung und ETL-Prozess.....	16

1 Aufgabenstellung und Ziel des Projektes

Im Rahmen des Studiengangs „Wirtschaftsinformatik“ soll in dem Kurs „Business Intelligence“ eine Prüfungsleistung in Form eines Prüfungsprojekt durchgeführt werden, welches die korrekte Durchführung eines BI-Projekts mit einem BI-Tool prüfen soll. In dem Projekt soll Ein Business Case selbst gefunden werden. Der Business Case soll anschließend auf Eignung für das Projekt von Frau Prof. Frey-Luxemburger geprüft werden. Folgende Quellen durften dafür angewendet werden:

- http://www.databaseanswers.org/data_models/index.htm
- <https://www.kaggle.com/>
- https://www.destatis.de/DE/Themen/Branchen-Unternehmen/_inhalt.html

Das BI-Tool war dabei von den Studierenden frei wählbar.

Für unsere Gruppe hielten wir die Business Intelligence Roadmap (Moss/Atre 2003) für sinnvoll, um einen besseren Überblick über das andauernde Projekt zu erhalten. Dieses wurde bei uns in folgenden fünf Bereichen eingeteilt:

1. Vorbereitung

Im Bereich Vorbereitung wurden alle für das Projekt relevanten Aufgaben abgearbeitet, wie z.B. Ziele setzen.

2. Initialisierung

In dem Bereich Initialisierung wurden die nötigen Zusammenhänge ermittelt.

3. Business Analyse

Im Bereich Business Analyse wurden alle Daten Aufbereitet und eine ordentliche Datenbasis geschaffen.

4. Sollkonzept

Im Sollkonzept wurde ein Prototyp erstellt und mit dessen Hilfe Die nachfolgenden Prozesse, bis zu Dashboard.

5. Abgabe

Hier wurden die Daten nochmals überprüft und alles zusammengetragen

Nr.	Task	Verantwortlicher	Start	Ende	Zeitaufwand/ Stunde	Status
1.	Vorbereitung					
2.	Ziele setzen	Adrian Coman/Eric Satorius	15.07.2020	15.07.2020	4	100%
3.	Aufgabenstellung Analysieren	Adrian Coman	16.07.2020	17.07.2020	2	100%
4.	Projektplan erstellen	Adrian Coman	17.07.2020	18.07.2020	11	100%
5.	Informationsbeschaffung	Adrian Coman/Eric Satorius	18.07.2020	20.07.2020	18	100%
6.	Initialisierung					
7.	Business Case erstellen	Adrian Coman	20.07.2020	23.07.2020	8	100%
8.	KPI definieren	Adrian Coman	23.07.2020	23.07.2020	4	100%
9.	Tool installieren	Adrian Coman/Eric Satorius	23.07.2020	23.07.2020	0,1	100%
10.	Business Analyse					
11.	Daten analysieren	Adrian Coman	23.07.2020	24.07.2020	2	100%
12.	Datenmodell anfertigen	Eric Satorius	24.07.2020	24.07.2020	4	100%
13.	Datenbank erstellen	Eric Satorius	24.07.2020	25.07.2020	8	100%
14.	Sollkonzept					
15.	Excel-Prototyp erstellen	Eric Satorius	25.07.2020	27.07.2020	6	100%
16.	ETL-Prozess erstellen	Eric Satorius	27.07.2020	30.07.2020	4	100%
17.	Dashboard erstellen	Eric Satorius	30.07.2020	03.07.2020	7	100%
18.	Abgabe					
19.	Prüfung	Adrian Coman/Eric Satorius	03.07.2020	04.07.2020	18	100%
20.	Dokumentation anfertigen	Adrian Coman/Eric Satorius	04.07.2020	07.07.2020	24	100%

Abbildung 1 Projektplan

2 Business Case

Die Dienstaufsichtsbehörde der Vereinigten Staaten „Federal Bureau of Investigation“, kurz auch FBI genannt, führt aufgrund von Beschwerden eine Untersuchung durch, die Eingriffe der Polizei mit Todesfolge im Hinblick auf Rassismus gegenüber schwarzfarbige Menschen Klarheit schaffen soll, ob schwarzfarbige Menschen öfters bei Polizeieinsätzen ums Leben kommen als andersfarbige. Teil dieser Untersuchung ist es alle Fälle von Todesopfern durch die Polizei in den USA zu analysieren und anhand der Verhältnismäßigkeit gegenüber andersfarbige Abweichungen festzustellen. In diesen Fällen werden Indianer, Asiaten, Weise, Latinos und Schwarze unterschieden. Die Gewerkschaft Erziehung und Wirtschaft (GEW) hat es sich zur Aufgabe gemacht gegen Rassismus vor zu gehen und möchte auch weitere Länder unterstützen. Daher beteiligt sich unsere Firma an der US-Amerikanischen Dienstaufsichtsbehördlichen Untersuchung.

2.1 Motivation und Ziele

Im Rahmen des Projekts in Kooperation mit der US-Amerikanischen Dienstaufsichtsbehördliche möchte unsere Firma einen Beitrag dazu leisten, mögliche Einsätze der US-Amerikanischen Polizei mit rassistischen Todesfolgen gegenüber schwarzfarbigen Menschen, mithilfe unserer angeeigneten BI-Kenntnisse aufzudecken. Dies soll über ein Dashboard grafisch mittels eines BI-Tools visualisiert werden und der Dienstaufsichtsbehörde als Hilfestellung angeboten werden. Dies hat auch zur Folge, dass unsere Firma einen Einblick darin gewinnt, ob schwarzfarbige Menschen mehreren rassistischen und tödlichen Übergriffen der US-Amerikanischen Polizei ausgeliefert sind, als andere.

2.2 Herausforderungen, Maßnahmen und Kennzahl(en)

Herausforderung	Maßnahme	Kennzahl	
Aufgrund der zunehmenden Todesfälle bei Polizeieinsätzen in den USA und Berichten aus den Medien, wird das Bild der US-Amerikanischen Polizei zunehmend rassistisch gegenüber schwarzfarbige Menschen dargestellt.	Um den Wahrheitsgehalt der Aussagen zu ermitteln wird eine BI-Analyse durchgeführt, die bestätigen oder widerlegen soll, dass die US-Amerikanische Polizei in Einsätzen mit Todesfolge weniger acht auf schwarzfarbige Menschen nimmt, als auf andersfarbige.	Anzahl schwarzfarbiger Menschen pro Staat	Anzahl andersfarbige Menschen pro Staat

2.3 Datenquelle

Der Datensatz wurde von der Webpage kaffle.com bezogen. Dieser stammt von Andriy Samoshin vom 05.2020. Das FBI protokollierte alle tödlichen Schießungen von Polizisten. Der umfasst 14 Spalten und 5417 Zeilen.

