

# MINOR DATAVISUALISATIE & INFOGRAPHIC

Afsluiting periode A. Tutorialdag 1.

TUTORIALDAG 1

## HANDLEIDING PDF NAAR CSV

Gepresenteerd door:  
Gijs Bol  
Rick van der Meijden  
Anne Waelput

Docenten:  
Daniela van Geenen  
Wouter van Dijke

# INHOUDSOPGAVE

*	3	INTRODUCTIE
*	4	INLEIDING
*	6	LEERDOELEN
*	8	CISDEM PDF CONVERTER ORC
*	10	ADOBE ACROBAT DC
*	12	EXEL
*	14	TOEPASBAARHEID
*	15	CONCLUSIE

# INTRODUCTIE

Vanuit onze Minor Datavisualisatie en Infographics werden door verschillende multidisciplinaire teams Tutorials gepresenteerd. De tutorials zijn gepresenteerd in de vorm van een workshop. Naast onze workshop is deze schriftelijke handleiding geschreven.

Het is geschreven voor iedereen die gebruik wil maken van onze gekozen softwareprogramma's om data mee te coverten.

# INLEDING HANDLEIDING

## PDF NAAR CSV

**DEZE HANDLEIDING  
IS GEMAAKT VOOR  
HET OMZETTEN VAN  
EEN PDF-BESTAND  
NAAR EEN  
CSV-BESTAND.**

### **VAN PDF NAAR CSV**

Als het aankomt op bestandenformaten zijn er veel verschillende soorten. Voor elke taak is een bijbehorend bestandformaat. Als je bijvoorbeeld een afbeelding wilt exporteren kijk je snel naar een JPEG of PNG. Bij een Word bestand is dit een DOCX-file. Een bestand soort die minder vaak voorkomt is een CSV-bestand. Het wordt vooral gebruikt bij het verplaatsen van grote hoeveelheden gegevens. Bijvoorbeeld om mailinglijsten te maken of om een begin van een assortiment te kunnen laden op de website binnen de webshop branche.

### **WAT IS EEN PDF?**

PDF staat voor Portable Document Format, In het Nederlands betekent dit een uitwisselbaar bestandsformaat. Het is in 1991 bedacht om alle gegevens die op papier stonden over te zetten naar een digitaal format. Het doel hiervan was om documenten gemakkelijk vast te kunnen leggen en te kunnen delen. Een PDF-bestand ziet er altijd uit zoals de maker het bedoeld heeft. Denk hierbij aan lettertype gebruik, lettertype grootte, afbeeldingen en kleuren. Ongeacht de manier van openen of afdrukken blijft het altijd hetzelfde. Het meest bekende programma om PDF bestanden te kunnen openen is Adobe Reader. Het PDF format is dan ook ontwikkeld door Adobe.



**CISDEM PDF  
CONVERTER OCR**



**ADOBE  
ACROBAT DC**



**MICROSOFT  
EXCEL**

## WAT IS EEN CSV?

Dit type bestand is niets anders dan een database indeling. CSV staat voor Comma Seperated Values ofwel komma gescheiden bestand in het Nederlands. Aan de hand van een spreadsheet kun je de informatie makkelijk indelen. Hierdoor is het een bestandsvorm die wordt gebruikt om gegevens te scheiden of in kaart brengen. Een CSV is er om bepaalde gegevens of informatie te transporteren van locatie A naar de locatie B. Het is vooral handig als het aankomt op eindeloos lange rijen of tabellen. Het is een van de meest gebruikte soorten formaten die je maar kunt vinden binnen de bedrijfsvoering. Zo wordt het ook gebruikt binnen de boekhouding en is het mogelijk om CSV-bestanden te uploaden bij Woocommerce of andere webshop plug-ins.

## WAAROM VAN PDF NAAR CSV?

Informatie of gegevens uit verslagen, rapporten of onderzoeken worden vaak in PDF-format gepubliceerd. Om deze informatie werkbaar te maken wordt het omgezet in een CSV-bestand. Door een PDF naar CSV-bestand te converteren kan de data uit PDF bestanden gemakkelijk bewerkt worden. Je kunt hierdoor informatie filteren, cijfers herberekenen of gegevens importeren. Zoals al eerder genoemd werd, wordt dit bestandstype het meest gebruikt binnen bedrijfsvoering en webshop branche.

