Wahlpflicht-Seminar Wintersemester 2023/24

Statistischer Bericht

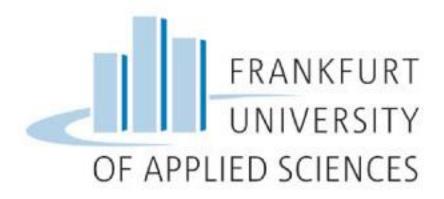
über die Umsetzung von Power BI in der Statistik durch die Semester-Aufgabe

Eingereicht durch:

Anh Quoc Nguyen 1397466

Betreut durch:

Dozent Andreas Macho



Frankfurt University of Applied Sciences

Fachbereich 2: Informatik- & Ingenieurwissenschaften

Studiengang: Informatik (B.Sc.)

Inhaltsverzeichnis

١.	Ε	inleitung	4
		Lösung der Aufgabe	
1		Verwendete Quellen	5
2	· •	Vorerarbeitung der Daten	6
3	3.	Datenmodell	8
4	.	Individuelle Spalten und Measures	9
5	.	Datenbasis zu den jeweiligen PowerBI Seiten	. 10
6	j.	Beschreibungen der Lösung	. 13
IV.		Zusammenfassung	. 14
V.	L	iteraturverzeichnis	. 15

Abbildungen

Abbildung 1: Inhaltsübersicht	6
Abbildung 2: Die Dateien von einer Tabelle	6
Abbildung 3: Power Query Editor	7
Abbildung 4: Das Ergebnis der Verarbeitung	7
Abbildung 5: Erstellen Sie eine Tabelle mit dem Tool in Excel	8
Abbildung 6: Modellansicht	8
Abbildung 7: Measure über DAX - Teil 1	9
Abbildung 8: Measure über DAX - Teil 2	9
Abbildung 9: Measure über DAX - Teil 3	9
Abbildung 10: Individuelle Spalten - Teil 1	9
Abbildung 11: Individuelle Spalten - Teil 2	9
Abbildung 12: Startseite	10
Abbildung 13: Zweite Seite	10
Abbildung 14: Dritte Seite	11
Abbildung 15: Vierte Seite	11
Abbildung 16: Fünfte Seite	12
Abbildung 17: Sechste Seite	12

I. Einleitung

Im Rahmen des WP-Seminars zum Thema "Lösungen mit MS Power BI" ist es unsere Aufgabe, ein Dashboard zu erstellen, das auf den Daten des Statistischen Bundesamtes basiert. Dieses Dashboard soll verschiedene demografische und statistische Daten visualisieren und analysieren. Die relevanten Daten umfassen:

- Sterbefallstatistik ab 2000 nach Region, Altersgruppe und Geschlecht
- Demografische Verteilung nach Altersgruppen
- Lebenserwartung und Sterbewahrscheinlichkeit
- Zusätzliche Sterblichkeit durch außergewöhnliche Ereignisse (z.B. Influenza)
- Gesamtwanderungen über die Grenzen der Bundesländer
- Bevölkerungsverteilung nach Bundesländern

Das Dashboard soll die Bevölkerungsverteilung und Sterbefälle nach Jahren darstellen und die Entwicklung dieser Zahlen analysieren. Es wird auch die Einflussfaktoren wie Wanderungsbilanz und außergewöhnliche Sterblichkeitsereignisse berücksichtigen. Die Daten werden so aufbereitet, dass sie in Power BI integriert und visualisiert werden können.

Zusätzlich sollen demografische Verteilungen und Lebenserwartungen bei der Analyse der Sterbefälle berücksichtigt werden, und Prognosen für zukünftige Jahre sollen erstellt werden. Alle Schritte der Datenverarbeitung und die verwendeten Quellen werden dokumentiert, um ein transparentes und nachvollziehbares Ergebnis zu gewährleisten.

II. Theorie

Um diese Aufgabe zu lösen, sind folgende theoretische Grundlagen relevant:

Demografie

Die Demografie untersucht die Struktur und Entwicklung von Bevölkerungen. Wichtige Aspekte sind Geburten- und Sterberaten sowie die Altersverteilung. Diese Daten sind entscheidend für die Erstellung der Bevölkerungsverteilung nach Altersgruppen und Regionen.