Attribut	Skalenniveau	Ausprägung
id	Nominal	1 - 5927
name	Nominal	Individuelle Personennamen
date	Nominal	Individueller Todeszeitpunkt 02.01.2015 – 16.06.2020
manner_of_death	Nominal	Shot, shot and Tasered
armed	Nominal	Individuelle Bewaffnung Gun, nail gun, shovel, unarmed...
age	Rational	6,12,13,14,15,16,17,18,19,20...
gender	Nominal	M = Man, F = Female
race	Nominal	A = Asian, B = Black, W = White, N = Native American, H = Hispanic
city	Nominal	Individuelle USA Städte
state	Nominal	Individuelle US-Amerikanische Staaten Aloha, Assaria, Evans, Fairfield...
signs_of_mental_illness	Nominal	Wahr, Falsch
threat_level	Nominal	Attack, undermined, other
flee	Nominal	Car, foot, not fleeing, other
Body_camera	Nominal	Wahr, Falsch

3 Sollkonzept

Die zuvor erfolgte IST-Analyse hat Schwachstellen und Verbesserungspotenzial gezeigt. In der Soll-Analyse folgt die Umsetzung der Lösungen zu der Ist-Analyse.

3.1 Beschreibung

Mittels BIA-Tool wollen wir eine Grafische Darstellung abbilden, die es einem ermöglicht auf einem Blick zu sehen ob sich der Rassismus gegenüber schwarzfarbigen Menschen bei Polizeieinsätzen bestätigt oder widerlegt. In diesem Fall entschieden wir uns für das Datenbankmanagementsystem MySQL, mit dessen Hilfe wir die Daten bearbeiten konnten und damit auch den ETL-Prozess umsetzen konnten. Dabei sollte die Kennzahl eine möglichst geringe Toleranz aufweisen. Die Werte sollen in einem Kartendiagramm, einem Säulendiagramm und einem Kreisdiagramm visualisiert werden. Der Prototyp und das Dashboard werden so aufgebaut, dass diese mit minimalen Änderungen auch für Analysen anderer Länder verwendet werden können.

3.2 Anwendungsarchitektur

Die Darstellung (Abbildung 2) zeigt links die allgemeine Archtiektur einer BI-Anwendung. Rechts ist die für das Projekt verwendete Architektur abgebildet. Der von uns verwendete Datensatz wird als .CSV-Datei auf der Webseite Kaggle.com bereitgestellt. Die .CSV-Datei ist über den sogenannten ETL-Prozess in das Data Warehouse geladen worden. Das Data Warehouse ist im Rahmen des Projekts das zentrale Datenbanksystem MySQL, in dem wir die externen Daten aufbereiten und speichern. Da das Data Warehouse Datenbanksystem groß und umfangreich werden kann, wird das Data Warehouse in sogenannte Data Marts aufgeteilt. Diese werden bei unserem Projekt ebenfalls über MySQL aufbewahrt. Die schlussendliche Auswertung der Daten findet über das Frontend Tool Tableau statt.

