
PDF naar CSV

Het omzetten van PDF-bestanden naar CSV kan handig zijn als je gegevens uit een PDF wilt bewerken of analyseren. PDF-bestanden zijn goed om informatie op een mooie en gestructureerde manier te presenteren, maar ze zijn minder geschikt voor het bewerken en verwerken van data. CSV-bestanden (Comma-Separated Values) zijn compacter en kunnen makkelijk worden geopend in programma's zoals Excel of Google Sheets voor verdere verwerking.

In deze handleiding leggen we stap voor stap uit hoe je PDF's kunt omzetten naar CSV met verschillende methodes. We bespreken zowel eenvoudige online tools als meer geavanceerde opties, zoals AI. Of je nu snel een PDF wilt omzetten of een geautomatiseerde oplossing zoekt, deze gids helpt je de juiste methode te kiezen.

Voor deze tutorial is het vereist om de volgende software te bemachtigen:

- Adobe Acrobat
- Microsoft Excel
- Microsoft OneNote
- ChatGPT Account
- Microsoft Account

01 Inhoudsopgave

Leerdoelen	02
1. Gebruik tijdens de minor	02
Introductie tool	03
1. Wat is PDF	03
2. Wat is CSV	03
Uitleg tool	04
1. Afbeelding PDF omzetten naar tekst PDF	04
a. Adobe Acrobat	04
b. OneNote	07
2. Opschonen van PDF met Acrobat	09
3. PDF omzetten naar CSV	11
a. Excel	11
b. Adobe Acrobat	13
c. AI	15
i. ChatGPT	15
ii. Copilot	17
d. Online Converters	18
4. Voor- en nadelen van de verschillende tools	19
a. Microsoft Excel	19
b. Adobe Acrobat	19
c. AI-tools	19
d. Online tools	19
Flowchart	20
Slotwoord	22
Bronnen	23

02 Leerdoelen

Tijdens deze tutorial ga je leren om PDF bestanden om te zetten naar CSV bestanden. Dit proces heeft diverse toepassingen, elk met een unieke aanpak. We zullen verschillende technieken en tussenstappen behandelen om je een grondige basis te geven in deze vaardigheid. De leerdoelen omvatten:

1. Begrijpen van de structuur van PDF en CSV bestanden
2. Leren werken met verschillende PDF-extractietools
3. Datamanipulatie en -opschoning na extractie
4. Optimaliseren van het conversieproces voor verschillende soorten PDF's
5. Begrijpen van de uitdagingen en beperkingen bij PDF naar CSV conversie

Gebruik tijdens de minor

Gegevens worden vaak in een PDF geplaatst, omdat deze onderdeel zijn van een rapport. De tabel is dan maar een klein onderdeel van het gehele verhaal. Daarnaast is het voor consumenten makkelijker om data uit een tabel af te lezen in een PDF dan vanuit een CSV. Tijdens onze minor, Datavisualisatie en Infographics, is het echter handig om de data in een CSV te gebruiken, omdat we dan met de gegevens kunnen rekenen en het kunnen bewerken/verwerken in verschillende software, wat tijdens deze minor vaak aan bod komt. Daarom leggen we in deze handleiding uit hoe je dit kan omzetten, en welke tools daarvoor gebruikt kunnen worden. Deze vaardigheid kan in verschillende fasen van je studie van pas komen:

1. Dataverwerking projecten: Bij het werken met grote datasets die in PDF-formaat zijn opgeslagen.
2. Onderzoeksopdrachten: Voor het extraheren van gegevens uit wetenschappelijke papers of rapporten in PDF-formaat.
3. Data-analyse: Voor het voorbereiden van data uit verschillende bronnen voor analyse.

03 Introductie tool

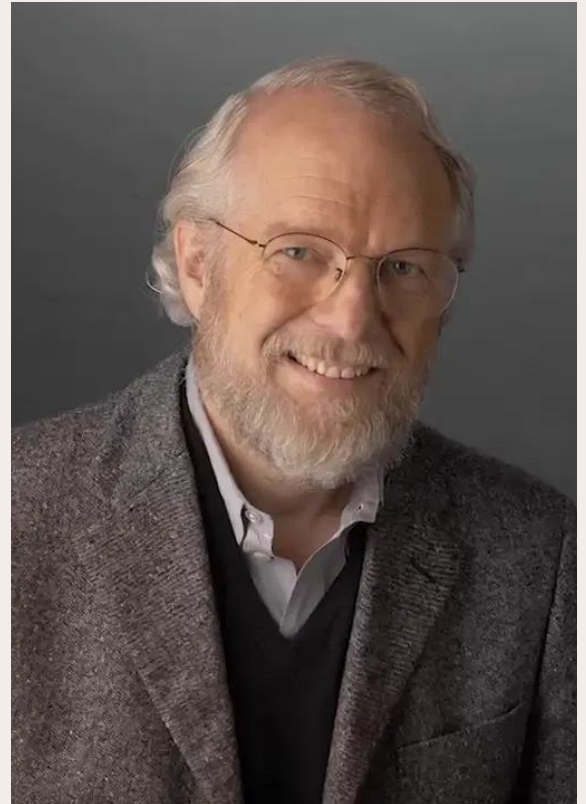
Wat is een PDF?

Het PDF (Portable Document Format) is ontwikkeld door Adobe Systems. Een van de oprichters van Adobe, Dr. John Warnock, startte in 1991 de ontwikkeling van dit formaat als onderdeel van zijn visie om papierloze communicatie mogelijk te maken (Adobe, z.d.).

Enkele belangrijke kenmerken van PDF:

- Platformonafhankelijk: PDF-bestanden kunnen op elk apparaat en besturingssysteem worden geopend zonder verlies van opmaak.
- Compact: PDF's kunnen tekst, afbeeldingen en vectorafbeeldingen efficiënt comprimeren.
- Beveiligingsopties: PDF's kunnen worden beveiligd met wachtwoorden en andere beveiligingsmaatregelen.
- Interactieve elementen: Moderne PDF's kunnen formulieren, knoppen en multimedia bevatten.

Het grote gebruik van de PDF heeft het tot een standaardformat gemaakt voor het delen van documenten, vooral in professionele en academische omgevingen.



Bron: Adobe

Wat is een CSV?

Het CSV-bestandsformaat (Comma-Separated Values) werd voor het eerst ondersteund door de IBM Fortran-compiler in 1972 ((ByteScout, 2021). Het is ontstaan om het gemakkelijker te maken om gegevens in te typen en op te slaan in een formaat dat gemakkelijk door computers kan worden gelezen en verwerkt (Microsoft, z.d.). Sindsdien is het formaat verder gedefinieerd en ontwikkeld.

Belangrijke eigenschappen van CSV:

- Eenvoud: CSV is een simpel tekstformaat dat gemakkelijk te begrijpen en te maken is.
- Universele compatibiliteit: Vrijwel alle spreadsheet- en databaseprogramma's kunnen CSV-bestanden lezen en schrijven.
- Compactheid: CSV-bestanden zijn doorgaans kleiner dan equivalente spreadsheetbestanden.
- Snelheid: Het lezen en schrijven van CSV-bestanden is vaak sneller dan complexere formaten.

Ondanks de opkomst van meer geavanceerde formaten zoals JSON en XML, blijft CSV populair vanwege zijn eenvoud en brede ondersteuning, vooral in data-analyse en -verwerking (ByteScout, 2021).

04 Uitleg tools en werkwijzen

Hieronder staan meerdere manieren beschreven om een PDF om te zetten naar een CSV-bestand. In sommige gevallen is het van belang om de PDF voor te bereiden op deze tools door het bestand op te schonen of om te zetten van een afbeelding naar een tekstbestand. Ook is het van belang om na te denken over het type gegevens waar je mee werkt. Zijn het gevoelige gegevens of niet? Bekijk welke situatie voor jouw PDF van toepassing is en volg dan de stappen van de uitleg hieronder.

