Data Analyst – analytical mind + coding

Data Scientist – solid tech background ML

BI Analyst – SQL + BI tools

Data Engineer – Tech infrastructure + Cloud

**DAY 2 – DATA EXERCISE**

<https://iclimate.au.dk/services> - actions to reduce Denmark pollution.

<https://iclimate.au.dk/index.php?id=60613> – artiles from Iclimate, focus in reduce pollution until 2030.

<https://www2.dmu.dk/1_viden/2_miljoe-tilstand/3_luft/4_maalinger/5_database/hentdata_en.asp> - air polution dataset from Copenhagen from 1983 to 2019.

<http://lpdv-en.spatialsuite.dk/spatialmap> - map with pollution since 2012

<https://dce2.au.dk/pub/SR234.pdf> - 2016 annual summary for the danish air quality monitoring programme

<https://www.iqair.com/us/denmark/capital-region/copenhagen> - hour and daily AQI – air quality index – but data available only for one month.

**18.02 DAY 2 – THEORY**

* Pandas is designed for 2 dimensions (row: axis =0 and columns: axis=1);
* <https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html> -- a guide for pandas library to use concat, merge, etc;
* Need to pay attention in data.info() – normally you can find some problems while merging or importing to jupyter. In class we could see that had 4028 entries, but the index was limited going to 0 to 1025. Losing almost 3000 of rows – we need to reset index pandas or we could just ignore index while doing the concat.
* Today we saw:
  + How to import csv to jupyer (also .txt)
  + Clean methods to do in the database
  + Filter methods