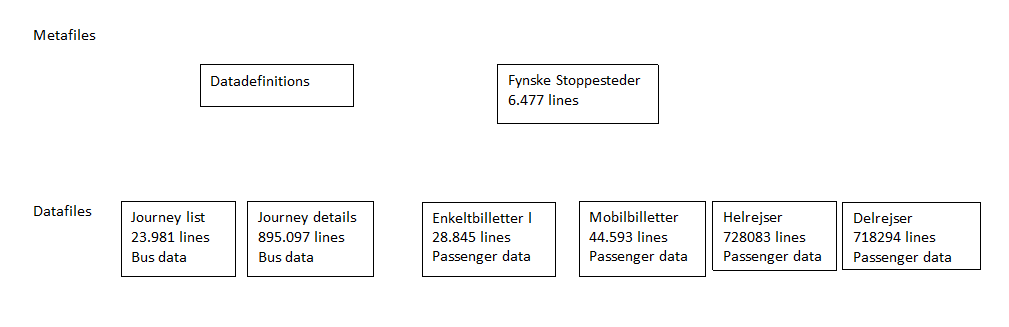
# First Data Run



## Which questions can you ask?

**External Data**

* Include weather data from DMI
* Correlation between weather and public transportation habits

**Segments**

* Segment – child, adult, retiree tickets, bikes, dogs, passengers with special needs.
* Correlation between delays and passengers with special needs – do we need to reserve more time in the bus stops where more handicapped passengers get on to decrease delays on average?
* Are there more forgotten check outs with personal travel cards vs. anonymous travel cards.

**Bus Delays**

* Correlation between number of passengers and delays
* Rush hour – øget rejsetid – hvad tid på dagen folk rejser og hvor mange forsinkelser der er.
* Hvornår begynder forsinkelser og hvornår stopper de?
* ”-1” kolonner og deres påvirkning på forsinkelser
* Visualisere med kortdata - evt. sammenligne data med DMI (tager folk mere bus, når det regner)
* Forskel mellem forskellige grupper og vejret - nogle tager bus uanset, andre efter vejret.
* Studerende og pensionister tager nok bus uanset.
* Forsinkelser: Sammenhænge mellem hvor forsinket bussen er, og parametre som kapacitet eller hvor mange, der stiger på ved de enkelte stop.
* Stoppesteder: Hvor langt ligger de fra hinanden? Er der så mange, der stiger på? Eller er dækningen af stoppestederne tilpas, når man ser på, hvor mange der stiger på.

## What kind of software can you deploy to analyze the data for which goal?

**Python**

Plotly

Plotly dash

Logistic regression

**R**