Afbeelding PDF omzetten naar tekst PDF

Over het algemeen zijn de meeste PDF's een bestand waarin de tekst ook te herkennen is als tekst. Je kan dan ook losse woorden of gegevens selecteren in het bestand door eroverheen te slepen. Soms werkt dat niet, omdat de PDF dan bestaat uit een afbeelding van de gegevens. In dat geval moet het bestand eerst omgezet worden van een afbeelding naar een tekst PDF. Hier worden 2 manieren beschreven waarop dit kan.

Adobe Acrobat

1. Sla eerst een back-up van dit bestand op als je de PDF met de afbeelding wil bewaren. Het bestand wat je omzet vervangt namelijk je originele bestand wanneer je deze opslaat.
2. Open de PDF met Adobe Acrobat. (In de verkenner rechtermuisklik op het bestand en kies voor 'openen met' en klik vervolgens op Adobe Acrobat).
3. Als het goed is staat het tabblad 'Alle tools' automatisch open. Anders moet je deze eerst openen. Klik vervolgens op de knop 'Scannen en OCR'. (OCR staat voor Optical Character Recognition en dat kan door patroonherkenning tekst uit een afbeelding halen (Van Muylder, 2019))

The screenshot shows the Adobe Acrobat interface. The 'Alle tools' menu is open, and 'Scannen & OCR' is highlighted with a red box. The main content area displays a table titled 'Hoger onderwijs; eerstejaarsstudenten, vooropleiding en opleidingsfase' with filters for 'Onderwijsoort: Hoger onderwijs', 'Periodes: 2023/24*', and 'Onderwerp: Eerstejaarsstudenten'. The table shows the number of students by education phase and gender.

	Opleidingsfase		Geslacht	
	Bachelo- en doctoraalopleidingen		Master- en vervolopleidingen	
	Mannen	Vrouwen	Mannen	Vrouwen
Vooropleiding	aantal			
Havo-diploma cultuur en maatschappij	1 110	4 220	0	0
Havo-diploma economie en maatschappij	10 790	7 860	0	0
Havo-diploma natuur en gezondheid	3 420	5 260	0	10
Havo-diploma natuur en techniek	1 980	480	0	0
Vwo-diploma cultuur en maatschappij	520	2 090	0	10
Vwo-diploma economie en maatschappij	4 080	2 780	10	10
Vwo-diploma natuur en gezondheid	2 250	4 000	0	10
Vwo-diploma natuur en techniek	2 450	880	0	0
Mbo/col-diploma	12 830	17 470	0	20

Bron: CBS

Tabeltoelichting
INHOUDOPGAVE
1. Toelichting
2. Drukken en versieren van symbolen
3. Kopieëren naar externe tabellen en artikelen
4. Bronnen en methoden
5. Meer informatie
1. TOELICHTING
Deze tabel bevat het aantal eerstejaarsstudenten aan de door de overheid bekostigde instellingen voor het hoger onderwijs naar onderwijssoort, opleidingsfase en studiejahr. Deze tabel bevat vanaf studiejahr 2 gegevens beschikbaar vanaf studiejahr 2011/12.
Status van de cijfers:
De cijfers over de studiejahren tot en met 2022/23 zijn definitief en de cijfers over het studiejahr 2023/24 zijn voorlopig.
Wijzigingen per 29 maart 2024:
De definitieve cijfers over het studiejahr 2022/23 zijn toegevoegd.
Waarom komen de nieuwe cijfers?
In het eerste kwartaal van 2025 worden de voorlopige cijfers over het studiejahr 2024/25 beschikbaar.

4. Aan de linkerkant wordt nu het menu Scannen en OCR geopend. Klik op 'In dit bestand.'

The screenshot shows the 'Scannen & OCR' menu on the left side of the PDF viewer. The 'In dit bestand' option is highlighted with a red box. The main content area displays a table of higher education statistics for first-year students.

Hoger onderwijs; eerstejaarsstudenten, vooropleiding en opleidingsfase

Gewijzigd op: 29 maart 2024

Onderwijssoort: Hoger onderwijs
Perioden: 2023/24*
Onderwerp: Eerstejaarsstudenten

Vooropleiding	Opleidingsfase		Geslacht	
	Bachelor- en doctoraalopleidingen		Master- en vervolopleidingen	
	Mannen	Vrouwen	Mannen	Vrouwen
	aantal			
Havo-diploma cultuur en maatschappij	1 110	4 220	0	0
Havo-diploma economie en maatschappij	10 790	7 880	0	0
Havo-diploma natuur en gezondheid	3 420	5 260	0	10
Havo-diploma natuur en techniek	1 980	480	0	0
Vwo-diploma cultuur en maatschappij	520	2 090	0	10
Vwo-diploma economie en maatschappij	4 080	2 780	10	10
Vwo-diploma natuur en gezondheid	2 250	4 000	0	10
Vwo-diploma natuur en techniek	2 450	860	0	0
Mbo/Isol-diploma	12 830	17 470	0	20

Bron: CBS

Tabeltoelichting

INHOUDSGEGAVE

- Toelichting
- Definitie en verklaring van symbolen
- Koppelingen naar relevante tabellen en artikelen
- Bronnen en methoden
- Meer informatie

1. TOELICHTING

Deze tabel bevat het aantal eerstejaarsstudenten aan de door de overheid bekostigde instellingen voor het hoger onderwijs naar onderwijssoort, vooropleiding, opleidingsfase en studiejaar. Deze tabel bevat vanaf studiejaar 2 gegevens beschikbaar vanaf studiejaar 2011/12.

Status van de cijfers:
De cijfers over de studiejaren tot en met 2022/23 zijn definitief en de cijfers over het studiejaar 2023/24 zijn voorlopig.

Wijzigingen per 29 maart 2024:
De definitieve cijfers over het studiejaar 2022/23 zijn toegevoegd.

Wanneer komen de nieuwe cijfers?
In het eerste kwartaal van 2025 komen de voorlopige cijfers over het studiejaar 2024/25 beschikbaar.

5. Kies voor 'Tekstherkenning'.

The screenshot shows the 'Scannen & OCR' menu on the left side of the PDF viewer. The 'In dit bestand' option is highlighted with a red box. A dropdown menu is open, showing the 'Tekstherkenning' option highlighted with a red box.

The dropdown menu shows the following options:

- Pagina's
- Huidige pagina
- Taal: Nederlands
- Annuleren
- Tekstherkenning**

6. Vanaf nu is de tekst ook echt als tekst te gebruiken. Sla het bestand op en je kan hiermee verder werken.

The screenshot shows a PDF viewer interface with a document titled "Hoger onderwijs; eerstejaarsstudenten, vooropleiding en opleidingsfase". The document content includes a table with student counts by education level and gender, and a table of contents.

Hoger onderwijs; eerstejaarsstudenten, vooropleiding en opleidingsfase
Gewijzigd op: 29 maart 2024

Onderwijssoort: Hoger onderwijs
Periode: 2023/24*
Onderwerp: Eerstejaarsstudenten

Vooropleiding	Opleidingsfase		Geslacht	
	Bachelor- en doctoraalopleidingen		Master- en vervolopleidingen	
	Mannen	Vrouwen	Mannen	Vrouwen
aantal				
Havo-diploma cultuur en maatschappij	1 110	4 220	0	0
Havo-diploma economie en maatschappij	10 790	7 880	0	0
Havo-diploma natuur en gezondheid	3 420	5 200	0	10
Havo-diploma natuur en techniek	1 980	490	0	0
Vwo-diploma cultuur en maatschappij	500	2 080	0	10
Vwo-diploma economie en maatschappij	4 080	2 760	10	10
Vwo-diploma natuur en gezondheid	2 290	4 000	0	10
Vwo-diploma natuur en techniek	2 490	860	0	0
Mbo/Bol-diploma	12 830	17 470	0	20

bron: CBS

Tabeloelichting
INHOUDSOVERZICHT

- Tsclitng
- Dtelsten en vektng van sytemen
- Koppngng naar reserare tabelen en artkelen
- Bronnen en methoden
- Meer informatie

1. TOELICHTING

Deze tabel bevat het aantal eerstejaarsstudenten aan de door de overheid bekostigde instellingen voor het hoger onderwijs naar onderwijssoort, vooropleiding, opleidingsfase en studiejaar. Deze tabel bevat vanaf studiejaar 2 gegevens beschikbaar vanaf studiejaar 2011/12.

