

Arbeidsmarktonderzoek 2022

Data scientist



Datum	19 januari 2023
Auteurs	Marcia den Uijl Pradeep Kumar
Versie	1.0
Classificatie	standaard



Uitgave

Centerdata
info@centerdata.nl
www.centerdata.nl

Contact

Marcia den Uijl
marcia.den.uijl@centerdata.nl

© **Centerdata, Tilburg, 2023**

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.



Inhoudsopgave

1	Inleiding	4
2	Data scientist	5
2.1	Aantal vacatures	5
2.2	Functietitels	5
2.3	Sector	5
2.4	Regio	6
3	Algemene vaardigheden	7
4	Digitale vaardigheden	9
5	Transities	12
5.1	Startberoepen	12
5.2	Skills gap	13
A	Categorieën vaardigheden	15





1 Inleiding

Dit rapport behoort bij het Arbeidsmarktonderzoek 2022 dat door Centerdata is uitgevoerd in opdracht van het Convenant Werkperspectief. Zie het hoofdrapport voor meer informatie over de achtergrond van het algehele onderzoek. Deze verdiepende analyse gaat specifiek in op het beroep data scientist. De data waarop deze analyse is gebaseerd zijn vacaturedata en de beroependatabase O*NET.

Dit rapport is als volgt opgebouwd. Hoofdstuk 2 begint met algemene informatie over vacatures voor data scientist, namelijk de functietitels, voorkomende sectoren en verschillen in regio's. Hoofdstukken 3 en 4 gaan vervolgens in op de gevraagde algemene en digitale vaardigheden in de vacatures voor data scientist. In hoofdstuk 5 worden transities naar data scientist onder de loep genomen. Welke startberoepen haalbaar en wenselijk om deze overstap te maken en wat voor soort skills gap moet dan worden overbrugd?





2 Data scientist

In de vacaturedata zijn 11.758 vacatures aanwezig voor data scientist over de periode van 2016 tot en met 2021. In de rest van dit hoofdstuk wordt ingegaan op een aantal veel voorkomende functietitels, de meest voorkomende sectoren en het regionale aandeel van deze vacatures.

2.1 Aantal vacatures

In 2016 was ongeveer 1 op de 2.500 vacatures voor data scientist, dit loopt op tot 1 op de 1.300 vacatures in 2021. We bekijken hier hoeveel vacatures voor data scientist er zijn geweest in verhouding tot het totale aantal vacatures in plaats van het absolute aantal vacatures over de jaren heen. Het is namelijk niet te zeggen of er meer vacatures zijn geplaatst of dat er meer zijn gevonden omdat de manier waarop online vacatures worden verzameld steeds wordt verbeterd.

2.2 Functietitels

Data scientist is een verzamelnaam voor verschillende soorten functies. Tabel 1 toont de meest voorkomende functietitels. Functietitels bevatten in sommige gevallen een niveau (junior/medior/senior), een specifieke bedrijfsnaam of een specifieke plaatsnaam. Deze gegevens worden genegeerd.

Tabel 1 Meest voorkomende functietitels voor data scientist

Functietitel	Aantal vacatures	Percentage
Data scientist	9713	83%
Lead Data Scientist	319	3%
Big data scientist/engineer/expert	245	2%
Machine learning scientist/engineer/expert	190	2%

2.3 Sector

In Tabel 2 worden de meest voorkomende sectoren getoond waar vacatures voor data scientist zijn geplaatst. Veel vacatures worden via een bemiddelaar geplaatst waardoor de sector niet te achterhalen is. Daarnaast ontbreekt ook voor een deel van de vacatures informatie over de sector. De meeste vacatures voor data scientist waarbij de sector wel bekend is komen van organisaties die dienstverlenende werkzaam zijn op het gebied van ICT advisering, ontwikkeling van software en management- en bedrijfsvoering advies.