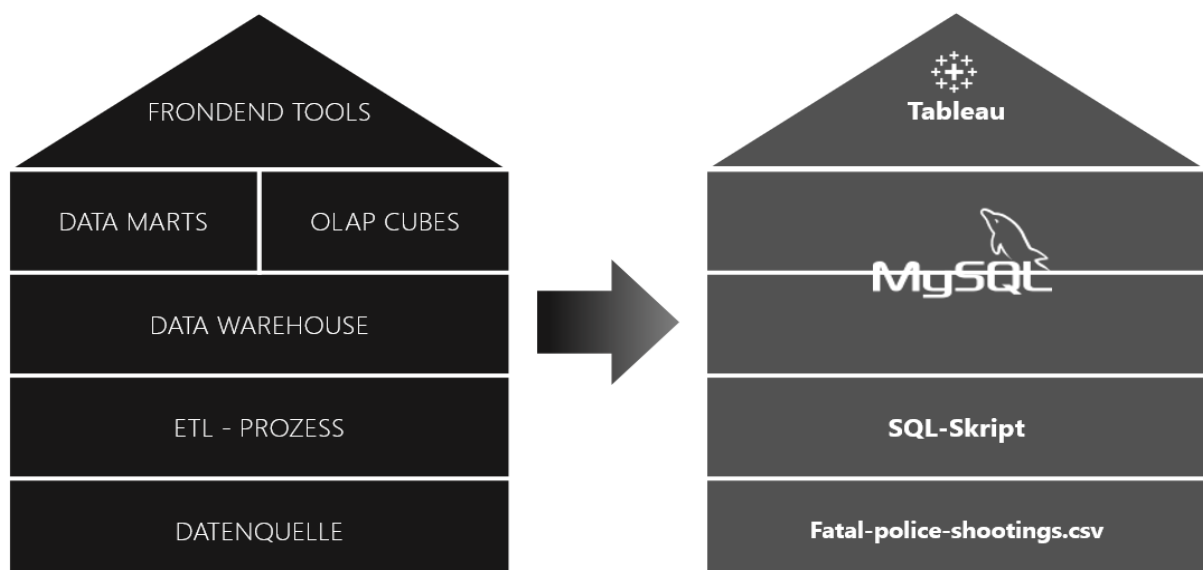


Abbildung 2 Anwendungsarchitektur

3.3 Kennzahl(en)-Definition

Todesfälle durch Polizisten		
Teil I: Beschreibung		
1	Bezeichnung	Klarer und eindeutiger Name (Codierung und/oder verbale Bezeichnung)
2	Beschreibung	Kurze Beschreibung von Zielsetzung und Kontext
3	Adressat	Empfänger und Nutzer der Kennzahl
4	Formel	Mathematische Formel der Kennzahl
5	Zielwert	Wert der Kennzahl, der erreicht werden soll
6	Einheit	Einheit, in der die Kennzahl gemessen wird
7	Periodizität	Angabe, ob es sich um Tages-, Wochen-, Monatswerte etc. handelt
8	Dimensionen	Beschreibende Dimensionen der Kennzahl (z.B. Zeit, Kunde, Region, Produkt...)
9	Aggregationsverhalten	Aggregationsverhalten bei Verdichtung der Daten (Summe, Durchschnitt etc.)
10	Planwerte	Zwischenplanwerte zu verschiedenen Zeitpunkten bis zu Erreichung des Zielwertes (Pkt. 5)
11	Toleranzwert	Erlaubte Abweichung bei Plan- und Zielwert (in Prozent)
12	Eskalationsregel	Festlegung der Reaktion auf Abweichung der Toleranzwerte
13	Gültigkeit/Periodizität	Gültigkeitszeitraum der Kennzahl (ab wann hat diese Definition Gültigkeit und bis wann?)
14	Abhängigkeit	Abhängigkeit der Kennzahl von anderen Kennzahlen oder Grundlage der beschriebenen Kennzahl für andere Kennzahlen
15	Verantwortlicher	Verantwortet die Definition der Kennzahl und Erreichung der Plan- und Sollwerte (ggf. zwei verschiedene Personen)
16	Beispiel	Beschreibendes Beispiel inkl. Zielwert, Abweichung und Interpretation der Zahl
Teil II: Datenermittlung		
1	Datenquellen	Beschreibung der Datenquellen (Berichtssysteme, Datenbestände etc)
2	Messverfahren	Häufigkeit, Aufwand und Kosten der Datenerhebung
3	Messpunkte	Terminierung der Datenerhebung
4	Verantwortlicher	Verantwortet die korrekte und termingerechte Ermittlung

Abbildung 3 Kennzahldefinition

3.4 Datenmodell

Der anfängliche Datenbestand beinhaltet eine flache Tabelle in der ersten Normalform, welche unverändert aus der Datenquelle erzeugt wird.

	id	name	datum	todesursache	bewaffnung	lebensalter	geschlecht	hautfarbe	stadt	staat	anzeichen_fuer_psychische_erkrankung	bedrohung	flucht	koerperkamera
▶	3	Tim Elliot	2015-01-02 00:00:00	shot	gun	53	M	A	Shelton	WA	1	attack	Not fleeing	0
	4	Lewis Lee Lembke	2015-01-02 00:00:00	shot	gun	47	M	W	Aloha	OR	0	attack	Not fleeing	0
	5	John Paul Quintero	2015-01-03 00:00:00	shot and Tasered	unarmed	23	M	H	Wichita	KS	0	other	Not fleeing	0
	8	Matthew Hoffman	2015-01-04 00:00:00	shot	toy weapon	32	M	W	San Francisco	CA	1	attack	Not fleeing	0
	9	Michael Rodriguez	2015-01-04 00:00:00	shot	nail gun	39	M	H	Evans	CO	0	attack	Not fleeing	0
	11	Kenneth Joe Brown	2015-01-04 00:00:00	shot	gun	18	M	W	Guthrie	OK	0	attack	Not fleeing	0
	13	Kenneth Arnold Buck	2015-01-05 00:00:00	shot	gun	22	M	H	Chandler	AZ	0	attack	Car	0
	15	Brook Nichols	2015-01-06 00:00:00	shot	gun	35	M	W	Assaria	KS	0	attack	Not fleeing	0
	16	Autumn Steele	2015-01-06 00:00:00	shot	unarmed	34	F	W	Burlington	IA	0	other	Not fleeing	1
	17	Leslie Sapp III	2015-01-06 00:00:00	shot	toy weapon	47	M	B	Knoxville	PA	0	attack	Not fleeing	0
	19	Patrick Wetter	2015-01-06 00:00:00	shot and Tasered	knife	25	M	W	Stockton	CA	0	attack	Not fleeing	0
	20	Jessie Hernandez	2015-01-26 00:00:00	shot	gun	17	F	H	Denver	CO	0	other	Not fleeing	0
	21	Ron Sneed	2015-01-07 00:00:00	shot	gun	31	M	B	Freeport	TX	0	attack	Not fleeing	0
	22	Hashim Hanif Ibn A...	2015-01-07 00:00:00	shot	knife	41	M	B	Columbus	OH	1	other	Not fleeing	0
	25	Nicholas Ryan Brick...	2015-01-07 00:00:00	shot	gun	30	M	W	Des Moines	IA	0	attack	Car	0
	27	Omarr Julian Maxim...	2015-01-07 00:00:00	shot	gun	37	M	B	New Orleans	LA	0	attack	Foot	1
	29	Loren Simpson	2015-01-08 00:00:00	shot	gun	28	M	W	Huntley	MT	0	undeterm...	Not fleeing	0
	32	James Dudley Barker	2015-01-08 00:00:00	shot	shovel	42	M	W	Salt Lake City	UT	0	attack	Not fleeing	1
	36	Artago Damon How...	2015-01-08 00:00:00	shot	unarmed	36	M	B	Strong	AR	0	attack	Not fleeing	0
	37	Thomas Hamby	2015-01-08 00:00:00	shot	gun	49	M	W	Syracuse	UT	0	attack	Not fleeing	1
	38	Jimmy Foreman	2015-01-09 00:00:00	shot	gun	71	M	W	England	AR	0	attack	Not fleeing	0
	42	Tommy Smith	2015-01-11 00:00:00	shot	gun	39	M	W	Arcola	IL	1	attack	Not fleeing	0
	43	Brian Barbosa	2015-01-11 00:00:00	shot	gun	23	M	H	South Gate	CA	0	attack	Not fleeing	0
	45	Salvador Figueroa	2015-01-11 00:00:00	shot and Tasered	gun	29	M	H	North Las Ve...	NV	0	attack	Foot	0
	46	John Edward O'Keefe	2015-01-13 00:00:00	shot	gun	34	M	W	Albuquerque	NM	0	attack	Foot	1
	48	Richard McClendon	2015-01-13 00:00:00	shot	knife	43	M	W	Jourdanton	TX	1	other	Not fleeing	0

Abbildung 4 Flache Tabelle

Das in Abbildung 5 dargestellte Datenmodell besitzt eine einfache Hierarchie bestehend aus den Dimensionen „dim_staat“ und „dim_stadt“. Dadurch steht die Tabelle „dim_staat“ nicht in direkter Verbindung mit der Faktentabelle. Daher spricht man bei diesem Modell von einem Snowflake-Schema. „Während beim Sternschema die umgebenden Dimensionstabellen alle direkt mit der Faktentabelle verbunden sind, gibt es beim Schneeflockenschema mehrere Ebenen.“ (Rouse, 2016) Relevant für unsere Kennzahlen sind ausschließlich die Dimensionen Personendetails und Staat.

