



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

NEUMANN JÁNOS
INFORMATIKAI KAR



SZAKDOLGOZAT

OE-NIK
2022

Hallgató neve:
Hallgató törzskönyvi száma:

Szócs Andrea
T/005489/F112904/N

Óbudai Egyetem
Neumann János Informatikai Kar
Kiberfizikai Rendszerek Intézet

SZAKDOLGOZAT
FELADATLAP

Hallgató neve: **Szócs Andrea**
Törzskönyvi száma: T/005489/FI12904/N

A dolgozat címe:

Adatharmonizáló, elemző és vizualizációs rendszer kialakítása a globális egészségügyi adatokra a klinikai prognózisok és döntések támogatásához

Development of a data harmonization, analysis and visualization system for global healthcare data to support clinical prognosis and decisions

Intézményi konzulens: :Dr. Komoróczy-Steiner Henriette

Külső konzulens:

Beadási határidő: 2022 Május 15

A záróvizsga tárgyai: Számítógép architektúrák
Big Data és üzleti intelligencia specializáció

A feladat

A szakdolgozat feladata, hogy találjon megfelelő adatharmonizáló megoldást, amellyel a globálisan elérhető, különféle struktúrájú és logikájú egészségi adatok tárolhatóak, azonos szerkezetre alakíthatóak, illetve ezekre alapozva a megfelelő vizualizációs eszközzel kimutatások és jelentések készíthetők. Fontos, hogy az elérhető adatokból és trendvonalakból következtetéseket lehessen levonni akár bizonyos betegségekre vagy kórképekre, akár egy ország egészséggel kapcsolatos helyzetére vonatkozóan, ezzel is segítve a rövid és hosszútávú döntések meghozatalát.

A dolgozatnak tartalmaznia kell:

- az elérhető egészségügyi adatok bemutatását, azok harmonizálását és egységes szerkezetre hozatalát,
- a megfelelő adattárház kiválasztásának szempontjait, ehhez kapcsolódóan a rendszertervet és adatmodellt,
- a vizualizáláshoz szükséges eszköz kiválasztásának szempontjait,
- globális és lokális trendek bemutatását az elmúlt évek adataira vonatkozóan,
- az adatvizualizációk jelentőségének kiemelését,
- továbbfejlesztés lehetőségeinek mérlegelése.



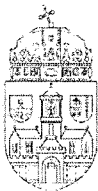
FC M
.....
Dr. Fleiner Rita
intézetigazgató

A szakdolgozat elévülésének határideje: **2023. december 15.**
(OE TVSz 55.§ szerint)

A dolgozatot beadásra alkalmasnak tartom:

.....
külső konzulens

Dr. Komoróczy-Steiner Henriette
.....
intézményi konzulens



HALLGATÓI NYILATKOZAT

Alulírott Szőcs Andrea kijelentem, hogy a szakdolgozat saját munkám eredménye, a felhasznált szakirodalmat és eszközöket azonosíthatóan közöltem. Az elkészült szakdolgozatomban található eredményeket az egyetem és a feladatot kiíró intézmény saját céljára térítés nélkül felhasználhatja.

Budapest, 2022. május 11.

hallgató aláírása

KONZULTÁCIÓS NAPLÓ

Hallgató neve:

Neptun Kód:

Tagozat: esti

Szócs Andrea

Szakdolgozat / Diplomamunka¹ címe magyarul:

Adatharmonizáló, elemző és vizualizációs rendszer kialakítása a globális egészségügyi adatokra a klinikai prognózisok és döntések támogatásához

Szakdolgozat / Diplomamunka² címe angolul:

Development of a data harmonization, analysis and visualization system for global healthcare data to support clinical prognosis and decisions

Intézményi konzulens:

Külső konzulens:

Dr. Komoróczy – Steiner Henriette

Kérjük, hogy az adatokat nyomtatott nagy betűkkel írja!

Alk.	Dátum	Tartalom	Aláírás
1.	2022. 02 8	bevezető, szerkezet, követelmények	Dr. K. Steiner
2.	2022. 03 8	Irodalomkutatás eredményei dolgozat vázlata	Dr. K. Steiner
3.	2022 04 5	Az adattáblaszerkezet	Dr. K. Steiner
4.	2022 04 19	A dolgozat átnézése javítása	Dr. K. Steiner

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4 ³ tantárgy követelményét teljesítette, beszámolóra / védésre ⁴bocsátható.

Dr. Komoróczy – Steiner Henriette

Intézményi konzulens

Budapest, 2022. 05.1.

¹ Megfelelő aláhúzendő!

² Megfelelő aláhúzendő!

³ Megfelelő aláhúzendő!

⁴ Megfelelő aláhúzendő!

KONZULTÁCIÓS NAPLÓ

Hallgató neve:

Neptun Kód:

Tagozat: esti

Szócs Andrea

Szakdolgozat / Diplomamunka¹ címe magyarul:

Adatharmonizáló, elemző és vizualizációs rendszer kialakítása a globális egészségügyi adatokra a klinikai prognózisok és döntések támogatásához

Szakdolgozat / Diplomamunka² címe angolul:

Development of a data harmonization, analysis and visualization system for global healthcare data to support clinical prognosis and decisions

Intézményi konzulens:

Külső konzulens:

Dr. Komoróczy – Steiner Henriette

Kérjük, hogy az adatokat nyomtatott nagy betűkkel írja!

Alk.	Dátum	Tartalom	Aláírás
1.	2021. 01 8	Irodalomkutatás módszertana	Dr. K. Steiner
2.	2021 03 17	Adatbázis építésének fő irányelvei	Dr. K. Steiner
3.	2021 04 6	Az egészségügyi adatok konvertálása adatbázisba	Dr. K. Steiner
4.	2021 04 20	A dolgozat átnézése javítása	Dr. K. Steiner

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4³ tantárgy követelményét teljesítette, beszámolóra / védésre⁴bocsátható.

Dr. Komoróczy – Steiner Henriette

Intézményi konzulens

Budapest, 2021. 05. 05

¹ Megfelelő aláhúzendő!

² Megfelelő aláhúzendő!

³ Megfelelő aláhúzendő!

⁴ Megfelelő aláhúzendő!

Tartalomjegyzék

1.	BEVEZETŐ.....	9
2.	AZ EGÉSZSÉGÜGYI ADATOK ELÉRHETŐSÉGE.....	11
2.1.	A feladathoz szükséges egészségügyi és egyéb adatok.....	11
2.2.	Források bemutatása	12
2.3.	Forrástáblák bemutatása	13
3.	ADATTÁRHÁZ TERVEZÉSE.....	19
3.1.	Adattárházak alapjai.....	19
3.1.1.	William H. Inmon definíciója.....	19
3.1.2.	Különböző adattárház szerkezetek.....	19
3.2.	Adattárházak jellemzői	20
3.3.	Adattárházak funkciói.....	20
3.4.	Az adattárház komponensei.....	20
3.5.	Adatbázis megoldások	21
3.5.1.	Adatbázisok alapfogalmai.....	21
3.5.2.	Adatmodellezés és tervezés	22
3.5.3.	Relációs adatbázis tervezése és kiválasztása	22
3.5.4.	Az egészségügyi adatokhoz EK Diagram megtervezése	23
3.6.	Az egészségügyi adatok betöltése, racionalizálása	23
4.	VIZUALIZÁCIÓS MEGOLDÁS TERVEZÉSE	25
4.1.	Big Data jellemzői	25
4.2.	Az adatvizualizáció jelentősége.....	26
4.3.	Vizualizációs eszköz tervezése	26
4.4.	A Power BI kiemelkedő előnyei.....	28
4.4.1.	Adatforrások csatlakoztatása és formázása.....	28
4.5.	Power BI folyamatai	30
4.6.	Az elkészült rendszer tervezete.....	34
5.	AZ EGÉSZSÉGÜGYI ADATOK VIZUALIZÁCIÓJA	35
6.	TRENDEK MEGFIGYELÉSE ÉS KÖVETKEZTETÉSEK	41
6.1.	A vizualizációk elemzése	41

6.2.	Következtetés és továbbfejlesztési lehetőségek.....	42
7.	MELLÉKLETEK.....	44
7.1.	Egyéb ábrák.....	44
7.2.	Táblák átalakítása Power BI dataflowban	47
7.3.	Táblák létrehozó scriptjei	56
8.	ÁBRAJEGYZÉK.....	61
9.	IRODALOMJEGYZÉK	63

1. BEVEZETŐ

Szakdolgozatom célkitűzése egy adatharmonizáló és vizualizációs rendszer megtervezése, amely a jelenleg globálisan elérhető egészségügyi adatokat dolgozza fel. A feladat magában foglalja az általános egészségi állapotok bemutatását, mely kórképek, betegségek és egészségügyi problémák a legelterjedtebbek. Valamint rávilágít olyan általános tényekre is, hogy mennyi a születéskor várható élettartam az egyes országokban, régiókban, milyen az egészségügy rendszere kormányzatok szintjén és esetlegesen milyen konzekvenciákat lehet levonni a jelenlegi helyzetből.

Az egészségügy és az orvostudomány mondhatni egyidős az emberiséggel. A folyamatos fejlődésének köszönhetően rengeteg betegség háttérbe szorult a történelem során. Valamint azoknak köszönhetően, akik feljegyezték a megfigyeléseiket és javaslatokat fogalmaztak meg a kezeléseket illetően, rengeteg egészségügyi szakirodalom jöhetett létre, amelyekre a mai modern orvostudomány is tudott építkezni és továbbfejleszteni.

A mai világban az emberek számára már nagyon sok minden –az egészséges életet támogató eszköz és kezelés-- elérhetővé vált, sok veszélyforrás már nem áll fenn. Ennek köszönhetően például egy gyermek születésekor már legtöbbször nem az a kérdés, hogy megélheti a felnőttkort vagy sem, hanem hogy milyen minőségben és milyen hosszan élvezheti majd. A születéskor várható élettartam a 21. századra a korábbiakhoz képest megduplázódott. [1] Ugyanakkor a várható élettartamot nagyban befolyásolja, hogy hol is születik a gyermek, milyen az egészségügy helyzete az adott országban, milyen az emberi és tárgyi felkészültség a rendszerben, milyen döntéseket hoznak az egészségügyre vonatkozóan, illetve mennyire veszik figyelembe azokat a trendeket, amelyek előre jelezhetnek egy vészhelyzetet esetleges egész népességet érintő egészségügyi problémát.

A Covid-19 2019 őszén elindult világijárvány is rávilágított arra a helyzetre, hogy akár mennyire is áll stabil lábakon egy ország egészségügyi rendszere, attól még nem lesz megingathatatlan és a váratlan helyzetekre való felkészültség és a gyors reakció elengedhetetlen.

Habár évek óta már-már divat szinten is hangoztatják, hogy a 21.-ik század legnagyobb erőforrása az adat, az adatok tárolása, előállítása, bányászata és feldolgozása, mégis ennek megfelelő és érdemi kihasználásával kevesen kezdtek el foglalkozni, különösen ezen a területen. Véleményem szerint az egészségügyi adatoknak és statisztikáknak is, kevésbé van a jelentőségük a helyén kezelve, nehezen tudunk globális képet kapni a jelenlegi és a várható epidemiológiai helyzetről. Pedig nagyon sok szervezet foglalkozik azzal, hogy a statisztikákat összegyűjtse az egyes országoktól, valamennyi következtetést is igyekeznek megfogalmazni, az elmúlt évek és adataira vonatkozóan, ugyanakkor jóslásokba vagy becslésekbe kevésbé mernek belemenni.

Ezzel a problémával kapcsolatban fogalmazódott meg bennem az a gondolat, hogy egy megfelelően átgondolt rendszerrel, amely felkészült arra, hogy a különféle forrásokból származó adatsorokat egybegyűjtse, egységesítse azok struktúráját, rugalmasan kezelje a változásokat, képes legyen a trendeket szinte „élőben” mutatni és ezek alapján megbecsülni a várható változásokat és eseményeket, egészen új távlatok nyílhatnak meg az egészségügyben és a betegségek előre jelzésében. Egy jól működő rendszer jelentősége megkérdőjelezhetetlen, minden és mindenki javára válhat, ha foglalkozunk ezzel a témakörrel.

Emellett mindig is foglalkoztatott annak a gondolata, hogy mennyire lehet előre tekintve vizsgálódni a múlt adatai alapján, különösképpen ilyen minden egyes embert érintő témákban, mint a minőségi élet, élettartam és a halál. A téma, amelyet kiválasztottam csupán egy nagyon vékony szelete ennek a kérdéskörnek, többféle irányba ki lehet szélesíteni a kutatásokat, akár a statisztikák mellett egyéni egészségi eredmények felhasználásával, mesterséges intelligencia, gép tanulás bevonásával, vagy egyéb technológiai haladások figyelembevételével. Már több olyan fejlesztést használnak aktívan, ami kihasználja a gépi tanulás és a neurális hálózatok előnyeit. Például már vannak algoritmusok, amelyek röntgenfelvételeket elemeznek ki, megmutatják mely hatóanyagok lehetnek hatásosak vírusok ellen vagy a tünetek alapján segítenek diagnosztizálni kardiológiai vagy bőrgyógyászati problémákat. [2]

A dolgozatban szeretném megvizsgálni annak is a lehetőségét, hogy a jelenleg elérhető informatikai megoldások közül melyek lehetnek a legalkalmasabbak egy ilyen jellegű projekt vagy feladat támogatásához, illetve, hogy egy modern adattároló és vizualizációs rendszernek milyen követelményeknek kell megfelelnie.

2. AZ EGÉSZSÉGÜGYI ADATOK ELÉRHETŐSÉGE

Többféle módszertan is létezik arra vonatkozóan, hogyan is érdemes egy ilyen projektbe belekezdeni. Mindegyik megpróbálja a lehető leghatékonyabban kezelni a felmerülő problémákat és gyors megoldásokat találni. Én hat pillére tervezem felépíteni a feladatot, amelyek mentén haladva remélhetőleg eljutok a keresett eredményemre:

1. Célkitűzés megfogalmazása
2. Adatok összegyűjtése
3. Adatok megfelelő előkészítése, tisztítása
4. Adatok alakítása a céloknak megfelelően
5. Összefüggések keresése és adatok vizualizációja
6. Visszacsatolás és a folyamat felülvizsgálata

A célkitűzésemet már bevezetőmben megfogalmaztam, a következő lépés átgondolni milyen adatokra van szükségem és hogyan juthatok hozzájuk.

2.1. A feladathoz szükséges egészségügyi és egyéb adatok

Mivel a cél egy globális rendszer kialakítása, országszintű adatokra lesz szükség, ezen belül mindenképpen olyan mutatókra és eredményekre kell építeni, amelyek a legtöbb országról elérhetőek vagy amelyekből lehet következtetni például az adott régió többi országának helyzetére is. A kutatásaim során kiderült számomra, hogy jelenleg nincs olyan szervezet, amely a világ minden államát rá tudná venni, hogy adatokat szolgáltatson az ország egészségügyi (vagy bármilyen egyéb) helyzetéről. Az országok többsége önkéntes alapon szolgáltatja az adatokat.

Emellett felmerül az a probléma is, hogy minden ország más és más alapokra építette fel az egészségügyi rendszerét, eltérőek a kötelezően gyűjtendő adatok a betegekről, vagy akár a teljes népességről szóló statisztikák. Például Magyarországon *„egy egészségügyi dokumentáció: a gyógykezelés során a betegellátó tudomására jutott egészségügyi és személyazonosító adatokat tartalmazó feljegyzés, nyilvántartás vagy bármilyen más módon rögzített adat, függetlenül annak hordozójától vagy formájától”*. [3] Ezen adatok megőrzésére többféle szabály is létezik, magát a dokumentációt az adatfelvételtől számolva 30 évig kell megőrizni, de a zárójelentéseket 50 évig. Ezen felül is megtarthatóak természetesen az adatok például kutatási vagy további gyógykezelési céllal. [4]

A rendszerem kialakításához szerencsére nem ezen konkrét kórházi dokumentációkat és zárójelentéseket szeretném felhasználni, mivel ezek nem hozzáférhetőek a nyilvánosság számára, adatvédelmi problémák is felmerülnek, valamint az ebben található adatokat elemezni sokkal komplexebb feladat lenne. (Különösen, ha a kézzel készített feljegyzésekre gondolok. Ugyanakkor egy külön fejlesztési terület megnyílik ezen adatok által, keresve a megoldást arra vonatkozóan, hogyan is lehetne ezeket egy egységes rendszerbe összegyűjteni és rövid és hosszútávú konzekvenciákat levonni országos, regionális, uniós és világ szinten. Az országos statisztikákkal foglalkozó szervezetek (mint a Központi Statisztikai Hivatal) feladata, hogy minden országban összegyűjtse az adott országra jellemző különféle adatokat és azokból kimutatásokat készítsen. *„A hivatalos statisztikai tevékenység célja, hogy statisztikai információk nyilvánosságra hozatalával valósághű, tárgyilagos képet*

adjon a társadalom, a gazdaság, a környezet állapotáról és annak változásairól az állami szervek, az önkormányzatok, a gazdasági szervezetek – beleértve a pénzügyi piacokat –, a civil szervezetek, a tudományos tevékenységet végzők, a közvélemény, a média szereplői, valamint a nemzetközi szervezetek, különösen az Európai Unió intézményei számára” [5]. Az Európai Unióhoz való csatlakozással minden ország vállalja, hogy a saját nemzeti statisztikáit az Európai Unió szervezetei részére is elérhetővé biztosítja. Ez egyben azt is jelentheti, hogy az ezen szervezetek által összegyűjtött adatok megbízhatóak és nyugodtan felhasználhatóak.

A következő adatsorokra lesz szükségem az egészségügyi rendszer felépítéséhez:

1. Országok népességi adatai, demográfiai megoszlások
2. GDP mutatók és az összjövedelemből egészségügyre fordított költségek
3. Egészségügyi ellátás mutatói (orvosok, nővérek, kórházi ágyak száma)
4. Születéskor várható élettartamok, életminőség
5. Általános egészségi állapotok, betegségek megoszlása
6. Környezetegészségügyi mutatók (légszennyezettség, tiszta ivóvíz)

A megfelelő következtetések levonásához és a trendek megfigyeléséhez több évnyi adatsorra lesz szükségem, illetve célszerű mindenből minél aktuálisabb adatokat is begyűjteni. Ugyanakkor a demográfiai, népességi és gazdasági adatok mindig csak 1-2 év csúszással válnak elérhetővé.

2.2. Források bemutatása

Az adataimat többféle forrásból szerezem be. Szerencsére a legtöbb online elérhető adatsor már szinte teljesen használatra kész formátumban vannak, így könnyedén lehet majd velük dolgozni.