Tabel 2 Meest voorkomende sectoren voor data scientist

Sector	Aantal vacatures	Percentage
Arbeidsbemiddeling, uitzendbureaus en personeelsbeheer	3829	33%
Onbekend	2281	19%
Advisering op het gebied van informatietechnologie	559	5%

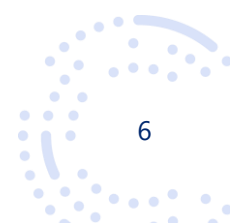
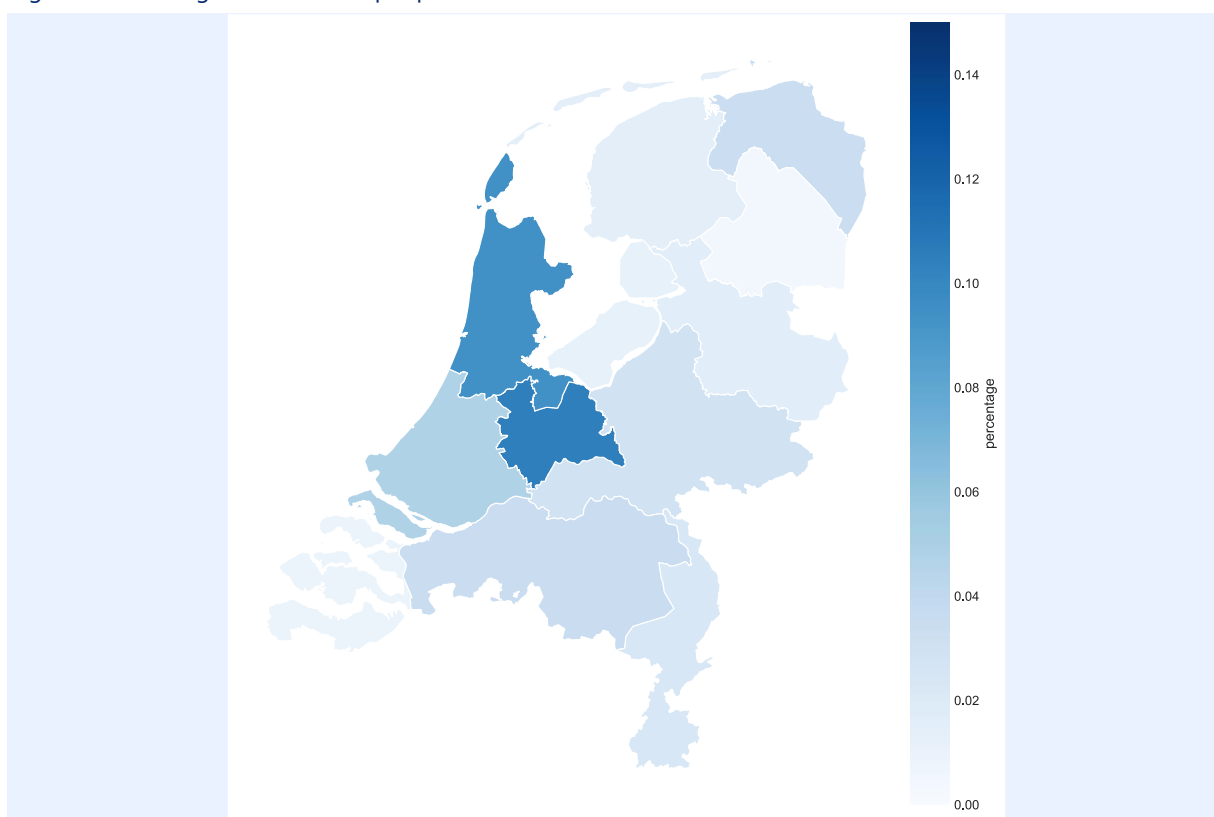


Ontwikkelen, produceren en uitgeven van software	534	5%
Advisering op het gebied van management en bedrijfsvoering	581	5%
Financiële instellingen (geen verzekeringen en pensioenfondsen)	375	3%
Openbaar bestuur en overheidsdiensten	213	2%

2.4 Regio

Van alle vacatures in 2016-2021 is ongeveer 0,06% een vacature voor data scientist, ongeveer 1 op 2.000. Regionaal zijn hier verschillen in. Het grootste aandeel data scientist is in Utrecht, met 1 op 1.000 vacatures en het kleinste aandeel in Drenthe, 1 op 21.000 vacatures. Zie ook Figuur 1.

