

Kunnskapsgrunnlag kompetanse

Dokumentasjon

Dokumentasjon på
datamodell i Power BI,
datalagring m.m.





Innhold

Sammendrag	3
Kilder	3
Datamodellen.....	4
Lagring av data	5
Oppdatering av data	5
Prosedyre for oppdatering	5
Løse kommentarer	6
API skript	7



Sammendrag

Dokumentet dokumenterer grunnlaget for dashboardet og har en veiviser til hvordan oppdatering av dashboardet skal oppdateres. Alt som er gjort i Power BI kan ikke dokumenteres her, da det blir for omfattende. Informasjon om handlinger i Power BI finner man ved å gå gjennom de ulike stegene i Power Query, og ved å studere datamodellens relasjoner. Det viktigste i dette dokumentet er kildene, og hvordan data er lagret. Dersom dashboardet skal oppdateres er det viktig at man setter seg inn i dette dokumentet.

Kilder

Tabellen under viser en oversikt over datakilder, hvor man finner dem, og hvordan data er hentet. Navn på kilden her og i Power Query kan avvike noe grunnet endring av navn etter at kilden er kommet inn i Query.

Kilder		Metode
SSB data		
SSB: 07459	https://data.ssb.no/api/v0/no/table/07459/	API
SSB: 09429	https://data.ssb.no/api/v0/no/table/09429/	API
SSB: 12869	https://data.ssb.no/api/v0/no/table/12869/	API
SSB: 12869-2	https://data.ssb.no/api/v0/no/table/12869/	API
SSB: 13563	https://data.ssb.no/api/v0/no/table/13563/	API
SSB: 13563	https://data.ssb.no/api/v0/no/table/13563/	API
SSB: 12971	https://data.ssb.no/api/v0/no/table/12971/	API
SSB: 13259	https://data.ssb.no/api/v0/no/table/13259/	API
SSB: 13259-2	https://data.ssb.no/api/v0/no/table/13259/	API
Panda Yrkesdata	Databank – Panda analyse 2021 Lønnstakere etter yrke, arbeidssted (4-siffer)	Manuelt uttrekk
SSB Yrkesdata Spesialbestilt	Spesialbestilling til SSB	Manuell bestilling Anders.Ekeland@ssb.no
NAV data		
NAV utenforskap	Mappe med 5 Excel-ark. En fil per region. Filene lastes inn i Query og samles til én felles tabell ved bruk av «append».	Manuell bestilling Gustav.Svane@nav.no
NAV arbeidsledige	Mappe med 5 Excel-ark. En fil per region. Filene lastes inn i Query og samles til én felles tabell ved bruk av «append».	Manuell bestilling Gustav.Svane@nav.no
Bedriftsundersøkelsen	Data fra rapport limt direkte inn i Power Query.	Manuelt uttrekk
Støttetabeller		
Regioninndeling	Manuelt laget i Power Query.	Trenger ikke oppdatering
Fylkesnummer	Manuelt laget i Power Query.	Trenger ikke oppdatering
År	Støttetabell med år fra 1980 til dagens År.	Oppdateres automatisk