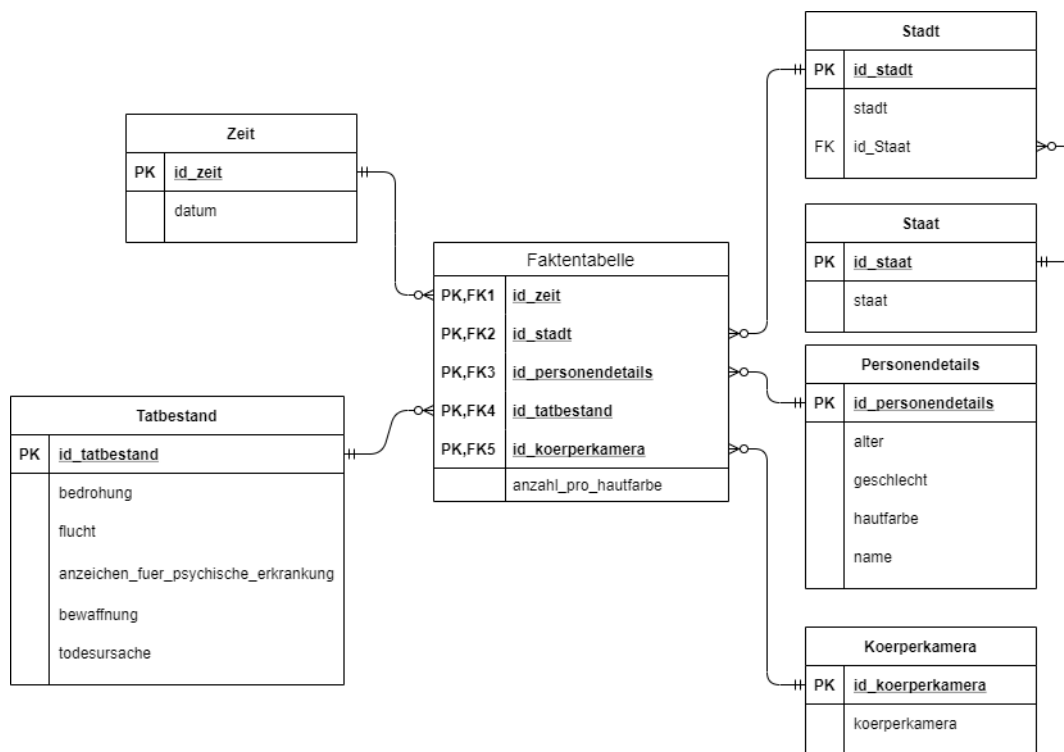


Abbildung 5 Snowflake Schema

3.5 Prototyp

Basierend auf dem Sollkonzept wird ein erster Prototyp erstellt, anhand dessen erste Datenexplorationen ermöglicht werden können. Das Ziel ist es, ein Prototyp zu erstellen, welcher dem künftigen Dashboard nahekommmt.

3.5.1 Datenbank

Als Datenbankmanagement-Tool wurde die MySQL-Workbench ausgewählt. Die Daten der .CSV-Datei werden über SQL-Skripte angepasst und in die Datenbank geladen. Nachdem die Daten den ETL-Prozess durchlaufen haben, kann mittels Reverse Engineering über das Tool MySQL Workbench das entwickelte Datenmodell grafisch dargestellt werden. (Siehe Abbildung 6)

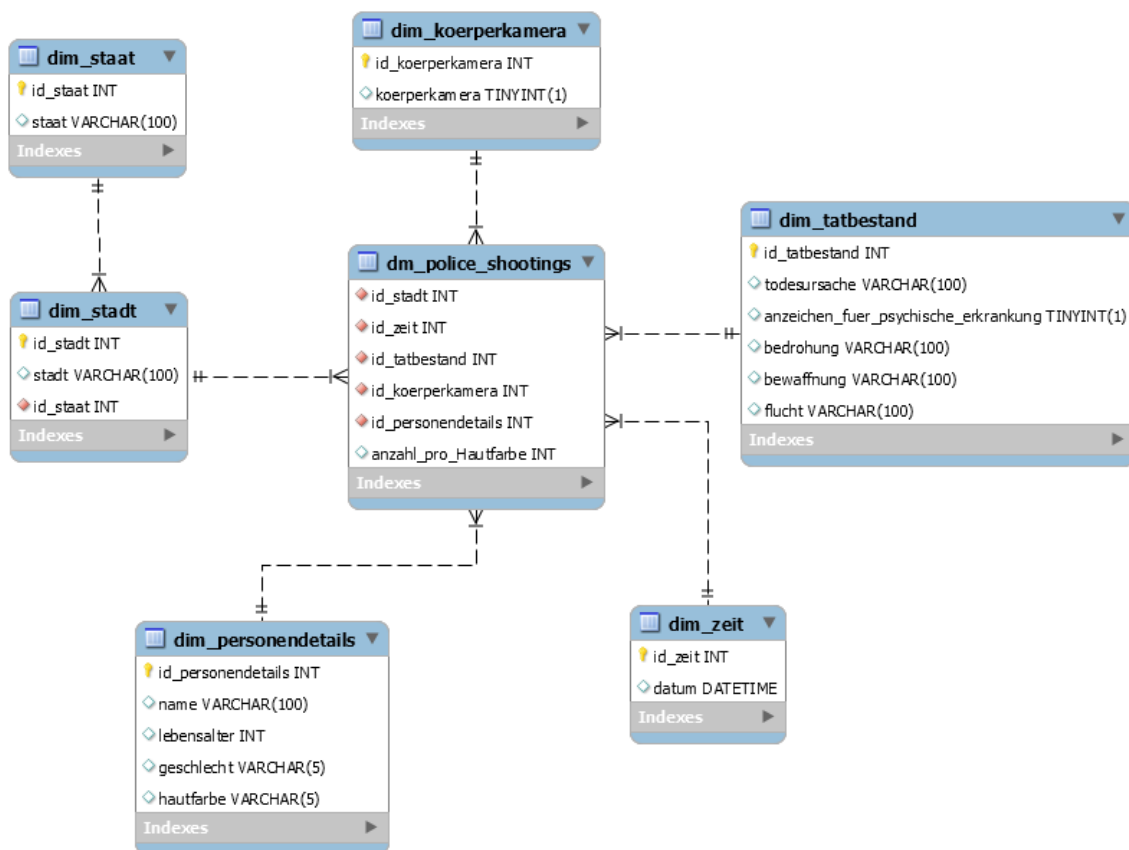


Abbildung 6 Datenbanktabellen (ERM)

