**Naam** Stefan de Jong  
**Student-nr** 625583

**Git URL** https://github.com/Stefan-de-jong/pdp\_opdrachten

**Uitleg**

Opdracht 3 bestaat uit een python script pdp\_assignment\_3.py en bijbehorende databestand titanic.csv

Ik ben onwijs lang bezig geweest om Spark überhaupt werkend te krijgen aangezien dit puur op basis van documentatie en zelfstudie moest. Uiteindelijk is het gelukt, maar dat lijkt mee geluk dan wijsheid te zijn geweest als ik eerlijk ben. Hoewel ik de demo heb proberen te volgen krijg ik het niet voor elkaar om echt gebruik te maken van ML algoritmes of modellen. Voor de uiteindelijk uitkomsten ben ik bij vraag 1, 2 en 3 dan ook maar uitgegaan van het zelf berekenen van de waardes..

Script kan gedraaid worden (mits titanic.csv in dezelfde folder staat) met:

python pdp\_assignment\_3.py

Om met spark te draaien: spark-submit pdp\_assignment\_3.py

**Screenshots**

**Via python:**

**Afbeelding met tekst

Automatisch gegenereerde beschrijving**

**Via spark:**

**Afbeelding met tekst

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst

Automatisch gegenereerde beschrijving**

**Afbeelding met tekst, schermafbeelding, elektronica, computer

Automatisch gegenereerde beschrijving**