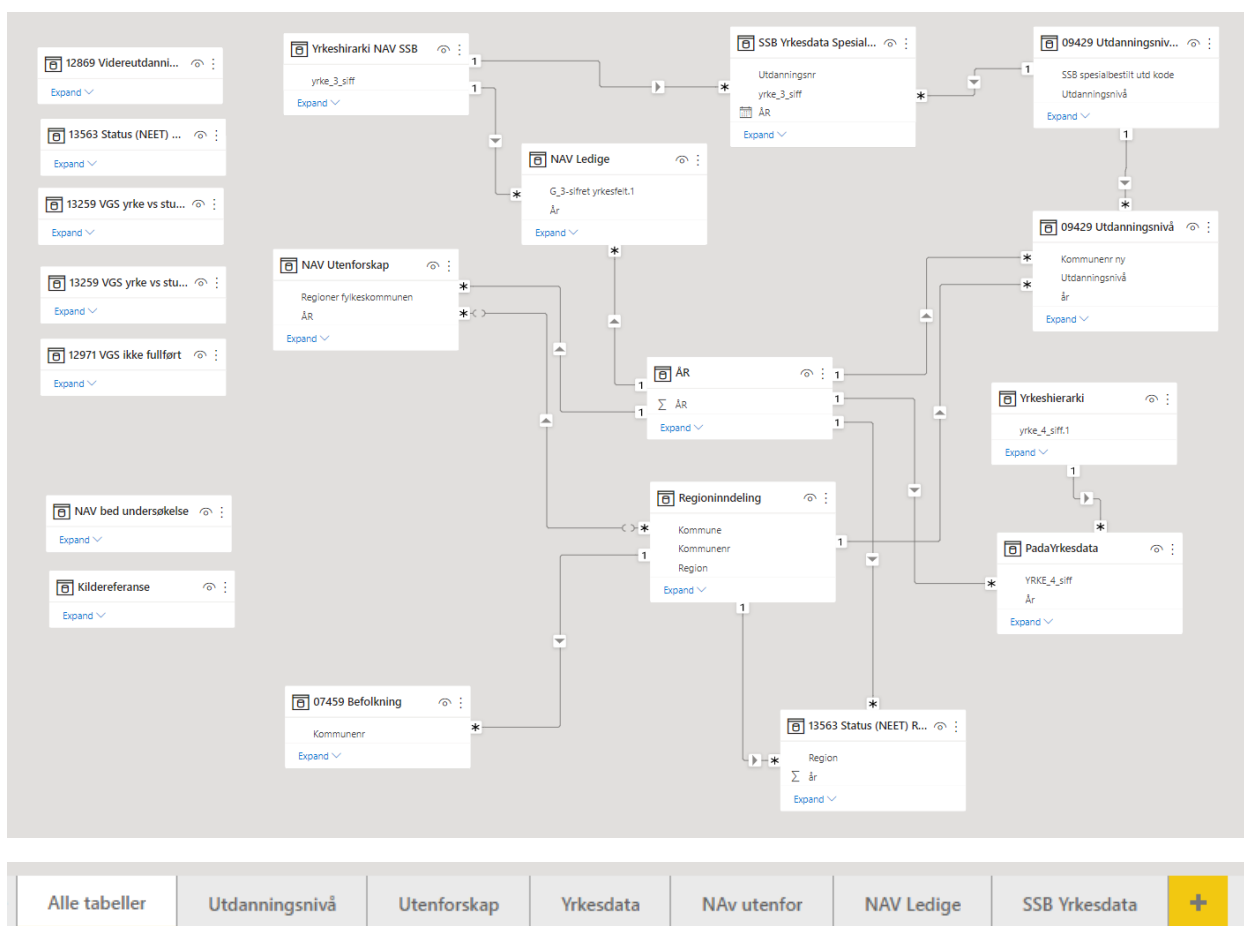


Yrkeshierarki/SSB Klassifikasjoner	SSB.no. Yrkeshierarki som ikke trenger å oppdateres.	Trenger ikke oppdatering
Yrkeshierarki NAV SSB	Hierarki som støtter spesialbestilt data	Trenger ikke oppdatering

Datamodellen

Datamodellen i Power BI er stor og kompleks. Grunnet et ønske om å lage en stor rapport med data fra veldig mange ulike kilder, blir strukturen uoversiktlig. På den annen side er det lite data i hver av tabellene, og kun én faktatabell har mer enn to relasjoner knyttet til seg. Region og År er dimensjonstabeller og har av den grunn flest relasjoner knyttet til seg. Region og År filtrerer andre tabeller. Det er ingen tabeller som filtrerer dem. Dette er viktig for at modellen skal fungere, og kan ikke endres på.

