

# Statistiek 5 – Practicumopdracht 1

Lees eerst het volgende:

- Dit is de eerste opdracht voor Statistiek 5. Deze telt voor 20 punten mee.
- Je mag deze opdracht maken en inleveren in groepen van maximaal 2 personen.
- De opdracht is op de volgende bladzijde te vinden.
- Voor deze opdracht is de dataset `college_math.csv` nodig, die je op Brightspace kan vinden. Een beschrijving van deze dataset vind je verderop.
- Je werkt de opdracht uit in een `.ipynb`-bestand; hierin is de code én de rapportage verwerkt.
- Je code is overzichtelijk, dat wil zeggen, goed leesbaar en voorzien van commentaar zodat een ander begrijpt wat je aan het doen bent. Onderbouw gemaakte stappen. Waarom kies je voor een bepaalde werkwijze? Waarom maak je een bepaalde aanname? Jouw onderbouwing laat jouw denkproces zien!
- De output van de code chunks moet in de ingeleverde notebook zichtbaar zijn; in principe gaat de beoordelaar dus **niet** de code runnen!
- De code moet in principe op elke computer kunnen runnen (alleen het pad naar het databestand mag veranderd worden), zonder foutmeldingen. Test dit van tevoren op een andere computer!
- Het `.ipynb`-bestand lever je in via het inleverpunt op Brightspace. Vermeld duidelijk de namen van jou en je practicumpartner, zowel in de bestandsnaam als in de notebook zelf.
- Gebruik je externe bronnen, dan verwijst je naar deze bronnen volgens APA.
- Deadline: **zondag 5 oktober om 23:59 uur**

## Opgave: De wiskundeklas

Op een wat afgelegen locatie in Italië is een school speciaal voor zij die de middelbare school niet hebben kunnen afmaken (om welke reden dan ook). Deze school noemen we vanaf nu de VAVO (ook al is dat niet de term die ze in Italië gebruiken).

De voortgang van studenten op de VAVO is wisselvallig. Hoewel iedereen gemotiveerd is begonnen, haakt een deel snel af of lopen ze toch weer tegen allerlei problemen aan waardoor het ze niet lukt om bepaalde vakken te halen.

Van de regering heeft VAVO middelen gekregen om te investeren. Doel: gerichte hulp kunnen aanbieden aan zij die dat nodig hebben, zodat ook zij de VAVO kunnen afmaken.

Voordat de VAVO een heel studentbegeleidingstraject inricht, wil zij iets meer inzicht in de studenten die zij onderwijzen. Welke soorten studenten komen er eigenlijk allemaal binnen? En welke hulpbehoeftes zijn er zoal binnen die groep?

De VAVO heeft, voor één specifiek vak (wiskunde), de grote groep studenten voor één semester gevolgd en daarvan allerlei gegevens bijgehouden (met medewerking van de studenten). De gegevens daarvan zijn te vinden in de dataset `college_math.csv`.

Gebruik je arsenaal aan data analyse-technieken om VAVO van een passend advies te voorzien.

## De dataset

Hieronder vind je een overzicht van de variabelen in de dataset college\_math.csv.

Variabele	Betekenis
age	De leeftijd van de student
alcohol_weekday	Hoeveel alcohol drinkt de student gemiddeld doordeweeks, per dag? (1 = geen; 2 = max 1 glas; 3 = max 2 glazen; 4 = 3-4 glazen; 5 = meer dan 4 glazen)
alcohol_weekend	Hoeveel alcohol drinkt de student gemiddeld in het weekend, per dag? (1 = geen; 2 = max 1 glas; 3 = max 2 glazen; 4 = 3-4 glazen; 5 = meer dan 4 glazen)
commute_minutes	Aantal minuten reistijd huis-school voor de student
extra_paid_classes	Of de student bijlessen voor het vak tegen betaling heeft (0 = nee; 1 = ja)
free_time	Hoeveel vrije tijd heeft de student buiten VAVO? (1 = nauwelijks, student helpt de hele tijd mee met klusjes en werk; 2 = beperkt, slechts 1-2 avonden per week; 3 = redelijk, op de meeste dagen wel een paar uurtjes vrije tijd; 4 = genoeg, elke dag is er wel een dagdeel vrije tijd; 5 = veel, naast school zijn er nauwelijks verplichtingen)
gender	Het geslacht van de student (M; F)
grade_final	Het eindcijfer voor wiskunde (hier zitten periodes 1 en 2 in verwerkt)
grade_period1	Het cijfer voor wiskunde na periode 1
grade_period2	Het cijfer voor wiskunde na periode 2 (hier zit periode 1 in verwerkt)
health_status	Wat is de gezondheid van de student? (1 = ernstig ziek; 2 = langdurig ziek en mogelijk ernstig; 3 = langdurig ziek maar niet ernstig; 4 = lichtjes ziek; 5 = gezond)
internet_access	Of de student thuis internet heeft (0 = nee; 1 = ja)
past_failures	Hoeveel keer eerder de student het vak bij de VAVO niet heeft gehaald
school_absences	Het aantal dagen dat de student niet naar colleges is gegaan
social_outings	Hoe vaak de student uitgaat met vrienden (1 = nooit; 2 = enkele weken 1 keer per week; 3 = 1 keer per week; 4 = 2-3 keer per week; 5 = vrijwel elke dag)
support_family	Heeft de student familie die kan helpen met studeren? (0 = nee; 1 = ja)
support_school	Gebruikt de student extra faciliteiten op de VAVO? (0 = nee; 1 = ja)
weekly_study_hours	Het gemiddeld aantal uren per week dat de student studeert