Project Big Data, Business Intelligence and NoSQL   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Teamleden:

Thomas Billiet

Ishan Ameel

Sven Depickere

Jonas Anseel

Inhoud

[1. Preprocessing van de dataset 3](#_Toc513390259)

[2. Tableau 4](#_Toc513390260)

[3. Machine learning met Scikit learn 4](#_Toc513390261)

[4. NetworkX 4](#_Toc513390262)

# 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 niet meer sinds 1986)
    - Bestaat uit:
      * Republic of the Marshall Islands
      * Federated States of Micronesia
      * Republic of Palau
      * Commenwealth of the Northern Mariana Islands
* PR staat voor:
  + Puerto Rico (is een territorium)
* VI staat voor:
  + U.S Virgin Islands (is een territorium)

Na het verwijderen van deze records zijn er nog: 10.642.322 over.

# Tableau

Inhoud:

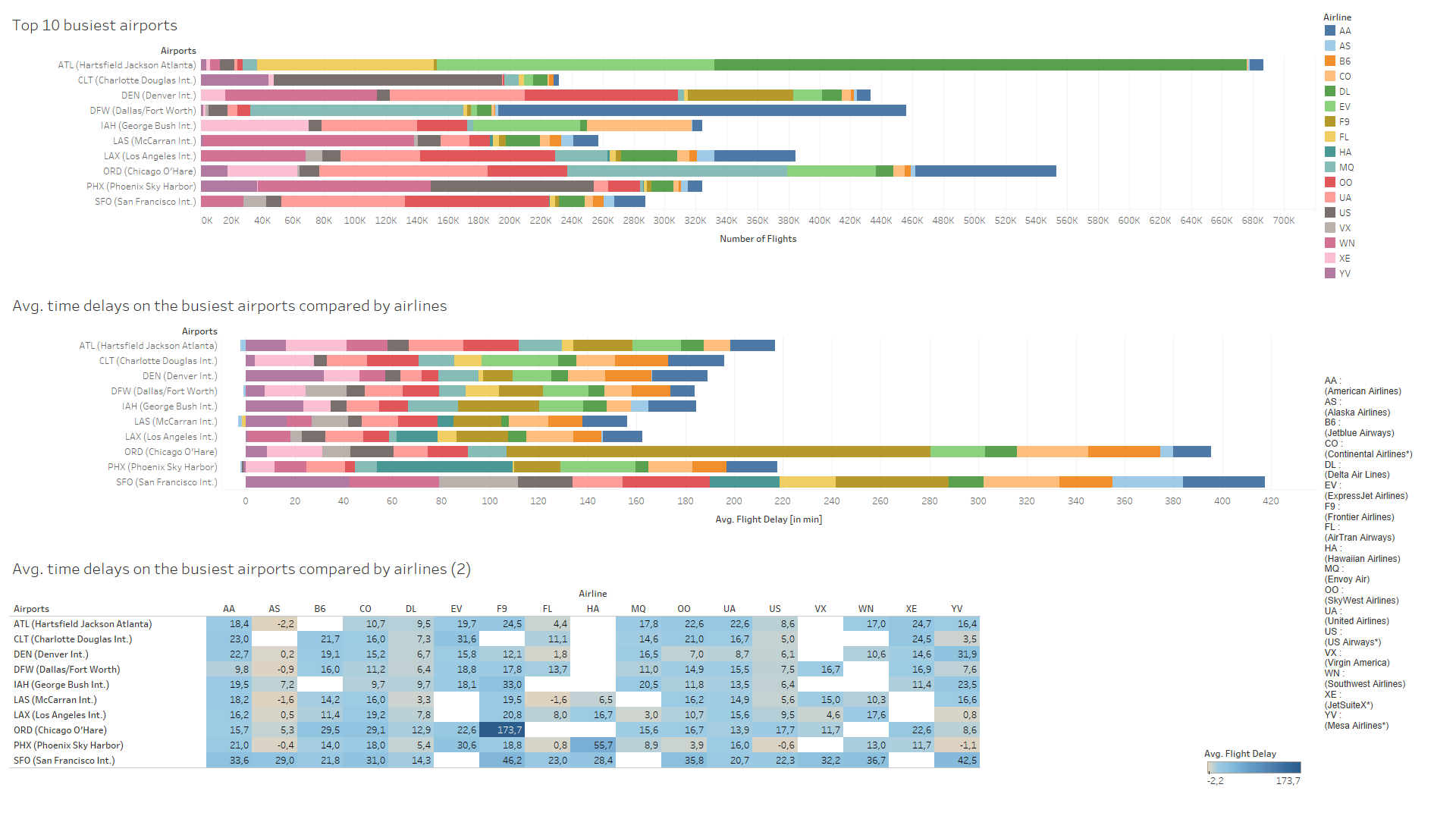
Grafiek 1

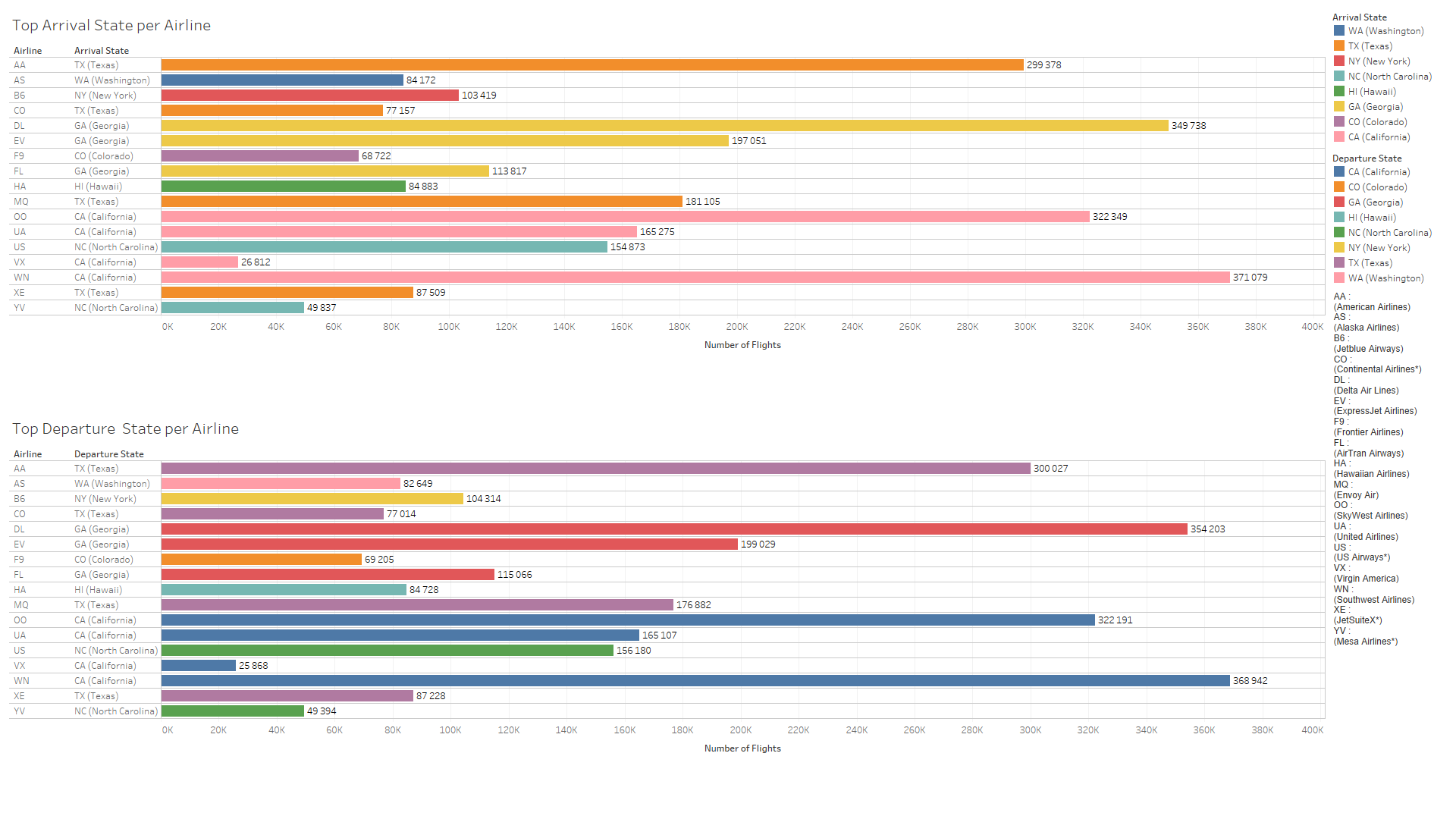
Grafiek 1

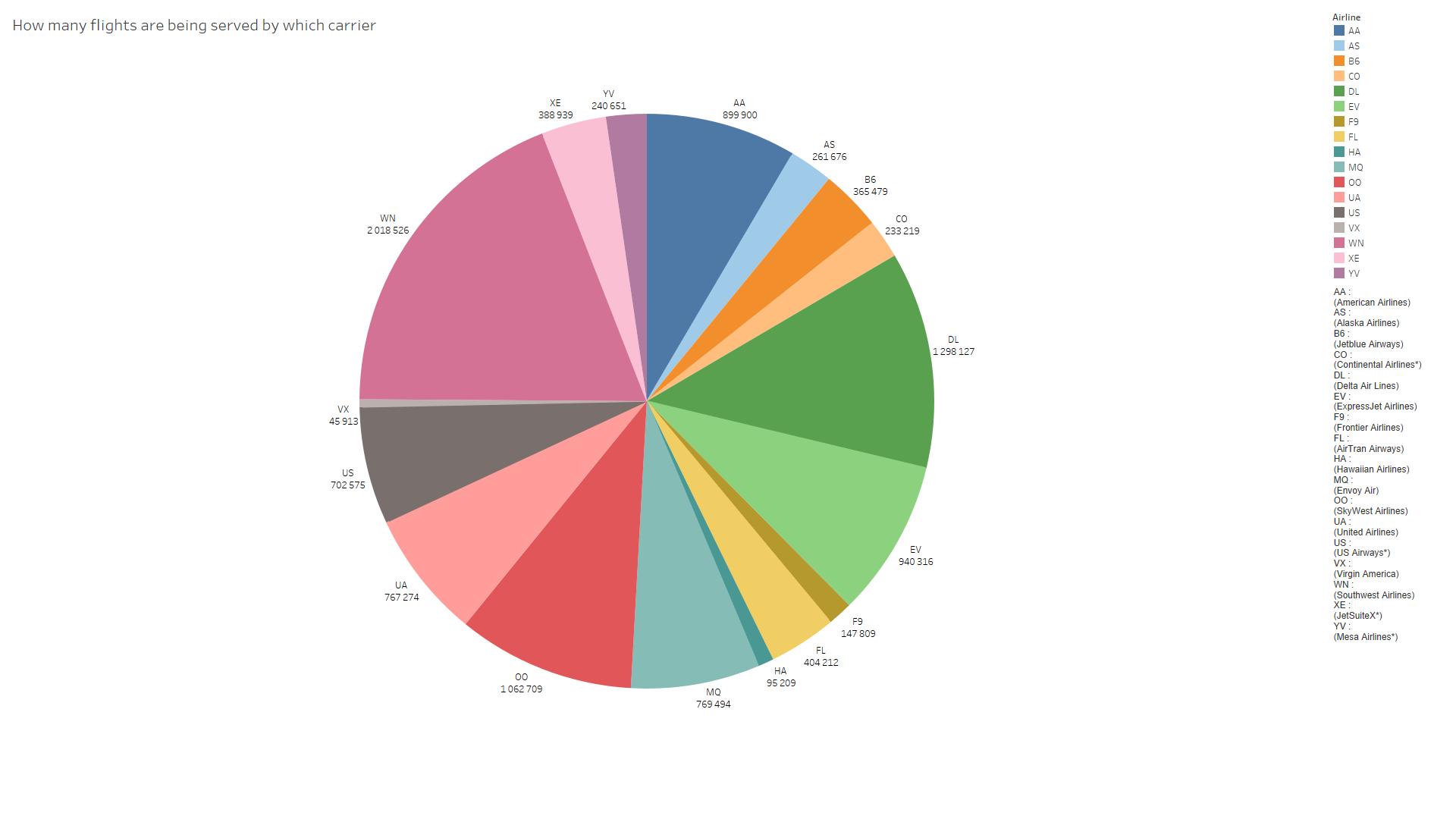
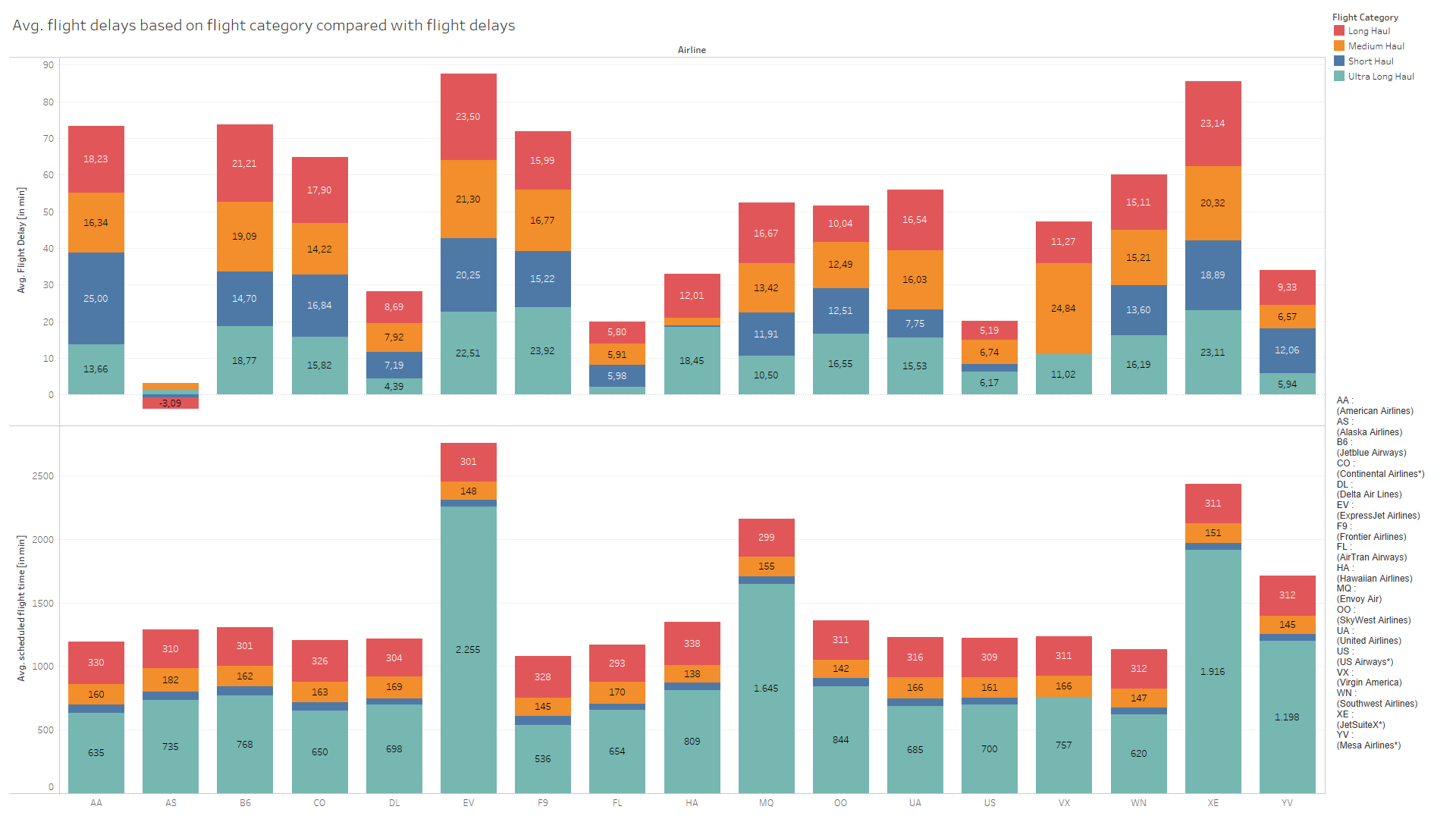
Grafiek 1