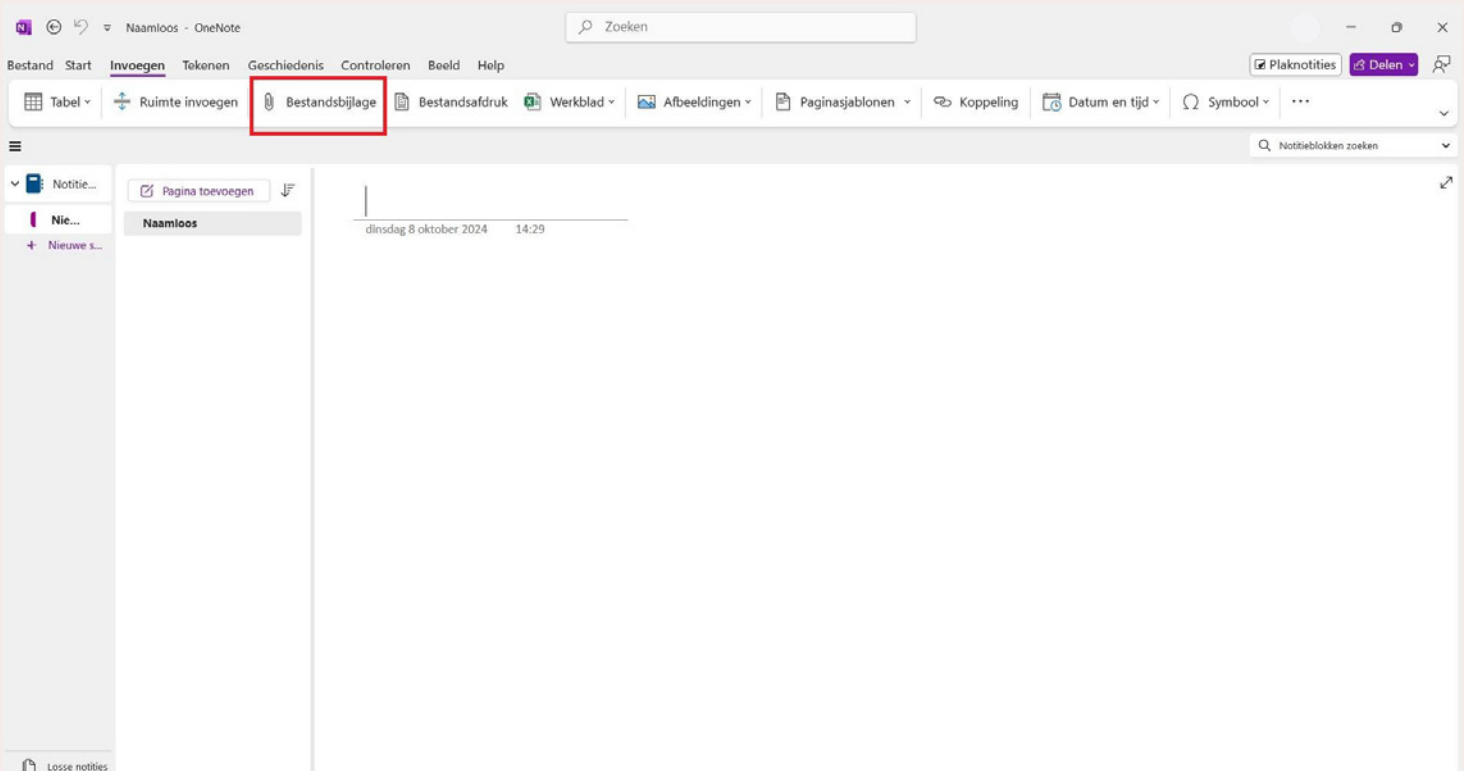
Status van de cijfers:
De cijfers over de studiejaar tot en met 2022/23 zijn definitief en de cijfers over het studiejaar 2023/24 zijn voorlopig.

Wijzigingen per 29 maart 2024:
De definitieve cijfers over het studiejaar 2022/23 zijn toegevoegd.

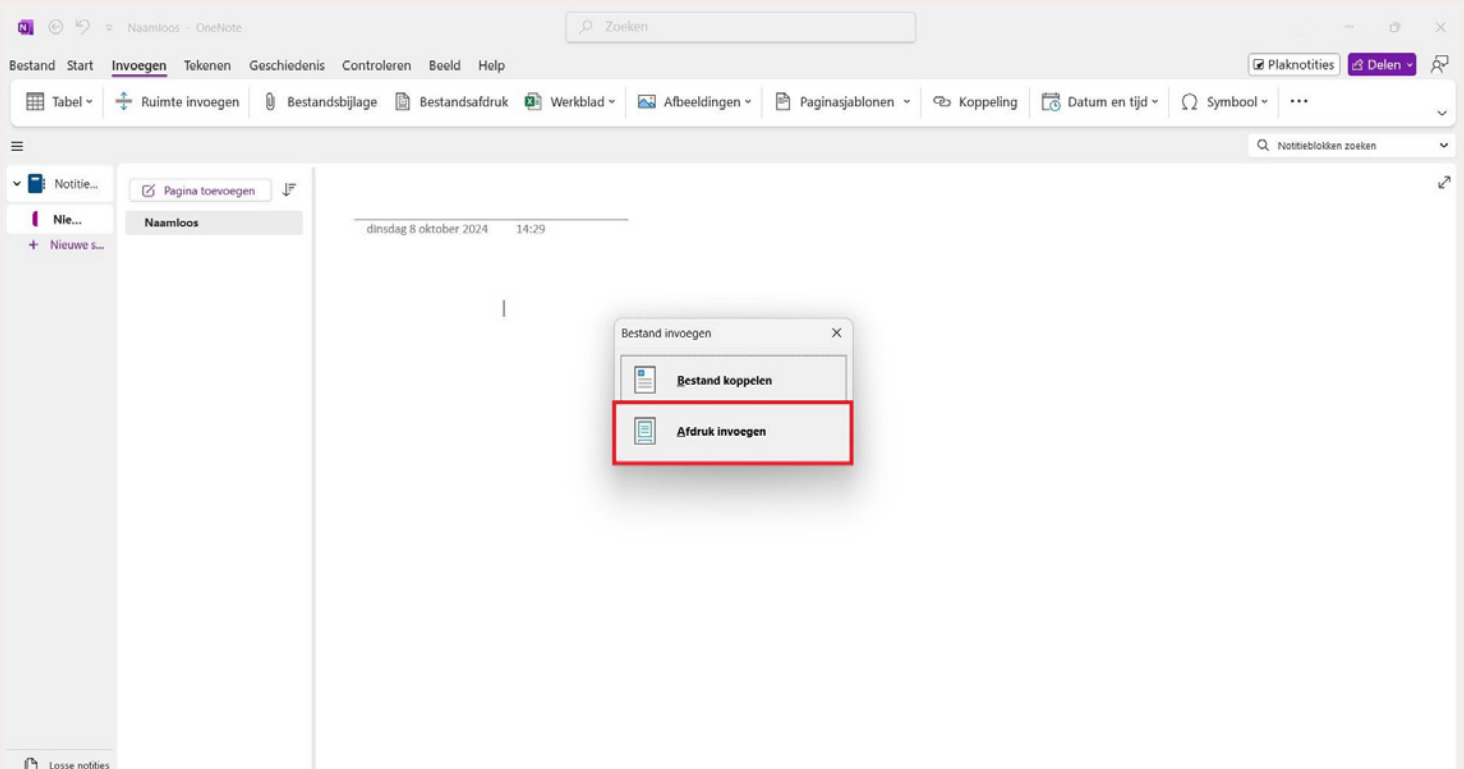
Wanneer komen de nieuwe cijfers?
In het eerste kwartaal van 2025 komen de voorlopige cijfers over het studiejaar 2024/25 beschikbaar.