Sterblichkeitsstatistik

Sterblichkeitsstatistiken erfassen Todesfälle nach verschiedenen Kriterien wie Alter, Geschlecht und Todesursachen. Diese Daten helfen, Trends zu erkennen und außergewöhnliche Sterblichkeitsereignisse zu analysieren. Sie bilden die Basis für die Darstellung der Sterbefälle und die Analyse der Sterblichkeitsentwicklung.

Datenvisualisierung

Datenvisualisierung ermöglicht die grafische Darstellung großer Datensätze, wodurch Muster und Trends sichtbar werden. Microsoft Power BI bietet leistungsfähige Tools zur Erstellung interaktiver Dashboards, die eine intuitive und detaillierte Analyse ermöglichen. Dies ist essenziell, um die Daten verständlich und nutzerfreundlich darzustellen.

Datenverarbeitung

Die Datenverarbeitung umfasst das Bereinigen, Transformieren und Aggregieren von Daten. Tools wie Power Query in Power BI unterstützen diese Schritte, um die Daten in eine nutzbare Form zu bringen. Eine sorgfältige Datenvorbereitung ist notwendig, um konsistente und aussagekräftige Analysen zu gewährleisten.

III. Lösung der Aufgabe

1. Verwendete Quellen

In diesem Projekt werde ich Informationen von destatis.de und dem Robert-Koch-Institut verwenden. Dies sind die Datensätze, die im Projekt enthalten sind (einschließlich Code):

- Gesamtwanderungen über die Grenzen der Bundesländer: Bundesländer, Jahre, Nationalität, Geschlecht (12711-0020)
- Gestorbene: Bundesländer, Jahre, Geschlecht (12613-0012)
- Sterbetafel (Periodensterbetafel): Deutschland, Jahre, Geschlecht, Vollendetes Alter (12621-0001)
- Gestorbene: Deutschland, Jahre, Todesursachen, Geschlecht, Altersgruppen (23211-0002)
- Statistischer Bericht Sterbefälle nach Tagen, Wochen und Monaten - endgültige Daten - 2000 bis 2019

 Statistischer Bericht - Sterbefälle nach Tagen, Wochen und Monaten - 2020 bis 2024

2. Vorerarbeitung der Daten

Tabellen	<u>Impressum</u>
Tabellen -	Life of the control o
Tabellen	Informationen zur Statistik
<u>12613-01</u>	Sterbefälle nach Tagen in Deutschland - Insgesamt
12613-02	Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Insgesamt
12613-03	Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Männlich
12613-04	Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Weiblich
12613-05	Sterbefälle nach Monaten und Altersgruppen in Deutschland - Insgesamt
12613-06	Sterbefälle nach Monaten und Altersgruppen in Deutschland - Männlich
12613-07	Sterbefälle nach Monaten und Altersgruppen in Deutschland - Weiblich
12613-08	Sterbefälle nach Tagen in den Bundesländern - Insgesamt
12613-09	Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Insgesamt
12613-10	Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Männlich
12613-11	Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Weiblich
12613-12	Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Insgesamt
12613-13	Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Männlich
12613-14	Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Weiblich
Daten zur Weiterv	verarbeitung verarbeitung
	Erläuterung zu CSV-Tabellen
csv-12613-01	Zu Tabelle 12613-01 : Sterbefälle nach Tagen in Deutschland - Insgesamt
sv-12613-02	Zu Tabelle 12613-02 : Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Insgesamt
sv-12613-03	Zu Tabelle 12613-03 : Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Männlich
sv-12613-04	Zu Tabelle 12613-04 : Sterbefälle nach Kalenderwochen und Altersgruppen in Deutschland - Weiblich
sv-12613-05	Zu Tabelle 12613-05 : Sterbefälle nach Monaten und Altersgruppen in Deutschland - Insgesamt
sv-12613-06	Zu Tabelle 12613-06 : Sterbefälle nach Monaten und Altersgruppen in Deutschland - Männlich
sv-12613-07	Zu Tabelle 12613-07 : Sterbefälle nach Monaten und Altersgruppen in Deutschland - Weiblich
sv-12613-08	Zu Tabelle 12613-08 : Sterbefälle nach Tagen in den Bundesländern - Insgesamt
sv-12613-09	Zu Tabelle 12613-09 : Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Insgesamt
sv-12613-10	Zu Tabelle 12613-10 : Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Männlich
sv-12613-11	Zu Tabelle 12613-11 : Sterbefälle nach Kalenderwochen und Altersgruppen in den Bundesländern - Weiblich
sv-12613-12	Zu Tabelle 12613-12 : Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Insgesamt
sv-12613-13	Zu Tabelle 12613-13 : Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Männlich
sv-12613-14	Zu Tabelle 12613-14 : Sterbefälle nach Monaten und Altersgruppen in den Bundesländern - Weiblich
	und Abkürzungen
	r = berichtigte Zahl
	. = Angabe fällt später an
	. = Zahlenwert unbekannt oder geheim zu halten
	X = Tabellenfach gesperrt, Aussage nicht sinnvoll
	e = endgültige Zahlen

Abbildung 1: Inhaltsübersicht

Als nächstes müssen wir die doppelten Daten bereinigen.