3.5.2 Excel-Tabellen

Die im Data-Warehouse befindlichen Daten werden über eine ODBC-Schnittstelle in ein Excel-Worksheet importiert. Die von uns verwendete Tabelle (in Abbildung 7 aufgeführt) beinhaltet die Daten der DWH-Ebene, da in dieser Ebene kleinste Änderungen am Datenbestand durchgeführt worden sind, welche zu einer zielführenden Analyse des Datenbestands beitragen. Um die großen Datenmengen auf eine überschaubare Größe zu reduzieren, wird die aufgeführte

Excel-Tabelle über eine Pivot-Tabelle ausgewertet. Die Pivot-Tabelle enthält die errechnete Kennzahl, die Spalte Hautfarbe, so wie die Spalte Staat.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	name	datum	todesursache	bewaffnung	lebensalter	geschlecht	hautfarbe	stadt	staat	anzeichen_fuer_psychische_erkrankung	bedrohung	flucht	koerperkamera
14	21	Ron Sneed	07.01.2015 00:00	shot	gun	31 M		Black	Freeport	TX		0 attack	Not fleeing	0
27	48	Richard McClendon	13.01.2015 00:00	shot	knife	43 M		Other	Jourdanton	TX		1 other	Not fleeing	0
32	54	Robert Edwards	14.01.2015 00:00	shot	gun	68 M		Other	Lake Jackson	TX		0 attack	Not fleeing	0
39	67	Scott Hall	16.01.2015 00:00	shot	gun	41 M		Other	Mabank	TX		0 attack	Car	0
42	78	Daniel Brumley	17.01.2015 00:00	shot	knife	27 M		Other	Fort Worth	TX		0 attack	Not fleeing	0
48	88	Miguel Angel de Santos-Rodriguez	21.01.2015 00:00	shot	gun	36 M		Other	Chapeno	TX		0 attack	Foot	0
51	99	Tiano Meton	22.01.2015 00:00	shot	toy weapon	25 M		Black	Sierra Blanca	TX		0 attack	Car	0
52	100	Kristiana Coignard	22.01.2015 00:00	shot and Tasered	knife	17 F		Other	Longview	TX		1 other	Not fleeing	0
54	102	Jose Antonio Espinoza Ruiz	23.01.2015 00:00	shot	knife	56 M		Other	Levelland	TX		0 other	Not fleeing	0
71	126	Jimmy Ray Robinson Jr.	04.02.2015 00:00	shot	gun	51 M		Black	Lorena	TX		0 attack	Car	0
77	132	Savvyer Flache	08.02.2015 00:00	shot	gun	27 M		Other	Austin	TX		0 attack	Not fleeing	0
80	135	Larry Hostetter	08.02.2015 00:00	shot	gun	41 M		Other	Nocona	TX		0 attack	Not fleeing	0
82	137	Desmond Luster	09.02.2015 00:00	shot	gun	45 M		Black	Dallas	TX		0 attack	Not fleeing	0
97	155	Roy Joy Day	14.02.2015 00:00	shot	gun	51 M		Other	Laredo	TX		0 attack	Not fleeing	0
104	162	Pedro Juan Saldivar	17.02.2015 00:00	shot		50 M		Other	Del Rio	TX		0 other	Not fleeing	0
106	165	Ruben Villalpando	20.02.2015 00:00	shot	unarmed	31 M		Other	Euless	TX		0 other	Not fleeing	0
132	195	Hung Trieu	01.03.2015 00:00	shot	gun	35 M		Other	Houston	TX		0 attack	Not fleeing	0
182	261	Benjamin Quezada	31.03.2015 00:00	shot	toy weapon	21 M		Other	Baytown	TX		0 attack	Not fleeing	0
205	287	Richard Castilleja	14.03.2015 00:00	shot	gun	29 M		Other	San Antonio	TX		1 attack	Not fleeing	0
208	290	Eugene Smith	17.03.2015 00:00	shot	gun	20 M		Other	Onalaska	TX		1 attack	Not fleeing	0
235	325	Andy Martinez	09.01.2015 00:00	shot	gun	33 M		Other	El Paso	TX		0 attack	Not fleeing	0
236	329	Douglas Faith	08.04.2015 00:00	shot	gun	60 M		Other	San Antonio	TX		0 attack	Car	0
242	336	Robert Francis Mesch	23.01.2015 00:00	shot	gun	61 M		Other	Austin	TX		1 attack	Not fleeing	0
243	338	Victor Manuel Reyes	31.01.2015 00:00	shot	gun	31 M		Other	Houston	TX		0 attack	Not fleeing	0
246	341	Wendell King	24.01.2015 00:00	shot	gun	40 M		Other	Fort Worth	TX		0 attack	Not fleeing	0
249	345	Alan Lance Alverson	28.01.2015 00:00	shot	gun	45 M		Other	Sunset	TX		0 attack	Not fleeing	0
251	347	Cody Karasek	28.01.2015 00:00	shot	gun	26 M		Other	Rosenberg	TX		0 attack	Not fleeing	0

Abbildung 7 Excell-Tabelle

3.5.3 Visualisierung in Excel

Der Excel-Prototyp gilt als erstes Entwurfsmuster, welches als Leitfaden für die Erstellung des eigentlichen Dashboards dient. In besagten Prototypen werden die Anzahl der von der Polizei getöteten Dunkelhäutigen pro Staat und die Anzahl der von der Polizei getöteten Nicht-Dunkelhäutigen pro Staat aufgeführt und gegenübergestellt. Für eine deutliche Gegenüberstellung bietet sich das Säulendiagramm an. Die Vertikale-Achse des Säulendiagramms gibt die Anzahl an erschossenen Menschen an. Die Horizontale-Achse unterteilt die aufgeführte Anzahl an erschossenen unter die entsprechenden US-Bundesstaaten und kategorisiert die Werte pro Staat unter den Hautfarben „Black“ und „Other“. Das Kreisdiagramm stellt die Summe aller erschossenen Dunkelhäutigen und die die Summe aller erschossenen Nicht-Dunkelhäutigen für alle US-Bundesstaaten gegenüber. Die Datenbeschriftung enthält den prozentualen Anteil des Gesamtwertes. Anhand des Kreisdiagramms ist bereits zu erkennen, dass der im Kapitel 3.3 aufgeführte „Idealwert“ von 1/5 überschritten wird und somit verhältnismäßig viele Menschen der Gruppe „Black“ von der US-Polizei erschossen werden.

