Mathias Caelenberghe, Nick Engels, Matthias Durivet, Matthias Bruynooghe

[E-mailadres]

Abstract

[Trek de aandacht van uw lezer met een interessante samenvatting. Dit is meestal een kort overzicht van het document.   
Wanneer u uw inhoud wilt toevoegen, klikt u hier en begint u te typen.]

Big Data Project

[Ondertitel van document]

Inhoudsopgave

[1 Data Cleaning 2](#_Toc513715125)

[2 Tableau 2](#_Toc513715126)

# Data Cleaning

Om te beginnen bij de data cleaning hebben we alle 6 sets ingeladen en gecombineerd in 1 grote set. We hebben de dubbele entries eruit verwijderd uit de tabel.

We hebben gecontroleerd of er onvolledige kolommen waren deze bleken de kolommen ‘Scheduled\_departure’ en ‘scheduled\_arrival’ incompleet te zijn. Aangezien de data uit de kolommen ‘actual\_departure’, ‘actual\_arrival’ en de elk respectievelijke delay kolommen volledig in gevuld zijn kunnen we hieruit de scheduled kolommen aanvullen.

Vervolgens hebben we de kolommen ‘departure\_delay’ en ‘arrival\_delay’ omgezet naar integers.

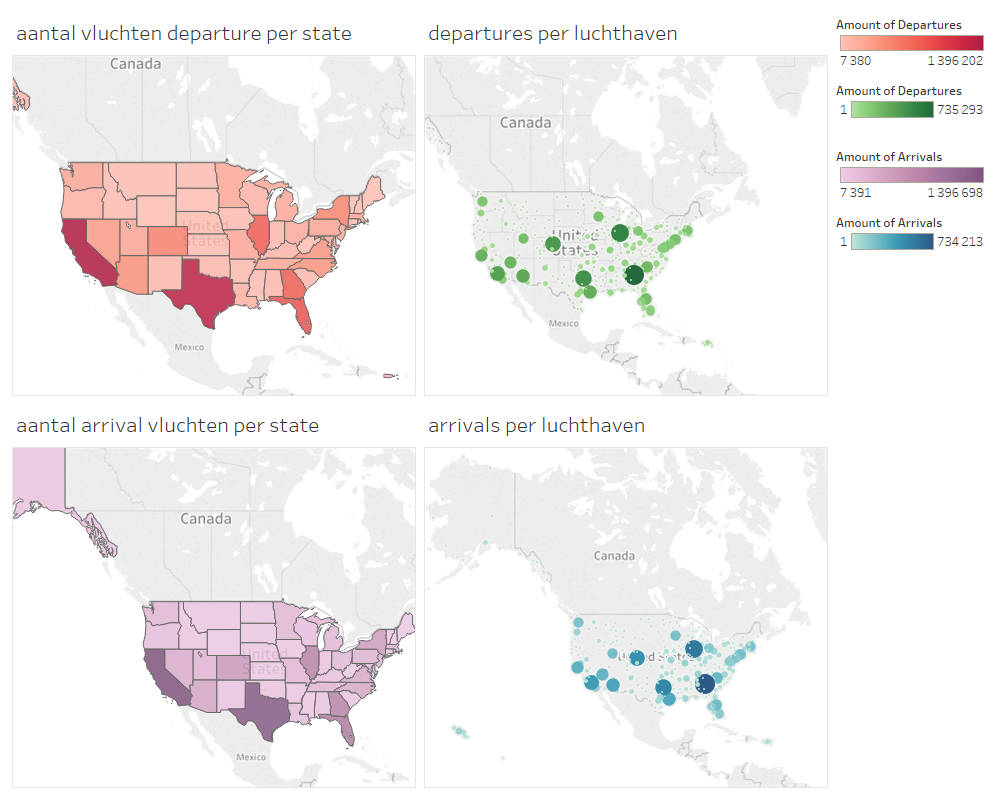
# Tableau

## Analyse aantal vluchten

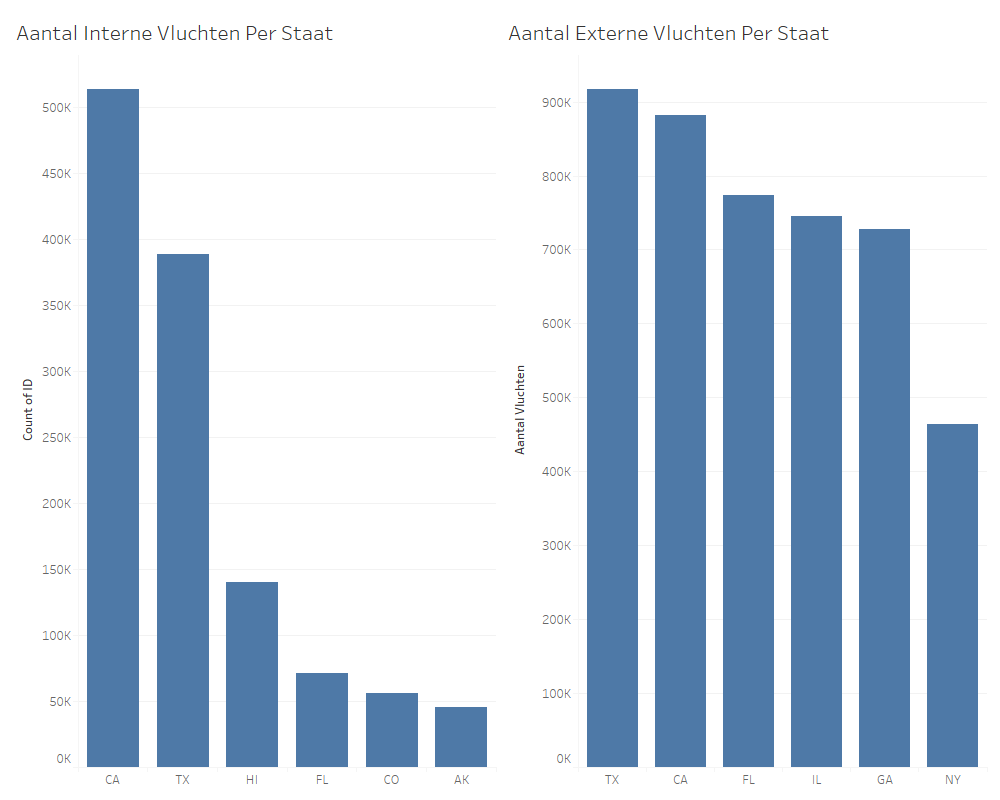
Uit de analyse via Tableau hebben we ontdekt dat het grootste aantal vluchten aankomen en vertrekken uit Atlanta Airport, maar de twee drukste staten van zijn Californië gevolgd door Texas.