			Baden-Württemberg		Bayern	Bayern	Berlin	Berlin	Berlin			Brandenburg		Bremen	Bremen	Hamburg	Hamburg	Hamburg
	männlich	weiblich	Insgesamt		weiblich	Insgesamt		weiblich		männlich	weiblich			weiblich	Insgesamt	männlich	weiblich	Insgesamt
31.12.2002 unter 1 Jahr	51482																	
31.12.2002 1-Jährige	52712																	
31.12.2002 2-Jährige	55144		4 108138					14096	28731	9738	9326	19064				795		
31.12.2002 3-Jährige	57079																	
31.12.2002 4-Jährige	58257																	
31.12.2002 5-Jährige	60485							13409	27706									
31.12.2002 6-Jährige	59492																	
31.12.2002 7-Jährige	58640	5533	1 113971	65849	62359	128208	13085	12526	25611	8115	7655	15770	3009	2960	5969	753	3 7029	9 145
31.12.2002 8-Jährige	59263	5595	4 115217	67344	63394	130738	13299	12290	25594	7735	7319	15054	3005	2848	585	767	3 7157	7 148
31.12.2002 9-Jährige	61333	5824	4 119577	70609	66850	137459	13256	12601	25857	7734	7299	15033	3089	2902	599:	748	4 7269	9 147
31.12.2002 10-Jährige	62252	5855	3 120805	70895	68191	139086	13955	12970	26925	854	7977	16513	3217	2988	620	778	6 7512	2 152
31.12.2002 11-Jährige	63283	6031	9 123602	72886	68471	141357	14363	13604	27967	7 10117	7 10036	20153	3277	3038	631	806	6 7706	6 157
31.12.2002 12-Jährige	65321	6230	5 127626	75216	71391	146607	17471	16569	34040	16231	15307			3064	644	844		
31.12.2002 13-Jährige	63382	6049	5 123877	72269	69050	141319	17941	17138	35076	17559	16843	34402	3267	3183	645	808	4 7624	4 157
31.12.2002 14-Jährige	64940	6202	3 126963	73546	69505	143051	18731	17600	36331	19078	17858	36936	3348	3137	6488	838	1 7818	8 161
31.12.2002 15-Jährige	62305	5887	2 121177	70505	66023	136528	18474	17760	36234	19621	18779	38400	3230	3207	643	791	0 7377	7 152
31.12.2002 16-Jährige	61905																	
31.12.2002 17-Jährige	59319	5641	4 115733	66488	63672	130160	18060	17468	35529	20128	18449	38577	3247	3098	634	788	3 7473	3 153
31.12.2002 18-Jährige	59823	5692	3 116746			131896			35786			37764			644			
31.12.2002 19-Jährige	59930	5935	2 119282	68924	66761	135685	19022	18720	37742	19477	7 17568	37042	3535	3634	716	847	6 8273	3 167
31.12.2002 20-Jährige	62935	6230	4 125239	71448	70820	142268	20476	20639	41115	19637	7 16460	36097	3698	4002	770	918	1 9609	9 187
31.12.2002 21-Jährige	63387		1 127448															
31.12.2002 22-Jährige	65057	6523	8 130295	73569	73441	147010	22883	23874	46757	7 18210	15300	33512	4085	4162	824	1054	0 11761	1 223
31.12.2002 23-Jährige	61948	6218	9 124137	71369	71110	142479	22683	23488	46171	17221	14400	31621	3970	4091	806	1078	1 11835	5 226
31.12.2002 24-Jährige	61811	6105	5 122866												8063			
31.12.2002 25-Jährige	61929	6163	8 123567	72018	71488	143506	23452	23912	47364	15447	7 13087	28534	4044	4040	808	1177	8 12446	6 242
31.12.2002 26-Jährige	63864	6198	8 125852	72926	71432	144358	23583	23605	47188	13791	11626	25417	4214	4133	834	1283	6 12757	7 255
31.12.2002 27-Jährige	62685	6124	0 123925	72204	70763	142967	23643	23198	46841	12836	10769	23605	4073	3946	801	1307	7 12609	9 256
31.12.2002 28-Jährige	64365	6348	0 127845	74374	72752	147126	23430	22788	46218	12494	10897	23391	4211	3950	816	1329	5 12695	5 259
31.12.2002 29-Jährige	64719	6391	7 128636	75573	73399	148972	23320	22622	45942	12660	11250	23910	4204	4110	8314	1291	0 12830	0 257
31.12.2002 30-Jährige	71515	6997	2 141487	82729	80421	163150	24754	23559	48313	14473	12564	27037	4503	4445	894	1413	2 13633	3 277
31.12.2002 31-Jährige	78110	7522	5 153335	90661	87675	178336	27972	26055	54027	16917	15025	31942	4954	4599	955	1508	6 14312	2 293
31.12.2002 32-Jährige	81911	7854	2 160453	95870	92159	188029	28638	26458	55096	17516	15868	33381	5195	4656	985	1577	5 14389	9 301
31.12.2002 33-Jährige	88501	8450	3 173004	104234	98977	203211	30173	2755	57724	18128	16927	35055	5495	5011	1050	1676	3 15035	5 317
31.12.2002 34-Jährige	93466	8883	9 182305	109191	103933	213124	31396	29010	60408	18948	18148	37096	5851	5251	1110	1802	4 15927	7 339