1. WHO központi honlapja

A WHO honlapja tartalmazza talán a legátfogóbb statisztikákat a globális egészségügyi adatokról, így ezen adatokat célszerű innen beszerezni. Ráadásul többféle lehetőség is van az adatok exportálására, emellett közvetlenül is lehet az adatokhoz kapcsolódni api-n keresztül, ha már nem végeznénk további átalakítást az adatokon.

2. Our World in Data honlapja

Ezen a honlapon nem csak egy szervezet adatai találhatóak meg, itt összegyűjtik a táblákat és adatokat több különböző forrásból és értelmezik azokat. Így találunk itt adatokat az ENSZ szervezeteitől, az OECD-től vagy a Világ Banktól [6]. Ebből következik, hogy innen többféle témában is kereshetünk adatokat.

3. OECD honlapja

A Gazdasági Együttműködési és Fejlesztési Szervezet, egy nemzetközi gazdasági szervezet, „*célja az, hogy segítse a tagállamok kormányait a lehető legjobb gazdasági és szociális politika kialakításában és értékelésében*” [7]. Itt főleg gazdasági mutatók találhatóak a 37 tagországról, de egyéb érdekességeket összefüggéseket is lehet találni. Sőt, néhány nem-OECD tagországról is elérhetőek adatok. Szintén többféle formátumban letölthetőek az adatok.

2.3. Forrástáblák bemutatása

Az alábbi táblákat .csv formátumban töltöttem le, majd, hogy egyszerűbb legyen a betöltés az adatbázisba .xlsx formátumba mentettem. A táblák itt még eredeti formátumban vannak, nem alakítok az adatokon ebben a lépésben.

Egyik központi tábla az **országok listája** névvel, kóddal és kontinens/földrajzi elhelyezkedési információval. Ez a tábla teszi lehetővé, hogy országokra vagy földrajzi régiókra szűrve vizsgáljuk meg az adatokat.

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
GeoLocation	nvarchar (255)	Az ország földrajzi elhelyezkedése

1. ábra Counties tábla [8]

Az egyes országokra vonatkozó **GDP adatokat** tartalmazó tábla országra és évekre lebontva 1960-tól 2020-ig.

Oszlop	Adat típus	Megjegyzés
Entity	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
Year	float	A sor adatára vonatkozó évszám
GDP per capita (US\$)	float	Az adott sorhoz tartozó GDP érték amerikai dollárban

2. ábra GDP per capita tábla [9]

Az országok **egészségügyre fordított költségeit** tartalmazó tábla GDP arányosan, országra és évekre lebontva 2000-től 2019-ig.

Oszlop	Adat típus	Megjegyzés
Year	float	A sor adatára vonatkozó évszám
Country	nvarchar (255)	Az ország neve
CHE per capita GDP	float	Az adott sorhoz tartozó egészségügyre fordított GDP érték amerikai dollárban

3. ábra Current Health Expenditure tábla [10]

Az országokban praktizáló **orvosok számának adatait** tartalmazó tábla, lebontva általános és specializálódott orvosokra, valamint országokra és évekre 1952-től 2020-ig

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Year	float	A sor adatára vonatkozó évszám
Medical doctors nr	float	Az összes orvos száma
Generalist medical practitioners nr	nvarchar (255)	Az általános orvosok száma
Specialist medical practitioners nr	nvarchar (255)	Specializálódott orvosok száma
Medical doctors not further defined nr	float	További orvosok száma

4. ábra Medical doctors tábla [11]

Az országokban dolgozó **nővérek, szülésznők és ápolók számának adatait** tartalmazó tábla, országokra és évekre lebontva 1960-tól 2020-ig.

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Year	float	A sor adatára vonatkozó évszám
Total person nr	float	Az összes ápoló száma
Nursing person nr	float	A nővérek száma
Midwifery person nr	float	A szülésznők száma

5. ábra Nursing, Midwifery tábla [12]

A **várható- és a várható egészséges élettartamokat** tartalmazó tábla országokra lebontva. Sajnos ebből a statisztikából csak nagyon kevés elérhető adatsor van, nem készítenek minden évben friss kimutatásokra erre vonatkozóan, így 2000, 2010, 2015 és 2019-es adat áll rendelkezésre.

Oszlop	Adat típus	Megjegyzés
Year	float	A sor adatára vonatkozó évszám
Country	nvarchar (255)	Az ország neve
Life expectancy at birth (years) Both sexes	float	Születéskor várható élettartam
Life expectancy at birth (years)_Male	float	Születéskor várható élettartam férfiak
Life expectancy at birth (years)_Female	float	Születéskor várható élettartam nők

Oszlop	Adat típus	Megjegyzés
Healthy life expectancy (HALE) at birth (years) Both sexes	float	Születéskor várható egészséges élettartam
Healthy life expectancy (HALE) at birth (years)_Male	float	Születéskor várható egészséges élettartam férfi
Healthy life expectancy (HALE) at birth (years)_Female	float	Születéskor várható egészséges élettartam nő

6. ábra Life expectancy tábla [13]

Talán a legfontosabb tábla, amely az adott országok 28 betegségre lebontott **halálozási adatait** tartalmazza 1990-től 2019-ig. A tábla elrendezése ebben a formátumban sajnos megnehezítené a halálokokra való szűrést vagy aggregációt, így ezen biztosan kell majd alakítanom.

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
Year	float	A sor adatára vonatkozó évszám
Outdoor air pollution - OWID	float	Kültéri légszennyezés okozta halálozások száma
High systolic blood pressure	float	Magas vérnyomás okozta halálozások száma
Diet high in sodium	float	Túlzott sófogyasztás okozta halálozások száma
Diet low in whole grains	float	Gabonaszegény étrend okozta halálozások száma
Alcohol use	float	Alkohol okozta halálozások száma
Diet low in fruits	float	Gyümölcszegény étrend okozta halálozások száma
Unsafe water source	float	Szennyezett ivóvíz okozta halálozások száma
Secondhand smoke	float	Másodlagos dohányzás okozta halálozások száma
Low birth weight	float	Alacsony születési súly okozta halálozások száma
Child wasting	float	Vetélés okozta halálozások száma
Unsafe sex	float	Nem biztonságos szex által okozott halálozások száma

Oszlop	Adat típus	Megjegyzés
Diet low in nuts and seeds	float	Olajos magvakban szegény étrend okozta halálozások száma
Household air pollution from solid fuels	float	Szilárd tüzelőanyagokból származó háztartási levegőszennyezés által okozott halálozások száma
Diet low in vegetables	float	Zöldségszegény étrend okozta halálozások száma
Low physical activity	float	Alacsony fizikai aktivitás okozta halálozások száma
Smoking	float	Dohányzás okozta halálozások száma
High fasting plasma glucose	float	Magas éhgyomori vércukorszint okozta halálozások száma
Air pollution	float	Légszennyezés okozta halálozások száma
High body-mass index	float	Magas testzsír százalék okozta halálozások száma
Unsafe sanitation	float	Nem megfelelő higiénia okozta halálozások száma
No access to handwashing facility	float	Kézmosási lehetőség hiánya által okozott halálozások száma
Drug use	float	Drogok okozta halálozások száma
Low bone mineral density	float	Alacsony csontsűrűség okozta halálozások száma
Vitamin A deficiency	float	A-vitamin hiány okozta halálozások száma
Child stunting	float	Fejlődési rendellenesség okozta halálozások száma
Discontinued breastfeeding	float	Szoptatás leghagyása által okozott halálozások száma
Non-exclusive breastfeeding	float	Csecsemő hozzátáplálás okozta halálozások száma
Iron deficiency	float	Vashiány okozta halálozások száma

7. ábra Death Number Riskfactors per country tábla [14]

A népességi adatok alakulását leíró tábla országokra és évekre lebontva 1970-től 2021-ig.

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
Year	float	A sor adatára vonatkozó évszám
Population (historical estimates)	float	A népesség számára vonatkozó adatsorok

8. ábra Population per country tábla [15]

A környezetegészségügyi mutatók egyik fontos adatsora az **elérhető ivóvízre** vonatkozó tábla országokra és évekre lebontva 2000-től 2020-ig.

Oszlop	Adat típus	Megjegyzés
Year	float	A sor adatára vonatkozó évszám
LocationType	nvarchar (255)	Teljes országra, városokra vagy vidékre vonatkozó adatok
Type	nvarchar (255)	A minőségi ivóvízhez jutás lehetőségei
Country	nvarchar (255)	Az ország neve
Value	float	Az ivóvízhez jutás százalékos aránya a lakosságra vetítve

9. ábra Drinking water tábla [16]

Szintén környezetegészségüggyel kapcsolatos adatokat mutat be a **káros anyagok kibocsátását** tartalmazó tábla, országokra és évekre lebontva 1990-től 2015-ig.

Oszlop	Adat típus	Megjegyzés
Entity	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
Year	float	A sor adatára vonatkozó évszám
Carbon Monoxide	float	Szén monoxid kibocsátás
Nox	float	Nitrogén oxid kibocsátás

Oszlop	Adat típus	Megjegyzés
PM10	nvarchar (255)	PM10 Szállópor
PM2.5	nvarchar (255)	PM2.5 Szállópor
Sulphur Oxides	float	Kén Oxid kibocsátás

10. ábra Emmissions tábla [17]

O3 vagy ózon szennyezettség mértékét bemutató tábla országokra és évekre lebontva 1990-től 2015-ig.

Oszlop	Adat típus	Megjegyzés
Country	nvarchar (255)	Az ország neve
Code	nvarchar (255)	Az ország kódja az adatbázisban
Year	float	A sor adatára vonatkozó évszám
Ozone concentration-StateofGlobalAir	float	Az ózon szennyezettség mértéke

11. ábra O3 concentration tábla [18]

Érdekességképpen egy extra táblát is letöltöttem az adatbázishoz, amely a föld népességének alakulását mutatja be, az adatok túlnyomó része becslés alapján, kerekítve van. (i.e 10.000-től 2022-ig)

Oszlop	Adat típus	Megjegyzés
Year	float	A sor adatára vonatkozó évszám
World Population (Spline Interpolation until 1950)	float	A becsült népesség száma

12. ábra Population world tábla [19]

3. ADATTÁRHÁZ TERVEZÉSE

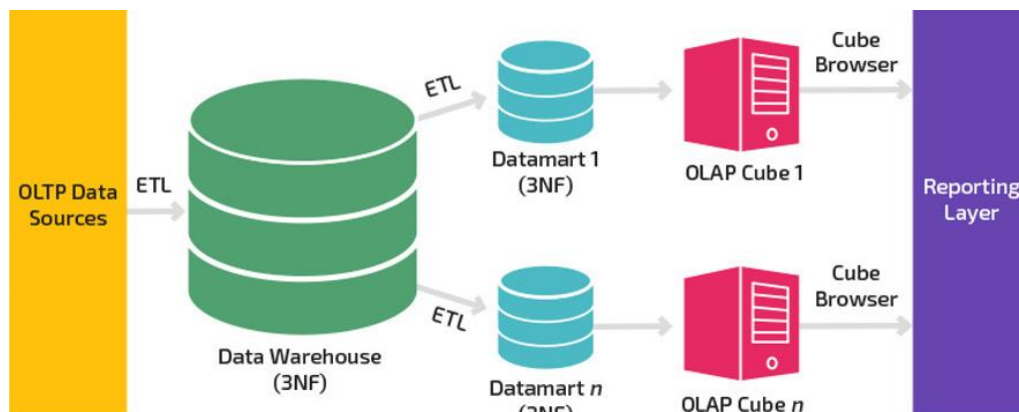
Az összegyűjtött adatokhoz elengedhetetlen, hogy megfelelő helyen tároljuk őket, ahol lehetőségünk van a célnak megfelelő átalakításra is akár. Ennek a feladatnak az eszközei lehetnek az adattárházak, melyek alapfogalmát, funkcióit és felépítését tekintsük röviden át.

3.1. Adattárházak alapjai

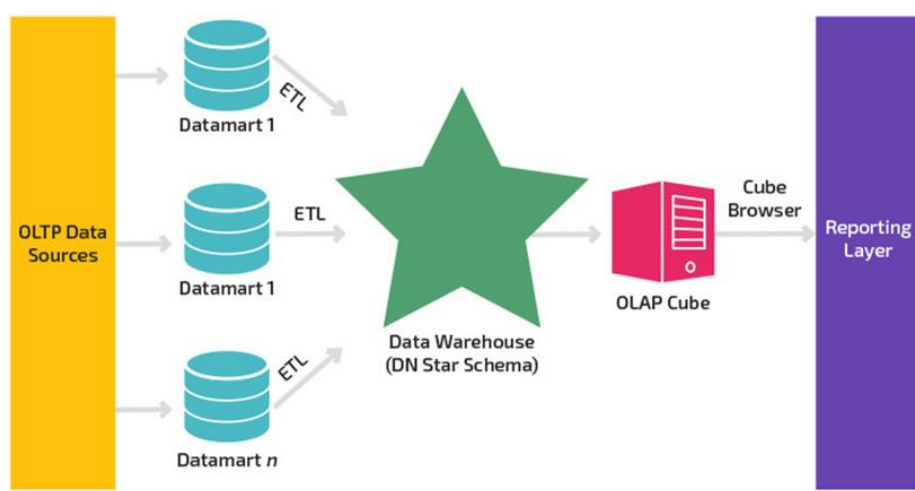
3.1.1. William H. Inmon definíciója

"A data warehouse is a subject oriented, integrated, nonvolatile, and time variant collection of data in support of management's decisions." - Az adattárház témakör-orientált, integrált, tartós, időfüggő, de időben nem változó adatok gyűjteménye, amelyet a cég/projekt stratégiai döntéshozatalának támogatására használnak [20]. A definíció nagy hangsúlyt fektet az adattárházakkal szembeni elvárásokra.

3.1.2. Különböző adattárház szerkezetek



13. ábra Inmon féle Data Warehouse struktúra [21]



14. ábra Kimball féle Data Warehouse struktúra [22]

3.2. Adattárházak jellemzői

A definíciójában megfogalmazott követelmények egyben tekinthetők a mai modern adattárházakkal szembeni elvárásoknak is, így vizsgáljuk meg őket részleteikben.

1. Témaorientált rendszer; az adattárházak tervezésénél az úgynevezett „data-driven” tematikát szokás követni, azaz az adatok határozzák meg, hogyan kerülnek az adatok csoportosításra. Gyakorlatban ez annyit jelent, hogy az adatok témák szerint vannak adathalmazokra bontva.
2. Integrált; szorosan kapcsolódva a témaorientáltsághoz, a forrás adatokat igyekezni kell egységes rendszerben összegyűjteni és egységbe rendezve eltárolni.
3. Tartós; arra utal, hogy az adattárházban eltárolandó adatok változatlanok. Csak abban az esetben kerülhet sor változásra, ha a forrásrendszerben is megváltoznak az adatok, ebben az esetben viszont verziókövetés elengedhetetlen.
4. Időfüggetlenség; a forrásrendszerekben az adatok rendszerint csak egy adott időpillanatról adnak képet. Ezzel szemben az adattárházban időről időre tároljuk az adatokat, megfelelő időbélyegzővel ellátva azokat, így lehetőség nyílik rá, hogy az adatsorok változásait és a trendeket megfigyeljük.

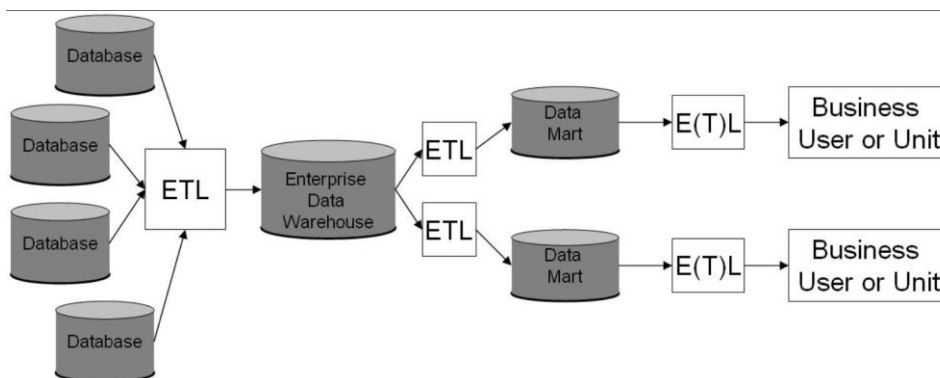
3.3. Adattárházak funkciói

Egy adattárháznak többféle funkciót is el kell tudnia látni:

1. Több forrásból származó, heterogén adatokat is kell tudnia integrálni.
2. A heterogén adatokat formailag és logikailag egyesítenie kell.
3. Képesnek kell lennie nagy adathalmazok befogadására, átalakítására és tárolására is egyaránt.
4. Az adattárházban tárolt adatok elsősorban döntéstámogató céllal vannak felhasználva, így nagyon fontos könnyen lehessen lekérdezéseket írni, akár több táblát összekapcsolva is.
5. A kezelő felületének relatíve könnyen kezelhetőnek kell lennie, ezzel is elősegítve az adattárház fejlesztők munkáját és a felhasználóit.
6. Kiemelten fontos továbbá az adatok időbélyegzése, és ehhez kapcsolódva a historikus adatok megőrzése is, verziókezelés. [23] [24]

3.4. Az adattárház komponensei

Ahhoz, hogy az adattárházak megfelelően tudjanak működni több komponensre is szükség van, amelyek összhangja szükséges a rendszer stabilitásához és folyamatos működéséhez (15. ábra).



15. ábra Az adattárház komponensei [25]

1. Adatforrás rendszerek, ahonnan az adatokat összegyűjtjük, ezeket általában nyersen, átalakítás nélkül töltjük be az adattárházba
2. ETL folyamat, Extract azaz „kinyerési” szakasza, ezzel szerezzük meg magunkat az adatokat és töltjük be az adattárházba (Data Warehouse a 15.ös ábrán)
3. ETL folyamat, Transform azaz „átalakítási” szakasza, amellyel a nyers adatokat átalakítjuk, megfelelő típusokat adunk a különböző adatoknak, esetleg mértékegységeket határozzunk meg, nyelvi szótárakat iktatunk be. Itt végezzük el az adatok ellenőrzését és tisztítását is. A folyamat végén a Data Mart rétegbe kerülnek az átalakított adatok.
4. ETL folyamat, Load azaz „betöltési” szakasza, amikor a már elkészült adatsorokat a végleges adatbázisokba töltjük be és azokat a felhasználók számára is könnyen kezelhetővé tesszük.

Már korábban is említettem, hogy a feladatom célja az egészségügyről elérhető adatok megfigyelése, a változások nyomon követése és ezekből megpróbáljak következtetéseket és becsléseket készíteni. Az adattárháza alapfogalmaiban megfogalmazott követelmények egy adattárházzal szemben tökéletesen támogatják ezen célom elérését, már csupán a megfelelő eszközt kell megtalálni a megvalósításhoz.