One Note

1. Open OneNote op je laptop. Dit is een programma van Microsoft wat je met je HU-account kan downloaden.
2. Open de PDF via de knop 'bestandsbijlage'.



3. Kies voor de optie 'afdruk invoegen'



4. Klik met je rechtermuis op de afbeelding, kies voor 'Tekst kopiëren van alle pagina's van de afdruk'.

The screenshot shows the OneNote interface with a PDF image of a table. The table is titled 'Hoger onderwijs; eerstejaarsstudenten, vooropleiding en opleidingsfase' and is dated 'Gewijzigd op: 29 maart 2024'. The table has columns for 'Opleidingsfase' and 'Geslacht', and sub-columns for 'Bachelor- en doctoraalopleidingen' and 'Master- en vervolopleidingen'. The data is as follows:

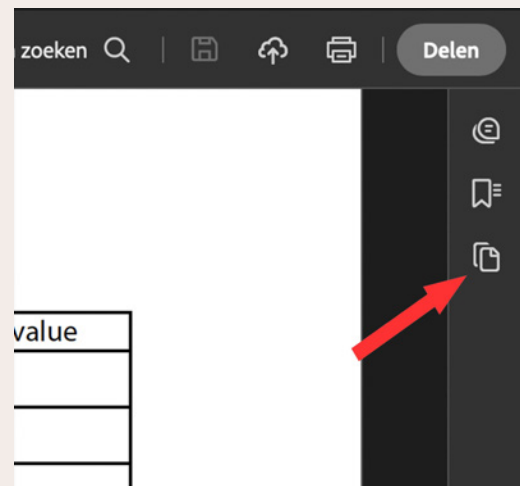
Vooropleiding	Bachelor- en doctoraalopleidingen		Master- en vervolopleidingen	
	Mannen	Vrouwen	Mannen	Vrouwen
Havo-diploma cultuur en maatschappij	1 110	4 220	0	0
Havo-diploma economie en maatschappij	10 790	7 880	0	0
Havo-diploma natuur en gezondheid	3 420	5 260	0	10
Havo-diploma natuur en techniek	1 880	490	0	0
Vvo-diploma cultuur en maatschappij	520	2 060	0	10
Vvo-diploma economie en maatschappij	4 080	2 780	10	10
Vvo-diploma natuur en gezondheid	2 220	4 000	0	10
Vvo-diploma natuur en techniek	2 450	860	0	0
Mbo/til-diploma	12 830	17 470	0	20

The context menu is open, and the option 'Tekst kopiëren van alle pagina's van de afdruk' is highlighted with a red box.

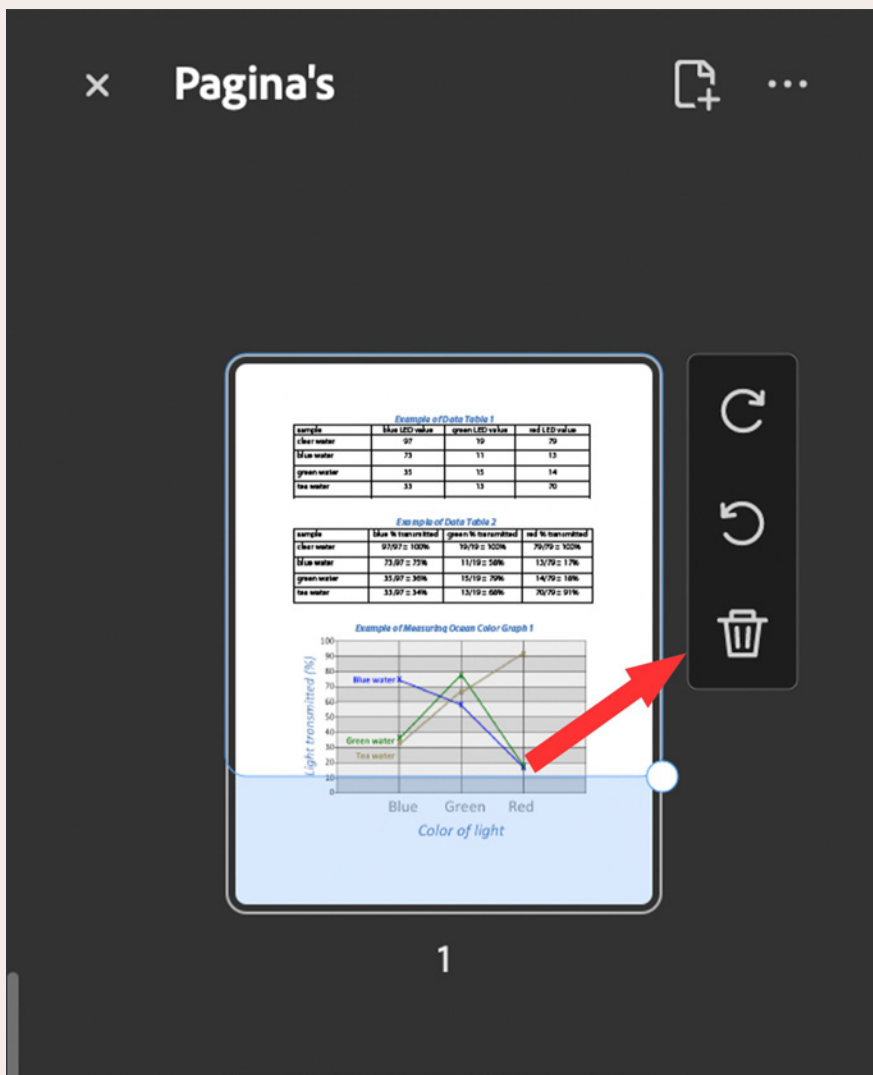
5. De tekst in de afbeelding is nu gekopieerd en kan je opslaan in een bestand naar keuze.

Opschonen van PDF met Acrobat

1. (Dataset van https://tos.org/oceanography/assets/images/content/29-1_schollaert-uz_examples.pdf)
 - a. Open je PDF in Acrobat
 - b. Als je PDF meerdere pagina's bevat, klik je eerst op de paginaknop aan de rechterkant van het scherm



3. Haal je muis over de pagina('s) die je wilt verwijderen en klik op het prullenbakje.



4. Selecteer alles behalve de tabel die je wilt converteren, en klik op backspace/delete.

Example of Data Table 3

sample	blue LED value	green LED value	red LED value
+ 1 tsp milk	13	14	12
+1tsp milk + blue	10	9	6
+1tsp milk +green	9	13	7
+1tsp milk + tea	7	9	9

Example of Data Table 4

Water sample	blue % transmitted	green % transmitted	red % transmitted
+ 1 tsp milk	13/13 = 100%	14/14 = 100%	12/12 = 100%
+1tsp milk + blue	10/13 = 77%	9/14 = 64%	6/12 = 50%
+1tsp milk +green	9/13 = 69%	13/14 = 93%	7/12 = 58%
+1tsp milk + tea	7/13 = 54%	9/14 = 69%	9/12 = 75%

Example of Measuring Ocean Color Graph 2

Water Sample	Blue (%)	Green (%)	Red (%)
Blue water	77%	64%	50%
Green water	69%	93%	58%
Tea water	54%	69%	75%

