PE Data Advanced

Team:

* Joachim Goris
* Jeroen Verwimp

Inhoud

[1 Opgaven 1a 1](#_Toc5881722)

[1.1 Hoe werd de dataset ingevoerd in python? 1](#_Toc5881723)

[1.2 Genereer zelf de geboortedatum 1](#_Toc5881724)

[1.3 Genereer de kolom inzet met 3 mogelijke categorieën 2](#_Toc5881725)

[1.4 Spreidingsdiagram: lengte en gewicht 2](#_Toc5881726)

[1.5 Staafdiagram van het aantal gemaakte goalen per positie 3](#_Toc5881727)

[1.6 Gemiddelde en de modus van kolom D (aantal gemaakte goalen) per positie 3](#_Toc5881728)

[1.7 Kwartiel 1 en standaardafwijking van kolom G (gewicht) 3](#_Toc5881729)

[1.8 Verband tussen positie op het veld en het aantal goals gescoord 4](#_Toc5881730)

[1.9 Cirkeldiagram: verdeling van de inzet 5](#_Toc5881731)

[1.10 Boxplot: aantal gemaakte goalen van posities linkervleugel, rechtervleugel en piloot 6](#_Toc5881732)

[1.11 Welke soort gegeven is ‘aantal gemaakte goalden’, ‘inzet’ en ‘gewicht’ 7](#_Toc5881733)

[2 Opgave 1b 7](#_Toc5881734)

[3 Opgave 2 7](#_Toc5881735)

# Opgave 1a

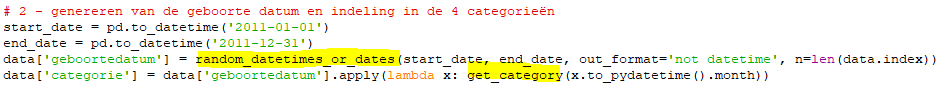
## Hoe werd de dataset ingevoerd in python?

Hiervoor hebben wij de python *library* “pandas” gebruikt. Deze heeft een ingebouwde functie voor het inladen van Excel documenten.

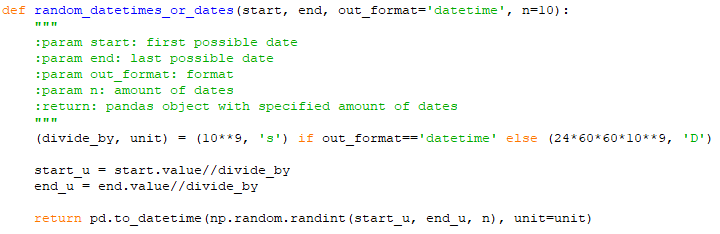


## Genereer zelf de geboortedatum

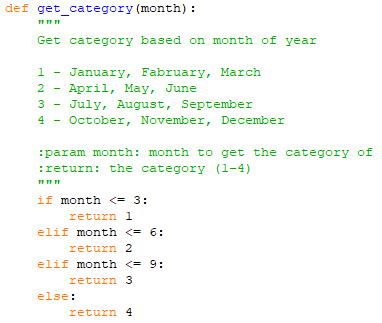
Om een willekeurige geboorte datum te genereren hebben we een functie gekozen die we hebben gevonden op StackOverflow. Wij hebben er ook voor gekozen om een extra kolom categorie in te voegen, omdat we deze later gemakkelijker kunnen gebruiker van het genereren van de kolom inzet.



Functie om een reeks van random datums te genereren:



Functie om de maand van de geboorte datum om te zetten naar de categorie:



(deze functie is zelf geschreven)

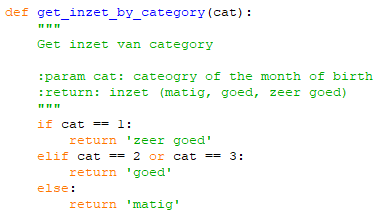
## Genereer de kolom inzet met 3 mogelijke categorieën

Om te kolom inzet te genereren hebben we een functie geschreven die de kolom categorie omzet in de inzet.

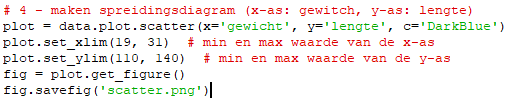
Dit word gedaan door een lambda functie die op elke rij word uitgevoerd.