Figuur 1 Percentage data scientist per provincie





3 Algemene vaardigheden

Tabel 4 toont de vraag naar algemene vaardigheden in vacatures voor data scientist in 2021. Vooral vaardigheden in computational thinking, gemiddeld bijna 3 keer per vacature. Ook communicatie, samenwerken en planning en organisatie worden vaker gevraagd dan in andere (ICT) vacatures.

Tabel 3 Gemiddeld aantal keer gevraagde vaardigheden per 100 vacatures voor data scientist per categorie algemene vaardigheden in 2021

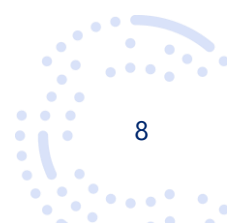
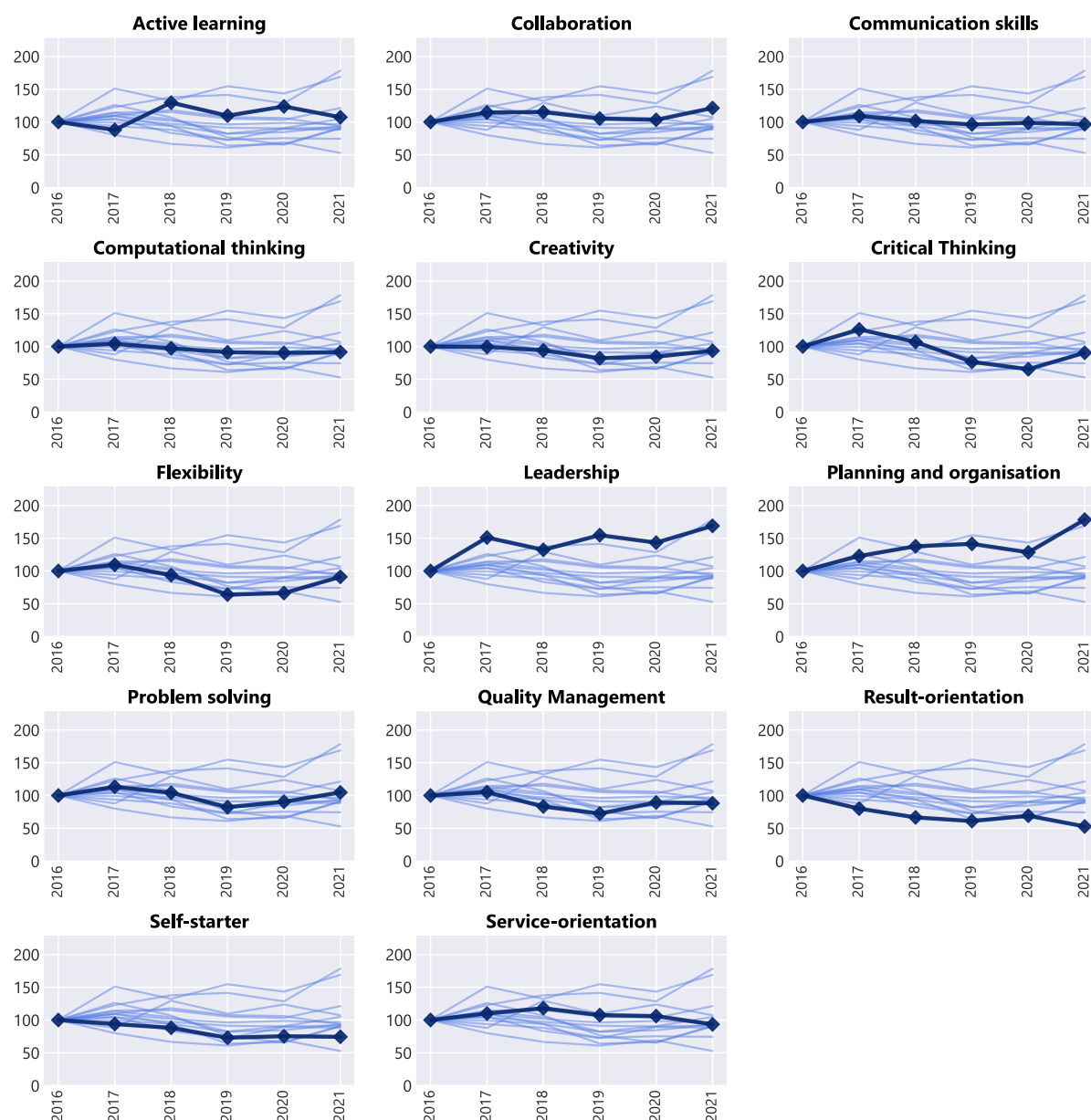
Categorie	Data scientist	ICT	Totaal
Computational thinking	274	48	19
Communication skills	112	106	84
Collaboration	97	85	55
Planning and organisation	66	89	40
Creativity	45	27	15
Self-starter	42	50	62
Leadership	30	29	25
Service-orientation	25	36	33
Problem solving	25	27	12
Critical Thinking	19	6	5
Flexibility	10	17	27
Active learning	10	13	12
Quality Management	7	11	13
Result-orientation	6	6	7