5. Sla de bewerkte PDF op. (tip: zorg dat je hem een andere naam geeft dan de originele pdf!)

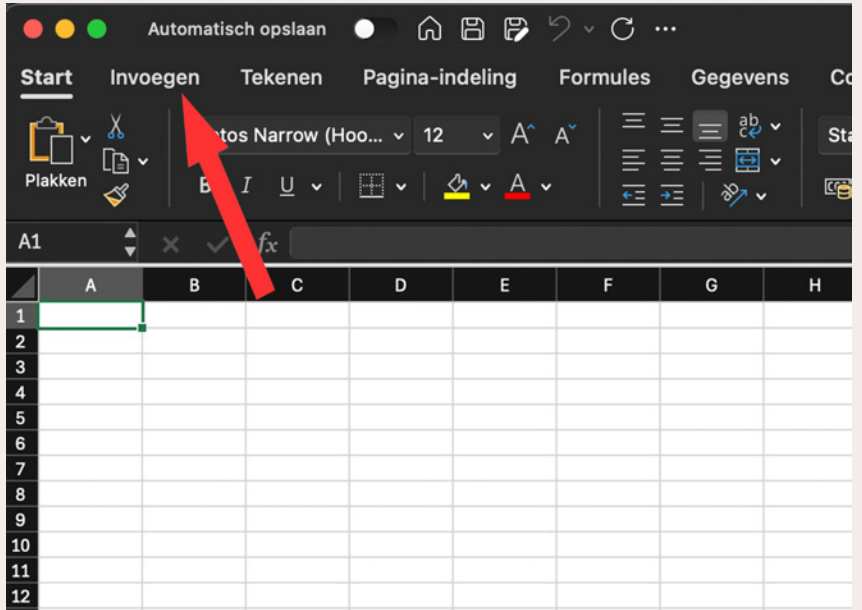
Opslaan als PDF

Bewaar als: -1_schollaert-uz_examples_EDITED

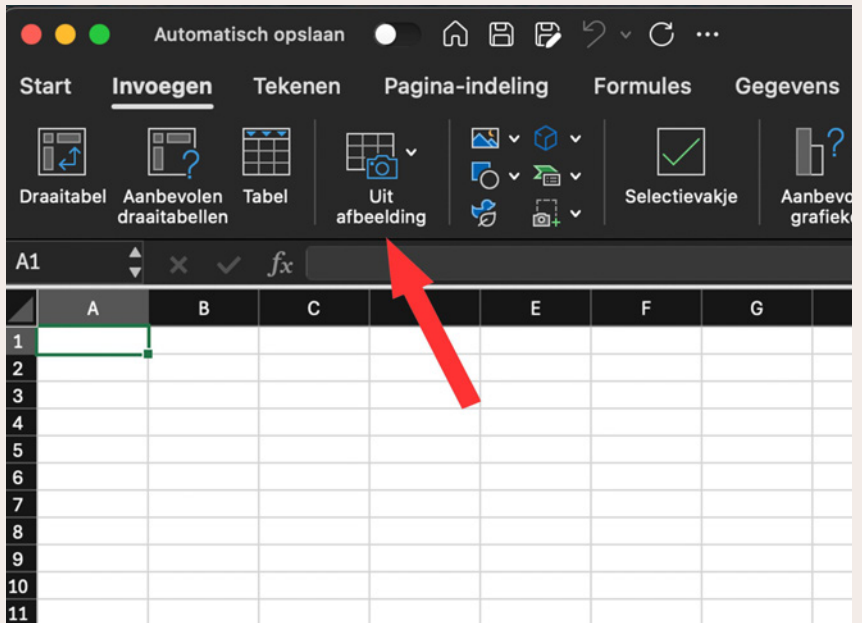
PDF omzetten naar CSV

Excel

1. Open een nieuw Excel-bestand.
2. Navigeer naar invoegen/insert.



3. Klik op van foto/from picture → foto uit bestand/picture from file.



4. Open je gekozen PDF.
5. Excel gaat het bestand nu uitlezen.
6. Klik op Insert data.

Als er data is waarvan Excel niet zeker weet wat er staat, zal je handmatig langs die cellen moeten gaan om ze vanuit de PDF over te schrijven. Gelukkig heeft Excel daar een semi-gebruikersvriendelijke methode voor. Als de data in Excel te zien is, kan je het bestand opslaan als .csv!

4. Open je gekozen PDF.
5. Excel gaat het bestand nu uitlezen.
6. Klik op Insert data.

Gegevens uit afbeelding

Example of Data Table 3

sample	blue LED value	green LED value	red LED value
+ 1 tsp milk	13	14	12
+1tsp milk + blue	10	9	6
+1tsp milk +green	9	13	7
+1tsp milk + tea	7	9	9

Gegevens invoegen Controleren

Example of Data Table 3

sample	blue LED value	green LED value	red LED value
+ 1 tsp milk	13	14	12
+1 tsp milk + blue	10	9	6
+1 tsp milk +green	9	13	7

Als er data is waarvan Excel niet zeker weet wat er staat, zal je handmatig langs die cellen moeten gaan om ze vanuit de PDF over te schrijven. Gelukkig heeft Excel daar een semi-gebruikersvriendelijke methode voor. Als de data in Excel te zien is, kan je het bestand opslaan als .csv!

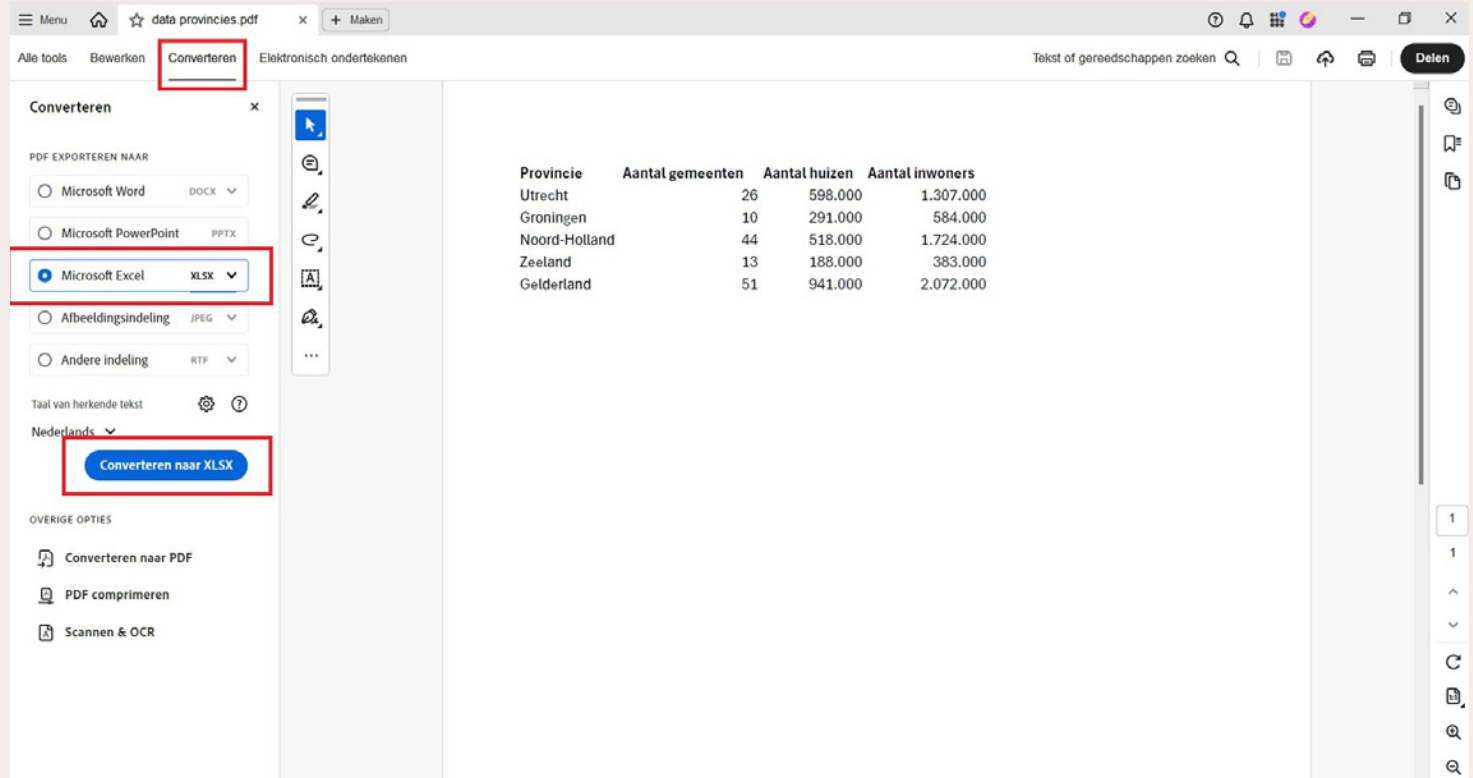
Adobe Acrobat

Open de PDF met Adobe Acrobat. (In de verkenner rechtermuisklik op het bestand en kies voor 'openen met' en klik vervolgens op Adobe Acrobat).