3.5. Adatbázis megoldások

A feladathoz szükséges elérhető forrásadatok viszonylag jól strukturáltak, könnyen alakíthatóak. Mivel szinte kivétel nélkül táblázatos formában kerülnek eltárolásra globálisan az egészségre vonatkozó adatok így célszerű ennek megfelelően egy relációs adatbázist választani az adattárház kialakításához, ráadásul a redundanciát így minimálisra lehet csökkenteni.

3.5.1. Adatbázisok alapfogalmai

Röviden tekintsük át, hogy mik egy adatbázis alapfogalmai, melyekkel érdemes tisztában lenni.

1. Egyed vagy entitás
2. Reláció vagy tábla, ennek oszlopainak neve a tulajdonságok, sorai a rekordok
3. Több táblázat alkot egy adatbázist
4. Elsődleges kulcs, idegen kulcs

5. Leírni egy relációt annak nevével és tulajdonságainak felsorolásával tudunk [26]

3.5.2. Adatmodellezés és tervezés

Célja a valóságban előforduló adatok leírása, az adatok összetevőinek reprezentálása, összefüggések, megszorítások megfogalmazása. Az adatmodell olyan eszközök összessége, amelyek segítségével leírhatóak az adattípusok, az adatok közötti kapcsolatok és ezek megszorításai. Ennek három lépése van:

1. Koncepcionális adatmodell készítése, amely független a valós fizikai végeredménytől, a felhasználók szempontjából reprezentálja az adatokat. Egyed-kapcsolat modell leírása.
2. Logikai adatmodell készítése, amely átalakítja a koncepcionális modellt, akár relációs modellé, emellett megjelenhetnek az adatokon végzett műveletek is.
3. Fizikai adatmodell készítése, ahol már meg kell adni formálisan is egy leíró programnyelvvvel az adatbázissémát. [27] [28]

3.5.3. Relációs adatbázis tervezése és kiválasztása

Az elérhető és legnépszerűbb relációs adatbázisok 2021 márciusában a teljesség igénye nélkül az Oracle, MySQL és a Microsoft Server SQL és a PostgreSQL [29]. A feladat végrehajtásához és az adatok tárolására elviekben bármelyik fenti eszköz megfelelhet, így tekintünk át őket röviden, melyiknek mi az előre vagy hátránya. [30]

1. Oracle

Az egyik legrégebbi és legelterjedtebben használt relációs adatbázis kezelő rendszer. Többféle programnyelvet is támogat, lekérdezőnyelve a PL/SQL. Lehetőség van tárolt eljárások és triggerok létrehozására, valamint többféle operációs rendszeren is futtatható. Kezelőfelülete ugyanakkor összehasonlítva a piacon lévő egyéb rendszerekkel, nem túl felhasználóbarát.

2. MySQL

Nyílt forráskódú adattároló rendszer, 2010-ben az Oracle-hez került, így már ők biztosítanak támogatást ehhez a rendszerhez is. A nyílt forráskód miatt nagyon népszerű lett a honlapok, az e-kereskedelem és a naplózási tevékenységet végző szolgáltatások körében.

3. PostgreSQL

Szintén egy nyílt forráskódú, talán az egyik legerősebb ingyenes rendszer. Nagy előnye bővíthetősége az egyedi igényeknek megfelelően. Emellett az adatbázis válasz idejét próbálják folyamatosan optimalizálni, hogy még gyorsabban elérhető legyen a rendszer.

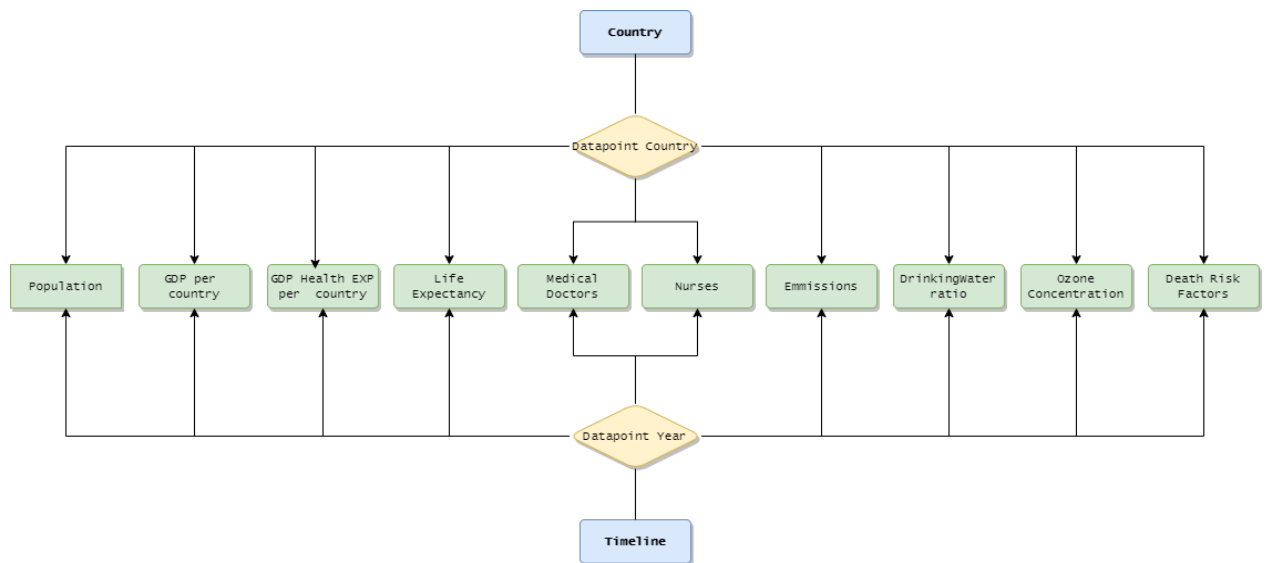
4. Microsoft Server SQL

A Microsoft saját adatbázis-kezelő szolgáltatása. Többnyire a globális, nagyvállalatok körében a legelterjedtebb, akik a Microsofttól kedvezményes csomagajánlat részeként hozzá tudnak jutni. Kiválóan tudja támogatni a nagyvállalati informatikai környezeteket a tranzakciók gyors feldolgozásával és üzleti intelligenciával. Lekérdező nyelve a Transact-SQL.

A feladatom megoldásához és az adatok tárolásához a Microsoft Server SQL-t fogom használni, melyre több okom is van. Egyrészt már elég nagy rutinra tettem szert ennek az eszköznek a használatában, másrészt úgy gondolom, hogy a Microsoft hosszú évek óta azon dolgozik, hogy az általa létrehozott eszközök egymással működjenek a leoptimalisabban. Emellett igazán felhasználóbarát és könnyen kezelhető a felülete és remekül hozzá lehet kapcsolni további adatmodellezéshez szükséges eszközöket is.

3.5.4. Az egészségügyi adatokhoz EK Diagram megtervezése

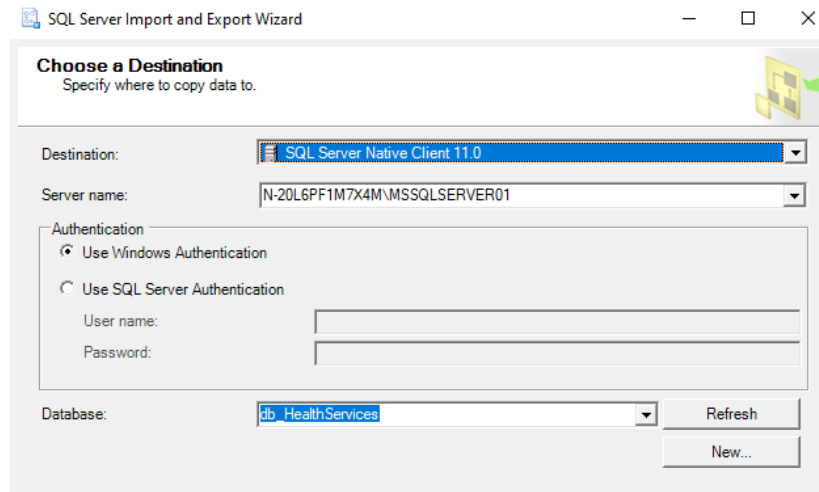
A forrás táblák kiválasztása után elkészítettem egy EK diagramot, amelynek segítségével már látható, mely táblák hogyan kapcsolódnak egymáshoz.



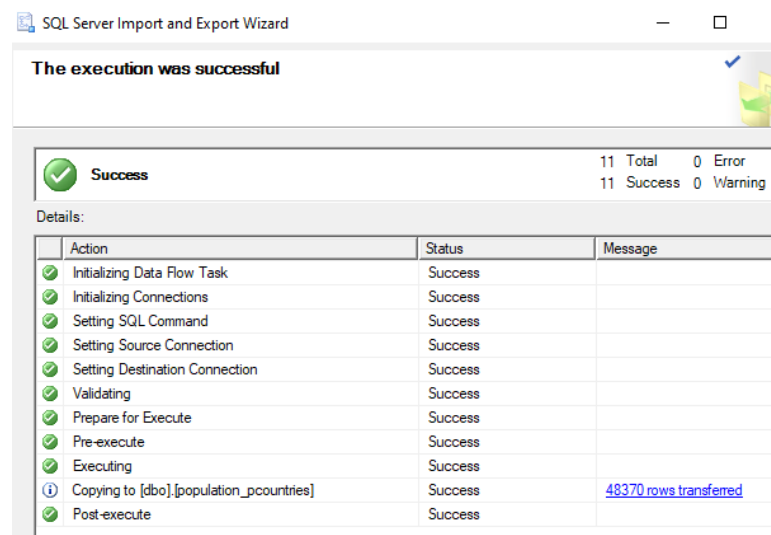
16. ábra EK Diagram az egészségügyi adatokhoz

3.6. Az egészségügyi adatok betöltése, racionalizálása

Első feladatként létrehoztam egy új adatbázist db_HealthServices néven az SQL Management Studio segítségével lokálisan. Majd az .xlsx formátumra mentett fájlokat betöltöttem az adatbázisba. A betöltési folyamatot a 17. és 18-as ábrákon mutatom be.



17. ábra Tábla betöltése SQL natív kliens segítségével



18. ábra Sikeres betöltési folyamat

Mivel ez az Extract folyamat része, így itt nem végzek átalakítást az adatokon és nem álltom át az adattípusokat. A sikeres betöltések után ellenőriztem, hogy minden folyamat rendben lefutott és a táblák valóban megjelennek az adatbázisban. (38-as ábra)

A következő folyamat az adatok megfelelő átalakítása, hogy majd a betöltési folyamatban könnyedén tudjak velük dolgozni. Ezt végrehajthatnám egy újabb adatbázis létrehozásával, majd SSIS package segítségével az adat konverziókat is könnyedén elvégezhetném. Azonban ezt a folyamat jól strukturált adatoknál akár a vizualizációs eszköz segítségével is elvégezhető, ezt a következő fejezetben be is mutatom.

Előtte tekintsük át, hogy miért is kiemelten fontos napjainkban a jó vizualizációs eszköz, nem csak fejlesztői, hanem felhasználói szempontokat is figyelembe véve.

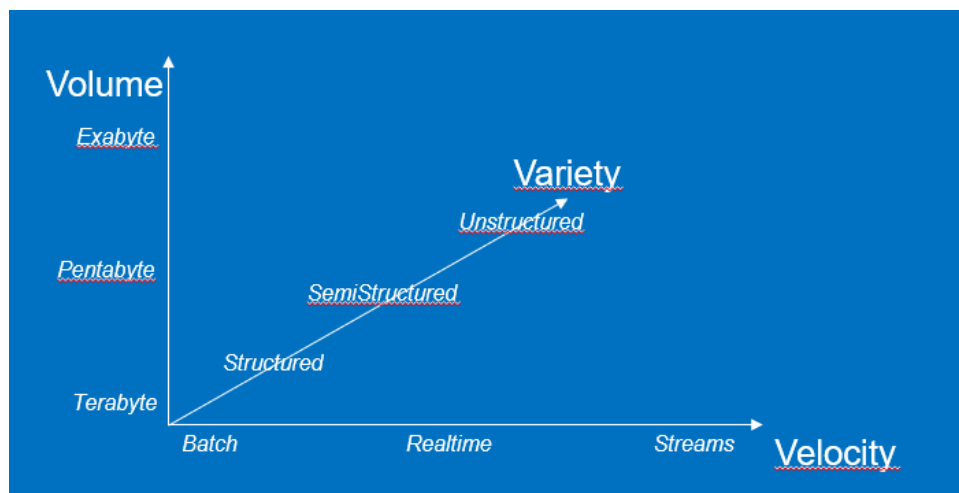
4. VIZUALIZÁCIÓS MEGOLDÁS TERVEZÉSE

A feladatom soron következő pillére az adattárház létrehozása után, amely talán az egyik legfontosabb lesz, az adatok harmonizációja majd vizualizálása, ahol értelmet tudnak a hosszú adatsorok nyerni és felfedezhetjük azon összefüggéseket, amelyek megtalálása az alapvető célkitűzésem volt.

Az adatok megjelenítése vagy vizualizálása mára egy elengedhetetlen követelmény lett a big data világában. Legyen szó óriás vállalatokról, statisztikai hivatalokról, kormányokról akár iskolai vagy egészségügyi intézményekről, mindegyik egyre több adatot állít elő, amelyek potenciálisan képet adhatnak a jelenlegi működési helyzetről és a körülményekről.

4.1. Big Data jellemzői

A big data-t nehezen lehet megfogalmazni, hogy mit is jelent pontosan vagy mit foglal magában. Talán az egyik legjobb és legerőteljesebb az úgynevezett 3-V pilléren alapuló megközelítés. [31]



19. ábra Big Data 3V pillére

1. Volume / Mennyiség; talán a legegyszerűbb jellemzője a big data-nak. Az emberiség soha nem állított elő annyi adatot, mint manapság, köszönhető ez az egyre jobban digitalizálódó világban, hogy lassan minden tevékenységünkhöz adatok előállítására van szükség, és ezeket tárolunk is kell akár fizikai vagy online/felhőben szolgáltatásban. Továbbá a Dolgok Internete (*Internet of Things*) elterjedésének is, nagy szerepe van az óriás mennyiségű adat keletkezésében. A strukturált adatok (excel tábla) mellett a félig-strukturált (html, json xml) vagy teljesen strukturálatlan (képek, szövegfájlok, videók) adatok mennyisége is exponenciálisan növekedik
2. Velocity / Sebesség; lényegében arra mutat rá, hogy mennyire gyorsan tudunk az adatokhoz jutni, illetve ezen igényünk is folyamatosan növekedik. Léteznek valós időben érkező adatok (real-time), illetve olyanok, amelyek lassabban több részletben állnak rendelkezésre (batch).

3. Variety / Sokféleség; az adatok, amelyeket gyűjtünk és feldolgozunk ma már nagyon sokfélék lehetnek és ezeknek feldolgozása ma már ugyanannyira fontos, mint a régi hagyományos fájloké.

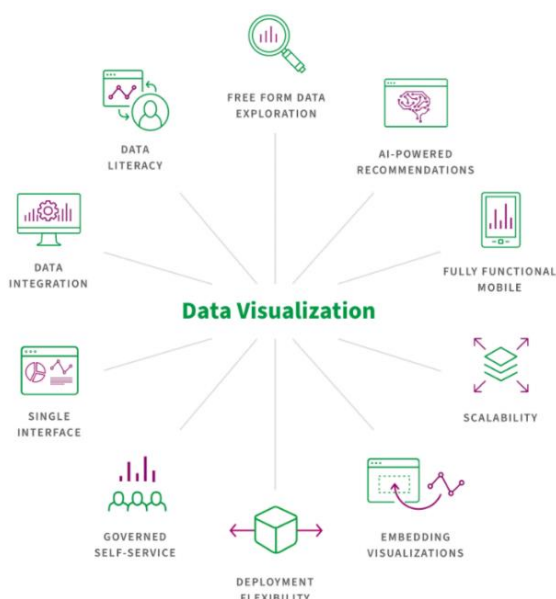
4.2. Az adatvizualizáció jelentősége

A big data korában az adatok prezentálása magában foglalja szinte bármely elemi adat vizualizálását és grafikus formátumban való megjelenítését, amely segíti a megértését és a döntéshozatalt is. A vizualizációk adnak igazi értelmet akár egy laikus számára is, hogy a mögöttes tartalom mit is jelent azáltal, hogy szemmel látható kontextusba helyezik. Egy grafikon, egy vonaldiagram vagy akár egy térképen elhelyezett jelölő sokkal közelebbivé, megfoghatóbbá és természetesebbé teszi mindenki számára az információ megértését és feldolgozását. Emellett, így könnyebb lehet az összefüggéseket realizálni, leolvasni a trendeket vagy szélsőértékeket.

A régi, klasszikus grafikonok és vonaldiagramok mellett ma már akár térképeken is tudunk területi megoszlásokat ábrázolni (ami a Covid-19 járvány miatt rendkívül népszerű lett), hőtérképeken láthatjuk a kritikus pontokat és azokat a területeket, amelyek elmaradtak.

4.3. Vizualizációs eszköz tervezése

A nagy adatmennyiséghez és a célok legoptimálisabb támogatásáért a vizualizációs eszköznek az alábbi követelményeknek érdemes megfelelnie [32] [33]:



20. ábra Adatvizualizációs eszközök funkciói

1. Többféle típusú és forrásból származó érkező adatot is fel kell tudnia dolgozni
2. Nagy mennyiségű adat feldolgozásával is el kell boldogulnia az eszköznek és ábrázolni az grafikonokat

3. Képesnek kell lennie a trendek bemutatására, hiszen azért is „gyűjtünk” időbélyegzővel ellátott historikus adatokat is, hogy a változásokat megfigyelhessük. Ez nehézséget okozott régebben, mikor többféle forrásnak kellett az adatait feldolgozni
4. Tudnia kell szűrőket alkalmazni, amelyekkel az eredményeket mindenki egyénileg be tudja állítani
5. Az adatforrást folyamatosan el kell tudja érni, hogy rögtön mutassa a változásokat
6. Egy jó vizualizációs eszközzel könnyedén felismerhetőek és átláthatóak az adathalmazok közötti összefüggések és kapcsolatok, míg, ha azok első ránézésre nem is tűnnek túl evidensnek. Az adathalmazok közötti korábban fel nem fedezett minták és kapcsolatok azonosítása hatalmas versenyelőnyt jelenthet a végfelhasználók számára.
7. Felhasználóbarátnak kell lenni, mind a riportkészítők és a végfelhasználók szempontjából is. A döntéshozatal elősegítéséhez gyorsan rá kell világítania és fókuszálnia azokra a területekre, amelyek beavatkozást igényelnek.