# LEERDOELEN

PDF bestanden kunnen relevante data bevatten, het is echter moeilijk om deze data aan een PDF bestand te onttrekken. Door het toepassen van een PDF naar CSV tool kan deze data wel aan een PDF bestand onttrokken worden. Aan de hand van drie tools: Excel, Adobe Acrobat en Cisdem bieden wij het handvat waarmee PDF bestanden gemakkelijk omgezet kunnen worden naar CSV.

## WIJ HOPEN DEELNEMERS AAN ONZE TUTORIAL EEN AANTAL SKILLS MEE TE KUNNEN GEVEN:

- 1.** De relevantie van de PDF naar CSV tools (waarom en wanneer zijn deze nuttig)
- 2.** Het aanmeten van de juiste tool per deelnemer, afhankelijk van zijn of haar laptop
- 3.** Het praktische gebruik van deze tool onder de knie krijgen door middel van een opdracht
- 4.** Door het gebruik van bovenstaande skills in staat stellen om sneller en gemakkelijker data uit grote PDF bestanden (rapporten ed.) te onttrekken.

# ONDERZOEK

## ONDERZOEK NAAR DE JUISTE TOOLS.

Wanneer je op het internet opzoek gaat naar een PDF to CSV converter, kom je een hoop mogelijkheden tegen. Om het makkelijk te maken hebben wij dit onderzoek al voor je gedaan.

Van de "beste" online converters tot betaalde versies. Alle mogelijkheden zijn onderzocht. Om de handleiding overzichtelijk te houden, hebben wij onze drie tools uitgewerkt.

Hieronder vind je per tool een handleiding.

# CISDEM PDF CONVERTER ORC

## TOOL 1

CISDEM PDF CONVERTER OCR IS EEN VEELZIJDIGE PDF CONVERTER. DEZE TOOL BIJDT DE MOGELIJKHEID OM PDF-BESTANDEN OP MAC TE CONVERTEREN, MAKEN EN COMPRIMEREN.

Met de conversiefunctie kan de gebruiker elk type PDF-bestand en afbeelding converteren naar 16 verschillende formaten. Zo heb je bijvoorbeeld Excel, Word, Powerpoint, Tekst en HTML als type bestand. De originele lay-out en opmaak van het bestand blijven hierin behouden.

Hieronder het stappenplan om de tool te downloaden en een PDF-bestand naar een CSV-bestand te converteren.

### VOOR –EN NADELEN

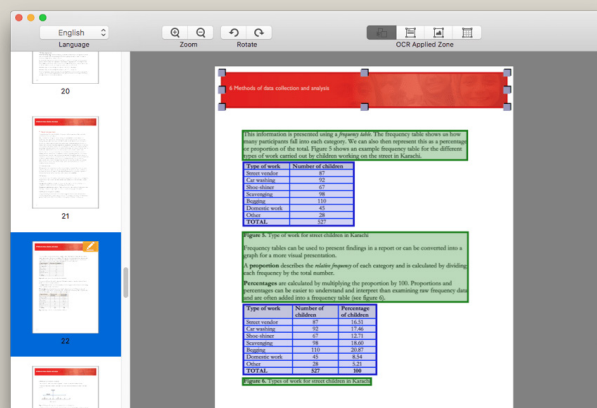
Bij het gebruik van deze tool zijn er een aantal grote voordelen. Zo is de tool gratis te downloaden en te gebruiken voor iedereen. Hier is geen account voor nodig. Na het downloaden is het meteen inzetbaar om PDF-bestanden naar CSV te converteren. Naast het converteren heeft het meerdere functies. Zo kan de tool PDF-bestanden converteren, maken en comprimeren op Mac. Ook handig is zoals eerder uitgelegd, de nauwkeurigere OCR-resultaten zelf nog handmatig aan te kunnen passen voordat je het bestand converteert.

Naast de voordelen van de tool zijn er ook een aantal nadelen, die belangrijk zijn om te benoemen. Wanneer je een Android gebruiker bent, is deze tool niet geschikt voor jou. De Cisdem PDF Converter OCR is namelijk speciaal voor Mac gebruikers ontworpen.