Vervolgens zijn er 3 opties om de PDF om te zetten:

Optie 1

1. Klik op het tabblad "converteren" en kies voor Microsoft Excel (xlsx). Klik daarna op 'Converteren naar XLSX'.



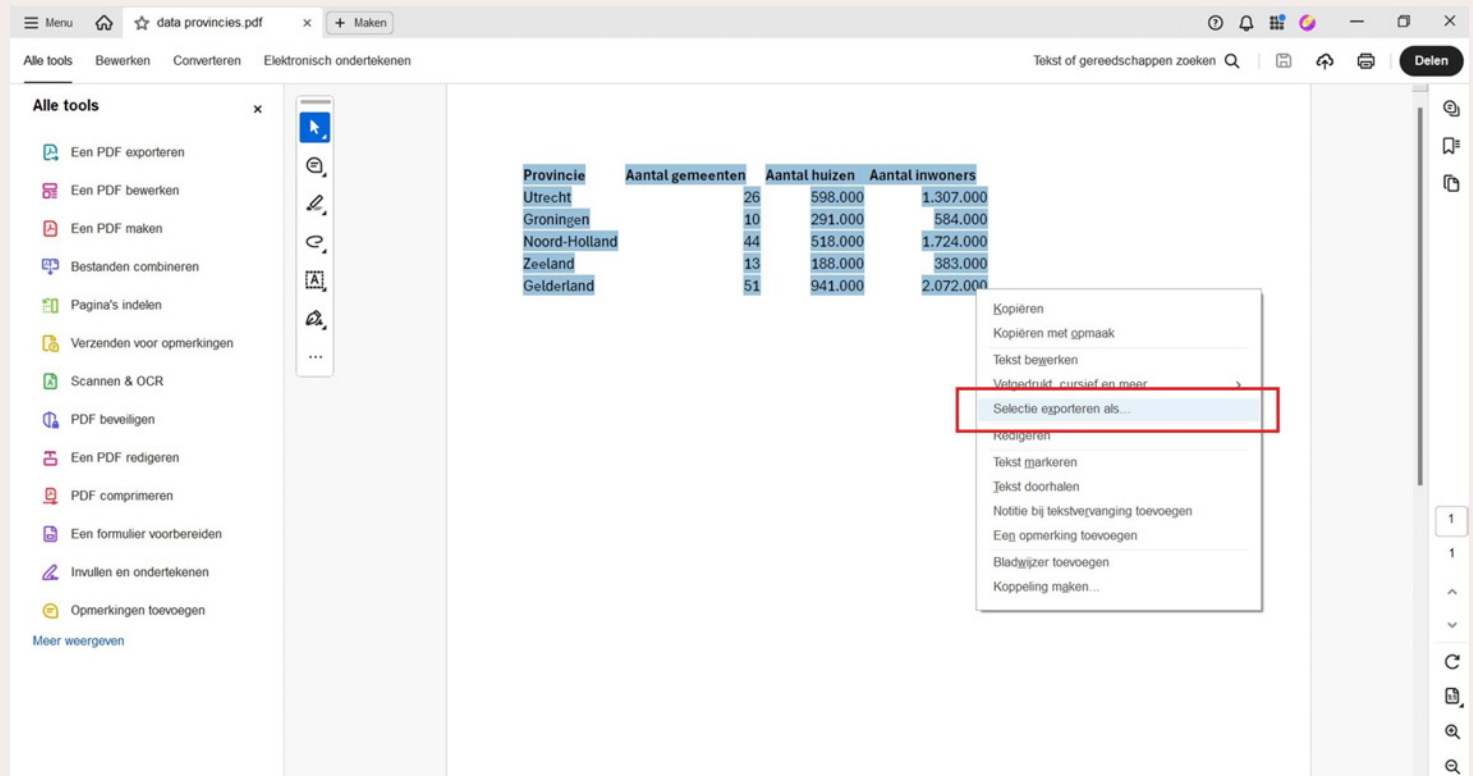
The screenshot shows the Adobe Acrobat interface with the 'Converteren' panel open. The 'Microsoft Excel' option is selected under 'PDF EXPORTEREN NAAR'. The 'Converteren naar XLSX' button is highlighted. The main content area displays a table with the following data:

Provincie	Aantal gemeenten	Aantal huizen	Aantal inwoners
Utrecht	26	598.000	1.307.000
Groningen	10	291.000	584.000
Noord-Holland	44	518.000	1.724.000
Zeeland	13	188.000	383.000
Gelderland	51	941.000	2.072.000

2. Kies een plek om je bestand op te slaan. Dit is nu een Excel-bestand, dus als je een CSV wilde maken om met Excel te kunnen werken aan de gegevens dan kan je nu aan de slag. Anders sla je nu het Excel-bestand op als CSV.

Optie 2

1. Selecteer de data in de PDF die je wil gebruiken voor je CSV-bestand.
2. Rechtermuisklik op de geselecteerde data en kies voor "Selectie exporteren als".



The screenshot shows a PDF editor window with a table of data. The table is as follows:

Provincie	Aantal gemeenten	Aantal huizen	Aantal inwoners
Utrecht	26	598.000	1.307.000
Groningen	10	291.000	584.000
Noord-Holland	44	518.000	1.724.000
Zeeland	13	188.000	383.000
Gelderland	51	941.000	2.072.000

The context menu is open over the data, and the option "Selectie exporteren als..." is highlighted in a red box. Other options in the menu include "Kopiëren", "Kopiëren met opmaak", "Tekst bewerken", "Veldruikt cursief en meer...", "Resurgeren", "Tekst markeren", "Tekst doorhalen", "Notitie bij tekstvervangning toevoegen", "Een opmerking toevoegen", "Bladzijzer toevoegen", and "Koppeling maken...".

3. Kies voor een CSV-bestand en sla het op.

Optie 3

Je kan ook het stappenplan volgen van optie 2 door de gegevens te selecteren, maar dan op te slaan als Excel-bestand. Bij sommige datasets komt het resultaat er dan anders uit te zien dan bij optie 1 en 2. Je kan deze data dan bewerken zoals je wilt en vervolgens het Excel-bestand weer opslaan als CSV-bestand.

Artificial intelligence

Chat GPT

(Let op!!: controleer goed of alles nog klopt nadat je je data door AI heen gooit!)


1. Navigeer naar <https://openai.com/chatgpt/>, klik op het knop 'Start now' en log in met je account. Heb je nog geen ChatGPT-account, registreer je dan gratis.
2. Start een nieuwe chat en voeg het PDF-bestand met data toe. Dit kan door het bestand rechtstreeks naar de chat te slepen of door via de onderstaande knop ('Bestand bijvoegen') het bestand op te zoeken en te selecteren.



3. Nadat het bestand is toegevoegd in de chat, vraag ChatGPT om het bestand voor je te converteren naar een CSV-bestand.
4. Mogelijke antwoorden die ChatGPT je kan geven zijn:
 - A. De gegevens zijn geconverteerd naar een CSV-bestand. Je kunt het bestand hier downloaden.
 - B. Hier is de CSV-content, kopieer en plak deze in een teksteditor en sla deze op als een .csv-bestand.
5. Indien je antwoord A krijgt, download het bestand en open deze.

6. Indien je antwoord B krijgt, kopieer de CSV-content door op 'Code kopiëren' te klikken en plak deze handmatig in Excel. Excel zal de gegevens automatisch in een juiste tabel-formaat zetten.


ChatGPT Delen


 Ik kan de gegevens van het PDF-bestand omzetten naar een CSV-bestand. Uit het document heb ik relevante data gevonden over studentenaantallen en inschrijvingen, zoals in het volgende voorbeeld:

Jaar	Master onbekostigd	Master bekostigd	Bachelor voltijd	Bachelor deeltijd en dual	Associate degree	Totaal	% verschil t.o.v. 2020
2017	54	3.726	27.873	2.247	638	34.538	-
2018	167	4.005	27.315	2.323	617	34.427	-
2019	283	4.498	26.540	2.228	623	34.172	-
2020	565	5.086	27.675	2.158	642	36.126	-
2021	868	5.762	28.063	2.393	64	37.731	4.4%

Ik zal deze gegevens in een CSV-bestand plaatsen. Eén moment, alsjeblieft.