A fenti követelményeknek manapság rengeteg szolgáltató igyekszik megfelelni, és folyamatosan fejlesztik a vizualizációs megoldásaikat, hogy jó eszközök álljanak rendelkezésre. A négy talán legismertebb és legelterjedtebb vizualizációs eszköz a következő, a teljesség igénye nélkül:

1. Jupyter

Egy nyílt forráskódú eszköz, amely több programozási nyelven kínálja a nagy mennyiségű adatok feldolgozására és vizualizálására a megoldást. A felületén egyszerű kódok megadásával tudjuk importálni az adathalmazainkat és azokon az átalakításokat elvégezni (35. ábra). Az elkészült kódokat és vizualizációkat ráadásul könnyedén meg lehet osztani a csapattagokkal, ezzel elősegítve az együttműködést a különböző belső folyamatok között. Az eleinte Python és R nyelvre épülő adatbevitel mára kiegészült Java, Ruby és C# nyelvvel is, illetve több keretrendszer használatát is lehetővé teszi, így mindenki egyénileg tudja eldönteni, hogy melyik a legszimpatikusabb egy adott feladat elvégzésére. [34]

2. Tableau

A Tableau talán jelenleg a piacvezetőnek számító eszköz, elég, ha csak valaki megnézi, hogy milyen követelmények vannak egy adatelemzővel szemben, a Tableau szinte biztosan az elvárások között lesz. Népszerűségére úgy tett szert, hogy hatékony és interaktív megoldásokat kínál a célok megvalósításához és kihasználja a mesterséges intelligencia és a gépi tanulás adta lehetőségeket is, hogy összefüggéseket keressen az adatok között. Az eszköz integrálható egyéb rendszerekbe is, így azokkal együtt még hatékonyabban kihasználhatóak az előnyei (Amazon AWS, MySQL, Hadoop). [35] Kezelőfelülete grafikus, amely nagyon megkönnyíti az elemzők munkáját, ezen felül nagyon sokféle megjelenítési eszköz közül lehet választani és egyénre szabni azokat, amelyeket nagyon látványos úgynevezett „dashboard” -ok tudnak létrejönni (36. ábra).

3. Google Data Studio

A Google az elmúlt évek során olyan népszerűségre tett szert, amellyel a világ vezető vállalatai közé emelkedett. (A Forbes Global 2000-es listáján az anyavállalata az

Alphabet Inc. a 13. helyen található.) [36] Talán ennek a népszerűségnek is köszönhető, hogy a Google Data Studio is egyre széleskörűbben használt vizualizációs eszköz, ráadásul a Google folyamatosan fejleszti és összekapcsolja a többi szolgáltatásával. Rengeteg megjelenítési típust kínál, az egyszerűbb kördiagramoktól és idősoroktól kezdve egészen a többdimenziós interaktív mátrixokig. (37. ábra).

4. Microsoft Power BI

A Microsoft saját adat vizualizációs eszköze melyre szintén igaz, ami az SQL Serverre, hogy többnyire a globális, nagyvállalatok körében a legelterjedtebb, akik csomagajánlat részeként hozzá tudnak jutni (a pro verziójához). Felhő alapú szolgáltatás is kapcsolódik az eszközhöz, amely nagyon rugalmasan alakítható az igényekhez, szintén nagyon sokféle adatot lehet importálni és elemezni. A kezelőfelülete „drag & drop” technológiára épül, amely nagyban megkönnyíti egy adatelemző munkáját. Az asztali alkalmazás mellett lehetőség van a riportokat az online Power BI Service-be is feltölteni, ahol azokat megoszthatjuk másokkal is, valamint a riportok optimalizálhatóak akár mobil eszközökre is.

A feladatomhoz szükséges elemzéseket és adatvizualizációkat a Power BI eszköz segítségével fogom elkészíteni, mivel ennek az eszköznek a használata az én egyéni és a feladat igényeinek tökéletesen megfelel, és a fent említett előnyökön kívül van még olyan funkciója is a Power BI-nak, amely miatt kiemelkedhet a többi vizualizációs eszköz közül. Ezeket a funkciókat a következő pontban mutatom be.

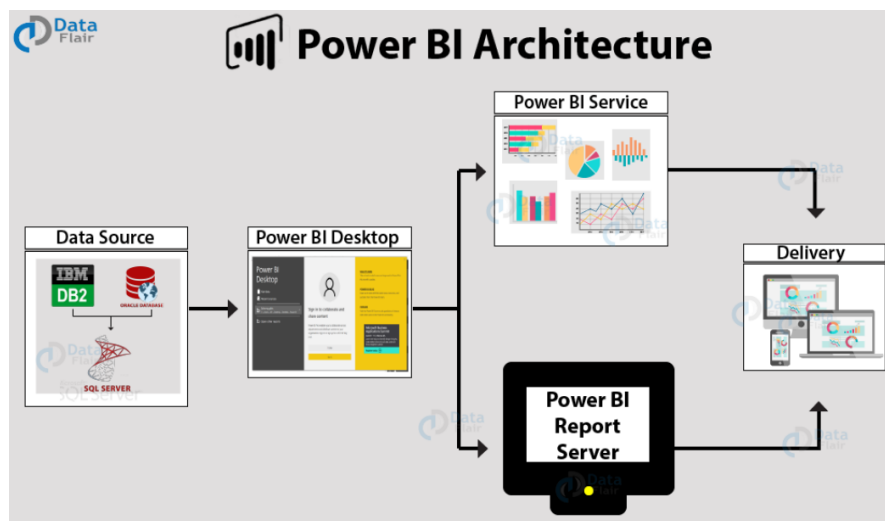
4.4. A Power BI kiemelkedő előnyei

A Power BI (továbbiakban PBI) három alapelemből áll, ezek a PBI Desktop (asztali alkalmazás), PBI Service (online alkalmazás) és a PBI Mobile Apps. A központi eleme a rendszernek a PBI Report Builder (riport készítő) eszköz, amely az asztali és az online eszközhöz kapcsolódik, ebben készíthetők el a riportok és jelentések.

Egy „átlagosnak” mondható munkafolyamatban az adatokat betöltjük az asztali alkalmazásba, ott, ha szükséges a query editor segítségével átalakítjuk őket, elkészítjük a vizualizációkat, majd az eredményeket megosztjuk a célközönséggel vagy döntéshozókkal. Ezt megtehetjük offline vagy a PBI Service-n keresztül online formában is. Természetesen a folyamat nem ennyire egyszerű minden esetben, lépésenként vizsgáljuk meg a fontosabb állomásokat.

4.4.1. Adatforrások csatlakoztatása és formázása

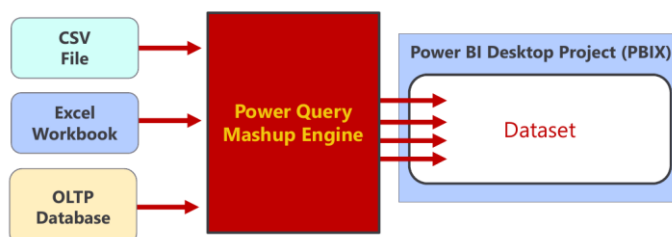
Szinte felsorolni is nehéz, hogy mennyiféle adatforrást képes a PBI kezelni és betölteni akár a helyi vagy a globális hálózatról. A teljesség igénye nélkül dolgozhatunk flat fájlokkal, excel fájlokkal, xml, json fájlokkal, kapcsolódhatunk adatbázis szerverekhez, CRM rendszerekhez, data lake-hez, python és r kódokkal létrehozhatunk saját queryket, illetve a PBI Power Platformjaihoz is tudunk kapcsolódni.



21. ábra Power BI munkafolyamata [37]

Az adatokhoz való csatlakozás után a query editor segítségével lehetőség van azok formázására, megfelelő típusra alakítására, sőt, akár teljesen meg is változtathatjuk egy adott forrás szerkezetét attól függően, hogy mi a célunk vele.

A PBI-ban található Power Query Editort értelmezhetjük akár egy ETL eszközként is, amit az adattárházak fejezetben kifejtettem. Ráadásul a PBI-ban található beépített eszköz nagyon könnyen kezelhető felülettel rendelkezik, az alap átalakításokhoz nem igényel magas szintű informatikai ismereteket, de látványos átalakításokat tud végezni.

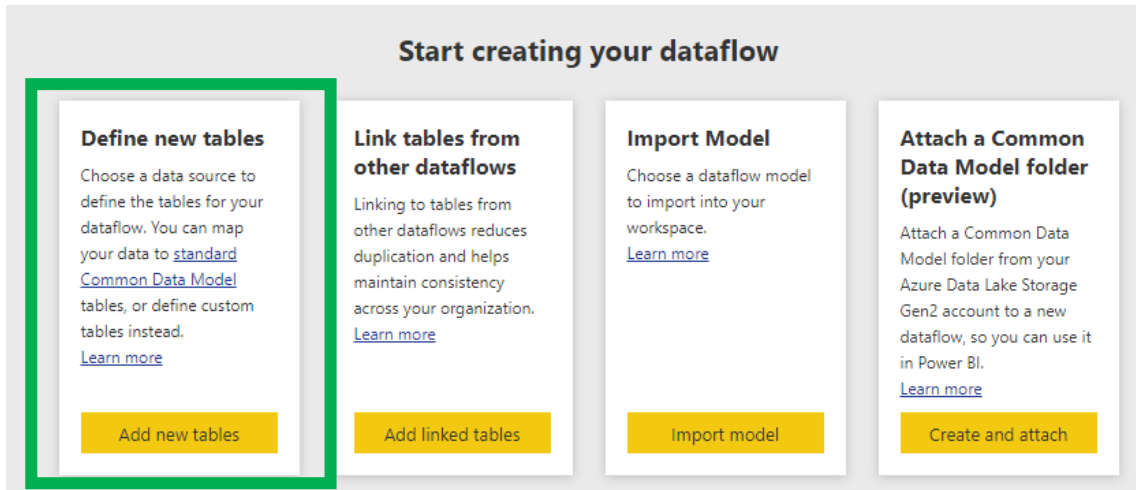


22. ábra A Power Query mint ETL eszköz [38]

4.5. Power BI folyamatai

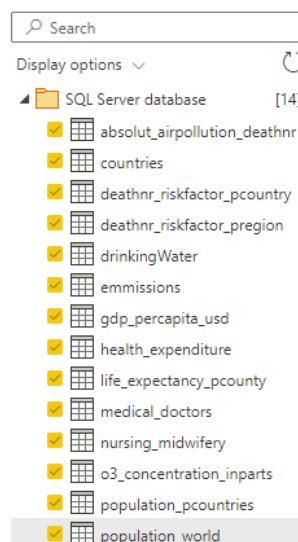
A következőkben bemutatom, hogy az adatbázisba betöltött tábláimat, hogyan is töltöm fel PBI-ba és milyen átalakításokat végzek rajtuk a Transform folyamat részeként.

1. Az adatátalakításokhoz létrehozok egy PBI Dataflow-t, melyet a Service részeként érek el egy munkaterületen belül (23. ábra). Amennyiben az adatbázisomon engedélyeztem a távoli elérését akkor meg tudom adni az elérési utat, majd a felhasználó nevem és jelszó megadásával már hozzá is férek az adatbázishoz (24. ábra).



23. ábra Power BI forrás választás

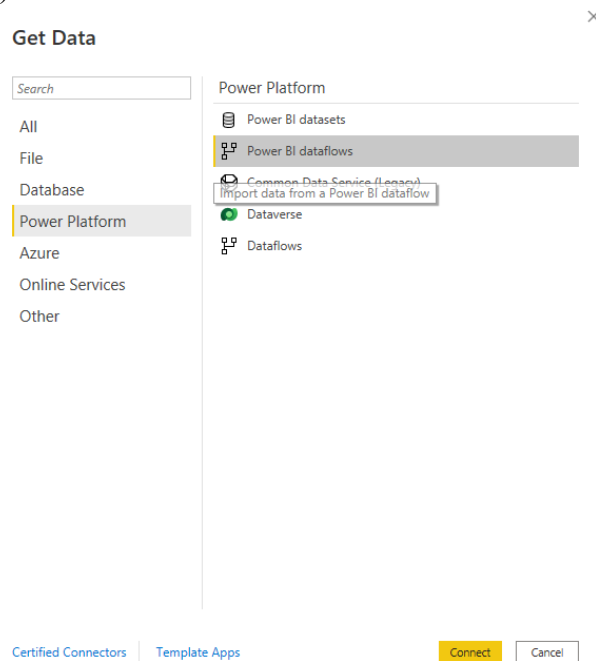
Choose data



24. ábra Adatbázisban elérhető táblák

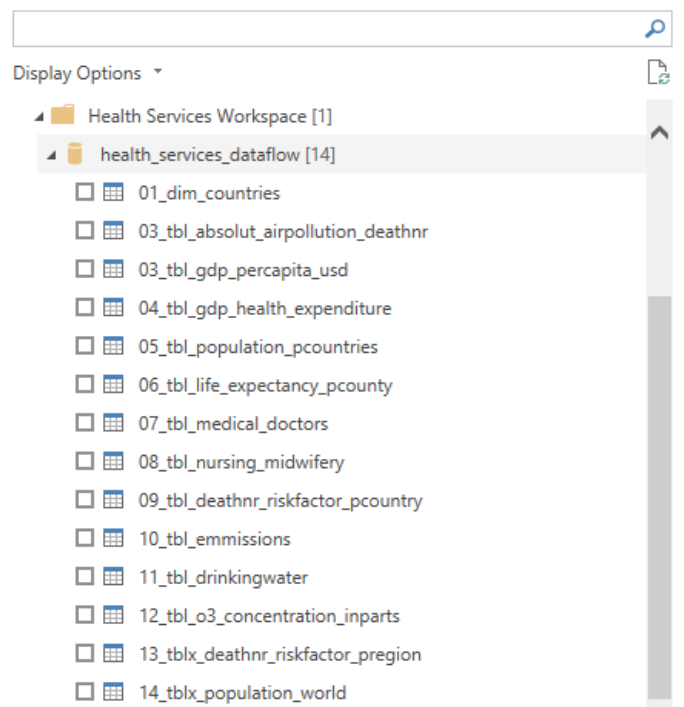
2. Miután kiválasztottam mely táblákra van szükségem hozzá is látok a táblák átnézésének és az adatok átalakításának. Az EK diagramból világossá vált számomra, hogy két tábla köré fogom csoportosítani a többi. A Countries tábla, amellyel az egyes országok adataira tudok majd szűrni a vizualizációkban. Valamint a Timeline tábla, amelyet nem internetes forrásból szereztem be, hanem majd a többi tábla szerkezete alapján és a célok figyelembevételével PBI-ban fogok létrehozni.
3. A következő lépésben mindegyik táblát sorszámmal láttam el és megkülönböztettem a dimenzió és ténytáblákat.
4. A Countries táblán nem végeztem átalakítást, egy szűrőt állítottam be, hogy az esetlegesen üresen érkező sorokat kiszűrje a dataflow. A tábla végleges szerkezete a 42. ábrán látható.
5. A GDP adatokat tartalmazó táblán már szükség volt némi átalakításra. Töröltem azokat a sorokat, amik összesített vagy régiós adatokat tartalmaznak, eltávolítottam a kód oszlopot és beállítottam az adattípusokat. Az átalakított tábla a 43-as ábrán látható, valamint alatta a kód az átalakításokról.
6. A legnagyobb átalakítást a Death Number Riskfactors per country táblán kellett végrehajtani, mivel itt az egyes halálokok oszlopként szerepeltek, de már tudtam hogy más csoportosításra lesz szükségem, hogy a vizualizációkban megfelelően tudjam használni. Így az *unpivot other columns* segítségével az ország és év adatok kivételével „kifejtettem” a táblát így a végleges szerkezete: County-Year-Type-Deceased N (49.ábra).
7. A mellékletben elhelyeztem a táblák végleges képét, valamint az M kódot az átalakításokról.
8. A dataflow mentése és frissítése után egy új PBI desktop verziót nyitottam és a források közül a Power BI dataflows-t választottam (25. ábra), majd megkerestem

a munkaterületen az elmentett dataflowt és kiválasztottam a használni kívánt táblákat (26. ábra).



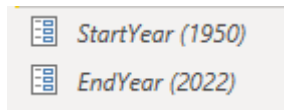
25. ábra Power BI dataflow forrás választása

Navigator



26. ábra A dataflow tábláinak kiválasztása

9. A hiányzó Timeline táblát az asztali verzióban hoztam létre 02_dim_timeline névvel egy új üres query-ben az alábbi paraméterek és kód segítségével:

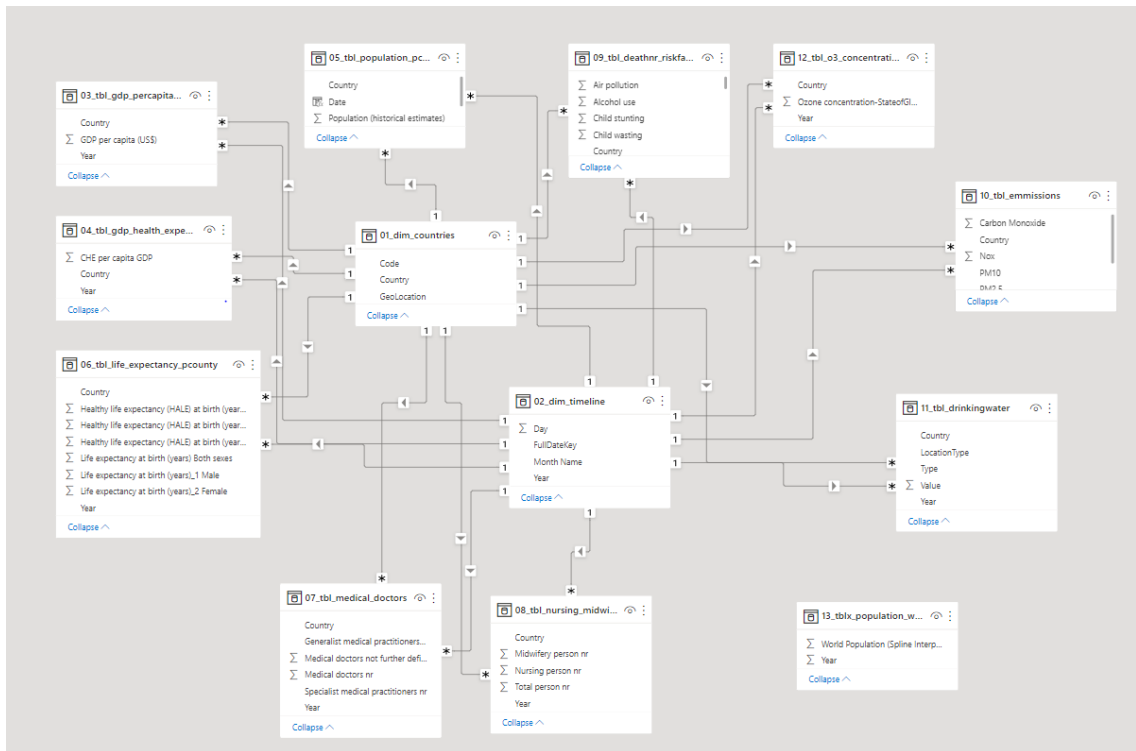


Paraméterek:

```
let
    StartDate = #date(StartYear, 1, 1),
    EndDate = #date(EndYear, 1, 1),
    NumberOfDays = Duration.Days (EndDate - StartDate),
    Dates = List.Dates(StartDate, NumberOfDays+1, #duration(1,0,0,0)),
    #"Converted to Table" = Table.FromList(Dates, Splitter.SplitByNothing(),
    null, null, ExtraValues.Error),
    #"Renamed Columns" = Table.RenameColumns(#"Converted to
    Table",{{"Column1", "FullDateKey"}}),
    #"Inserted Year" = Table.AddColumn(#"Renamed Columns", "Year", each
    Date.Year([FullDateKey]), Int64.Type),
    #"Inserted Month Name" = Table.AddColumn(#"Inserted Year", "Month
    Name", each Date.MonthName([FullDateKey]), type text),
    #"Inserted Day" = Table.AddColumn(#"Inserted Month Name", "Day", each
    Date.Day([FullDateKey]), Int64.Type),
    #"Filtered Rows" = Table.SelectRows(#"Inserted Day", each ([Month Name]
    = "January") and ([Day] = 1)),
    #"Reordered Columns" = Table.ReorderColumns(#"Filtered Rows",{"Year",
    "Month Name", "Day", "FullDateKey"})
in
    #"Reordered Columns"
```

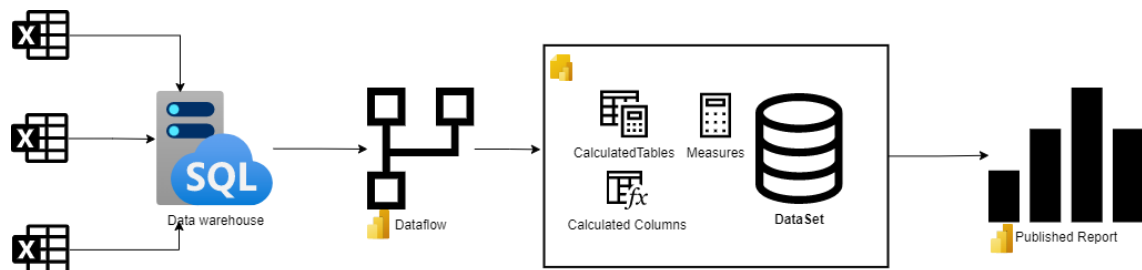
10. Utolsó lépésben beállítottam a kapcsolatokat a táblák között (27.ábra).