I prinsippet er dette mange datamodeller slått sammen til en stor modell. Derfor er det laget egne faner i modellvisningen som presenterer de enkelte modellene. Tabeller som ikke har relasjon til andre tabeller har ikke fått egen fane.

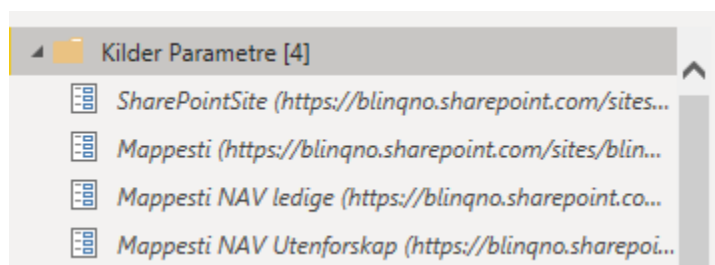




Lagring av data

Lagring av data skjer i SharePoint. Alle kilder er koblet til Power BI ved bruk av Sharepoint-mappe tilkobling. Destinasjonen (altså hvor filene er lagret) kan endres gjennom bruk av parameterne i Power Query. Dersom man flytter kilde-filene til et sted utenfor Sharepoint-området som parameteret peker til, vil ikke oppdatering av data fungere. Power Query vil miste kontakt med kilden og man vil få en feilmelding. Dette fører til at forhåndsvisning av data bryter ned.

I skrivende stund er det ikke gjennomført migrering av data over til Rogaland fylkeskommune sitt miljø. Vi kan derfor ikke vise til den korrekte lagringsplassen, men kan vise til parameterne i Power Query som svar på hvor filene skal ligge.



SharePointSite viser til selve området der filen ligger. Mappesti viser til en mappe som heter Rådata. Her ligger alle kilder til data hentet ved bruk av API. Mappesti NAV ledige og Mappesti NAV Utenforskap viser til en undermappe i Rådata. Når disse parameterne byttes ut med nye stier, oppdateres koden til hver enkelt tabell automatisk. Det skal derfor ikke være nødvendig å gå inn i koden til hver enkelt tabell for å endre stien til kilden.

Oppdatering av data

Fordi det i rapporten brukes flere kilder man må henta data fra manuelt, kan ikke rapporten oppdateres helautomatisk.

Prosedyre for oppdatering

- Data fra API oppdateres gjennom Python (se overskrift API skript).
- Spesialbestilt data fra SSB oppdateres gjennom ny bestilling av data fra SSB. Man bestiller da for det nye året og limer ny data inn i bunnen av dagens fil. En alternativ løsning er å laste inn den nye filen, og *appende/tilføye* inn i dagens fil. Det er viktig at SSB følger nøyaktig sammen skript for uttrekk som året før. Endringer i skriptet kan føre til feil i gjennomføring av eksisterende databehandling i Power Query. MERK! Det er gjort en manuelle operasjoner på regnearket før det ble lagret i Sharepoint. Nivå 1 navnet til Restgruppe 3 og Restgruppe 2 er døpt om til 9 Ukjente yrker. Ukjente yrker er i stor grad



selvstendig næringsdrivende, men vi kan ikke skrive dette da det ikke er SSB's definerte definisjon. At Restgruppe 2 og 3 ligger under Renholdere m.m. i den opprinnelige leveransen fra SSB, er en feil, en feil som trolig har oppstått fordi Restgruppe og Renholdere m.m. begge har kode nr. 9. Man må endre på bestillingen eller fortsette å rette dette opp i fremtiden for at nye og gamle tall skal kunne brukes sammen.