Grafiek 1

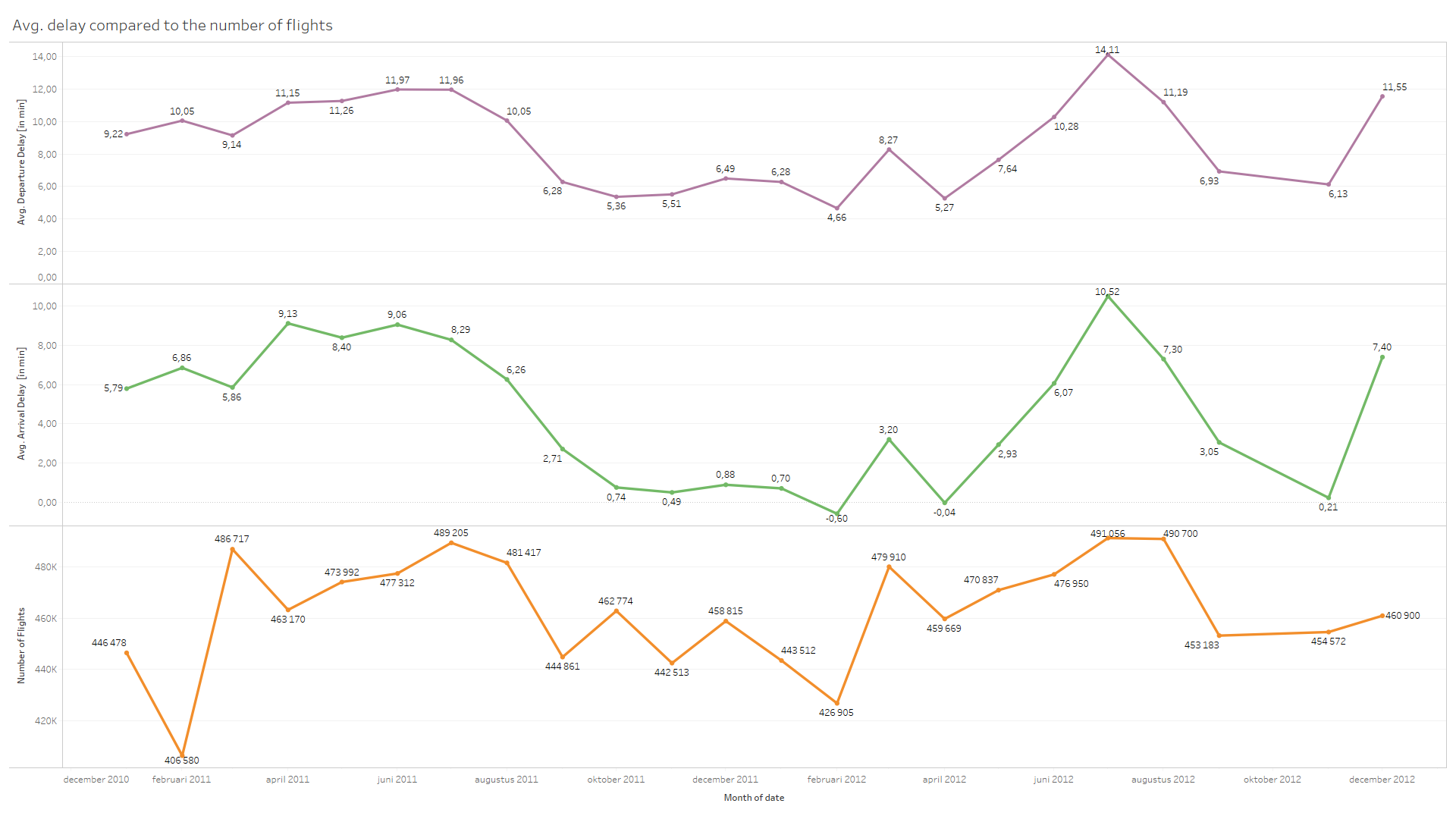
Grafiek 1

Grafiek 1,2 en 3:

Grafiek 4 en 5:

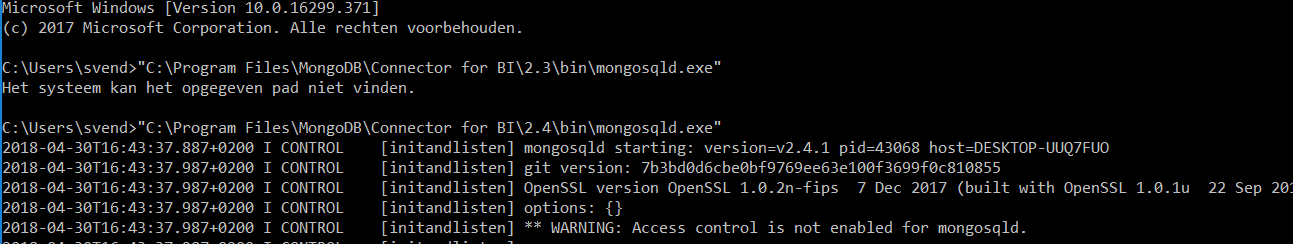
Grafiek 6: Grafiek 7: 