Ezzel a Transform és a Load folyamat is a végére ért. A PBI nyújtotta lehetőségekkel sikerült a táblákat megfelelően előkészíteni további használatra.



27. ábra A betöltött táblák és kapcsolataik Power BI-ban

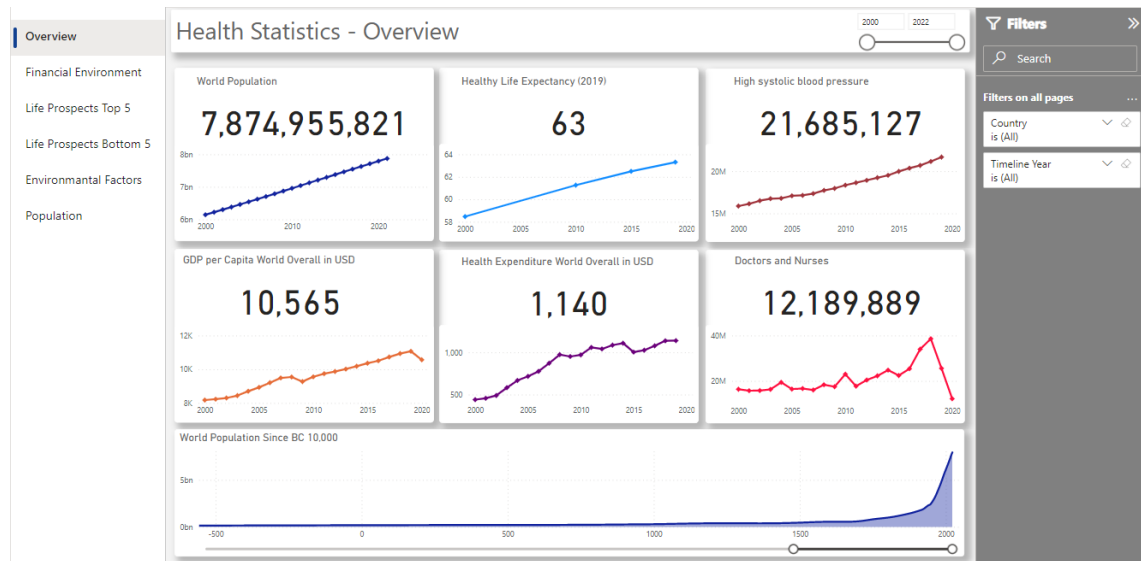
4.6. Az elkészült rendszer tervezete



28. ábra Rendszerterv az egészségügyi adatokhoz

5. AZ EGÉSZSÉGÜGYI ADATOK VIZUALIZÁCIÓJA

1. Áttekintő oldal az egészségüghöz kapcsolódó legfontosabb adatokról (29.ábra).



29. ábra Overview page

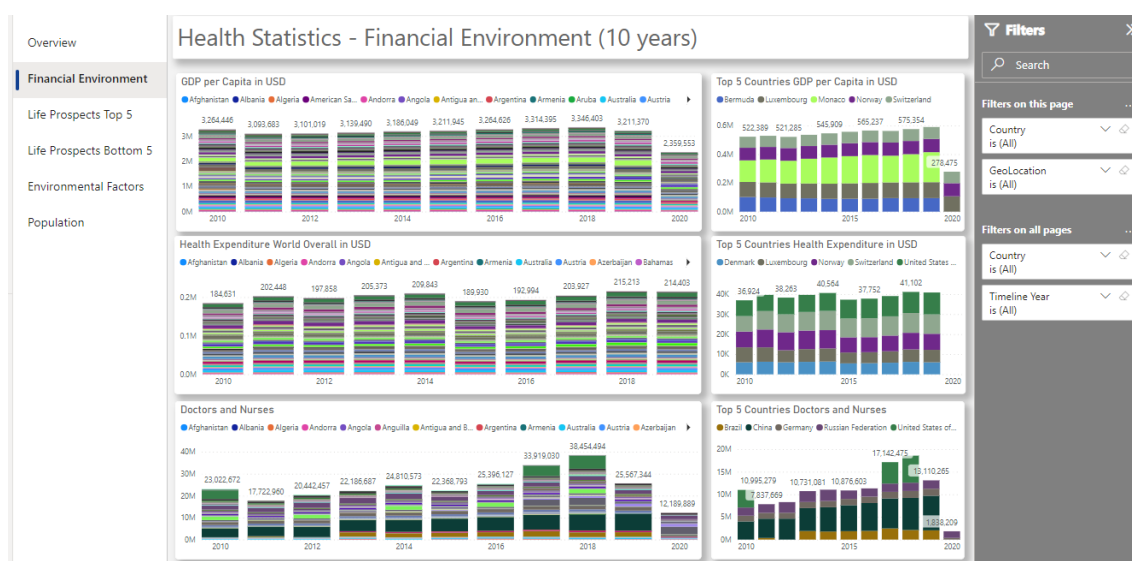
Az oldal megvalósításához vonal és területi diagramokat használtam, valamint a szűrést az elmúlt 20 évre állítottam be, hogy a legrelevánsabb adatok kerüljenek a fő táblára. Illetve a kiemelt számok mindig a legutolsó adatpontot mutatják az elérhető adatsorokból. A számok alatti trendvonalak pedig kiválóan mutatják a változásokat az eltelt évekről.

Az oldalon látható értékek közül az orvosok és nővérek számának kiszámításához szükségem volt measure-k használatára:

```
Doctors = sum ('07_tbl_medical_doctors'[Medical doctors nr])
Nurses = sum ('08_tbl_nursing_midwifery'[Total person nr])
Medical Person = '07_tbl_medical_doctors'[Doctors] + '08_tbl_nursing_midwifery'[Nurses]
```

Ezen kívül elérhetővé tettem szűrőket is a felhasználók számára, hogy egy konkrét országra és évre tudják az adatokat szűkíteni.

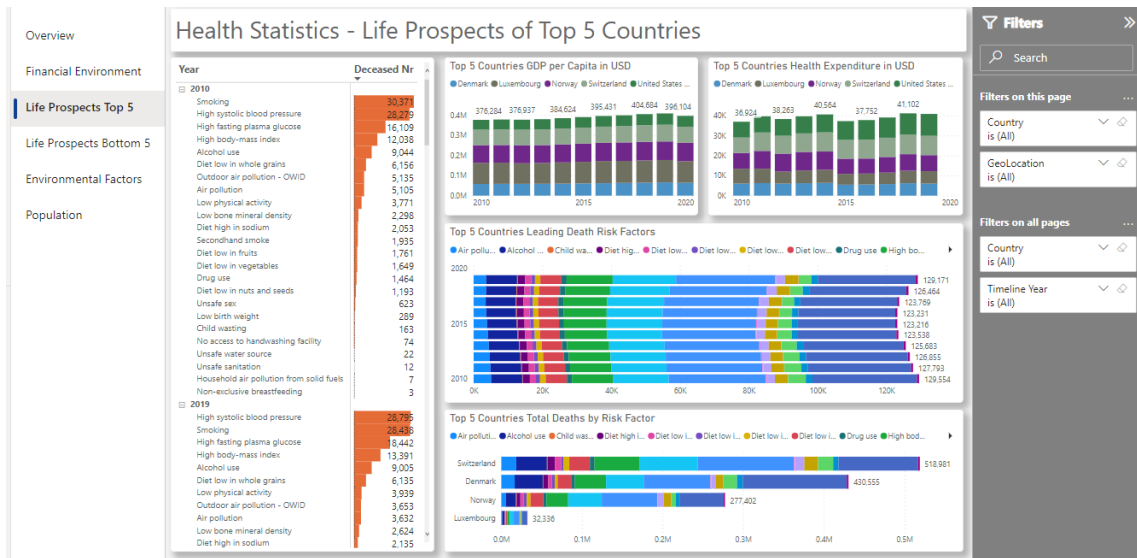
2. Az egészségügy anyagi környezetét bemutató oldal (30.ábra).



30. ábra Financial environment page

Az oldalon kizárólag oszlopdigramokat használtam, a szűrést az elmúlt 10 évre állítottam be, amennyiben még több évnnyi adatsort mutatnék be az végképp rontana a felhasználói élményen. Minden diagramon az x-tengely mutatja az éveket, a különböző színek az oszlopokon a különböző országokat jelölik. Minden kategóriából készítettem egy szűrt diagramot is a top 5 országra az adott kategóriában, de ettől függetlenül is van lehetőség akár az ország nevére kattintással, akár szűréssel egyéni országok adatait vizsgálni (39.ábra), sőt a *GeoLocation* szűrő segítségével földrajzi régiókra szűrt adatokat is láthatunk (40.ábra).

3. Az élet kilátásai a Top 5 országban, egészségügyre fordított költségek alapján (31.ábra).

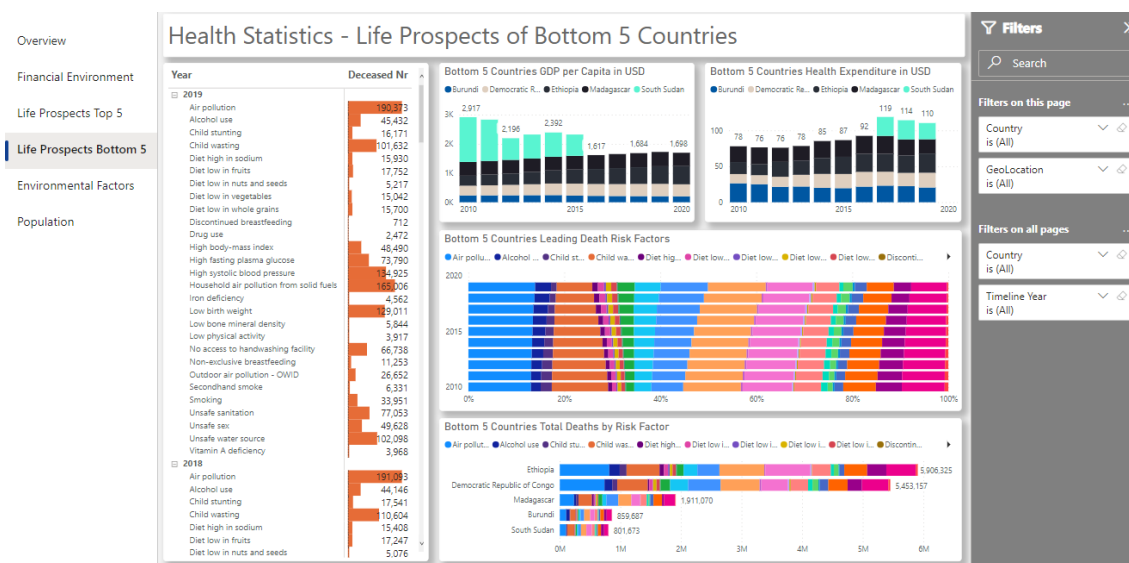


31. ábra Life prospects of Top 5 Countries page

Az oldalon sokféle diagramot használtam fel, hogy minél komplexebb képet mutassak a top 5 ország (egészségügyre fordított költségek alapján) életkilátásairól. Ismét megjelenítettem a GDP és Egészségügyi kiadások alapján a top 5 országot, valamint a hozzájuk tartozó vezető halálozási okokat és a hozzájuk tartozó halálozások számát, szintén az elmúlt 10 évet vizsgálva.

Az oldalon ismét van lehetőség országra vagy régióra szűrni, esetlegesen egy országra vagy halálokra kattintással megvizsgálni annak részletesebb adatait. A bal oldali táblázatban egyéni formázást állítottam be, hogy minden éven belül adatsávokkal is ki legyen emelve, hogy melyik a vezető halálok.

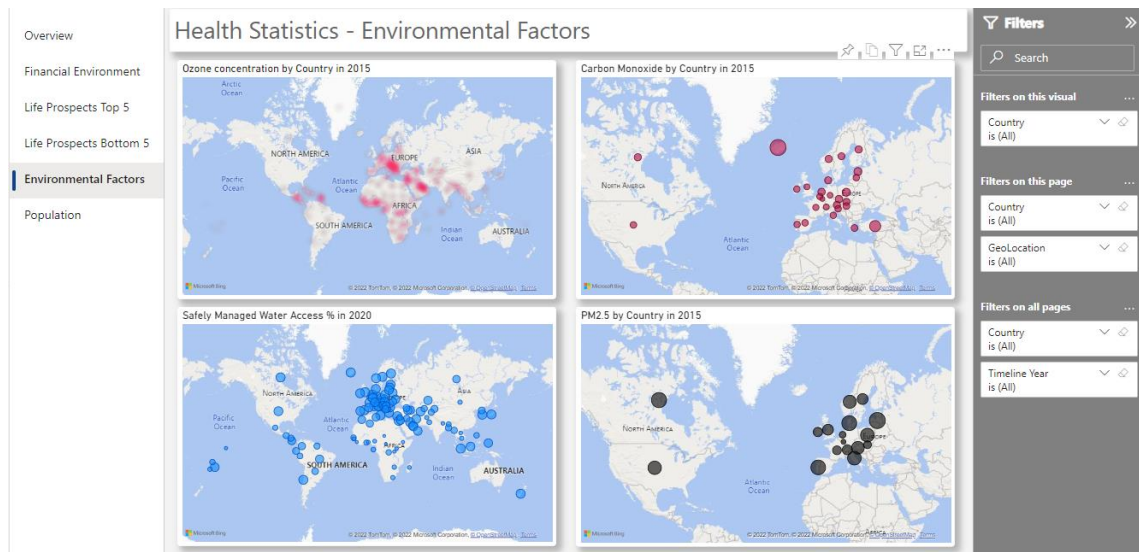
4. Életkilátások a sereghajtó országokban (32.ábra).



32. ábra Life prospects of Bottom 5 Countries page

Az oldal előzőnek a másolata azzal a különbséggel, hogy itt a sereghajtó országok kerültek bemutatásra az egészségügyre fordított költségek alapján. Ugyanazok a szűrési és vizsgálódási opciók érhetőek el.

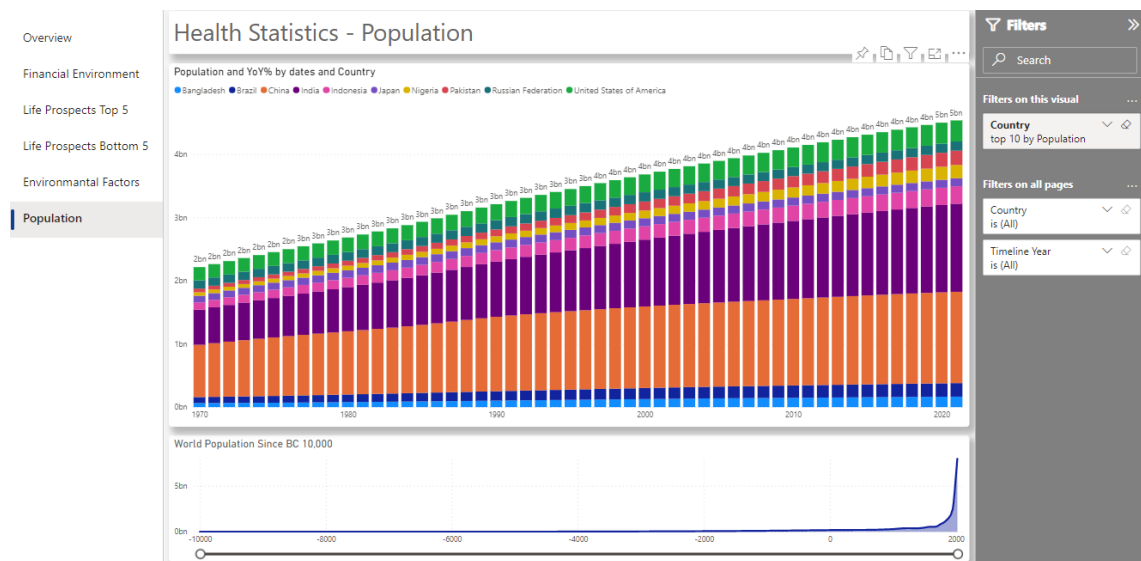
5. Az egészségügy környezeti tényezői (33.ábra).