De gegevens zijn omgezet naar een CSV-bestand. Je kunt het bestand via de onderstaande link downloaden:

[Download het CSV-bestand \[-\]](#) 



7. Controleer of de gegevens in het bestand kloppen met de originele data. Indien je fouten ziet kun je ervoor kiezen om deze handmatig te verbeteren, ChatGPT nog eens te vragen of die er beter naar kan kijken of alsnog Adobe Acrobat gebruiken voor de conversie.

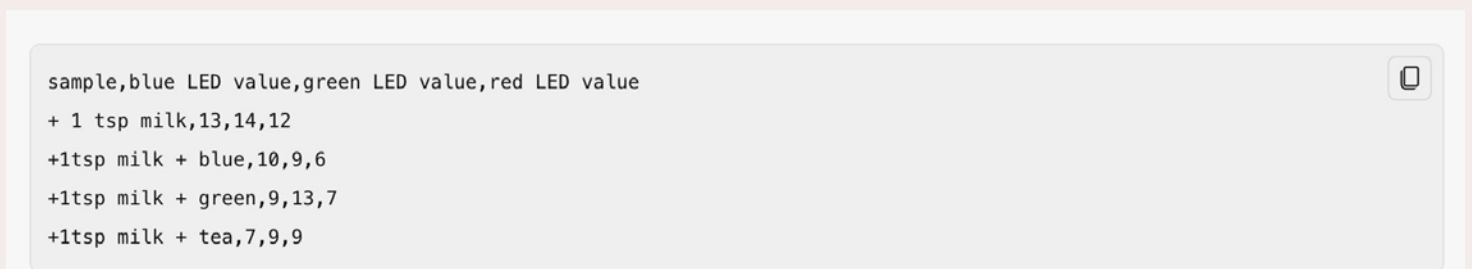
8. Als laatst, sla het bestand op als .csv.

Copilot

1. Navigeer naar <https://copilot.cloud.microsoft/> en log in met je HU-account.
2. Open een nieuwe chat, en klik op de bijlageknop.
3. Selecteer je PDF.
4. Je kunt proberen om te vragen of Copilot de tabel kan converteren naar een CSV-bestand.



5. Het antwoord wat Copilot geeft, moet handmatig in een tekstbestand worden gezet en moet opgeslagen worden als .csv-bestand (dus GEEN .txt) Het nadeel hiervan is dat je het niet direct kan downloaden van Copilot. Mocht je dat wel willen, lees dan verder.

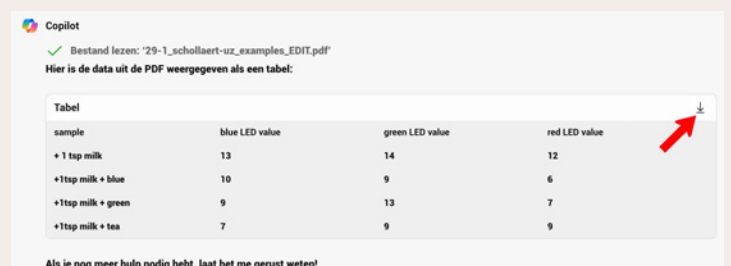


Als je nog meer hulp nodig hebt, laat het me gerust weten!

6. Als de prompt van hierboven niet werkt, of je wilt een downloadbaar bestand, kan je nog vragen of Copilot de data vanuit de PDF overschrijft.



7. Vanuit hier kun je de tabel als Excel-bestand downloaden. Klik op de downloadknop aan de rechterkant.



Als je nog meer hulp nodig hebt, laat het me gerust weten!

8. Open het bestand in Excel, en sla het opnieuw op als .csv-bestand.

Online converters

Er zijn online veel verschillende converters die het werk voor je kunnen doen. Vaak hoef je alleen maar je bestand te uploaden en aan te geven in wat voor bestandstype je het wil omzetten. De converters leveren soms verschillende resultaten op, welke lang niet allemaal even betrouwbaar zijn. Het is daarom belangrijk om altijd je CSV-bestand te vergelijken met je PDF nadat je een converter hebt gebruikt. Daarnaast is het handig als je de PDF eerst opschoont, voordat je het bestand uploadt. Dan is de kans op fouten het kleinst.

Hier zijn meerdere converters die we goed vonden werken:

- [Groupdocs](#)
- [Convertio](#)
- [Aspose](#)
- [Zamzar](#)
- [Nanonets](#)

Let op! Gebruik geen converter wanneer je bezig bent met gevoelige data. Het is onbekend hoe de sites omgaan met de data van je bestand.

Voor- en nadelen van de verschillende tools

Bij het converteren van een PDF naar een CSV-bestand zijn er dus verschillende tools beschikbaar, elk met specifieke voor- en nadelen. Lees hieronder de voor- en nadelen van de verschillende tools.

Microsoft Excel

Microsoft Excel is een eenvoudige en gebruiksvriendelijke optie om PDF-bestanden naar CSV te converteren. Het biedt de mogelijkheid om fouten direct te herstellen, wat handig is bij het handmatig controleren en bewerken van gegevens. Echter, wanneer het gaat om complexere of onregelmatige tabellen, kan Excel minder accuraat zijn. Bovendien is Excel geen gratis tool (wanneer je stopt bij de HU), wat een nadeel kan zijn voor gebruikers die een kosteloze oplossing zoeken.

Adobe Acrobat

Adobe Acrobat staat bekend om zijn veilige en betrouwbare conversies naar verschillende formaten, zoals van PDF naar CSV. Het grootste voordeel van deze tool is de hoge mate van nauwkeurigheid en de betrouwbaarheid van de conversie. Het nadeel is echter dat Adobe Acrobat een dure optie is voor mensen die niet in beschikking zijn van een licentie bij de Creative Cloud, wat het minder aantrekkelijk maakt voor gebruikers om deze aan te schaffen. Aangezien studenten binnen deze minor wel over deze licentie beschikken, is Adobe Acrobat een hele geschikte optie om PDF's te converteren naar CSV-bestanden.