- Panda yrkesdata oppdateres ved å skifte ut gammel med ny fil. Per i dag bestiller Panda Analyse data for hele perioden 2015 – d.d. hvert år. Dersom de går over til å bestille kun for det nye året, må man supplere ny data til kilden gjennom ny kilde og «append», eller manuell innliming av data direkte i filen.
- Spesialbestilte data fra NAV oppdateres gjennom ny bestilling av data fra NAV. Det er viktig at man ikke gjør endringer i skriptet, men kjører det samme uttrekket som tidligere. Man kan velge om man ønsker å ta ut data for hele perioden (2015 – d.d.), eller kun det nye året.
- Bedriftsundersøkelsen er en kilde som er manuelt limt inn i Power Query. Årsaken til at man har gjort det slik, er at det per i dag ikke er noe godt sted å hente dataene fra. Data.Norge har data fra Bedriftsundersøkelsen, men kun frem til 2018. I påvente av en bedre metode lastes Excel-arket ned manuelt, og man klipper ut data fra korrekt tabell. Studer tabellene for det foregående årene for å gjøre deg kjent med hvilke data vi er på jakt etter. Du kan redigere kilden i Power Query og lime en ny kolonne rett inn i tabellen.

Løse kommentarer

- Selv om mål og beregninger er laget for å være dynamiske, kan det ved oppdatering av data oppstå feil som vi ikke har forutsett. Fordi all data oppdateres årlig, har vi ikke hatt anledning til å teste hvordan en oppdatering påvirker visualiseringene.
- Det er god praksis å lagre pbix-filen som en ny fil før man gjennomfører oppdateringen av datasettet første gangen. På den måten kan man teste hvordan oppdatering av data påvirker visualiseringene, og dersom en visualisering bryter sammen kan man studere visualiseringen i filen med utdaterte data.
- Avhengig av oppdragsmengde kan det ta 2-4 uker å få oppdatert data fra SSB. Vis til tidligere gjennomført bestilling, og presiser at man ønsker data for nytt år (2022 tall i 2023 osv.) for Rogaland.



API skript

Henting av data fra SSB Skript

Dette skriptet er skrevet i Python 3 og henter inn data fra utvalgte SSB tabeller med angitte spørringer i JSON.

Main.py

```
import json
import requests
import ssbqueries as ssbq
import pandas as pd
from io import StringIO

def query_ssb(table):
    """
    Query SSB using the given table and query.
    """
    ssbtable = table
    if "-" in table:
        ssbtable = table.split("-")[0]
    headers = {"Content-Type": "text/csv; charset=Windows-1252"}
    url = f"https://data.ssb.no/api/v0/no/table/{ssbtable}"
    r = requests.post(url = url, headers=headers, data =
json.dumps(ssbq.ssb_post_queries[f"{table}"]))
    return r.text

#Loop through dictionary and query SSB for each table

for keys in ssbq.ssb_post_queries:
    print(keys)
    csvString = query_ssb(keys)
    csvStringIO = StringIO(csvString)
    df = pd.read_csv(csvStringIO)
    df.to_csv(f"{keys}.csv", index=None, encoding='utf-8')
```



ssbueries

#SSB JSON POST Queries

```
ssb_post_queries = {
  "09429": {
    "query": [
      {
        "code": "Region",
        "selection": {
          "filter": "agg_single:Komm2020",
          "values": [
            "1101",
            "1103",
            "1106",
            "1108",
            "1111",
            "1112",
            "1114",
            "1119",
            "1120",
            "1121",
            "1122",
            "1124",
            "1127",
            "1130",
            "1133",
            "1134",
            "1135",
            "1144",
            "1145",
            "1146",
            "1149",
            "1151",
            "1160"
          ]
        }
      },
      {
        "code": "Nivaa",
        "selection": {
          "filter": "item",
          "values": [
            "00",