Abbildung 2: Die Dateien von einer Tabelle

Mit der Leistungsfähigkeit des Power Query Editors können wir die Daten leicht neu ordnen.

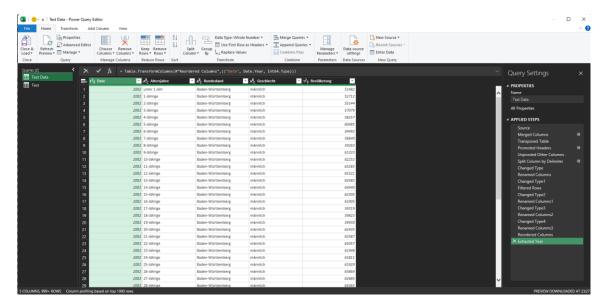


Abbildung 3: Power Query Editor

Dies ist das Ergebnis nach der Datenverarbeitung.

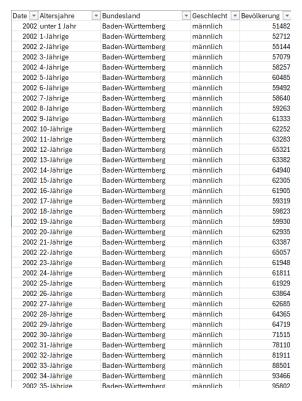


Abbildung 4: Das Ergebnis der Verarbeitung

Jetzt werden wir die Tabellen aus der Excel-Datei extrahieren und in einzelne Dateien umwandeln, um sie in Power Bl zu importieren.

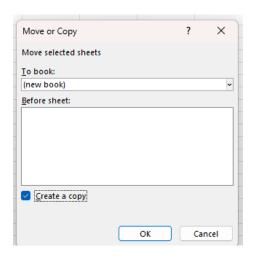


Abbildung 5: Erstellen Sie eine Tabelle mit dem Tool in Excel

3. Datenmodell

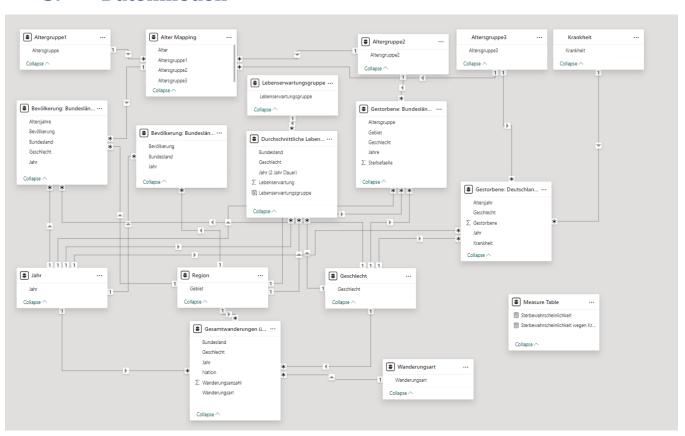


Abbildung 6: Modellansicht

4. Individuelle Spalten und Measures



Abbildung 7: Measure über DAX - Teil 1

1 Sterbewahrscheinlichkeit wegen Krankheit = 2 sum('Gestorbene]) / sum('Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre'[Bevölkerung])