Daarnaast wordt alles meegenomen (tekst, afbeeldingen etc.) naar een Excel werkblad. Dit houdt in dat je na het converteren zelf nog de taak hebt om het bestand op te schonen. Wel staan alle tabellen afbeeldingen etc. correct in het geconverteerde bestand.

### NAUWKEURIGERE OCR RESULTATEN

Voor nauwkeurigere OCR-resultaten kun je op geavanceerde instellingen klikken. Elk onderdeel is gemarkeerd in een gekleurd vak, ze worden herkend als tekst, tabel of afbeelding. Je kunt de OCR-zone handmatig aanpassen om ervoor te zorgen dat elk deel van het bestand voldoende gemarkeerd is. Als je bijvoorbeeld een deel wilt herkennen als tekst in plaats van een tabel, verwijder je het originele markeringsvak, klik op "Tekst" en sleep vervolgens een vak om dat specifieke deel te markeren. Dit deel wordt vervolgens uitgevoerd als tekst zonder originele opmaak

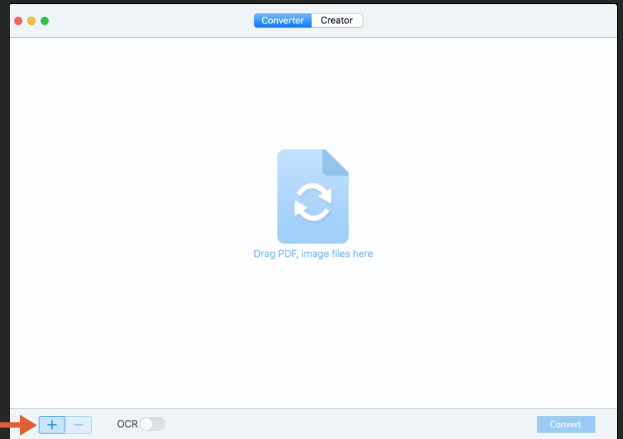




# HIERONDER HET STAPPENPLAN OM DE TOOL TE DOWNLOADEN EN EEN PDF-BESTAND NAAR EEN CSV-BESTAND TE CONVERTEREN.

**Stap 1.** Download en installeer Cisdem PDF Converter OCR. Via deze link <https://www.cisdem.com/pdf-converter-ocr-mac.html>

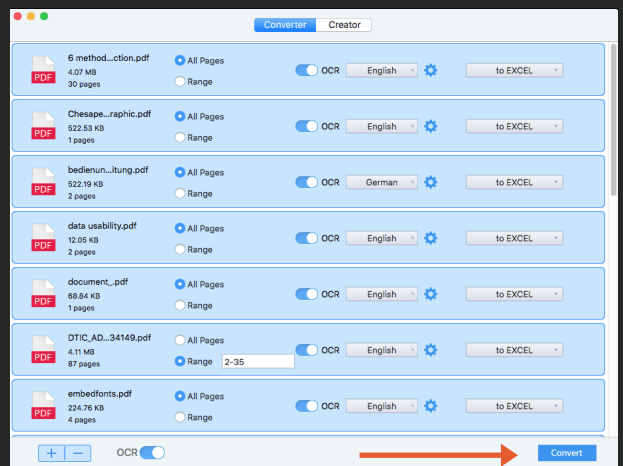
**Stap 2.** Voeg een of meerdere PDF-bestanden toe. Dit kun je doen door bestanden te slepen en neerzetten om te importeren. Je kunt ook door op “+” te klikken een PDF-bestand toe te voegen.



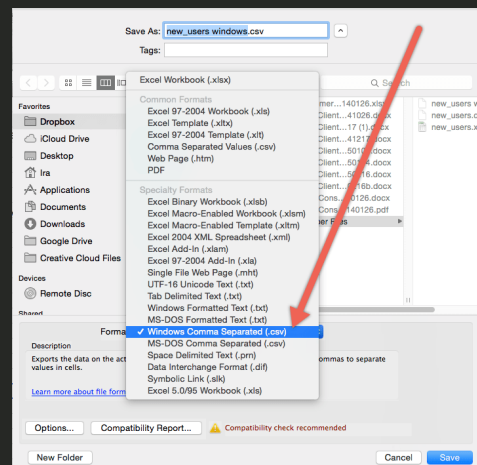
**Stap 3.** Nadat de PDF-bestanden zijn geïmporteerd, zet je de OCR-knop AAN en kiest je vervolgens uitvoer als Excel. (Het Excel uitvoerformaat is alleen beschikbaar wanneer OCR is ingeschakeld)



**Stap 4.** Klik op “Converteren” om PDF naar Excel te wijzigen.



**Stap 5.** Open het Excel-bestand in Microsoft Excel, ga naar Bestand > Opslaan als, kies uitvoer als CSV en sla het bestand op.