```




```
        "01",
        "02a",
        "11",
        "03a",
        "04a",
        "09a"
    ]
}
},
{
    "code": "Kjonn",
    "selection": {
        "filter": "item",
        "values": [
            "1",
            "2"
        ]
    }
},
{
    "code": "ContentsCode",
    "selection": {
        "filter": "item",
        "values": [
            "Personer"
        ]
    }
},
{
    "code": "Tid",
    "selection": {
        "filter": "item",
        "values": [
            "1980",
            "1981",
            "1982",
            "1983",
            "1984",
            "1985",
            "1986",
            "1987",
            "1988",
            "1989",
            "1990",
            "1991",
            "1992",
            "1993",
            "1994",
            "1995",
            "1996",
            "1997",
            "1998",
            "1999",
            "2000",
            "2001",
            "2002",
            "2003",
```



```
        "2004",
        "2005",
        "2006",
        "2007",
        "2008",
        "2009",
        "2010",
        "2011",
        "2012",
        "2013",
        "2014",
        "2015",
        "2016",
        "2017",
        "2018",
        "2019",
        "2020"
    ]
}
},
"response": {
    "format": "csv2"
}
},
"12869":{
    "query": [
        {
            "code": "Region",
            "selection": {
                "filter": "agg_single:FylkerGjeldende",
                "values": [
                    "11"
                ]
            }
        },
        {
            "code": "UtdanningOppl",
            "selection": {
                "filter": "item",
                "values": [
                    "01",
                    "02"
                ]
            }
        },
        {
            "code": "Kjonn",
            "selection": {
                "filter": "item",
                "values": [
                    "0"
                ]
            }
        },
        {
            "code": "ArbStyrkStatus",
```



```
      "selection": {
        "filter": "item",
        "values": [
          "1uS"
        ]
      }
    },
    ],
    "response": {
      "format": "csv2"
    }
  },
  "12869-2" :{
    "query": [
      {
        "code": "Region",
        "selection": {
          "filter": "vs:Landet",
          "values": [
            "0"
          ]
        }
      },
      {
        "code": "UtdanningOppl",
        "selection": {
          "filter": "item",
          "values": [
            "01",
            "02"
          ]
        }
      },
      {
        "code": "Kjonn",
        "selection": {
          "filter": "item",
          "values": [
            "0"
          ]
        }
      },
      {
        "code": "ArbStyrkStatus",
        "selection": {
          "filter": "item",
          "values": [
            "1uS"
          ]
        }
      }
    ],
    "response": {
      "format": "csv2"
    }
  }
}
```



```
},
"13563":{
  "query": [
    {
      "code": "Region",
      "selection": {
        "filter": "agg:KommSummer",
        "values": [
          "K-1101",
          "K-1103",
          "K-1106",
          "K-1108",
          "K-1111",
          "K-1112",
          "K-1114",
          "K-1119",
          "K-1120",
          "K-1121",
          "K-1122",
          "K-1124",
          "K-1127",
          "K-1130",
          "K-1133",
          "K-1134",
          "K-1135",
          "K-1144",
          "K-1145",
          "K-1146",
          "K-1149",
          "K-1151",
          "K-1160"
        ]
      }
    }
  ],
  {
    "code": "HovArbStyrkStatus",
    "selection": {
      "filter": "vs:ArbStatus2018niva4a",
      "values": [
        "TOT3",
        "NEET2",
        "A.01xU.01xU.03"
      ]
    }
  },
  {
    "code": "Alder",
    "selection": {
      "filter": "item",
      "values": [
        "15-29",
        "30-61",
        "62+"
      ]
    }
  },
  {

