



POWER BI - SENAWE ALEX

INLEIDING

1. oorsprong van de data

<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

2. verwerking van de data

2.1. probleem: inladen van data: '.' tov ','

pb: data is een csv file => ',' als veld-seperator en de '.' wordt gebruikt als decimale seperator

dus: ',' ingeven als decimale seperator gaat niet

<https://www.windowcentral.com/how-adjust-region-settings-windows-11>

<https://community.powerbi.com/t5/Desktop/Thousand-Separator-and-Decimal-Separator/m-p/194928>

maw extra bewerkingen nodig na instellingen juist te hebben gezet

1) BMI: type wijzigen naar number

2) bijvoegen: ",en-US"

```
= Table.TransformColumnTypes("#Headers met verhoogd niveau",{{"HeartDisease", type text}, {"BMI", type number}, {"Smoking", type text}, {"AlcoholDrinking", type text}, {"Stroke", type text}, {"PhysicalHealth", Int64.Type}, {"MentalHealth", Int64.Type}, {"DiffWalking", type text}, {"Sex", type text}, {"AgeCategory", type text}, {"Race", type text}, {"Diabetic", type text}, {"PhysicalActivity", type text}, {"GenHealth", type text}, {"SleepTime", Int64.Type}, {"Asthma", type text}, {"KidneyDisease", type text}, {"SkinCancer", type text}}, "en-US")
```

2.2. kolommen

- extra kolom 'Leeftijdsklasse' toegevoegd voor goede weergave van de leeftijdscategorieën in de 'slaap' visual

- andere kolommen laten staan in de data, voor indien men verdere extra analyse zou willen doen

- er is ook een meting 'kolom' toegevoegd:

Gemiddelde van Totaal van Mentale gezondheid per Leeftijdscategorie =

```
AVERAGEX(  
    KEEPFILTERS(VALUES('hartziekte_2020'[Leeftijdscategorie])),  
    CALCULATE(COUNTA('hartziekte_2020'[Mentale gezondheid]))  
)
```

2.3. importeren histogram template

Power BI-visuals


Door op Nu downloaden te klikken en een visual te downloaden, gaat u akkoord met de voorwaarden en het privacybeleid van de provider op de pagina van de visual. Het gebruik van AppSource van Microsoft is onderhevig aan de [Voorwaarden voor Commerciële Marketplace van Microsoft](#) en [Privacyverklaring](#).


Alle visuals Organisatorische visuals AppSource-visuals

Verken alle beschikbare visuals om uw zakelijke inzichten te vergroten [Meer informatie](#)

Filteren op: Alle

Sorteren op: Populariteit




Histogram Chart 

Microsoft Corporation

★★★★★ (80)


Visualises the distribution of data over a continuous interval or certain time period




Histogram by PQ S...

PQ Systems

★★★★★ (5)



Histogram with poi... 

MAQ LLC

★★★★★ (13)

1. groepering

uit visuals volgt: vnl een probleem voor 'white' en 'american indian / alaskan native'

grootste groep: white => focus op deze

met: visuals ifv geslacht

=> vnl een probleem voor blanke mannen

maar ook in verdere visuals: selectie man/vrouw mogelijk

2. factoren die we niet of slechts deels kunnen beïnvloeden

2.1. invloed van leeftijd

- beide geslachten sámen: hoe groter de leeftijd, hoe meer met hartziekten
- mannen: grootste groep: 70-74: de 'probleemleeftijd', maar begint reeds van rond de 60-65
opm: is groter dan groep 80-en-ouder -> doordat mannen minder oud worden?
- vrouwen: categorieën volgen elkaar (+-) 'chronologisch' op, maar met een geleidelijke overgang tussen de categorieën

2.2. invloed van suikerziekte

- eerste visual: analoge resultaten cfr vorige
- correlatie-visuals: in feite zijn gelijklopend: hartziekte ifv leeftijd en ifv suikerziekte
(maw leeftijd en suikerziekte zijn gerelateerd wat betreft hartziekte)
- ofwel: leeftijd geeft suikerziekte, dus suikerziekte is een indicatie van leeftijd
- ofwel: suikerziekte heeft zelfde effect op hartziekte als leeftijd
- verder te onderzoeken: correlatie-3: suikerziekte heeft +- zelfde invloed als leeftijd > 80 (m en/of v)

3. factoren die we wel kunnen beïnvloeden

3.1. invloed van roken

- zowel voor m als v: negatieve invloed van roken
- voor m: grote invloed, opnieuw het grootst vanaf de 'probleemleeftijd'
- voor v: minder grote invloed (zelfs omgekeerde invloed voor 80-en-ouder: outlier?)

3.2. invloed van alcohol

- zowel voor m als v: weinig tot geen invloed (en soms zelfs een licht positieve invloed)

3.3. invloed van fysieke gezondheid

er van uitgaan dat hartziekte beïnvloed wordt door fysieke gezondheid: zie visual

- BMI: eerste deel-visual (links): (BMI>60 er uit gefilterd)

v: fysieke gezondheid is het grootst voor BMI tussen 26,7-28,4

m: fysieke gezondheid is het grootst voor BMI tussen 24,4-29,5 (dus groter interval)

tweede deel-visual (rechts):

v: voor bovenvermeld interval: 2^e grootste staafkolom (dus oké)

m: voor bovenvermeld interval: idem, maar opnieuw veel data buiten deze staaf

=> huidige gemiddelde BMI: geen pb voor de fysieke gezondheid?

distributies zouden moeten bestudeerd worden:

omgekeerd werken: groot interval nemen voor BMI (rechts)

en nagaan of voor deze de fysieke gezondheid goed is (links)

- slaap: zowel voor m als voor v: best ofwel veel ofwel weinig slapen (aan u de keuze ;-))

maar minimum duidt waarschijnlijk op leeftijd waarop mensen de beste fysieke gezondheid

hebben: zie volgende visual

tweede visual: gemiddeld aantal uren slaap per leeftijdscategorie

4. opmerking

- invloed van hartziekte op mentale gezondheid:

v: vooral van belang op heel hoge leeftijd (>80)

m: reeds van belang vanaf een leeftijd van 60-64j

BESLUIT

hartziekte: vnl bij blanken: en hiervoor vnl voor mannen

- voor blanke mannen, vanaf een leeftijd van 70-75: de 'probleemleeftijd'
 - maar begint reeds rond de 60-65,
dan ook reeds invloed op de mentale gezondheid
- voor blanke vrouwen: geleidelijke overgang van risico ifv leeftijd
 - maw groter risico is +- linear ifv leeftijd

advies: roken vermijden, suikerziekte vermijden

opmerking: alcohol: weinig of geen invloed op hartziekte