# ADOBE ACROBAT DC

## TOOL 2

OM PDF-BESTANDEN OM TE ZETTEN NAAR CSV-BESTANDEN  
BESTAAN VERSCHILLENDE MOGELIJKHEDEN. HIER WORDT  
KORT BESPROKEN HOE DAT IN ZIJN WERK GAAT BIJ HET  
PROGRAMMA 'ADOBE ACROBAT (PRO) DC

Vaak worden bestanden, verslagen, onderzoeken en rapporten in PDF gepubliceerd. In deze verslagen zijn gegevens te vinden. Om die gegevens uit een PDF-bestand te filteren kun je werken met het programma: Adobe Acrobat (Pro). In plaats van meteen van PDF naar CSV over te gaan, maken we eerst de stap om het bestand naar een Excel werkblad over te zetten. De reden waarom je het bestand over moet zetten naar Excel: omdat CSV een gevoelig bestand. Om mogelijke fouten in het CSV-bestand te voorkomen is deze omweg een goede keuze.

Wanneer het document omgezet is naar Excel kan je de gegevens selecteren die je wilt gebruiken en opslaan als CSV-bestand.

### VOOR -EN NADELEN

Een groot nadeel van deze applicatie is dat het een betaalde applicatie is. De prijs voor het gebruik van deze app ligt rond de 200 euro per jaar, afhankelijk van de versie die ervoor gebruikt wordt.

Een voordeel van deze applicatie is dat het werkzaam is op alle apparaten en dat documenten veilig en vertrouwd bewaard worden. Bestanden hoeven niet omgezet te worden zodat het werkzaam is op een andere computer/laptop.

### EXTRA TOEPASSINGEN

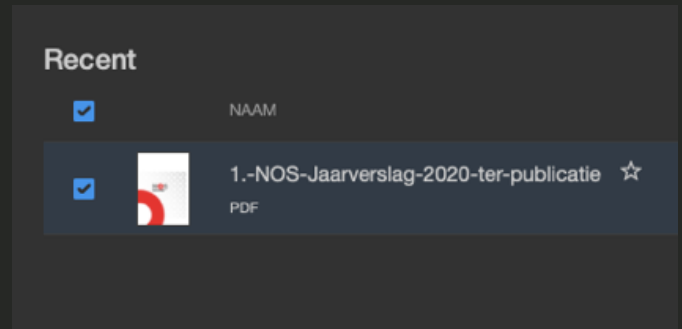
Naast het converteren van bestanden, kan Adobe Acrobat voor meer zaken gebruikt worden. Het bewerken van PDF's is een van deze functies. Een verslag waar aanpassingen aan gedaan moeten worden, kan zo snel gedaan worden.

Er kunnen ook PDF's gemaakt worden met deze applicatie. Word-documenten en Power Point Presentaties kunnen omgezet worden naar PDF's. Verder kunnen er ook bestanden ondertekend worden in Adobe Acrobat met een digitale handtekening.

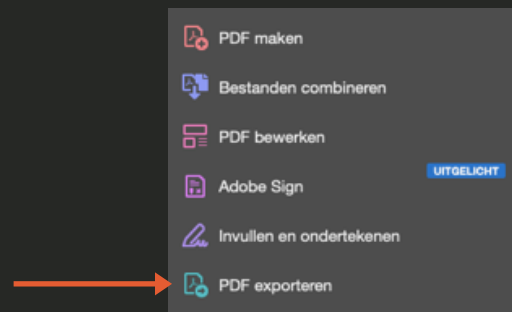
# HIERONDER HET STAPPENPLAN OM DE TOOL TE DOWNLOADEN EN EEN PDF-BESTAND NAAR EEN CSV-BESTAND TE CONVERTEREN.