Anzahl von name

Auswertung einzelner Staaten

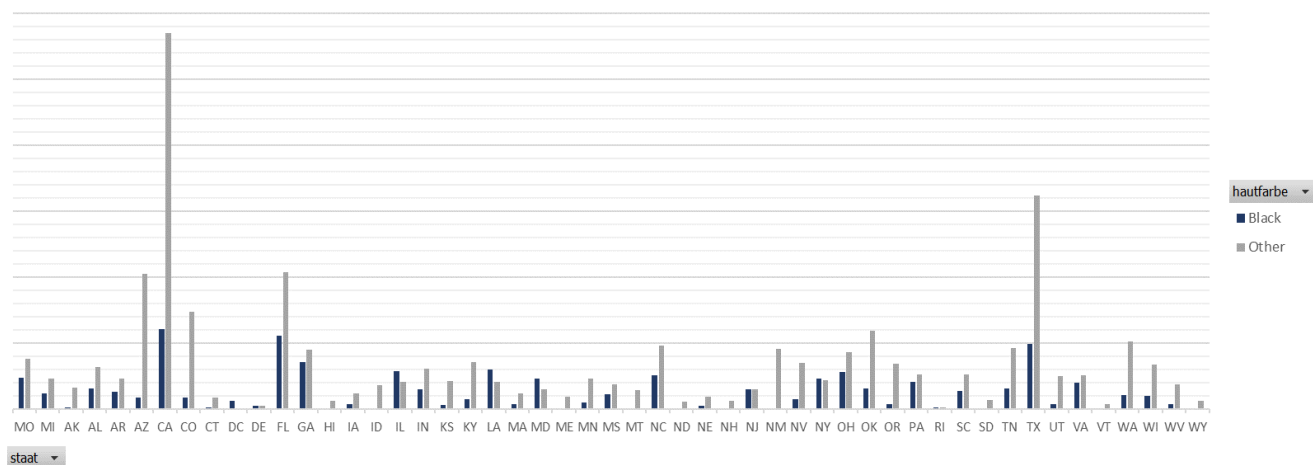


Abbildung 8 Prototyp des Dashboards in Excel (1)

Anzahl von name

Auswertung aller Staaten

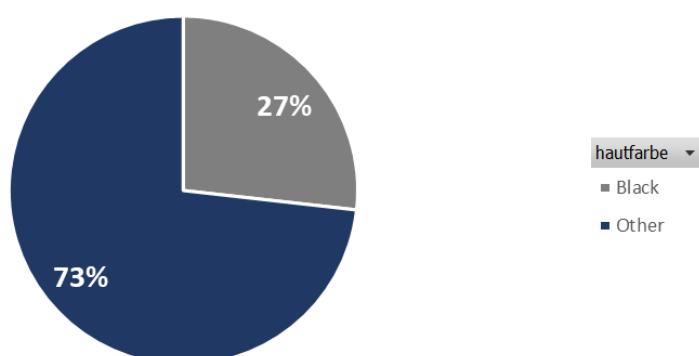


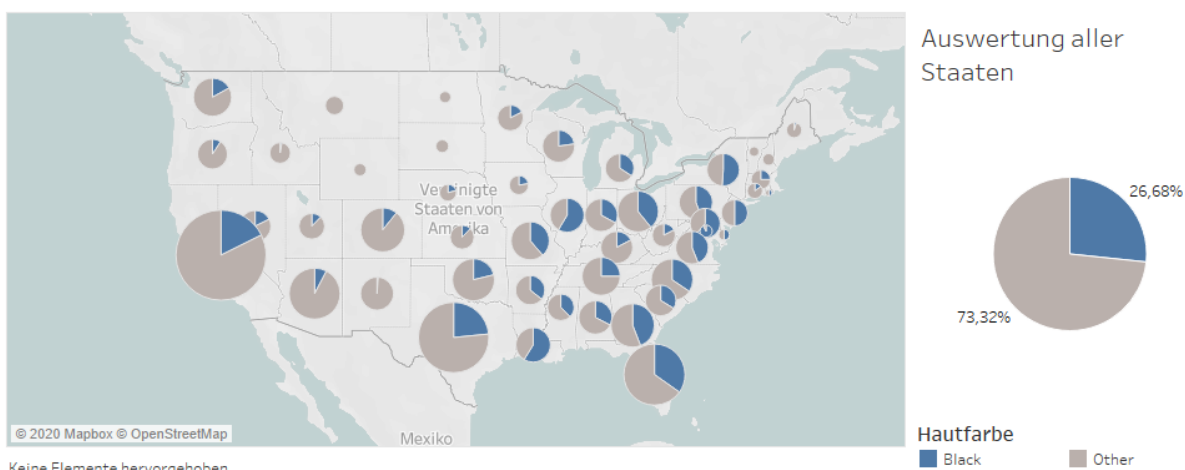
Abbildung 9 Prototyp des Dashboards in Excel (2)

4 BI Anwendung

Die Entwicklung des Dashboards wird über die Software Tableau realisiert. Dabei wird sich an das Muster des von uns erstellten Excel-Prototypen, in Kapitel 3.5.3, gehalten.

4.1 Dashboard und Datenanalyse

Anhand einer Visualisierung in Form eines Dashboards können wir mit unserem Auftraggeber kommunizieren. Damit diese Kommunikation auch einwandfrei gelingt, ist es essenziell wichtig, die Visualisierung lesbar und in ansprechender Form zu gestalten. Die Entwicklung unseres Dashboards erfolgt über die Visualisierungs-Software Tableau. Anhand der vorliegenden geografischen Daten lässt sich eine Karte für eine Analyse der einzelnen amerikanischen Bundesstaaten anfertigen. Zu jedem Bundesstaat gibt es jeweils ein Kreisdiagramm, welches einen Vergleich zwischen der Anzahl an erschossenen Dunkelhäutigen und der Anzahl an erschossenen Nicht-Dunkelhäutigen anstellt. Die Summe der beiden Werte wirkt sich auf den Durchmesser des Kreisdiagramms aus. Wie sich aus der Kennzahldefinition (Kapitel 3.3) entnehmen lässt, werden unter dem Wert „Other“ 4 verschiedene „Gruppen“ (White, Asian, Hispanic, Native American) zusammengefasst. Unter dem Wert „Black“ wird lediglich eine „Gruppe“ zusammengefasst. Auf die in Kapitel 2.2 gelisteten Maßnahmen wird zurückgegriffen, sobald der Kreissektor für den Wert „Black“ den ausgeglichenen Wert von $\frac{1}{5}$ des Gesamten Kreisumfangs übersteigt. Das Säulendiagramm stellt noch einmal die Werte - Anzahl an erschossenen Dunkelhäutigen und Anzahl an erschossenen Nicht-Dunkelhäutigen für den jeweiligen US-Bundesstaat gegenüber und gewährt eine Veranschaulichung der Größenverhältnissen über besagte Werte. Damit stets die Übersicht bewahrt werden kann, besitzt das Dashboard außerdem eine Filter-Funktionalität, welche die Möglichkeit bietet, gewünschte Werte eines beliebigen Staates im Kartendiagramm so wie im Säulendiagramm herauszufiltern und hervorzuheben, um sie so besser kenntlich zu machen. In einem weiteren Kreisdiagramm wird die Auswertung aller US-Bundesstaaten grafisch dargestellt. Das Kreisdiagramm beinhaltet zwei Teilwerte – die Summe aller erschossenen Dunkelhäutigen und die Summe aller erschossenen Nicht-Dunkelhäutigen für alle sich im Datensatz befindlichen US-Bundesstaaten. Die anhand der Teilwerte erstellten Kreissektoren sind mit einem Prozentwert beschriftet, welcher den Prozentualen Anteil des Gesamtwertes wiedergibt. Auch hier sollte der Kreissektor für den Wert „Black“ Idealerweise nicht mehr als $\frac{1}{5}$ des Kreisumfangs in Anspruch nehmen. Die im Dashboard behandelten Werte sind mit jeweils einer einheitlichen Farbe versehen, welche sich in einer Legende „Hautfarbe“ auslesen lassen.