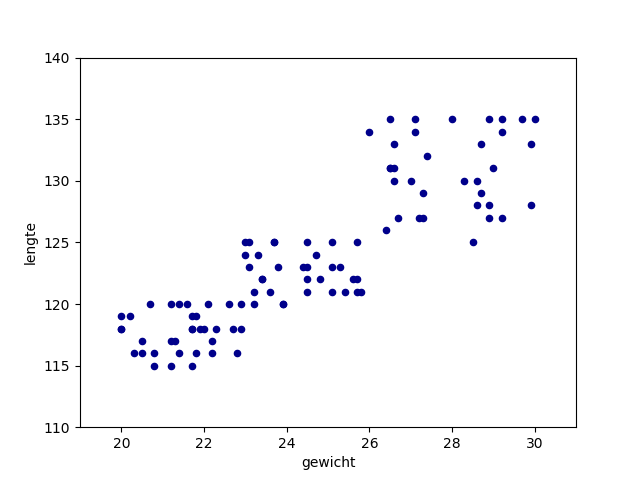


Functie om de categorie om te zetten naar inzet:

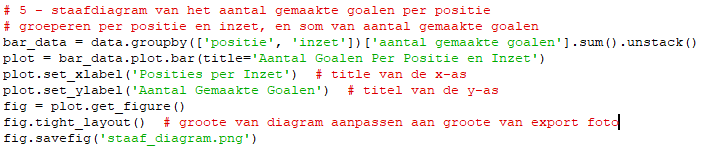


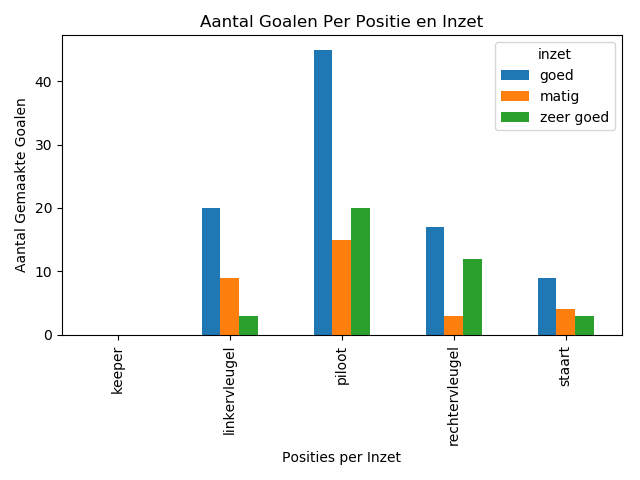
## Spreidingsdiagram: lengte en gewicht





## Staafdiagram van het aantal gemaakte goalen per positie en inzet

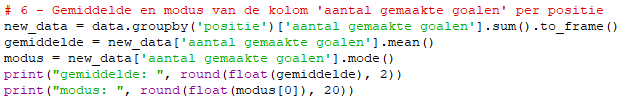




Uit deze staaf diagram kunnen we afleiden dat het aantal gemaakte goals gelinkt is aan de positie op het veld. Hoe dichten een speler’s positie bij de goal van de tegenstanders hoe meer goals hij/zij kan maken.

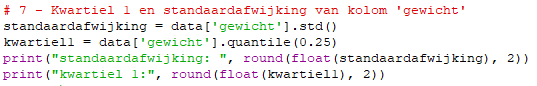
We zien ook dat speler met een goede inzet meer goals maken dan spelers met een zeer goede inzet. Dit is echter niet noodzakelijk correct, er zijn ook gewoon meer spelers met een goede inzet dan een zeer goede inzet.