33. ábra Environmental Factors page

A környezeti tényezők oldalon kizárólag térképek segítségével mutatom be az értékeket, még hozzá mindből a legutolsó elérhető adatokat kiválasztva (Top 1 szűrés az évszámra). A bal felső térképen az ózon koncentrációs értékeket hőtérkép segítségével ábrázoltam, a magasabb értékek magasabb koncentrációt jeleznek. A szén-monoxid és szállópor, valamint a biztonságos ivóvíz értékeket buborékokkal ábrázoltam, a nagyobb buborék nagyobb értéket jelent.

6. Érdekességképpen népességadatok, akár Kr.e. 10,000-től számítva (34.ábra)



34. ábra Population page

Az oldalon oszlop és területi diagramon mutatom be a népesség alakulását. A felső diagramon szűrést is alkalmaztam a top 10 legnépesebb országra 1970-től nézve. Az alsó diagram érdekességképpen bemutatja, hogy az elmúlt összesen 12,000 évben, hogyan alakult a népesség növekedése (az adatok nagy része itt becsléseken alapul). Ezen egy csúszka segítségével lehet az éveket beállítani, hogy közelebbről vizsgáljunk rövidebb időszakokat is.

A kurzor bármelyik oszlopra húzásával elérhetővé válik egy YOY% érték (41.ábra), amely évről-évre megmutatja, hogy hány százalékkal változott az adott ország népessége. Ehhez ismét egy *measure*-re volt szükségem, amelyet a Tooltip-ben helyeztem el:

```
YoY% =  
var prevyear = CALCULATE(SUM('05_tbl_population_pcountries'[Population]),  
PREVIOUSYEAR('05_tbl_population_pcountries'[dates]))  
var actualyear = CALCULATE(sum('05_tbl_population_pcountries'[Population]))  
return (actualyear - prevyear) / prevyear
```

6. TRENDEK MEGFIGYELÉSE ÉS KÖVETKEZTETÉSEK

6.1. A vizualizációk elemzése

Az elkészült vizualizációk, melyeket a Power BI Online felületére is megosztottam elérhetővé válhatnak a webes felületen keresztül egy szélesebb közönség számára is, akik hozzáértő szemmel is vizsgálhatják a bemutatott eredményeket. Habár úgy vélem nem szükséges egészségügyi végzettséggel rendelkezni, hogy levonjunk bizonyos következtetéseket a látottak alapján.

1. Ami a 34-es ábra alsó diagramjáról nagyon szembetűnő, hogy az elmúlt 50-70 évben óriási népességrobbanásról beszélhetünk. Ez a folyamat a fejlődő és főleg afrikai országokban a legjellemzőbb, itt a legmagasabb évről-évre a népességnövekedés aránya. Emellett az olyan hatalmas népességgel rendelkező országok, mint Kína és India is továbbra is növekedő tendenciát mutatnak, India lakossága majdnem 1%-ot emelkedik minden évben.
2. A GDP adatait figyelembe véve a legelső grafikonon a 30-as ábrán már megfigyelhető a 2019 év végén elindult COVID-19 járvány hatása, sőt a 2020-as évre majdnem harmadával csökkent a világ teljes GDP előállítása. Ez a tendencia feltehetően fokozódott a 2021-es évben, így érdemes lesz a friss adatokat áttekinteni amint végelegek és elérhetőek lesznek.
3. Az egészségügyre fordított költségeket megfigyelve véleményem szerint nem láthatunk túl nagy meglepetéseket, legalábbis a Top 5 ország tekintetében semmiképpen. Sajnálatos, hogy itt még nem elérhetőek a 2020 és 2021-es adatsorok, érdemes lesz majd rátekinteni ismét, hogy a világi járvány, hogyan változtatta meg az arányokat. Mindenesetre globálisan nézve 2015-től kezdve határozottan emelkedő tendenciát mutatnak az egészségügyre fordított költségek, ami nagyon biztató hosszútávon.
4. A 30-as ábra legelső sorában az egészségügyi személyzet száma sajnos épp a vártakkal/szükségessé ellentétes tendenciát mutat. 2018-tól kezdve határozottan csökken az egészségügyben dolgozók száma. Ez persze felvetheti a hiányzó adatoknak is a kérdését, feltételezve, hogy az elmúlt 2 évben nem az adatszolgáltatásra helyezték az országok a hangsúlyt.
5. A 31-es ábrán a leggazdagabb országoknál megfigyelhető az a trend, hogy a dohányzás és a magas vérnyomás, cukorbetegség és elhízás tartozik a vezető halálokok közé. Ezekben az országokban a direkt környezeti tényezők, mint akár a légszennyezés nem jelentenek akkor rizikó faktort, mint az előbb felsoroltak.
6. A 32-es ábrát elemezve teljesen más kép tárul elénk. Az egészségügyre fordított költségek szempontjából sereghajtó országokban nem az elhízás vagy a cukorbetegség a fő probléma. Ezekben az országokban a legtöbben a légszennyezés miatt,

korszerűtlen fűtés miatt, illetve általános szegénységi problémák miatt halnak meg (alacsony születési súly, tiszta víz hiánya, általános higiéniai hiányosságok).

7. A környezeti faktorok térképei, habár nagyon hasznosak lehetnének, az elérhető adatok alapján úgy vélem ezeket az adatokat az országok nem mindig rögzítik és osztják meg másokkal vagy nem készül elegendő mérés globálisan. Az elérhető biztonságos ivóvíz térképen megfigyelhető az a probléma, hogy a legszegényebb országokban nem állrendelkezésre sem a megfelelő higiéniahoz vagy tisztálkodáshoz szükséges víz. A szálló por buborékok pedig mutatják, hogy pont a fejlett országokban jelent ez nagyobb problémát.
8. Az összesítő oldalon, 29-es ábra kiemelt értékek és trendvonalak igen beszédesek. A Föld népesség konstans növekedésben van, emellett a várható egészséges élettartam is, ha lassabb tendenciában ugyan, de növekedik. A világ GDP értékei, és az abból egészségügyre fordított összegek is emelkednek, ami jelentheti, hogy az egészséges élet jelentőségét nem kérdőjelezzük meg, de jelentheti, hogy egyre drágábbak ezen költségek és az arányosan magasabb lakosságszám nagyobb ráfordításokat igényel. Érdekesség ugyan, de látszik az ábrán a 2008-as válság hatása is a GDP-re és az onnan emelkedő egyenletes tendencia a COVID-19 világjárvány kirobbanásáig. Ami mindenképpen elgondolkodtató és további vizsgálatokat igénylő adatsor az orvosok és nővérek számának csökkenése, valamint a vezető halálozási okok változása.

6.2. Következtetés és továbbfejlesztési lehetőségek

A dolgozatom célkitűzése egy adatharmonizáló rendszer létrehozása volt, amellyel az egészségügy adatait tudjuk megfigyeli és levonni a következtetéseket, emellett ráfókuszál a globális problémákra.

Úgy gondolom, hogy ebben a témában kétségkívül rengeteg adatsor elérhető és így amit bemutattam csupán egy szelete annak a komplex lehetőséghalmaznak, amelyet az egészségügy rejt. Igyekeztem olyan adatsorokat kiválasztani, amelyek egymással is jól, harmonikusan tudnak kapcsolódni, ehhez elengedhetetlen volt, hogy a forrásaim száma alacsony legyen. Az elérhető adatoknál is említettem, hogy amennyi szervezet annyi féle adatsor, számítás és adatszolgáltatási hajlandóság. Mégis ezeket a táblákat sikerült, olyan formátumra alakítani, hogy egymással is kapcsolódjanak és a szűrési opciók is működjenek. Mivel nem használtam indexeket vagy sorszámozást a soroknál ez probléma is lehetett volna. Így a rendszert a későbbiekben mindenképpen kiegészíteném indexekkel az országokhoz, mivel így az esetleges elnevezésből eredő különbségek könnyedén kiküszöbölhetőek.

A Power BI remekül működött nem csak, mint vizualizációs eszköz, hanem mint adat átalakító is. Ugyanakkor tapasztalatból tudom, hogy a Power BI 5 millió sornál nagyobb

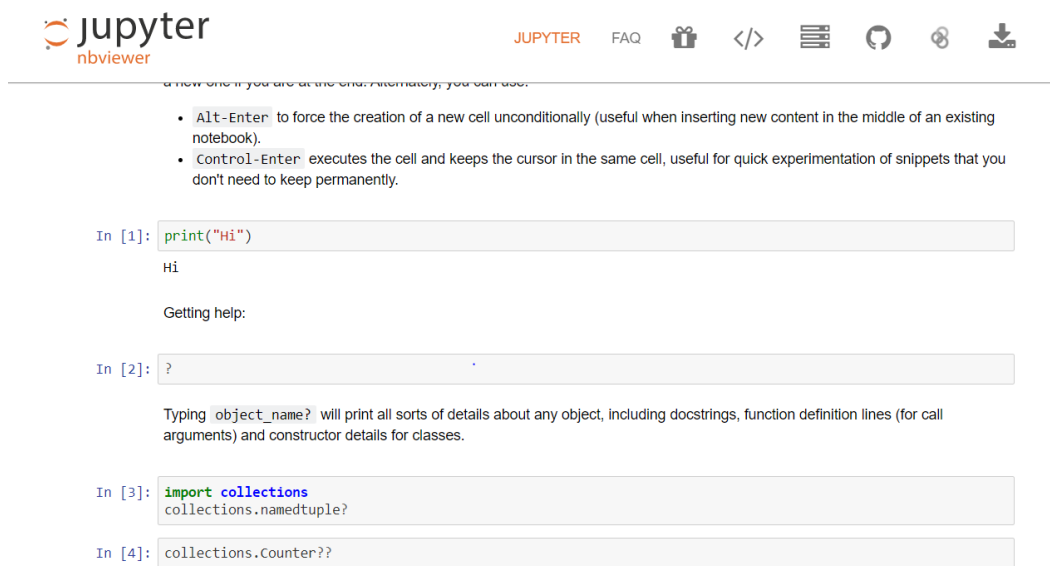
komplex táblákkal már nehezebben boldogul és lassabban tudja elvégezni az átalakításokat. Így a rendszerben ezt ki lehet cserélni a hagyományosabb adattárház megoldásokra, amennyiben az adatok ezt igénylik. A Power BI-ban elérhető funkciók nagyban megkönnyítik a bemutatók létrehozását, emellett az Online App-ba való feltöltéssel az riportok könnyedén eljutnak a megrendelőkhöz vagy a vezetőséghez. A *tool* a bemutatott vizualizációkon kívül még rengeteg potenciállal rendelkezik, így amennyiben további elemzésekre vagy mélyebb vizsgálatokra lenne szükség véleményem szerint továbbra is használható lenne.

Mindenképpen megfontolandó és továbbfejlesztendő lenne, hogy az adatokhoz direkt kapcsolódjunk, akár a forrásoldalakon akár a publikáló szervezetekhez, így mindig naprakészen lehet az adatbázist és az elemzéseket tartani. Az egészségügy egy sokrétű és érzékeny témakör, nagyon kis dolgok is képesek óriási hatásokat kiváltani a világukban, így a gyors és friss adatok elengedhetetlenek. Emellett a közelmúlt adatai mellett érdemes néha még messzebbre tekinteni.

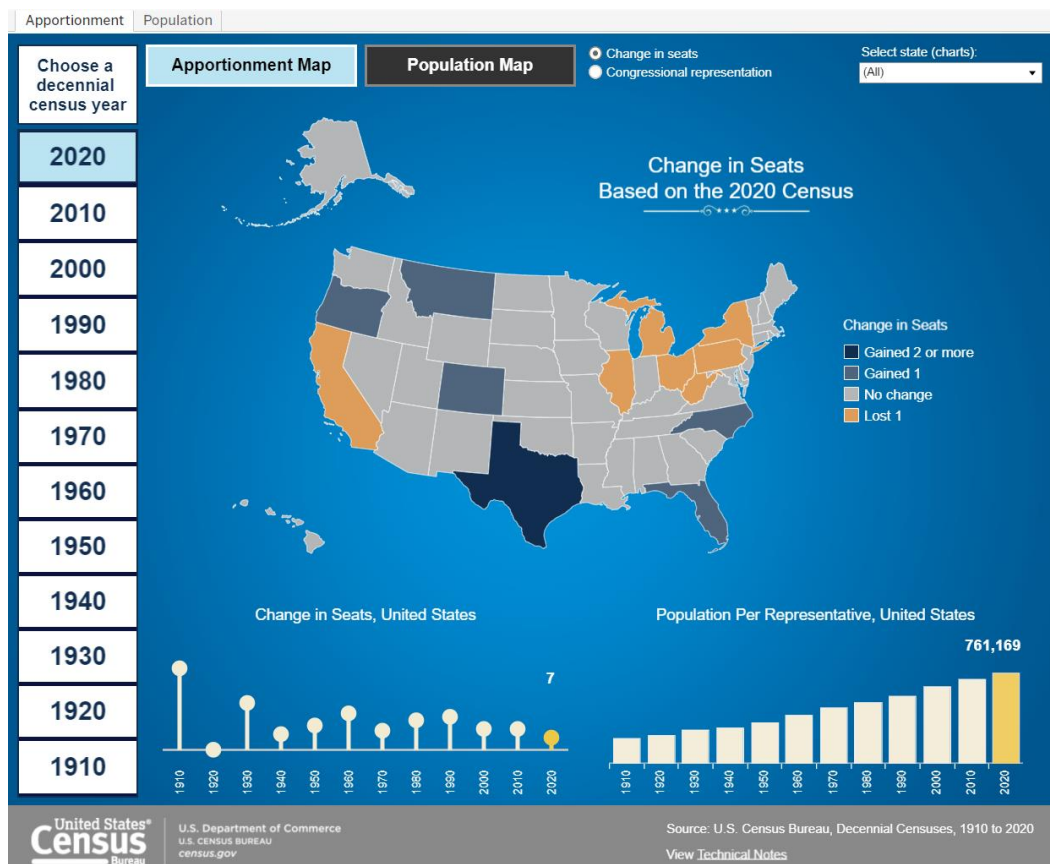
Összességében úgy vélem, hogy a dolgozatom elérte a célját, a kiválasztott adatokat „átvezettem” a rendszeren, felhasználtam mindazt az ismeretet, ami szükséges egy ilyen feladat megvalósításához. Emellett az elkészült riportot kielemezve sikerült néhány összefüggésre rávilágítani. Ezek alapján megbecsülni, hogy mi is várható ezen a területen szerintem nem lehetetlen feladat, azonban ennek a vizsgálata mindenképpen további és mélyebb kutatásokat igényel mind egészségügyileg mind szakmailag.

7. MELLÉKLETEK

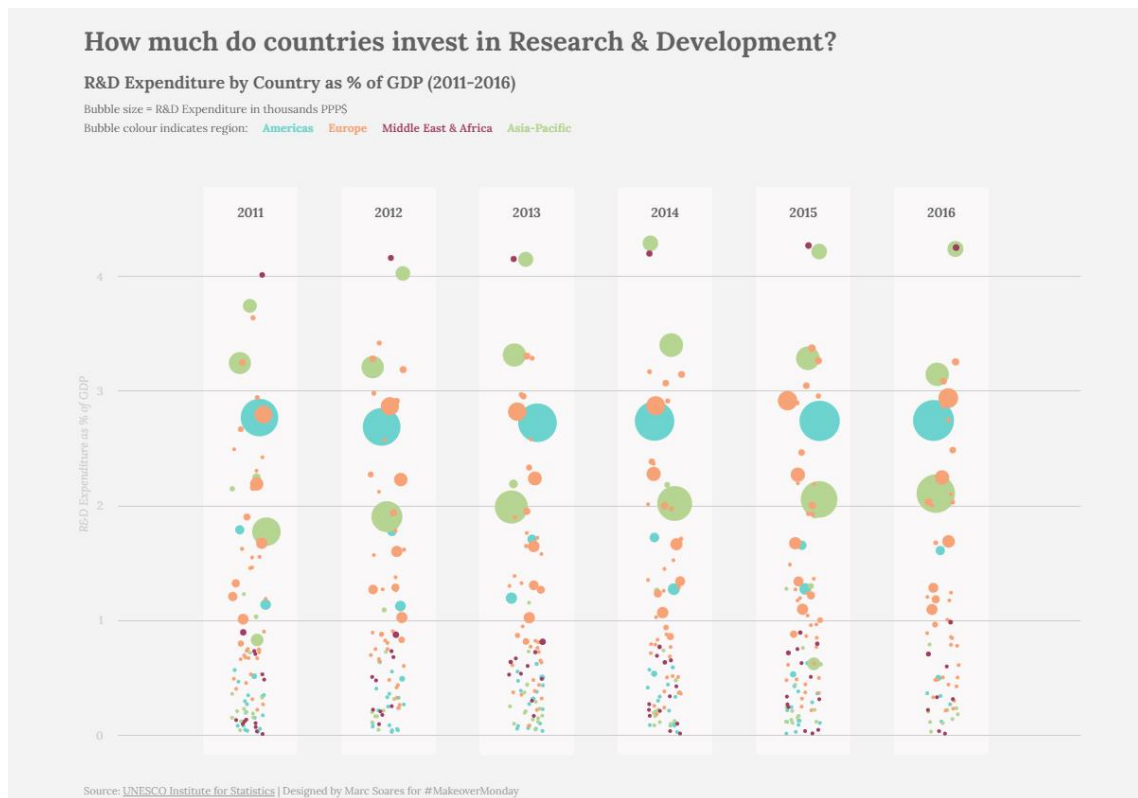
7.1. Egyéb ábrák



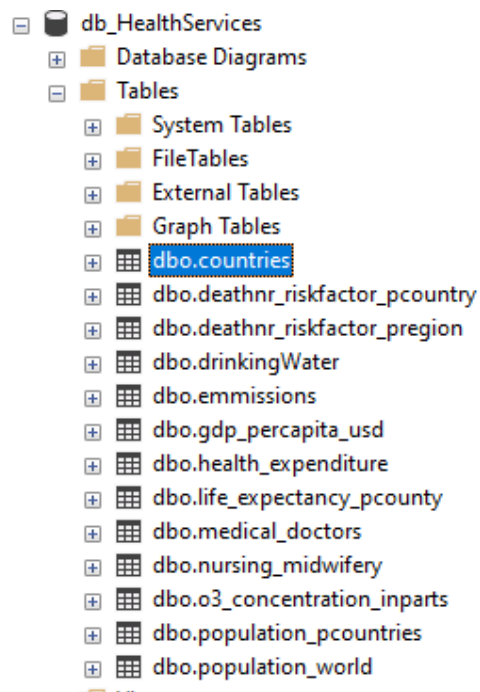
35. ábra JupyterR kezelőfelülete



36. ábra Példa Tableau dashboard-ra az USA szenátusának alakulásáról



37. ábra Google Data Studio példa vizualizáció



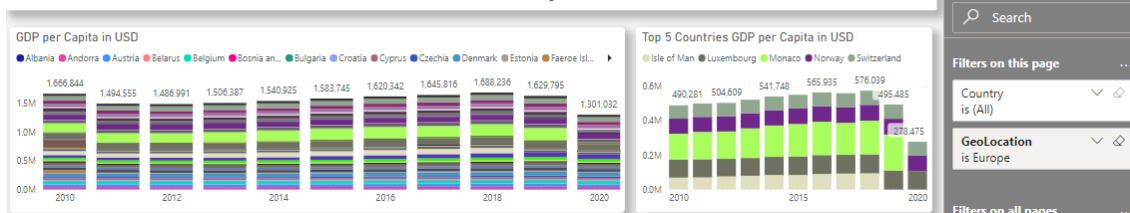
38. ábra Táblák sikeres betöltése az Extract folyamat végén