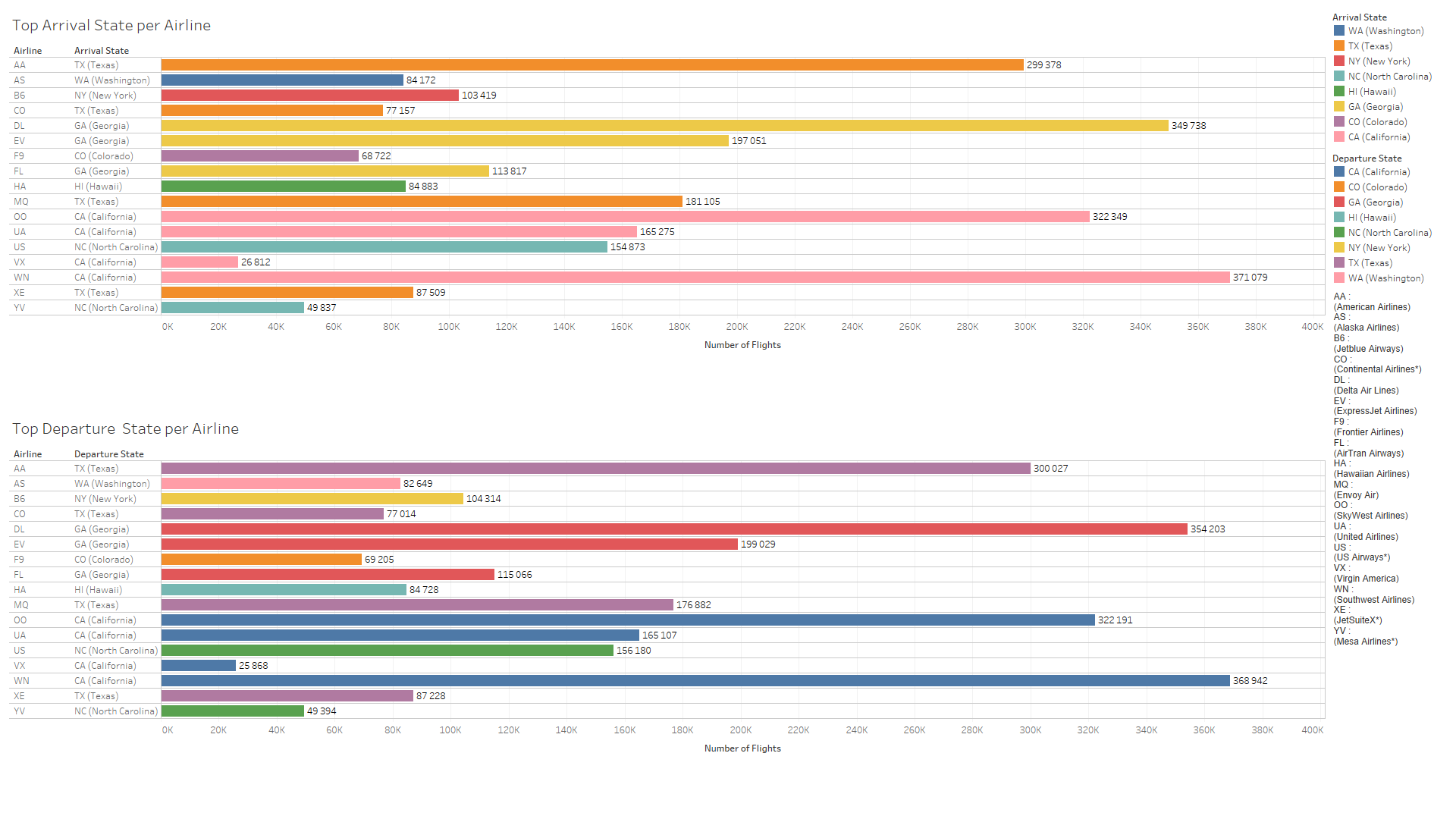
Grafiek 8:



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

cd C:\Program Files\MongoDB\Server\3.6\bin





# Machine learning met Scikit learn

# NetworkX