AI-tools

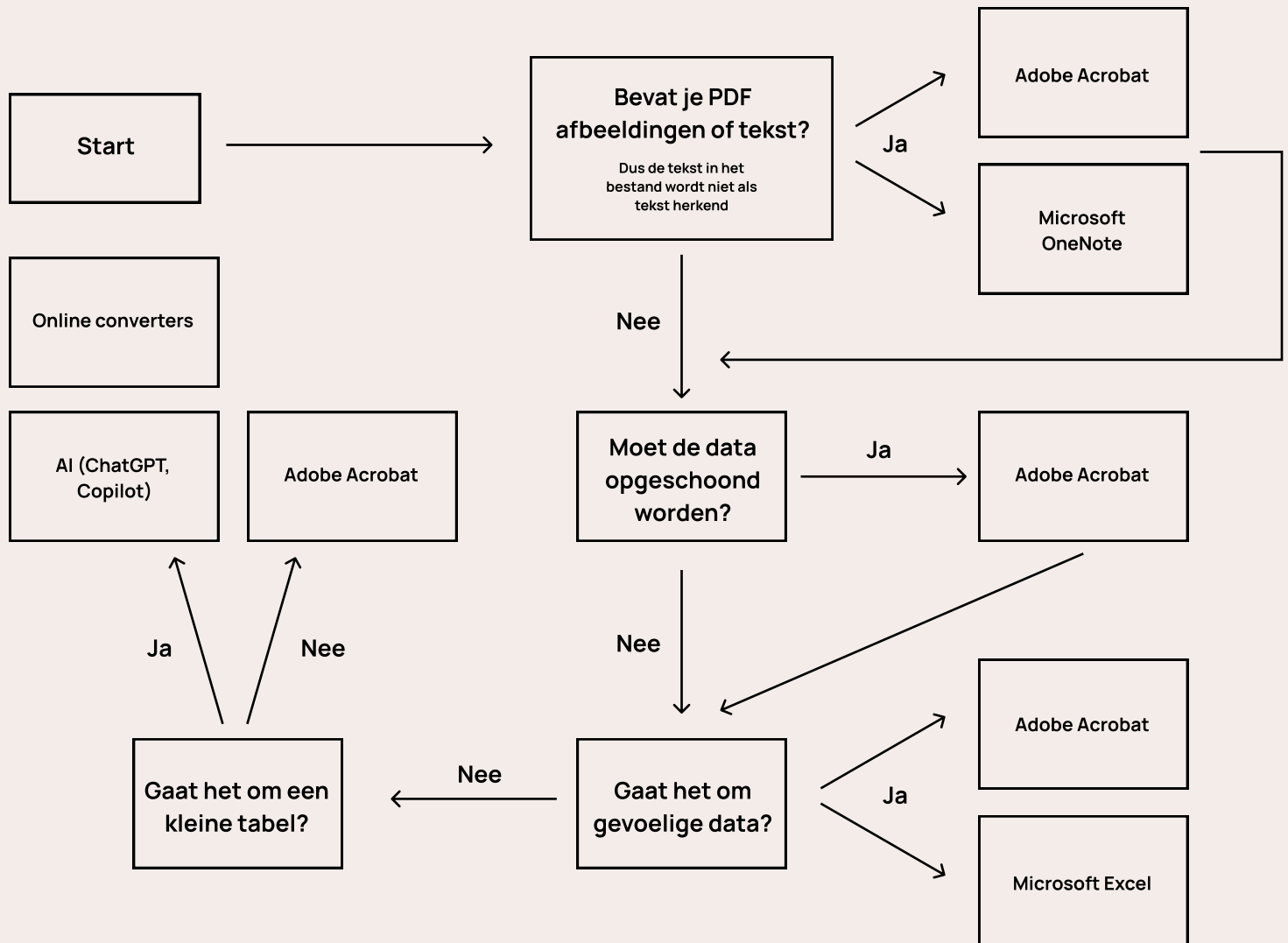
AI-tools bieden een moderne en snelle manier om PDF's naar CSV te converteren. Deze tools zijn vaak eenvoudig te gebruiken en kunnen helpen bij het opschonen en structureren van de data, wat de conversie efficiënter maakt. Een nadeel van AI-tools is echter dat ze moeite kunnen hebben met onregelmatige of zeer complexe tabellen en het is altijd noodzakelijk om handmatig langs de geconverteerde data te gaan en te controleren of alles goed is gegaan. Daarnaast zijn deze tools vaak in beperkte mate gratis. Zo kan er bij te veel gebruik gevraagd worden om je account te upgraden naar een betaalde versie.

Online tools

Tot slot zijn er diverse online tools beschikbaar die PDF-conversies snel en toegankelijk maken. Ze zijn vaak gratis en vereisen geen installatie, wat het een handige optie maakt voor mensen die er eenmalig gebruik van willen maken. Toch hebben deze tools beperkingen, zoals gelimiteerde bestandsgroottes en een beperkt aantal conversies per dag. Daarnaast kunnen er privacyrisico's zijn bij het uploaden van gevoelige bestanden naar een externe server en zijn deze tools vaak minder nauwkeurig bij complexe tabellen.

05 Flowchart

Om alles samen te vatten hebben we een simpele flowchart gemaakt die stap voor stap het proces van het converteren van een PDF-bestand naar een CSV-bestand uitlegt en de gebruiker meeneemt in welke tool handig is in de gegeven situatie.



Stap 1: Bevat je PDF afbeeldingen of tekst?

De eerste vraag die je jezelf moet stellen is of de PDF om een afbeelding gaat, waarbij de tekst in het bestand niet als echte tekst wordt herkend. Als dat het geval is, kun je kiezen tussen Adobe Acrobat of Microsoft OneNote. Deze tools zijn geschikt voor het herkennen van tekst (OCR) en het omzetten van een afbeelding-PDF naar een tekst-PDF. Als het geen afbeelding-PDF is, ga je door naar de volgende stap.

Stap 2: Moet de data opgeschoond worden?

Als de PDF geen afbeelding bevat of succesvol omgezet is naar een tekst-PDF, vraag je jezelf af of het bestand nog opgeschoond dient te worden. Dit is het geval als een bestand naast de data nog veel andere informatie bevat die niet van belang zijn om opgenomen te worden in het CSV-bestand. Als dit het geval is, kan Adobe Acrobat handig van pas komen voor het opschoonen. Als de data niet opgeschoond hoeft te worden, ga je verder naar de volgende stap.

Stap 3: Gaat het om gevoelige data?

Als de PDF gevoelige data bevat, zoals persoonlijke of vertrouwelijke informatie, is het belangrijk om een veilige tool te gebruiken. In dit geval zijn Adobe Acrobat en Microsoft Excel beide geschikt. Deze tools bieden meer controle en veiligheid bij het verwerken van gegevens, wat belangrijk is bij gevoelige informatie. Microsoft Excel zal meer geschikt zijn voor simpele data en Adobe Acrobat voor wat complexere data. Als de data niet gevoelig is, kun je verdergaan naar de volgende stap.

Stap 4: Gaat het om een kleine tabel?

In het geval dat het niet om gevoelige data gaat, kan je nagaan of de data in het PDF-bestand in een kleine tabel staat. Als het antwoord ja is, zal een AI-tool een snellere en gemakkelijker manier kunnen zijn voor het converteren. Tools zoals ChatGPT of Copilot zijn ideaal voor eenvoudige en snelle conversies zonder al te veel complexiteit. Voor de mensen die niet zo bekend zijn met AI-tools, kunnen online converters ook een oplossing zijn voor het converteren van kleine tabellen. Controleer echter altijd nog handmatig of het geconverteerde CSV-bestand klopt.

Gaat het om een grotere of complexere tabel, dan zal het toch efficiënter werken om deze zelf via Adobe Acrobat te converteren, aangezien AI-tools vaak minder accuraat kunnen omgaan met complexe data. Adobe Acrobat is een meer betrouwbare tool bij gebruik van grotere en complexere datasets.

06 Slotwoord

Deze handleiding heeft je een reeks methoden uitgelegd om PDF-bestanden naar CSV om te zetten, van eenvoudige online converters tot meer slimme oplossingen zoals Adobe Acrobat en AI-tools. Elk van deze opties biedt voordelen, afhankelijk van wat voor soort data je gebruikt.

Voor eenvoudige, niet-gevoelige bestanden kunnen online tools vaak snel en doeltreffend zijn, terwijl je voor complexere datasets het gebruik van krachtigere programma's kunt gebruiken, zoals Adobe Acrobat of Microsoft Excel, voor meer nauwkeurigheid, of zelfs AI-oplossingen, voor een sneller proces.

Door de tools en technieken die in deze handleiding zijn besproken te volgen, ben je nu goed voorbereid om PDF-bestanden effectief en nauwkeurig naar CSV om te zetten.

07 Bronnen

Adobe. (z.d.). Wat is een PDF? Portable Document Format. adobe.com. Geraadpleegd op 22 oktober 2024, van <https://www.adobe.com/nl/acrobat/about-adobe-pdf.html>

ByteScout. (2021, 20 september). CSV Format: History, Advantages and Why It Is Still Popular. Geraadpleegd op 24 oktober 2024, van <https://bytescout.com/blog/csv-format-history-advantages.html>

Microsoft. (z.d.). CSV-bestanden maken of bewerken om in Outlook te importeren - Microsoft Ondersteuning. Geraadpleegd op 22 oktober 2024, van <https://support.microsoft.com/nl-nl/office/csv-bestanden-maken-of-bewerken-om-in-outlook-te-importeren-4518d70d-8fe9-46ad-94fa-1494247193c7>

Van Muylder, E. (2019, 26 september). De logica achter factuurherkenning met OCR-software. De Digitale Versnelling. Geraadpleegd op 24 oktober 2024, van <https://www.digitaleversnelling.be/inspiratie/digitaal-advies/de-logica-achter-factuurherkenning-met-ocr-software>