**Stap 1.** Download en installeer Adobe Acrobat DC, via Creative Cloud.

**Stap 2.** Open het document of start Adobe Acrobat op en kies het bestand dat je wilt gebruiken.



**Stap 3.** Wanneer het document geopend is, kies je voor 'PDF exporteren', wat aan de rechterkant te zien is.

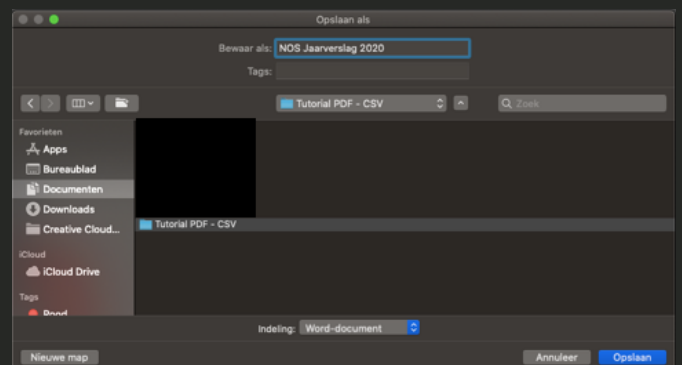


**Stap 4.** Kies vervolgens het kopje 'Werkblad' en selecteer 'Microsoft Excel-werkmap'.



**Stap 5.** Open het Excel-bestand in Microsoft Excel, ga naar Bestand > Opslaan als, kies uitvoer als CSV en sla het bestand op.

Nu is het document omgezet naar een Excel bestand. Controleer of de data correct weergegeven worden. Selecteer vervolgens de data die behouden moet blijven en filter weg wat niet nodig of van toepassing is. Wanneer dat gedaan is en de gegevens correct zijn, kan het Excel werkblad omgezet worden naar een CSV-bestand. Dat kan bij Excel zelf omgezet worden. Hoe dat werkt, is te zien in de Excel handleiding.



# MICROSOFT EXCEL

## TOOL 3

EEN MANIER OM EEN PDF OM TE ZETTEN NAAR EEN CSV-BESTAND IS VIA HET PROGRAMMA EXCEL. EXCEL WORDT GEBRUIKT DOOR ZOWEL BEDRIJVEN ALS PARTICULIEREN.

Excel wordt gebruikt door zowel bedrijven als particulieren. Dit kan gebruikt worden voor administratie, maar ook voor persoonlijke doeleinden. Verder is Excel een programma waar veelal met grafieken, diagrammen en tabellen wordt gewerkt. Grote hoeveelheden data kunnen in Excel geanalyseerd en gevisualiseerd worden. Daarnaast kan er ook een contactenlijst geïmporteerd worden in een mailprovider. Voor deze data moeten de gegevens verwerkt worden in een CSV-bestand. Via Excel is dat goed te doen.

### **VOOR -EN NADELEN**

Het is voordelig aan Excel dat er veel mogelijkheden zijn in het programma. Het kan een gegevens opslaan, tabellen kunnen gemakkelijk gemaakt worden, het maken van grafieken kan eenvoudig, en het werkt ook goed als tussenstation voor documenten die gebruikt kunnen worden in andere programma's.

Een ander voordeel is dat de meeste computer/laptops werken met een Microsoft-pakket. In dit pakket zit Excel automatisch het programma is hierdoor voor veel mensen toegankelijk

Wat betreft het omzetten van een PDF naar CSV bestand, hebben wij ondervonden dat Excel het beste werk. Dit komt door de Navigator waardoor je vooraf al instaat bent informatie te filteren. Ook heb je hierdoor geen extra tools nodig.

Een nadeel van Excel is dat het soms lastig in gebruik is. Vooral een beginner zal in de eerste fase problemen ondervinden om goed te werken met Excel. CSV-bestanden zijn dan ook lastig te lezen, laat staan om te zetten en in te zetten bij andere programma's.