Abbildung 8: Measure über DAX - Teil 2

1 Sterbewahrscheinlichkeit = 2 sum('Gestorbene: Bundesländer, Jahre, Geschlecht'[Sterbefaelle]) / sum('Bevölkerung: Bundesländer, Stichtag, Geschlecht, Altersjahre'[Bevölkerung])

Abbildung 9: Measure über DAX - Teil 3



Abbildung 10: Individuelle Spalten - Teil 1



Abbildung 11: Individuelle Spalten - Teil 2

5. Datenbasis zu den jeweiligen PowerBI Seiten



Abbildung 12: Startseite

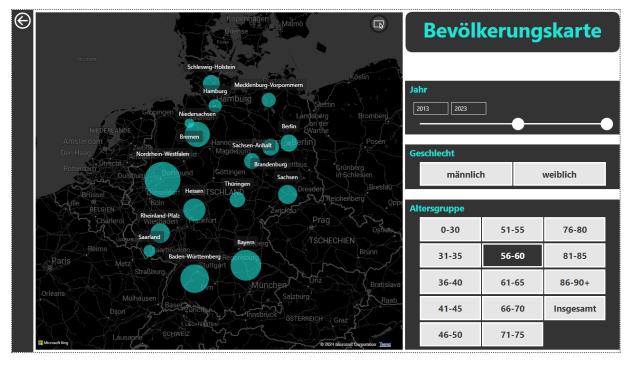


Abbildung 13: Zweite Seite

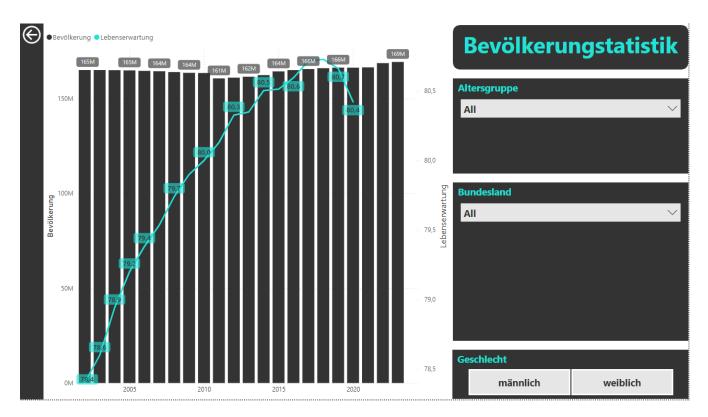


Abbildung 14: Dritte Seite

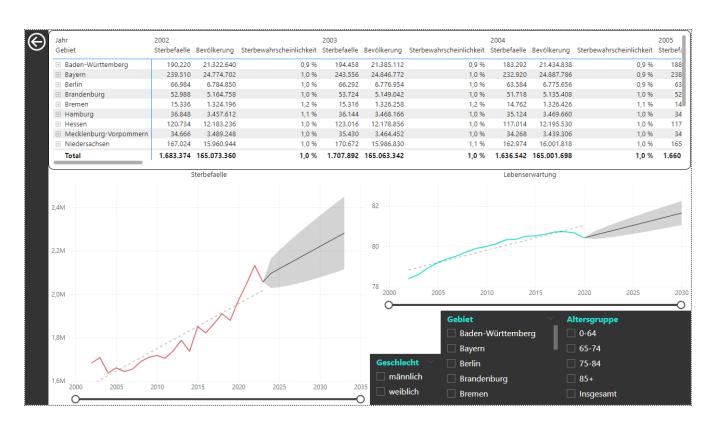


Abbildung 15: Vierte Seite

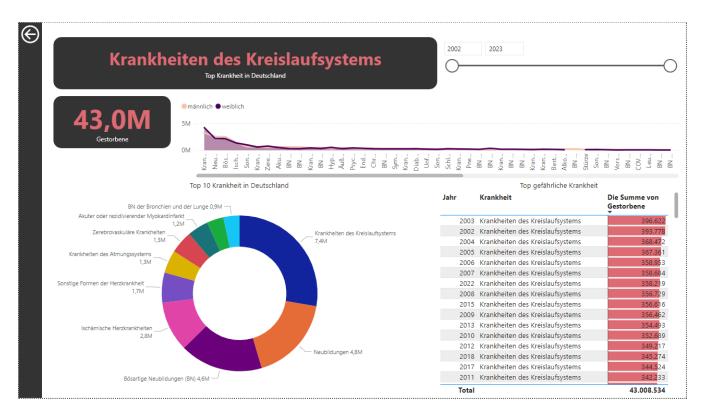


Abbildung 16: Fünfte Seite



Abbildung 17: Sechste Seite

6. Beschreibungen der Lösung

Zuerst lade ich die notwendigen Daten auf meinen Computer herunter. Danach verwende ich die Software Excel, um die Daten zu erstellen und zu bereinigen.

Die nach der Verarbeitung erhaltenen Daten werden im Ordner "source" gespeichert, um eine konsistente Verarbeitung beim Hinzufügen zu Power BI zu ermöglichen und unnötige Fehler zu vermeiden.

Darüber hinaus verwende ich Measures, um die Sterblichkeitsrate im Verhältnis zur Bevölkerung in der Region mit dem DAX-Tool zu berechnen.

Nachdem alle Daten vorliegen, stelle ich über die Modellansicht Beziehungen zwischen den Tabellen her.

Nachdem alle Daten vollständig vorhanden sind, verwende ich die verfügbaren Tools in Power BI, um die Aufgabe abzuschließen. Zu den hauptsächlich verwendeten Tools gehören: Slicer und verschiedene Visualisierungsdiagramme.

Abschließend entwerfe ich die Startseite mit Navigationsschaltflächen.

IV. Zusätzliche Antwortungen

Aus dem Diagramm kann man erkennen, dass Kreislauferkrankungen eine der schwerwiegendsten Krankheiten in Deutschland sind.

Je näher wir der Gegenwart kommen, desto höher wird die durchschnittliche Lebenserwartung der Bevölkerung. Außerdem nimmt die Anzahl der Todesfälle durch Krankheiten wie Kreislauferkrankungen ab.