## Gemiddelde en de modus van kolom D (aantal gemaakte goalen) per positie



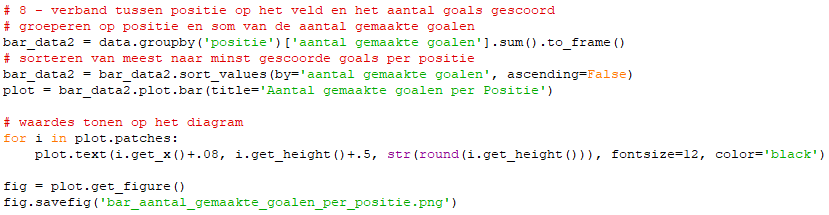


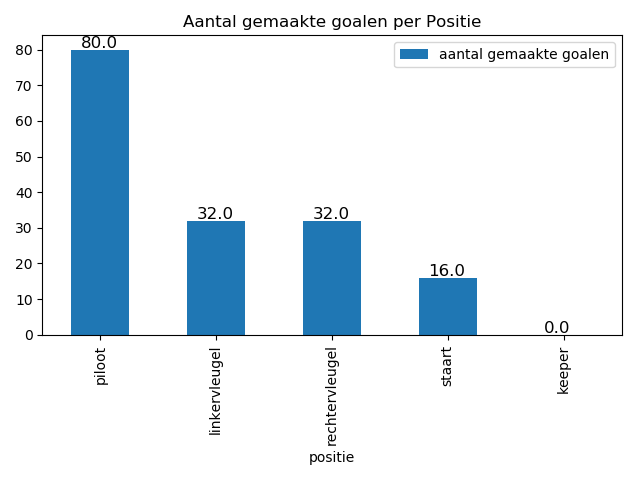
## Kwartiel 1 en standaardafwijking van kolom G (gewicht)





## Verband tussen positie op het veld en het aantal goals gescoord



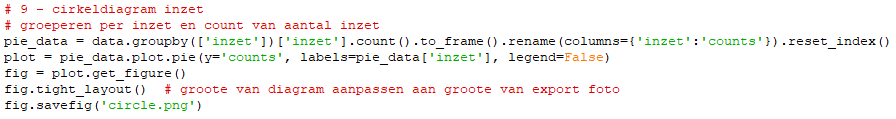


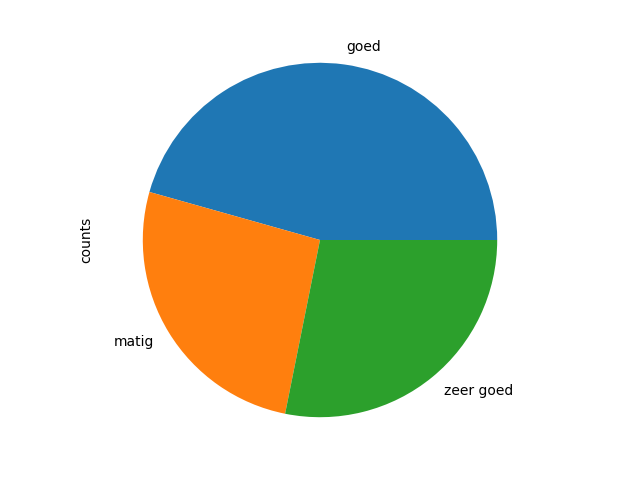
We zien in deze staafdiagram dat de positie op het veld gelinkt is aan het aantal goals dat er gemaakt word. De spelers dichter bij het doel van de tegenspelers maken meer goals dan de speler dichter bij hun eigen goal.

Dit is zeker ook logische sinds een keeper niet uit zijn goals zal komen om bij de tegenstanders te scoren. Hetzelfde geld voor speler op de staart positie, zij houden zich meer bezig met verdedigen ipv aanvallen.

Om dit te bekomen hebben we het aantal gemaakt goals per positie opgeteld en gesorteerd op het aantal gemaakte goals. Zo hebben we gezien dat de grafiek in volgorde van positie op het veld staat. Piloten vooraan en keeper achteraan.

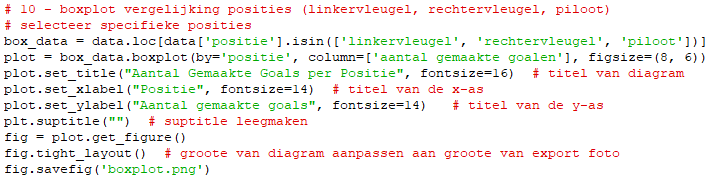
## Cirkeldiagram: verdeling van de inzet