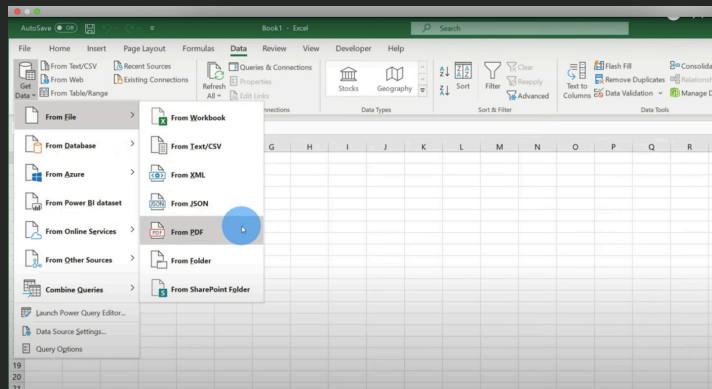
Een ander nadeel is dat Excel geld kost om in beheer te hebben. Hoewel het in het Microsoft-pakket zit, is er wel geld gemoeid met de aanschaf van dit pakket.

Deze manier van PDF-bestanden converteren is alleen mogelijk op een Windows aangestuurde laptop, alle Mac-gebruikers kunnen hier dus niet mee aan de slag. Dit is nadelig want het is dus niet voor iedereen toegankelijk.

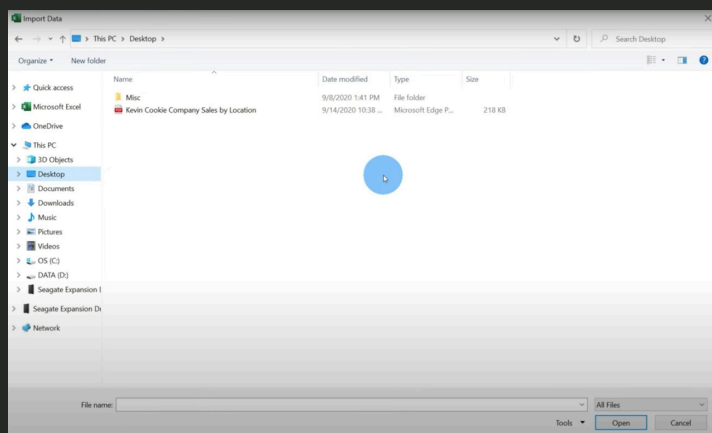
# HIERONDER HET STAPPENPLAN OM DE TOOL TE DOWNLOADEN EN EEN PDF-BESTAND NAAR EEN CSV-BESTAND TE CONVERTEREN.

**Stap 1.** Open een nieuw werkblad in Exel.

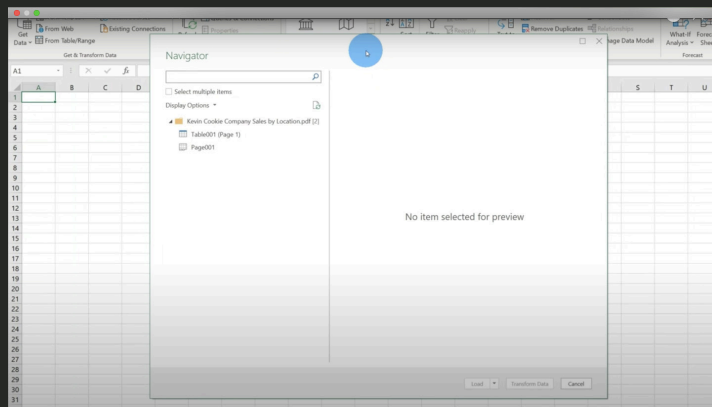
**Stap 2.** Ga naar het kopje Data en klik links in de hoek op Get Data. Hier krijg je een dropdown menu te zien. Klik vervolgens op From File, hiernaast komen verschillende opties te voorschijn. Klik dan op de optie From PDF.



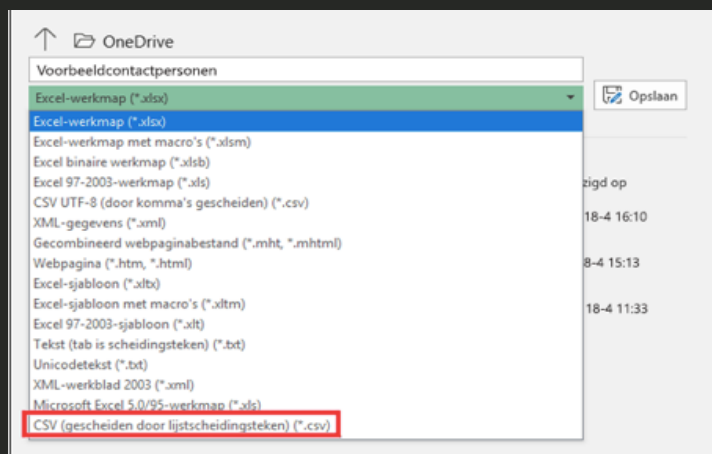
**Stap 3.** Klik in het scherm op het bestand dat je wilt openen.