Health Statistics - Financial Environment (10 years)



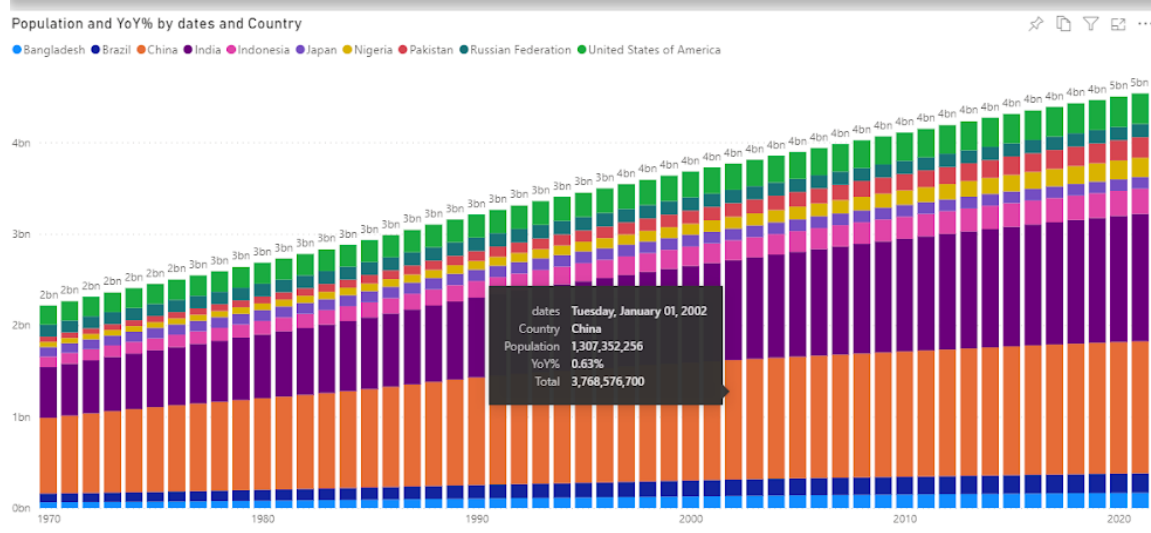
39. ábra Financial environment page - országra szűréssel

Health Statistics - Financial Environment (10 years)



40. ábra Financial environment page - régióra szűréssel

Health Statistics - Population



41. ábra YOY% bemutatása a népesség adatokra

7.2. Táblák átalakítása Power BI dataflowban

1. Countries

	AB_C Country	AB_C Code	AB_C GeoLocation
1	Abkhazia	OWID_ABK	Asia
2	Afghanistan	AFG	Asia
3	Akrotiri and Dhekelia	OWID_AKD	Asia
4	Albania	ALB	Europe
5	Algeria	DZA	Africa
6	American Samoa	ASM	Oceania
7	Andorra	AND	Europe
8	Angola	AGO	Africa
9	Anguilla	AIA	North America
10	Antarctica	ATA	Antarctica
11	Antigua and Barbuda	ATG	North America
12	Argentina	ARG	South America
13	Armenia	ARM	Asia
14	Aruba	ABW	North America
15	Australia	AUS	Oceania
16	Austria	AUT	Europe
17	Austria-Hungary	OWID_AUH	Europe
18	Azerbaijan	AZE	Asia
19	Baden	OWID_BAD	Europe
20	Bahamas	BHS	North America
21	Bahrain	BHR	Asia
22	Bangladesh	BGD	Asia

42. ábra 01_dim_countries átalakított tábla

2. GDP per country

	AB_C Country	1 ² 3 Year	1.2 GDP per capita (US\$)
1	Afghanistan	2002	330.3034944
2	Afghanistan	2003	343.0808897
3	Afghanistan	2004	333.2166171
4	Afghanistan	2005	357.2347617
5	Afghanistan	2006	365.2843706
6	Afghanistan	2007	405.5488893
7	Afghanistan	2008	412.0142124
8	Afghanistan	2009	488.3003713
9	Afghanistan	2010	543.3029674
10	Afghanistan	2011	528.7366831
11	Afghanistan	2012	576.1900885
12	Afghanistan	2013	587.5650355
13	Afghanistan	2014	583.6560179
14	Afghanistan	2015	574.1841143
15	Afghanistan	2016	571.0753885
16	Afghanistan	2017	571.4452409
17	Afghanistan	2018	564.6170395
18	Afghanistan	2019	573.2875058
19	Afghanistan	2020	549.3917871
20	Albania	1980	1992.291931

43. ábra 03_gdp_percapita_usd átalakított tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source{[Schema = "dbo", Item = "gdp_percapita_usd"]}[Data],
    #"Filtered rows" = Table.SelectRows(#"Navigation 1", each [Code] <> null and [Code]
    <> ""),
    #"Changed column type" = Table.TransformColumnTypes(#"Filtered rows", {"Year", Int
    64.Type}),
    #"Removed columns" = Table.RemoveColumns(#"Changed column type", {"Code"}),
    #"Renamed columns" = Table.RenameColumns(#"Removed columns", {"Entity", "Country"}
    })
in
    #"Renamed columns"

```

3. GDP health expenditure

	Country	Year	CHE per capita GDP
1	Afghanistan	2014	59.01
2	Afghanistan	2005	23.89
3	Afghanistan	2018	70
4	Afghanistan	2016	60.19
5	Afghanistan	2012	51.44
6	Afghanistan	2003	17.04
7	Afghanistan	2006	27.84
8	Afghanistan	2007	31.27
9	Afghanistan	2004	20.41
10	Afghanistan	2009	41.72
11	Afghanistan	2011	50.85
12	Afghanistan	2015	58.91
13	Afghanistan	2017	65.71
14	Afghanistan	2008	38.1
15	Afghanistan	2019	65.81
16	Afghanistan	2010	44.99
17	Afghanistan	2002	15.8
18	Afghanistan	2013	55.03
19	Albania	2008	220.46
20	Albania	2018	274.91

44. ábra 04_gdp_health_expenditure végleges tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source{[Schema = "dbo", Item = "health_expenditure"]}[Data],
    #"Reordered columns" = Table.ReorderColumns(#"Navigation 1", {"Country", "Year", "C
    HE per capita GDP"}),
    #"Changed column type" = Table.TransformColumnTypes(#"Reordered columns", {"Year",
    Int64.Type}),
    #"Sorted rows" = Table.Sort(#"Changed column type", {"Country", Order.Ascending})
in
    #"Sorted rows"

```

4. Population

	Country	Year	Population (historical estimates)
1	Afghanistan	1970	11173654
2	Afghanistan	1971	11475450
3	Afghanistan	1972	11791222
4	Afghanistan	1973	12108963
5	Afghanistan	1974	12412960
6	Afghanistan	1975	12689164
7	Afghanistan	1976	12943093
8	Afghanistan	1977	13171294
9	Afghanistan	1978	13341199
10	Afghanistan	1979	13411060
11	Afghanistan	1980	13356500
12	Afghanistan	1981	13171679
13	Afghanistan	1982	12882518
14	Afghanistan	1983	12537732
15	Afghanistan	1984	12204306
16	Afghanistan	1985	11938204
17	Afghanistan	1986	11736177
18	Afghanistan	1987	11604538
19	Afghanistan	1988	11618008
20	Afghanistan	1989	11868873

45. ábra 05_tbl_population_pcountries tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "population_pcountries"][Data],
    #"Changed column type" = Table.TransformColumnTypes(#"Navigation 1", {"Year", Int64.Type}),
    #"Filtered rows" = Table.SelectRows(#"Changed column type", each [Code] <> null and [Code] <> ""),
    #"Filtered rows 1" = Table.SelectRows(#"Filtered rows", each [Year] >= 1970),
    #"Removed columns" = Table.RemoveColumns(#"Filtered rows 1", {"Code"})
in
    #"Removed columns"

```

5. Life Expectancy

	Country	Year	1.2 Life expectancy at birth (years) Both sexes	1.2 Life expectancy at birth (years)_1 Male
1	Afghanistan	2019	63.2	63.3
2	Albania	2019	78	76.3
3	Algeria	2019	77.1	76.2
4	Angola	2019	63.1	60.7
5	Antigua and Barbuda	2019	76.5	74.9
6	Argentina	2019	76.6	73.5
7	Armenia	2019	76	72.5
8	Australia	2019	83	81.3
9	Austria	2019	81.6	79.4
10	Azerbaijan	2019	71.4	68.8
11	Bahamas	2019	73.2	69.9
12	Bahrain	2019	75.8	75
13	Bangladesh	2019	74.3	73
14	Barbados	2019	76	74.3
15	Belarus	2019	74.8	69.7
16	Belgium	2019	81.4	79.3
17	Belize	2019	74.4	71.4
18	Benin	2019	63.4	61.2
19	Bhutan	2019	73.1	72
20	Bolivia (Plurinational State...	2019	72.1	71.1

46. ábra 16_tbl_life_expectancy_pcountry tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[[Schema = "dbo", Item = "life_expectancy_pcountry"]][Data],
    #"Reordered columns" = Table.ReorderColumns(#"Navigation 1", {"Country", "Year", "Life expectancy at birth (years) Both sexes", "Life expectancy at birth (years)_1 Male", "Life expectancy at birth (years)_2 Female", "Healthy life expectancy (HALE) at birth (years) Both sexes", "Healthy life expectancy (HALE) at birth (years)_3 Male", "Healthy life expectancy (HALE) at birth (years)_4 Female"}),
    #"Changed column type" = Table.TransformColumnTypes(#"Reordered columns", {{ "Year", Int64.Type }})
in
    #"Changed column type"

```

6. Medical doctors

	Country	Year	1.2 Medical doctors nr	Generalist medical practitioners nr	Specialist medical practitioners nr	1.2 Medical doctors not further defined nr
1	Afghanistan	2001	4104	null	null	4104
2	Afghanistan	2006	4220	null	null	4220
3	Afghanistan	2007	4723	null	null	4723
4	Afghanistan	2008	4834	null	null	4834
5	Afghanistan	2009	6037	null	null	6037
6	Afghanistan	2010	6901	null	null	6901
7	Afghanistan	2011	7584	null	null	7584
8	Afghanistan	2012	7521	null	null	7521
9	Afghanistan	2013	9184	null	null	9184
10	Afghanistan	2014	9954	null	null	9954
11	Afghanistan	2015	9808	null	null	9808
12	Afghanistan	2016	9842	null	null	9842
13	Afghanistan	2017	8744	null	null	8744
14	Afghanistan	2018	11026	null	null	11026
15	Afghanistan	2019	8082	null	null	8082

47. ábra 07_tbl_medical_doctors tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "medical_doctors"][Data],
    #"Changed column type" = Table.TransformColumnTypes(#"Navigation 1", {{"Year", Int64.Type}}),
    #"Sorted rows" = Table.Sort(#"Changed column type", {"Country", Order.Ascending}, {"Year", Order.Ascending})
in
    #"Sorted rows"

```

7. Nursing and midwifery

	Country	Year	1.2 Total person nr	1.2 Nursing person nr	1.2 Midwifery person nr
1	Afghanistan	2005	14930	14930	null
2	Afghanistan	2006	11631	11631	null
3	Afghanistan	2007	13430	11809	1621
4	Afghanistan	2008	13780	12113	1667
5	Afghanistan	2009	17257	17257	null
6	Afghanistan	2013	8052	4552	3500
7	Afghanistan	2014	4925	4925	null
8	Afghanistan	2015	4471	4471	null
9	Afghanistan	2016	5243	5243	null
10	Afghanistan	2017	6370	6370	null
11	Afghanistan	2018	16581	11483	5098
12	Albania	1990	14785	14785	null
13	Albania	1993	15812	15812	null
14	Albania	1994	16209	14318	1891
15	Albania	1995	14318	14318	null
16	Albania	1996	14284	14284	null
17	Albania	1997	14284	14284	null
18	Albania	1998	12730	12730	null
19	Albania	1999	12730	12730	null
20	Albania	2000	12570	12570	null

48. ábra 08_tbl_nursing_midwifery tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "nursing_midwifery"][Data],
    #"Changed column type" = Table.TransformColumnTypes(#"Navigation 1", {{"Year", Int64.Type}}),
    #"Sorted rows" = Table.Sort(#"Changed column type", {"Country", Order.Ascending}, {"Year", Order.Ascending}),
    #"Filtered rows" = Table.SelectRows(#"Sorted rows", each [Country] <> null and [Country] <> "")
in
    #"Filtered rows"

```

8. Death Risk factors

	AB Country	123 Year	AB Type	1.2 Deceased Nr
1	Afghanistan	1990	Outdoor air pollution - OWID	3169
2	Afghanistan	1990	High systolic blood pressure	25633
3	Afghanistan	1990	Diet high in sodium	1045
4	Afghanistan	1990	Diet low in whole grains	7077
5	Afghanistan	1990	Alcohol use	356
6	Afghanistan	1990	Diet low in fruits	3185
7	Afghanistan	1990	Unsafe water source	3702
8	Afghanistan	1990	Secondhand smoke	4794
9	Afghanistan	1990	Low birth weight	16135
10	Afghanistan	1990	Child wasting	19546
11	Afghanistan	1990	Unsafe sex	351
12	Afghanistan	1990	Diet low in nuts and seeds	2319
13	Afghanistan	1990	Household air pollution from s...	34372
14	Afghanistan	1990	Diet low in vegetables	3679
15	Afghanistan	1990	Low physical activity	2637
16	Afghanistan	1990	Smoking	5174
17	Afghanistan	1990	High fasting plasma glucose	11449
18	Afghanistan	1990	Air pollution	37231
19	Afghanistan	1990	High body-mass index	9518
20	Afghanistan	1990	Unsafe sanitation	2798

49. ábra 09_tbl_deathnr_riskfactory_pcountry tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "deathnr_riskfactor_pcountry"][Data],
    #"Renamed columns" = Table.RenameColumns(#"Navigation 1", {{"Outdoor air pollution - OWID ", "Outdoor air pollution - OWID"}, {"High systolic blood pressure ", "High systolic blood pressure"}, {"Diet high in sodium ", "Diet high in sodium"}, {"Diet low in whole grains ", "Diet low in whole grains"}, {"Alcohol use ", "Alcohol use"}, {"Diet low in fruits ", "Diet low in fruits"}, {"Unsafe water source ", "Unsafe water source"}, {"Secondhand smoke ", "Secondhand smoke"}, {"Low birth weight ", "Low birth weight"}, {"Child wasting ", "Child wasting"}, {"Unsafe sex ", "Unsafe sex"}, {"Diet low in nuts and seeds ", "Diet low in nuts and seeds"}, {"Household air pollution from solid fuels ", "Household air pollution from solid fuels"}, {"Diet low in vegetables ", "Diet low in vegetables"}, {"Low physical activity ", "Low physical activity"}, {"Smoking ", "Smoking"}, {"High fasting plasma glucose ", "High fasting plasma glucose"}, {"Air pollution ", "Air pollution"}, {"High body-mass index ", "High body-mass index"}, {"Unsafe sanitation ", "Unsafe sanitation"}, {"No access to handwashing facility ", "No access to handwashing facility"}, {"Drug use ", "Drug use"}, {"Low bone mineral density ", "Low bone mineral density"}, {"Vitamin A deficiency ", "Vitamin A deficiency"}, {"Child stunting ", "Child stunting"}, {"Discontinued breastfeeding ", "Discontinued breastfeeding"}, {"Non-exclusive breastfeeding ", "Non-exclusive breastfeeding"}, {"Iron deficiency ", "Iron deficiency"}}),

```

```

#"Filtered rows" = Table.SelectRows("#Renamed columns", each [Code] <> null and
[Code] <> ""),
#"Changed column type" = Table.TransformColumnTypes("#Filtered rows", {"Year",
Int64.Type}),
#"Sorted rows" = Table.Sort("#Changed column type", {"Country", Order.Ascending},
{"Year", Order.Ascending}),
#"Removed columns" = Table.RemoveColumns("#Sorted rows", {"Code"}),
#"Unpivoted other columns" = Table.UnpivotOtherColumns("#Removed columns",
{"Country", "Year"}, "Attribute", "Value"),
#"Renamed columns 1" = Table.RenameColumns("#Unpivoted other columns",
{"Attribute", "Type"}, {"Value", "Deceased Nr"})
in
#"Renamed columns 1"

```

9. Emissions

	Country	Year	1.2 Carbon Monoxide	1.2 Nox	PM10	PM2.5	1.2 Sulphur Oxides
1	Australia	1990	100	100	null	null	100
2	Australia	1991	100.33	98.48	null	null	102.74
3	Australia	1992	102.93	101.31	null	null	109.76
4	Australia	1993	106.87	103.97	null	null	112.97
5	Australia	1994	109.33	104.08	null	null	116.97
6	Australia	1995	106.62	104.49	null	null	109.12
7	Australia	1996	109.66	107.67	null	null	112.99
8	Australia	1997	109.93	110.96	null	null	116.21
9	Australia	1998	103.83	113.26	null	null	112.38
10	Australia	1999	100.53	115.11	null	null	117.8
11	Australia	2000	94.78	115.82	null	null	151.73
12	Australia	2001	90.46	121.16	null	null	163.13
13	Australia	2002	85.25	125.58	null	null	173.74
14	Australia	2003	77.26	130.49	null	null	173.24
15	Australia	2004	78.52	133.67	null	null	157
16	Australia	2005	68.94	135.24	null	null	162.05
17	Australia	2006	66.01	135.26	null	null	155.14
18	Australia	2007	58.08	136.87	null	null	157.18
19	Australia	2008	56.06	140.04	null	null	168.46
20	Australia	2009	54.97	138.68	null	null	167.16

50. ábra 10_tbl_emissions

```

let
Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
#"Navigation 1" = Source[Schema = "dbo", Item = "emissions"][Data],
#"Filtered rows" = Table.SelectRows("#Navigation 1", each [Code] <> null and [Code]
<> ""),
#"Changed column type" = Table.TransformColumnTypes("#Filtered rows", {"Year", Int
64.Type}),
#"Sorted rows" = Table.Sort("#Changed column type", {"Entity", Order.Ascending}, {"
Year", Order.Ascending}),
#"Removed columns" = Table.RemoveColumns("#Sorted rows", {"Code"}),

```

