Verslag project Big Data en BI

Teamleden: Thomas Billiet, Ishan Ameel, Sven Depickere, Jonas Anseel

# Preprocessing van de dataset

In het begin hebben we nog 11.401.196 records.

De eerst stap is de dubbele records verwijderd

1. copy.drop\_duplicates();

Na deze code uitgevoerd te hebben zijn er 649.275 records verwijderd.

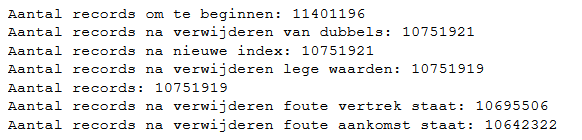
Daarna controleren over lege waardes tussen zitten.

1. copy.dropna()

Daardoor zijn 2 records verwijderd. Er zijn momenteel nog 10.751.919 records.

Er zijn bepaalde vertrek/aankomst staten die geen officiële staten zijn van de United States Of America. Deze staten zijn:

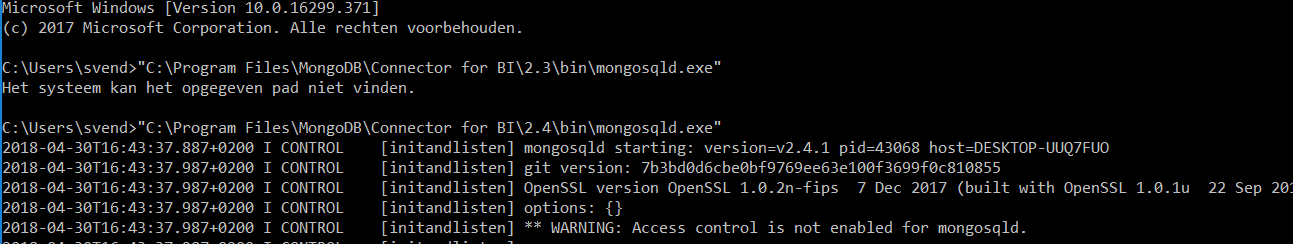
* TT staat voor:
  + Trust Territory of the Pacific Islands
    - Bestaat uit:
      * Republic of the Marshall Islands
      * Federated States of Micronesia
      * Republic of Palau
      * Commenwealth of the Northern Mariana Islands
* PR staat voor:
  + Dddd
* VI staat voor:
  + ddddd





# Tableau

"C:\Program Files\MongoDB\Connector for BI\2.4\bin\mongosqld.exe"



# Machine learning met Scikit learn

# NetworkX