```



```
    "code": "InnvandrKat",
    "selection": {
      "filter": "item",
      "values": [
        "A_C-G",
        "B"
      ]
    }
  },
  "response": {
    "format": "csv2"
  }
},
"13563-2":{
  "query": [
    {
      "code": "Region",
      "selection": {
        "filter": "vs:Landet",
        "values": [
          "0"
        ]
      }
    },
    {
      "code": "HovArbStyrkStatus",
      "selection": {
        "filter": "vs:ArbStatus2018niva4a",
        "values": [
          "TOT3",
          "NEET2",
          "A.01xU.01xU.03"
        ]
      }
    },
    {
      "code": "Alder",
      "selection": {
        "filter": "item",
        "values": [
          "15-29",
          "30-61",
          "62+"
        ]
      }
    }
  ],
  {
    "code": "InnvandrKat",
    "selection": {
      "filter": "item",
      "values": [
        "A_C-G",
        "B"
      ]
    }
  }
}
```



```
    }
  },
  ],
  "response": {
    "format": "csv2"
  }
},
"12971":{
  "query": [
    {
      "code": "Region",
      "selection": {
        "filter": "item",
        "values": [
          "01-99",
          "11"
        ]
      }
    },
    {
      "code": "FullforingVGO",
      "selection": {
        "filter": "item",
        "values": [
          "1a",
          "2a",
          "4b",
          "003a",
          "8b",
          "9"
        ]
      }
    },
    {
      "code": "UtdProgram",
      "selection": {
        "filter": "item",
        "values": [
          "01",
          "02"
        ]
      }
    },
    {
      "code": "Kjonn",
      "selection": {
        "filter": "item",
        "values": [
          "1",
          "2"
        ]
      }
    },
    {
      "code": "ContentsCode",
      "selection": {
```



```
        "filter": "item",
        "values": [
            "Personer"
        ]
    },
    {
        "code": "Tid",
        "selection": {
            "filter": "item",
            "values": [
                "2010-2016",
                "2011-2017",
                "2012-2018",
                "2013-2019",
                "2014-2020"
            ]
        }
    },
    ],
    "response": {
        "format": "csv2"
    }
},
"13259":{
    "query": [
        {
            "code": "Region",
            "selection": {
                "filter": "agg_single:FylkerGjeldende",
                "values": [
                    "11"
                ]
            }
        },
        {
            "code": "Kurstrinn",
            "selection": {
                "filter": "item",
                "values": [
                    "1"
                ]
            }
        },
        {
            "code": "StudieRetn",
            "selection": {
                "filter": "item",
                "values": [
                    "01",
                    "02"
                ]
            }
        }
    ],
    "response": {
        "format": "csv2"
    }
}
```



```
    }

},
"13259-2":{
  "query": [
    {
      "code": "Region",
      "selection": {
        "filter": "vs:Landet",
        "values": [
          "0"
        ]
      }
    },
    {
      "code": "Kurstrinn",
      "selection": {
        "filter": "item",
        "values": [
          "1"
        ]
      }
    },
    {
      "code": "StudieRetn",
      "selection": {
        "filter": "item",
        "values": [
          "01",
          "02"
        ]
      }
    }
  ],
  "response": {
    "format": "csv2"
  }
},
"07459":{
  "query": [
    {
      "code": "Region",
      "selection": {
        "filter": "agg:KommSummer",
        "values": [
          "K-1101",
          "K-1103",
          "K-1106",
          "K-1108",
          "K-1111",
          "K-1112",
          "K-1114",
          "K-1119",
          "K-1120",

```




```
        "K-1121",
        "K-1122",
        "K-1124",
        "K-1127",
        "K-1130",
        "K-1133",
        "K-1134",
        "K-1135",
        "K-1144",
        "K-1145",
        "K-1146",
        "K-1149",
        "K-1151",
        "K-1160"
    ]
}
},
{
    "code": "Alder",
    "selection": {
        "filter": "agg:FemAarigGruppering",
        "values": [
            "F00-04",
            "F05-09",
            "F10-14",
            "F15-19",
            "F20-24",
            "F25-29",
            "F30-34",
            "F35-39",
            "F40-44",
            "F45-49",
            "F50-54",
            "F55-59",
            "F60-64",
            "F65-69",
            "F70-74",
            "F75-79",
            "F80-84",
            "F85-89",
            "F90-94",
            "F95-99",
            "F100G5+"
        ]
    }
}
},
{
    "code": "Tid",
    "selection": {
        "filter": "item",
        "values": [
            "2015",
            "2016",
            "2017",
            "2018",
            "2019",
            "2020",
```



```
        "2021",
        "2022"
    ]
}
],
"response": {
    "format": "csv2"
}
},
}
```