Basierend auf den Daten von 2002 bis 2023 können wir schnell vorhersagen, dass die Zahl der Todesfälle in Deutschland weiter steigen wird. Laut statistischen Prognosen wird im Jahr 2030 die Zahl der Todesfälle voraussichtlich 2.219.814 betragen.

V. Zusammenfassung

Diese Aufgabe und das Erstellen eines PowerBI-Dashboards vermitteln wichtige Lernerfahrungen:

Datenanalyse und -visualisierung

- Datenverarbeitung: Erlernen, wie man große Datensätze sammelt, bereinigt und transformiert.
- Datenvisualisierung: Fähigkeit, komplexe Daten grafisch darzustellen und Trends sichtbar zu machen.

Demografische und statistische Analysen

- Demografische Einblicke: Verständnis der Bevölkerungsdynamik und demografischer Trends.
- Sterblichkeitsstatistiken: Analyse der Einflussfaktoren auf Sterblichkeitsraten und außergewöhnliche Ereignisse.

Technische Fertigkeiten

- Power BI-Nutzung: Kompetenzen im Umgang mit Power BI und Power Query zur Datenvorbereitung und Dashboard-Erstellung.
- Dokumentation: Systematisches und transparentes Vorgehen bei der Datenanalyse.

Praktische Anwendung

- Reale Datenanalyse: Praxisnahe Erfahrung mit echten Daten des Statistischen Bundesamtes.
- Prognosen und Szenarien: Erstellung von Prognosen basierend auf historischen Daten.

VI. Literaturverzeichnis

- Bevölkerung. (kein Datum). Von https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de &query=12411#abreadcrumb abgerufen
- Destatis. (kein Datum). *Datenbank durchsuchen*. Von https://www-genesis.destatis.de/genesis/online abgerufen
- Durchschnittliche Lebenserwartung. (kein Datum). Von https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de &query=12621#abreadcrumb abgerufen
- Gesamtwanderungen. (kein Datum). Von https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de &query=12711#abreadcrumb abgerufen
- Gestorbene. (kein Datum). Von https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de &query=12613#abreadcrumb abgerufen
- Gestorbene. (kein Datum). Von https://www-genesis.destatis.de/genesis/online?operation=find&suchanweisung_language=de &query=23211#abreadcrumb abgerufen
- Microsoft. (kein Datum). *Power BI documentation*. Von https://learn.microsoft.com/en-us/power-bi/ abgerufen