Keine Elemente hervorgehoben

Auswertung einzelner Staaten

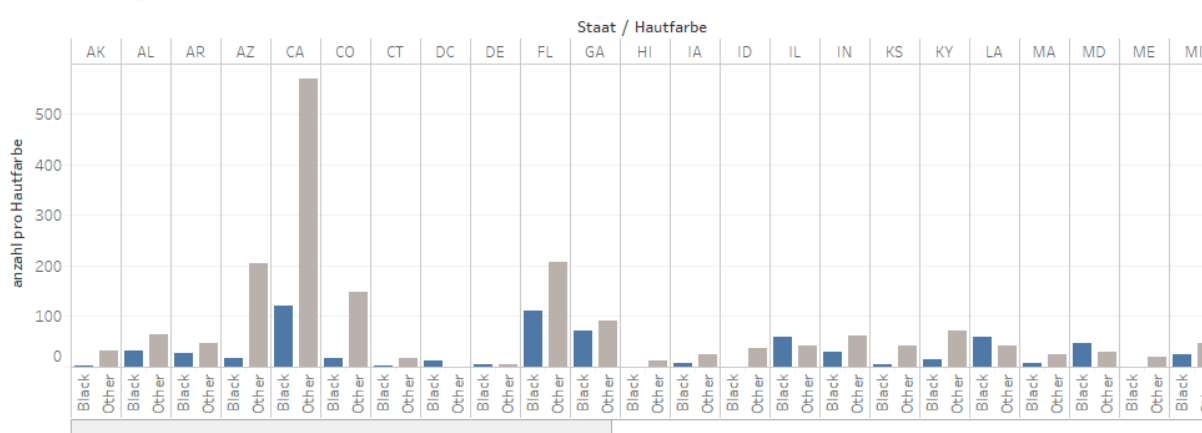


Abbildung 10 Dashboard im Überblick

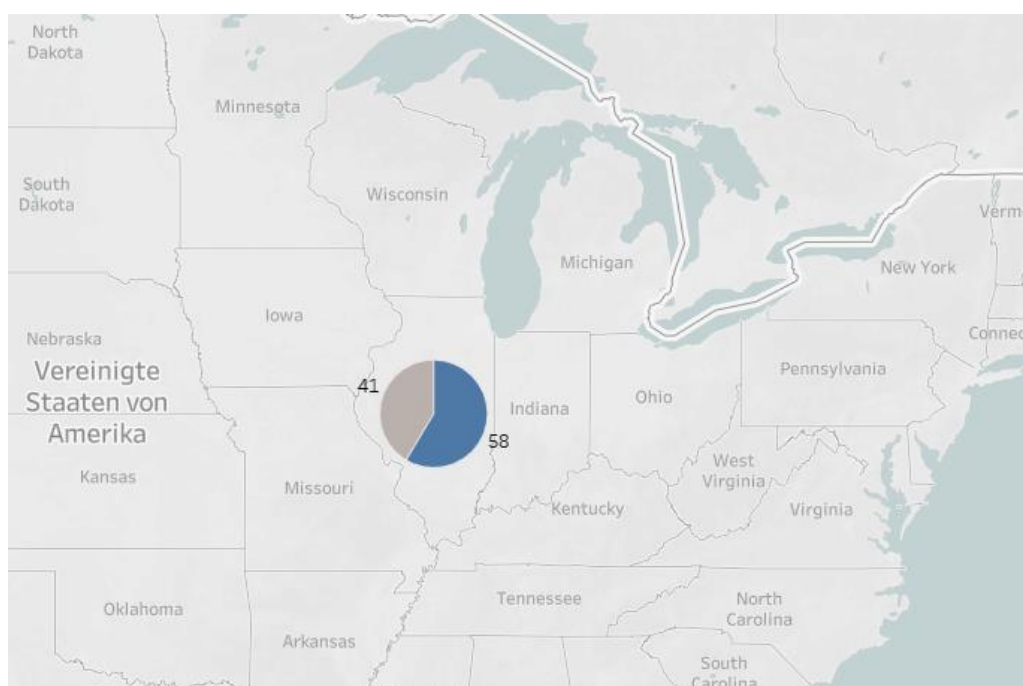


Abbildung 11 Datenanalyse unter Einsatz des Dashboards

4.2 Datenbereitstellung und ETL-Prozess

Für den Datenbereitstellungs- und ETL-Prozess des Projekts werden SQL-Skripte angefertigt, welche den folgenden beschriebenen Prozess umsetzen:

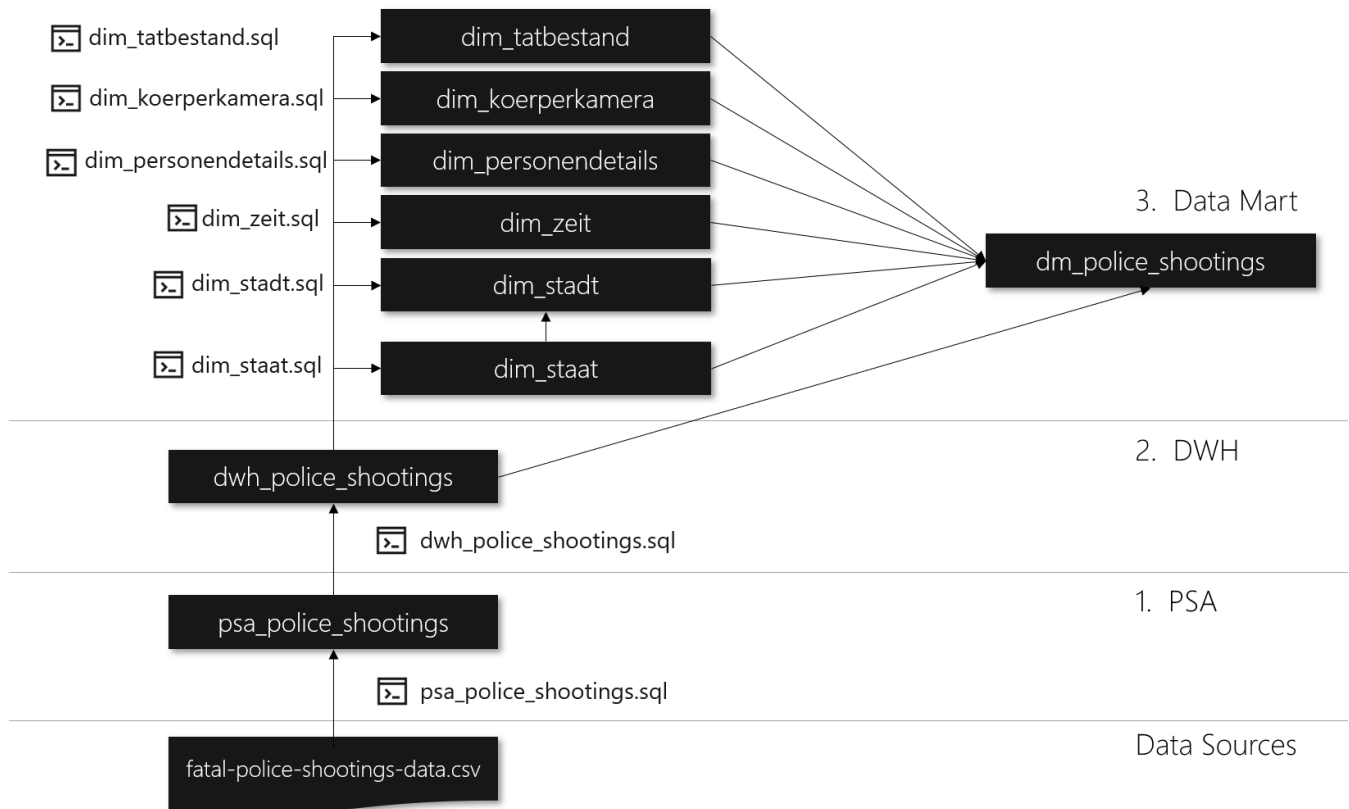


Abbildung 12 Datenbereitstellung und ETL-Prozess

- Die unveränderte CSV-Datei des Data Source Bestands wird in die PSA-Ebene übertragen. Dort wird sie lediglich zwischengespeichert.
- Damit in der DHW-Ebene mit den Daten weitergearbeitet werden kann, werden diese zunächst aus der PSA-Ebene geladen und gespeichert. In der DWH-Ebene werden einige Anpassungen an den vorliegenden Daten vorgenommen. Zeilen, welche einen NULL-Wert in der Spalte „hautfarbe“ aufweisen, sind für unsere Analyse nicht Zielführend und müssen daher aussortiert werden. Des Weiteren wird die Kategorisierungen der Spalte Hautfarbe für unsere Analyse eingegrenzt.
- Aus den Daten, der im Datawarehouse-Ebene befindlichen Flachen Tabelle, werden in der DIM-Ebene die einzelnen Dimensionen des Snowflake-Schemas erstellt. Den einzelnen Dimensionstabellen wird außerdem ein Primärschlüssel beigefügt.
- Abschließend wird die Faktentabelle erstellt. Damit alle Daten der Dimensions-Ebene zusammengeführt werden können, beinhaltet die Faktentabelle als Fremdschlüssel die Primärschlüssel der Dimensionstabellen.

Zusätzlich wird über eine Berechnung der in der DWH-Ebene befindlichen Daten die Kennzahl der Faktentabelle erstellt.

5 Resümee und Ausblick

Durch unsere Analyse im Rahmen des Projektes mussten wir offensichtlich feststellen, dass der Rassismus in den vereinigten Staaten von Amerika mehr als präsent ist und eine Dienstaufsicht für die Polizeigewalt intensiviert werden sollte. Das Projekt konnte unter strikter Aufgabenaufteilung und guter Kommunikation mit letztlich hoher Zufriedenheit unsererseits durchgeführt werden. Der vorliegende Projektplan als Hilfsmittel war für uns dennoch unverzichtbar. Er diente uns durchwegs als stützender Leitfaden. Zudem ist ein gegenseitiges Vertrauen im Arbeiten als Team maßgebend.

Lessons Learned:

- Die Arbeit in einem 2er Team wird als deutlich lockerer empfunden, da weitere Mitglieder auch höhere Risiken für einen Ausfall bergen und jeder weitere Sprachkanal eine gute Kommunikation erschweren kann.
- Da die Kommunikation aufgrund der derzeitigen Corona-Pandemie zwingendermaßen auf Online-Plattformen stattfinden musste, haben wir gelernt auch die Vorteile aus einer Online-Projektarbeit zu ziehen.
- Wir haben die Möglichkeit bekommen, im Rahmen des Studiums den Umgang mit der Software Tableau weiter zu intensivieren und uns neues Wissen anzueignen.

Ausblick:

Eine Weiterentwicklung unserer BI-Anwendung könnte durch das Einspeisen der Daten anderer Länder erfolgen. Die von uns entwickelte BI-Anwendung ist über wenige Abänderungen durchaus dazu ausgelegt tödliche Polizeischießereien anderer Länder auszuwerten.

Literaturverzeichnis

Samoshin, A. (2020). Data Police shootings [Datensatz].

Abgerufen von <https://www.kaggle.com/mrmorj/data-police-shootings>

Rouse, M. (2016, Juni 4). Schneeflockenschema. Abgerufen 6. August 2020, von [https://www.computerweekly.com/de/definition/Schneeflockenschema#:~:text=Ein%20Schneeflockenschema%20\(auch%20Snow-flake%2DSchema,das%20Schneeflockenschema%20das%20verwandte%20Stern-schema.](https://www.computerweekly.com/de/definition/Schneeflockenschema#:~:text=Ein%20Schneeflockenschema%20(auch%20Snow-flake%2DSchema,das%20Schneeflockenschema%20das%20verwandte%20Stern-schema.)