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](#_Toc513707669)

# Data Cleaning

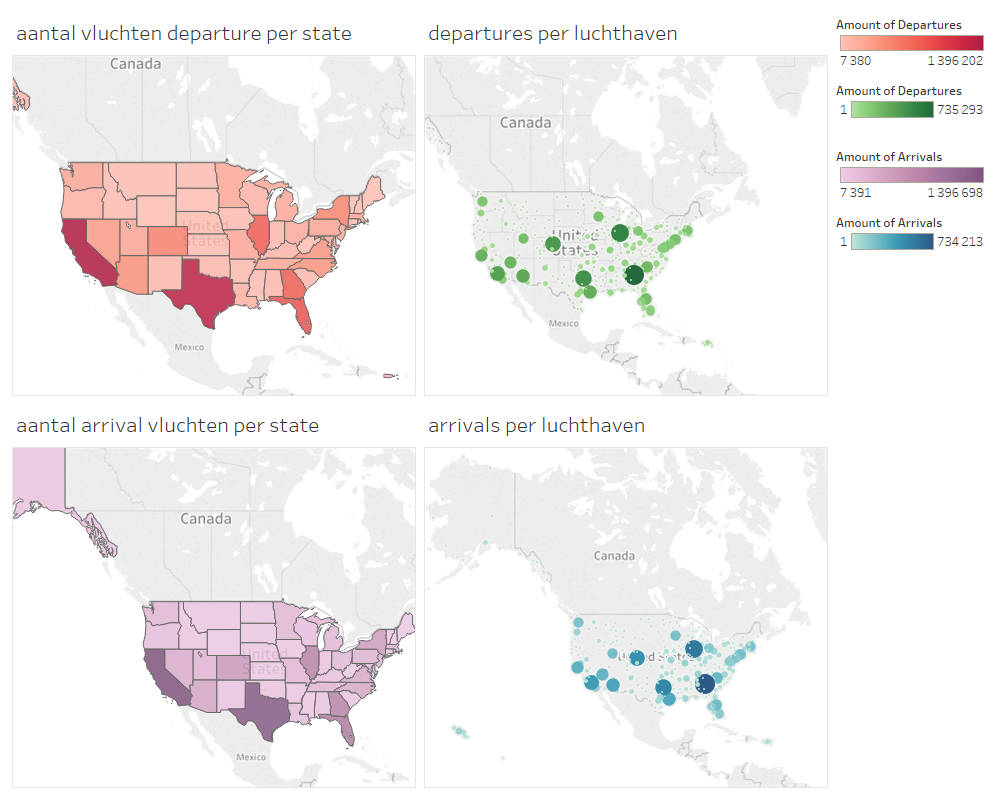
Om te beginnen bij de data cleaning hebben we alle 6 sets ingeladen en gecombineerd in 1 grote set. We hebben de dubbele entries eruit verwijderd uit de tabel.

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

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.



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.