**Stap 4.** Na de vorige stap opent Exel Navigator zich. Op dit scherm zie je de verschillende onderdelen die in het PDF bestand staan. Door de Navigator kun je informatie gaan filteren als je bijvoorbeeld een specifieke tabel wilt overnemen. Dit doe je door ze te selecteren.



**Stap 5.** Wanneer het bestand goed te zien is, kies je 'opslaan als'. Gebruik de vervolgkeuzelijst, selecteer daar 'CSV'. Het bestand is omgezet naar een CSV-bestand



# TOEPASBAARHEID

## TOP 3 TOOLS

### CISDEM PDF CONVERTER OCR

Wanneer je gebruikt maakt van Cisdem PDF Converter OCR, om van een PDF-bestand naar CSV-bestand te gaan, gaat dat niet zonder Excel. Je moet in Excel het geconverteerde bestand opslaan als .csv. Door het converteren via deze tool staan alle tabellen en informatie makkelijk te bewerken in Excel. Via de tool kun je door de handige OCR-functie de markeringsvakken vooraf bewerken om het filteren in Excel achteraf gemakkelijker te maken. Het filteren van informatie is vooraf dus niet mogelijk.

### ADOBE ACROBAT PRO

Als je van een PDF naar een CSV-bestand wilt gaan, dan lukt dat niet alleen met Adobe Acrobat. Je komt wel een heel eind, maar de stap om van een Excel-werkblad naar een CSV-bestand te gaan, is alleen te doen bij het programma Excel.

Het filteren van de data is in Adobe Acrobat lastig uit te voeren. In principe is deze applicatie alleen bedoeld om bestanden om te zetten, het filteren van gegevens behoort echter niet tot de directe een van de directe functies van Acrobat. Adobe Acrobat is meer een tussenprogramma waar je bestanden instopt, die er vervolgens anders uit komen.

### EXCEL

In Excel kan je gemakkelijk een werkblad omzetten naar een CSV-bestand. Echter, wanneer je er een PDF-bestand in plaatst, bestaat de kans dat er veel tekst of afbeeldingen in staan die je niet nodig hebt. Het filteren kan dan redelijk snel gebeuren. Door de data eruit te pikken die je nodig hebt, kan je daarna makkelijk werken met de geselecteerde data.

Verder kan je ook gemakkelijk bepaalde tekens aanpassen. Staan er bijvoorbeeld komma's in plaats van punten dan kan je dat gemakkelijk aanpassen.

Wanneer je CSV-documenten opent in Excel, staan er veel komma's en punten tussen de data. Hoewel dat er vreemd uitziet, is dat wel nodig voor de CSV-bestanden.

## CONCLUSIE

Er zijn een hoop online gratis PDF naar CSV-converters te vinden, maar de resultaten van deze tools zijn vaak onbruikbaar.

We hebben in ons onderzoek een hoop onlineprogramma's getest waaronder: Zamzar, Convertio, Easypdf en Aconvert. Ze beweren allemaal hetzelfde: een PDF naar CSV te kunnen converteren. Het resultaat hiervan was dat, in het geconverteerde bestand, een hoop rare symbolen of vraagtekens te zien waren. Niets wat daadwerkelijk bewerkbaar was.

We kunnen dus concluderen dat er geen goede online gratis PDF naar CSV-converter is om PDF-bestand rechtstreeks naar CSV-bestand te exporteren.

Met dit onderzoek hebben we uiteraard ook onze drie verschillende toegelichte programma's getest. Hiermee kunnen we dan ook concluderen dat Adobe Acrobat, Cisdem PDF Converter OKR en Excel het beste uit onze test zijn gekomen. Daarom raden we deze dan ook aan om te gebruiken. Iedere tool heeft voor- en nadelen, maar al met al dienen zij allemaal hetzelfde doel: een CSV-bestand uit een PDF-bestand halen.