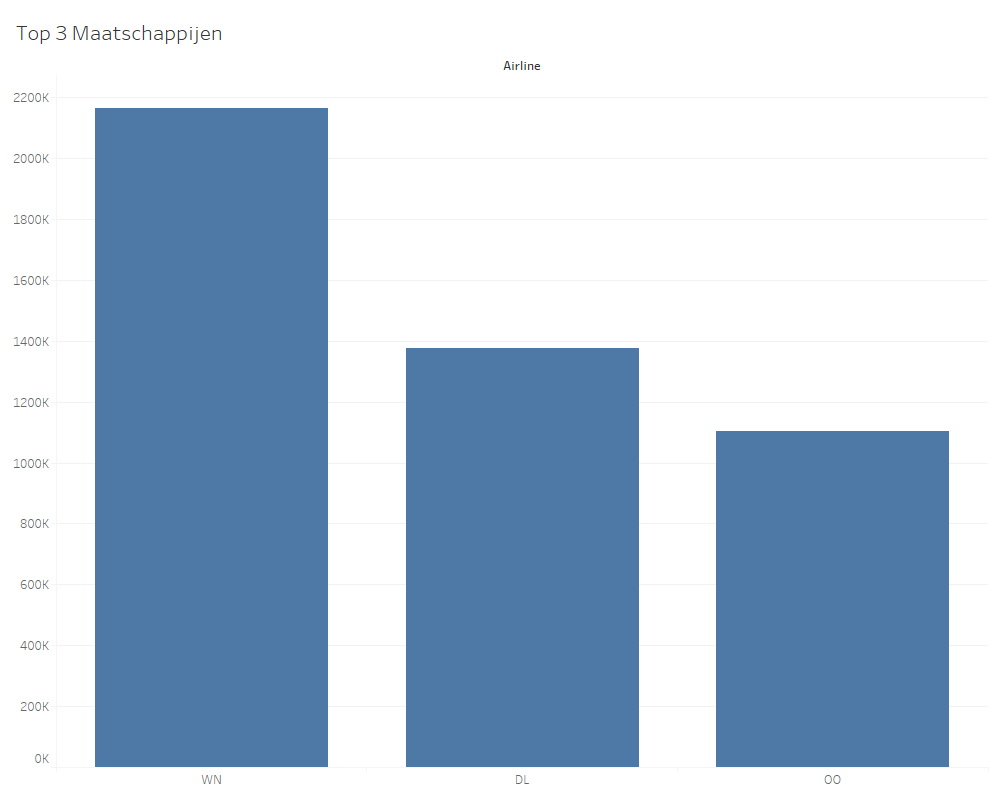
hoewel Atlanta de luchthaven is met de meeste vluchten heeft is dit de enige grote luchthaven in de staat Georgia Atlanta met 735 293 vertrekkende vluchten per jaar en de tweede grootste luchthaven is Savannah Hilton Head International Airport met slechts 17 550 vluchten per jaar.

Californië en Texas daarentegen hebben meerdere grote luchthavens waar de top 3 meer dan 90 000 vertrekkende Vluchten heeft jaar.



Bij Californië is het ons opgevallen dat iets meer dan een derde van de vluchten die vertrekken vanuit een luchthaven van Californië ook landen op een luchthaven van Californië. Texas heeft een gelijkaardig fenomeen maar hier gaat het slechts om iets meer dan een vijfde van de vluchten.



Het aantal vluchten per luchtvaartmaatschpij blijkt dat de top 3 maatschappijen de low-cost maatschappij Southwest Airlines voor 2 165 114 vand de vluchten verantwoordlijke zijn gevolgd door Delta Airlines voor 1 376 347 vluchten en SkyWest Airlines op de derde plaats met 1 103 192 vluchten.

## Analyse Vertragingen

Om over te gaan naar wat niemand graag heeft op zijn vlucht de vertragingen. We hebben uit de data kunnen afleiden dat er in de zomermaanden, juni, juli en augustus, De meeste vertragingen heeft