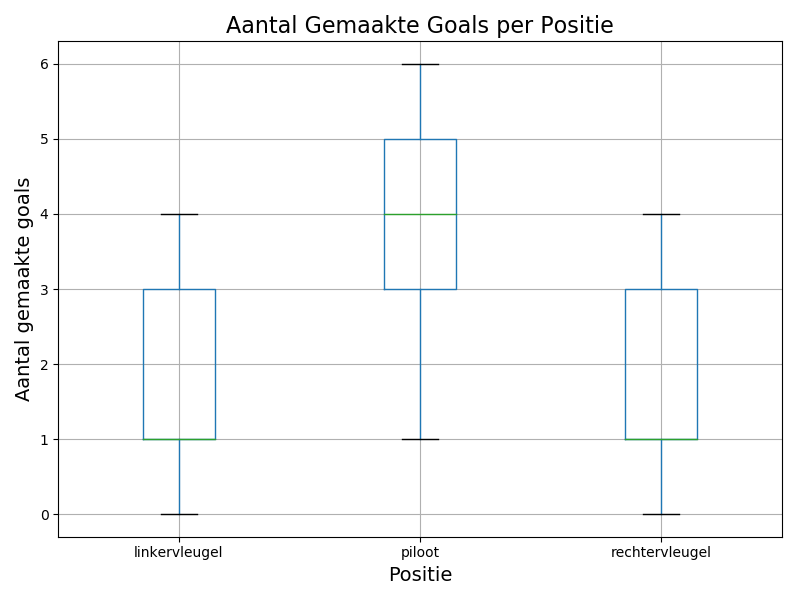




We zien in de bovenstaande diagram dat er bijna evenveel goede spelers zijn als matige en zeer goede bijeengeteld. Dit toont dus ook wel aan dat we het diagram van 1.5 niet kunnen gebruiker om het aantal gemaakte goals per inzet te bespreken, omdat deze geen rekening houd met het feit dat er meer goede spelers zijn dan matige en zeer goede.

## Boxplot: aantal gemaakte goalen van posities linkervleugel, rechtervleugel en piloot





Aan de hand van het bloxplot kunnen we zien dat voor de posities linker- en rechtervleugel vrij gelijk zijn in hoeveel goals elke persoon maakt 50% van deze speler maken 1-4 goals.

Wanneer we kijken naar de piloot positie, zien we dat zij meestal meer goals maken. 50% van de spelers op deze positie maakt 3-5 goals terwijl de onderste 25% maar 1-3 goals maakt. We zien dat ze in het algemeen meer goals maken dan de spelers op andere posities.

## Welke soort gegeven is ‘aantal gemaakte goalen’, ‘inzet’ en ‘gewicht’

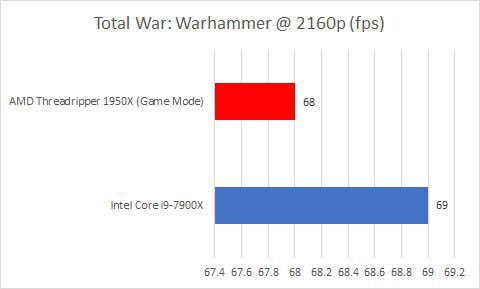
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Kwalitatief | | Kwantitatief | |
| **Nominaal** | **Ordinaal** | **Discreet** | **Continue** |
| Aantal gemaakte goalen |  |  |  | X |
| Inzet |  | X |  |  |
| Gewicht |  |  |  | X |

# Opgave 1b

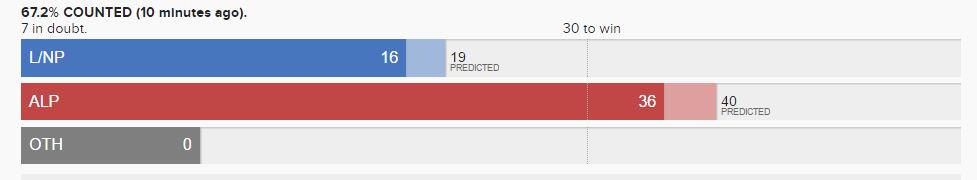
Wij hebben beide de cursus ‘Python: Getting Started’ gevolgd. In deze cursus werd de basis van python behandeld. Vooral hoofdstuk 3 en 4 waren behulpzaam om te leren hoe er in python gewerkt word.

Om te leren hoe we de library pandas moeten gebruiken, hebben we de pandas documentatie en stackoverflow gebruikt.

# Opgave 2



Dit is gevonden in een TechRadar artikel waarin de nieuwe processors van AMD en Intel werden vergeleken. Op het eerste zicht is het meest in orde met dit. De labels een beetje inkorten of anders plaatsen maar het grootste probleem is dat het start op 67 en niet op 0. Waardoor het lijkt alsof het product van Intel vele malen beter is dan het product van AMD.

 Deze grafiek mist een titel en het veld OTH(other) is 0 maar lijkt 8 met hoe de grafiek is voorgesteld. Deze grafiek gaat over de Australische verkiezings resultaten, deze informatie hadden we graag bij de grafiek zelf gehad. Een legende voor de grafiek was ook zeer goed geweest.