```

#"Renamed columns" = Table.RenameColumns("#Removed columns", {"Entity", "Country"}
})
in
#"Renamed columns"

```

10. Drinking water

	AB_C Country	1_1	1_2_3 Y...	2_1	AB_C Type	AB_C LocationType	1_2 Value
1	Afghanistan		2000		Population using at least basic drinking-wat...	Rural	21
2	Afghanistan		2000		Population using safely managed drinking-wat...	Rural	8
3	Afghanistan		2000		Population using safely managed drinking-wat...	Total	11
4	Afghanistan		2000		Population using at least basic drinking-wat...	Urban	52
5	Afghanistan		2000		Population using safely managed drinking-wat...	Urban	21
6	Afghanistan		2000		Population using at least basic drinking-wat...	Total	28
7	Afghanistan		2001		Population using safely managed drinking-wat...	Rural	8
8	Afghanistan		2001		Population using at least basic drinking-wat...	Total	28
9	Afghanistan		2001		Population using at least basic drinking-wat...	Urban	52
10	Afghanistan		2001		Population using safely managed drinking-wat...	Urban	21
11	Afghanistan		2001		Population using at least basic drinking-wat...	Rural	21
12	Afghanistan		2001		Population using safely managed drinking-wat...	Total	11
13	Afghanistan		2002		Population using at least basic drinking-wat...	Rural	23
14	Afghanistan		2002		Population using safely managed drinking-wat...	Rural	9
15	Afghanistan		2002		Population using at least basic drinking-wat...	Total	30
16	Afghanistan		2002		Population using safely managed drinking-wat...	Total	12
17	Afghanistan		2002		Population using at least basic drinking-wat...	Urban	54
18	Afghanistan		2002		Population using safely managed drinking-wat...	Urban	21
19	Afghanistan		2003		Population using safely managed drinking-wat...	Urban	22
20	Afghanistan		2003		Population using at least basic drinking-wat...	Urban	56

51. ábra 11_tbl_drinkingwater tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "drinkingWater"][Data],
    #"Reordered columns" = Table.ReorderColumns("#Navigation 1", {"Country", "Year", "LocationType", "Type", "Value"}),
    #"Changed column type" = Table.TransformColumnTypes("#Reordered columns", {"Year", Int64.Type}),
    #"Reordered columns 1" = Table.ReorderColumns("#Changed column type", {"Country", "Year", "Type", "LocationType", "Value"}),
    #"Sorted rows" = Table.Sort("#Reordered columns 1", {"Country", Order.Ascending}, {"Year", Order.Ascending}),
in
    #"Sorted rows"

```

11. Ozone Concentration

	Country	Year	Ozone concentration-StateofGlobalAir
1	Afghanistan	1990	62
2	Afghanistan	1995	62
3	Afghanistan	2000	62
4	Afghanistan	2005	62
5	Afghanistan	2010	63
6	Afghanistan	2013	63
7	Afghanistan	2015	63
8	Albania	1990	72
9	Albania	1995	71
10	Albania	2000	70
11	Albania	2005	69
12	Albania	2010	68
13	Albania	2013	68
14	Albania	2015	68
15	Algeria	1990	67
16	Algeria	1995	68
17	Algeria	2000	69
18	Algeria	2005	69
19	Algeria	2010	70
20	Algeria	2013	70

52. ábra 12_tbl_o3_concentration_inparts tábla

```

let
    Source = Sql.Database("N-20L6PF1M7X4M\MSSQLSERVER01", "db_HealthServices"),
    #"Navigation 1" = Source[Schema = "dbo", Item = "o3_concentration_inparts"][Data]
,
    #"Changed column type" = Table.TransformColumnTypes(#"Navigation 1", {{"Year", Int64.Type}}),
    #"Filtered rows" = Table.SelectRows(#"Changed column type", each [Code] <> null and [Code] <> ""),
    #"Removed columns" = Table.RemoveColumns(#"Filtered rows", {"Code"}),
    #"Sorted rows" = Table.Sort(#"Removed columns", {"Country", Order.Ascending}, {"Year", Order.Ascending})
in
    #"Sorted rows"

```

7.3. Táblák létrehozó scriptjei

1. [dbo].[countries]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[countries](
    [Country] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [GeoLocation] [nvarchar](255) NULL
) ON [PRIMARY]
GO
```

2. [dbo].[gdp_percapita_usd]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[gdp_percapita_usd](
    [Entity] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [GDP per capita (US$)] [float] NULL
) ON [PRIMARY]
GO
```

3. [dbo].[health_expenditure]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[health_expenditure](
    [Year] [float] NULL,
    [Country] [nvarchar](255) NULL,
    [CHE per capita GDP] [float] NULL
) ON [PRIMARY]
GO
```

4. [dbo].[medical_doctors]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[medical_doctors](
    [Country] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Medical doctors nr] [float] NULL,
    [Generalist medical practitioners nr] [nvarchar](255) NULL,
    [Specialist medical practitioners nr] [nvarchar](255) NULL,
    [Medical doctors not further defined nr] [float] NULL
) ON [PRIMARY]
GO
```

5. [dbo].[nursing_midwifery]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[nursing_midwifery](
    [Country] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Total person nr] [float] NULL,
    [Nursing person nr] [float] NULL,
    [Midwifery person nr] [float] NULL
) ON [PRIMARY]
GO
```

6. [dbo].[life_expectancy_pcounty]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[life_expectancy_pcounty](
    [Year] [float] NULL,
    [Country] [nvarchar](255) NULL,
    [Life expectancy at birth (years) Both sexes] [float] NULL,
    [Life expectancy at birth (years)_1 Male] [float] NULL,
    [Life expectancy at birth (years)_2 Female] [float] NULL,
    [Healthy life expectancy (HALE) at birth (years) Both sexes] [float] NULL,
```

```

        [Healthy life expectancy (HALE) at birth (years)_3 Male] [float] NULL,
        [Healthy life expectancy (HALE) at birth (years)_4 Female] [float] NULL
    ) ON [PRIMARY]
GO

```

7. [dbo].[deathnr_riskfactor_pcountry]

```

USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[deathnr_riskfactor_pcountry](
    [Country] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Outdoor air pollution - OWID ] [float] NULL,
    [High systolic blood pressure ] [float] NULL,
    [Diet high in sodium ] [float] NULL,
    [Diet low in whole grains ] [float] NULL,
    [Alcohol use ] [float] NULL,
    [Diet low in fruits ] [float] NULL,
    [Unsafe water source ] [float] NULL,
    [Secondhand smoke ] [float] NULL,
    [Low birth weight ] [float] NULL,
    [Child wasting ] [float] NULL,
    [Unsafe sex ] [float] NULL,
    [Diet low in nuts and seeds ] [float] NULL,
    [Household air pollution from solid fuels ] [float] NULL,
    [Diet low in vegetables ] [float] NULL,
    [Low physical activity ] [float] NULL,
    [Smoking ] [float] NULL,
    [High fasting plasma glucose ] [float] NULL,
    [Air pollution ] [float] NULL,
    [High body-mass index ] [float] NULL,
    [Unsafe sanitation ] [float] NULL,
    [No access to handwashing facility ] [float] NULL,
    [Drug use ] [float] NULL,
    [Low bone mineral density ] [float] NULL,
    [Vitamin A deficiency ] [float] NULL,
    [Child stunting ] [float] NULL,
    [Discontinued breastfeeding ] [float] NULL,
    [Non-exclusive breastfeeding ] [float] NULL,
    [Iron deficiency ] [float] NULL
) ON [PRIMARY]
GO

```

8. [dbo].[population_pcountries]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[population_pcountries](
    [Country] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Population (historical estimates)] [float] NULL
) ON [PRIMARY]
GO
```

9. [dbo].[drinkingWater]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[drinkingWater](
    [Year] [float] NULL,
    [LocationType] [nvarchar](255) NULL,
    [Type] [nvarchar](255) NULL,
    [Country] [nvarchar](255) NULL,
    [Value] [float] NULL
) ON [PRIMARY]
GO
```

10. [dbo].[emmissions]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[emmissions](
    [Entity] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Carbon Monoxide] [float] NULL,
    [Nox] [float] NULL,
    [PM10] [nvarchar](255) NULL,
    [PM2.5] [nvarchar](255) NULL,
    [Sulphur Oxides] [float] NULL
)
```

```
) ON [PRIMARY]
GO
```

11. [dbo].[o3_concentration_inparts]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[o3_concentration_inparts](
    [Country] [nvarchar](255) NULL,
    [Code] [nvarchar](255) NULL,
    [Year] [float] NULL,
    [Ozone concentration-StateofGlobalAir] [float] NULL
) ON [PRIMARY]
GO
```

12. [dbo].[population_world]

```
USE [db_HealthServices]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[population_world](
    [Year] [float] NULL,
    [World Population (Spline Interpolation until 1950)] [float] NULL
) ON [PRIMARY]
GO
```

8. ÁBRAJEGYZÉK

1. ábra Counties tábla [21]	10
2. ábra GDP per capita tábla [22]	10
3. ábra Current Health Expenditure tábla [23]	10
4. ábra Medical doctors tábla [24]	11
5. ábra Nursing, Midwifery tábla [25]	11
6. ábra Life expectancy tábla [26]	12
7. ábra Death Number Riskfactors per country tábla [27]	14
8. ábra Population per country tábla [28]	14
9. ábra Drinking water tábla [29]	14
10. ábra Emmissions tábla [30]	15
11. ábra O3 concentration tábla [31]	15
12. ábra Population world tábla [32]	15
13. ábra Inmon féle Data Warehouse struktúra	16
14. ábra Kimball féle Data Warehouse struktúra	16
15. ábra Az adattárház komponensei	18
16. ábra EK Diagram az egészségügyi adatokhoz	20
17. ábra Tábla betöltése SQL natív kliens segítségével	21
18. ábra Sikeres betöltési folyamat	21
19. ábra Big Data 3V pillére	22
20. ábra Adatvizualizációs eszközök funkciói	23
21. ábra Power BI munkafolyamata []	26
22. ábra A Power Query mint ETL eszköz []	26
23. ábra Power BI forrás választás	27
24. ábra Adatbázisban elérhető táblák	27
25. ábra Power BI dataflow forrás választása	29
26. ábra A dataflow tábláinak kiválasztása	29
27. ábra A betöltött táblák és kapcsolataik Power BI-ban	31
28. ábra Rendszerterv az egészségügyi adatokhoz	31
29. ábra Overview page	32
30. ábra Financial environment page	33
31. ábra Life prospects of Top 5 Countries page	34
32. ábra Life prospects of Bottom 5 Countries page	35
33. ábra Environmental Factors page	36
34. ábra Population page	37
35. ábra JupyterR kezelőfelülete	43
36. ábra Példa Tableau dashboard-ra az USA szenátusának alakulásáról	43
37. ábra Google Data Studio példa vizualizáció	44
38. ábra Táblák sikeres betöltése az Extract folyamat végén	44
39. ábra Financial environment page - országra szűréssel	45
40. ábra Financial environment page - régióra szűréssel	45

41. ábra YOY% bemutatása a népesség adatokra.....	45
42. ábra 01_dim_countries tábla.....	46
43. ábra 03_gdp_percapita_usd átalakított tábla.....	46
44. ábra 04_gdp_health_expenditure végleges tábla	47
45. ábra 05_tbl_population_pcountries tábla.....	48
46. ábra 16_tbl_life_expectancy_pcountry tábla	49
47. ábra 07_tbl_medical_doctors tábla	49
48. ábra 08_tbl_nursing_midwifery tábla.....	50
49. ábra 09_tbl_deathnr_riskfactory_pcountry tábla.....	51
50. ábra 10_tbl_emissions.....	52
51. ábra 11_tbl_drinkingwater tábla	53
52. ábra 12_tbl_o3_concentration_inparts tábla	54

9. IRODALOMJEGYZÉK

- [1] Józán Péter: Halálozási viszonyok és életkilátások a 21. század kezdetén a világ, Európa és Magyarország népességében (<http://www.matud.iif.hu/09okt/11.htm>) utoljára megtekintve: 2021.05.11.
- [2] Meskó, B., Hetényi, G. & Györffy, Z. Will artificial intelligence solve the human resource crisis in healthcare?. *BMC Health Serv Res* 18, 545 (2018).
<https://doi.org/10.1186/s12913-018-3359-4>
- [3] 1997. évi XLVII. törvény az egészségügyi és a hozzájuk kapcsolódó személyes adatok kezeléséről és védelméről. (3. § e pont) (2021.05.11.)
- [4] Társaság a Szabadságjogokért - Hány évig őrzik az egészségügyi intézmények a betegek adatait? (<https://tasz.hu/egeszsegugyi-adatok-gyik>) utoljára megtekintve: 2021.05.11.
- [5] 2016. évi CLV. törvény a hivatalos statisztikáról. (3. § e pont) (2021.05.11.)
- [6] Our World in Data – Where do you get your data from (<https://ourworldindata.org/faqs#where-do-you-get-your-data-from>) utoljára megtekintve: 2021.05.11.
- [7] Wikipédia
https://hu.wikipedia.org/wiki/Gazdasagi_Egyuittmuikodesi_es_Fejlesztési_Szervezet) utoljára megtekintve: 2021.05.11
- [8] Our World in Data, Continents according to Our World in Data 2015
(<https://ourworldindata.org/world-region-map-definitions?fbclid=IwAR2kqKUdB4YoOeiNfbOfe9-2pKkCnUR0hutE24DlkF2cJBx5VgpZSTMHmFE>) utoljára megtekintve 2022.04.05
- [9] Our World in Data, GDP per capita in US\$, 2020
(<https://ourworldindata.org/grapher/gdp-per-capita-in-us-dollar-world-bank?fbclid=IwAR3ODt51f5DO3auECVbxtoS09Io7oHhK4PGjkV0utiW4Zxkd1TqPk79euv4>) utoljára megtekintve 2022.04.01
- [10] WHO, Current health expenditure (CHE) per capita in US\$ Data by country
(https://apps.who.int/gho/data/node.main.GHEDCHEpcUSSHA2011?lang=en&fbclid=IwAR2BC6TMIBs_DCgRX1qBmXStHAWgWleo166vs86HW03uwTKZqnSNuJxaFm4) utoljára megtekintve 2022.04.01
- [11] WHO, Medical doctors
(https://apps.who.int/gho/data/node.main.HWFGRP_0020?lang=en/) utoljára megtekintve 2022.04.01

-
- [12] WHO, Nursing and midwifery personnel (https://apps.who.int/gho/data/node.main.HWFGRP_0040?lang=en) utoljára megtekintve 2022.04.01
- [13] WHO, Life expectancy and Healthy life expectancy (<https://apps.who.int/gho/data/view.main.SDG2016LEXv?lang=en>) utoljára megtekintve 2022.04.01
- [14] Our World in Data, Risk factors for death (<https://ourworldindata.org/causes-of-death#risk-factors-for-death>) utoljára megtekintve 2022.04.05.
- [15] Our World in Data, Population 1800 to 2021, (<https://ourworldindata.org/grapher/population-since-1800?country=CHN~IND~ZAF~NGA~USA~IRN~BRA>) utoljára megtekintve 2022.04.01
- [16] WHO, Basic and safely managed drinking water services (<https://apps.who.int/gho/data/node.main.WSHWATER?lang=en>) utoljára megtekintve 2022.04.01
- [17] Our World in Data, Emissions of air pollutants (<https://ourworldindata.org/grapher/emissions-of-air-pollutants-oecd?tab=table&time=1990..latest&country=~FRA>) utoljára megtekintve 2022.04.05
- [18] Our World in Data, Ozone (O3) concentration (<https://ourworldindata.org/grapher/ozone-o3-concentration-in-ppb?tab=chart>) utoljára megtekintve 2022.04.05.
- [19] Our World in Data, World population from 10,000 BC to today (<https://ourworldindata.org/world-population-growth#how-has-world-population-growth-changed-over-time>) utoljára megtekintve 2022.04.01
- [20] William H. Inmon: Building the Data Warehouse, Third Edition, Wiley, 1992
- [21] Rusznák Attila: Adattárházak és Üzleti Intelligencia, 2. előadás jegyzet, 2021
- [22] Rusznák Attila: Adattárházak és Üzleti Intelligencia, 2. előadás jegyzet, 2021
- [23] Dr Kovács László: OLAP rendszerek alapjai (<http://users.iit.uni-miskolc.hu/~kovacs/hirdetes/olapdoc.pdf>) utoljára megtekintve: 2021.04.20.
- [24] Sidló Csaba: Összefoglaló az adattárházak témaköréről, ELTE, 2004 (<http://scs.web.elte.hu/Work/DW/adattarhazak.htm#3>) utoljára megtekintve: 2021.03.25.
- [25] Wikipédia, Data Warehouse (https://en.wikipedia.org/wiki/Data_warehouse) utoljára megtekintve 2021.03.25.
- [26] Ullman J.D., Widom J.: Adatbázisrendszerek; alapvetés, 2. kiadás, PANEM Kiadó, Budapest, 2008
- [27] Dr. Fleiner Rita: Adatbázisok, 2. előadás jegyzet, 2018

-
- [28] E.F.Codd: The relational Model for Database Management, version 2, *Addison-Wesley*, 1990
- [29] DB-Engines Ranking of Relational DBMS (<https://db-engines.com/en/ranking/relational+dbms>) utoljára megtekintve 2021.03.20.
- [30] BoostHigh - Relational Database Comparison (<https://boosthigh.com/relational-database-comparison/>) utoljára megtekintve 2021.03.25.
- [31] Big data algoritmusok és programozás, 1. előadás, 2021
- [32] TowardsDataScience – Top 4 Popular Big Data Visualizazion tools (<https://towardsdatascience.com/top-4-popular-big-data-visualization-tools-4ee945fe207d>) utoljára megtekintve: 2021.05.12
- [33] Datamation – Big Data Visualization (<https://www.datamation.com/big-data/big-data-visualization/>) utoljára megtekintve: 2021.05.12
- [34] Jupyter documentation (<https://jupyter.org/>) utoljára megtekintve: 2021.05.13.
- [35] Tableau documentation (<https://public.tableau.com/en-us/gallery/historical-population-and-apportionment-data-map?tab=viz-of-the-day&type=viz-of-the-day>) utoljára megtekintve: 2021.05.13.
- [36] Forbes (<https://www.forbes.com/global2000/#1c5676d335d8>) utoljára megtekintve: 2021.05.13.
- [37] Data-Flair.Training.com – Power BI Architeckture (<https://data-flair.training/blogs/power-bi-architecture/>) utoljára megtekintve: 2021.05.13.
- [38] Power BI Dev Camp – Session 8: Intro to M Programming 2021.03.25 (<https://www.powerbidevcamp.net/sessions/session08/>) utoljára megtekintve: 2021.03.30.