De groei in de categorie planning en organisatie komt voornamelijk door een groei in de vraag naar agile vaardigheden. Het is gegroeid van 7 vacatures in 2016 naar 15 op de 100 vacatures in 2021. Vergelijkbaar was er in de periode tussen 2016 en 2021 ook een groei in de vraag naar leiderschapsvaardigheden. Het is gestegen van 17 in 2016 naar 30 in 2021 per 100 vacatures. Ook samenwerken wordt vaker gevraagd in vacatures.





Figuur 2 Ontwikkeling in vraag naar overige vaardigheden voor data scientist tussen 2016 en 2021 (basisjaar 2016=100), per categorie



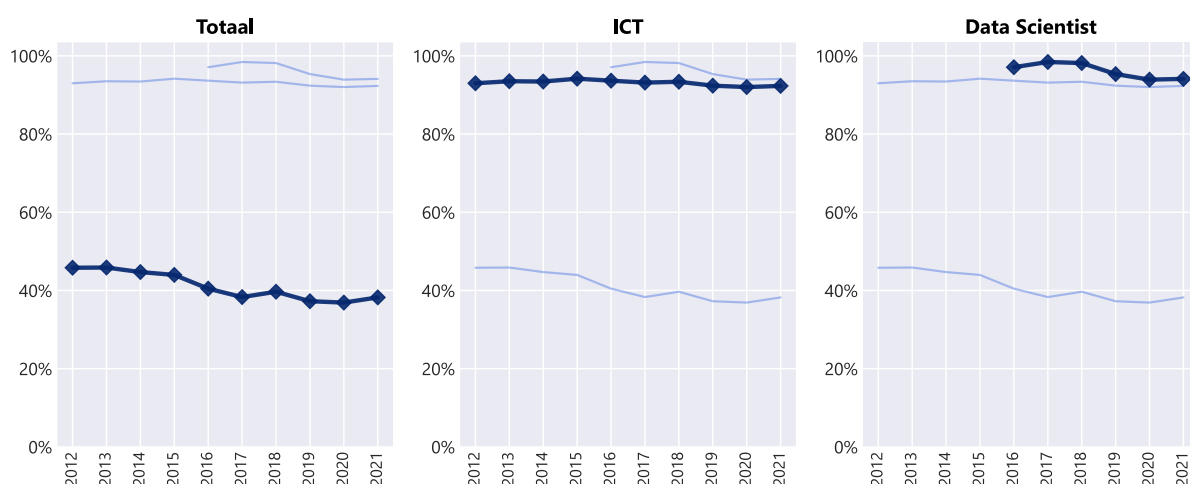


4 Digitale vaardigheden

We hebben gekeken welke digitale vaardigheden in alle vacatures en specifiek in de vacatures naar data scientist worden gevraagd.

Figuur 3 geeft de fractie vacatures weer waarin 1 of meerdere digitale vaardigheden gevraagd worden, voor alle vacatures, ICT vacatures en voor data scientist vacatures. Bij data scientist is dit ongeveer vergelijkbaar met alle ICT vacatures, 94% in 2021.

Figuur 3 Fractie vacatures met 1 of meer digitale vaardigheden, totaal, ICT en ICT Data Scientist



In Tabel 4 wordt het aantal keer dat een digitale vaardigheid wordt gevraagd getoond, per 100 vacatures in 2021, per categorie. Zie bijlage A voor voorbeelden van welke vaardigheden in welke categorie vallen. De meest gevraagde categorie was big data & analytics, gemiddeld 3,7 keer per vacature.

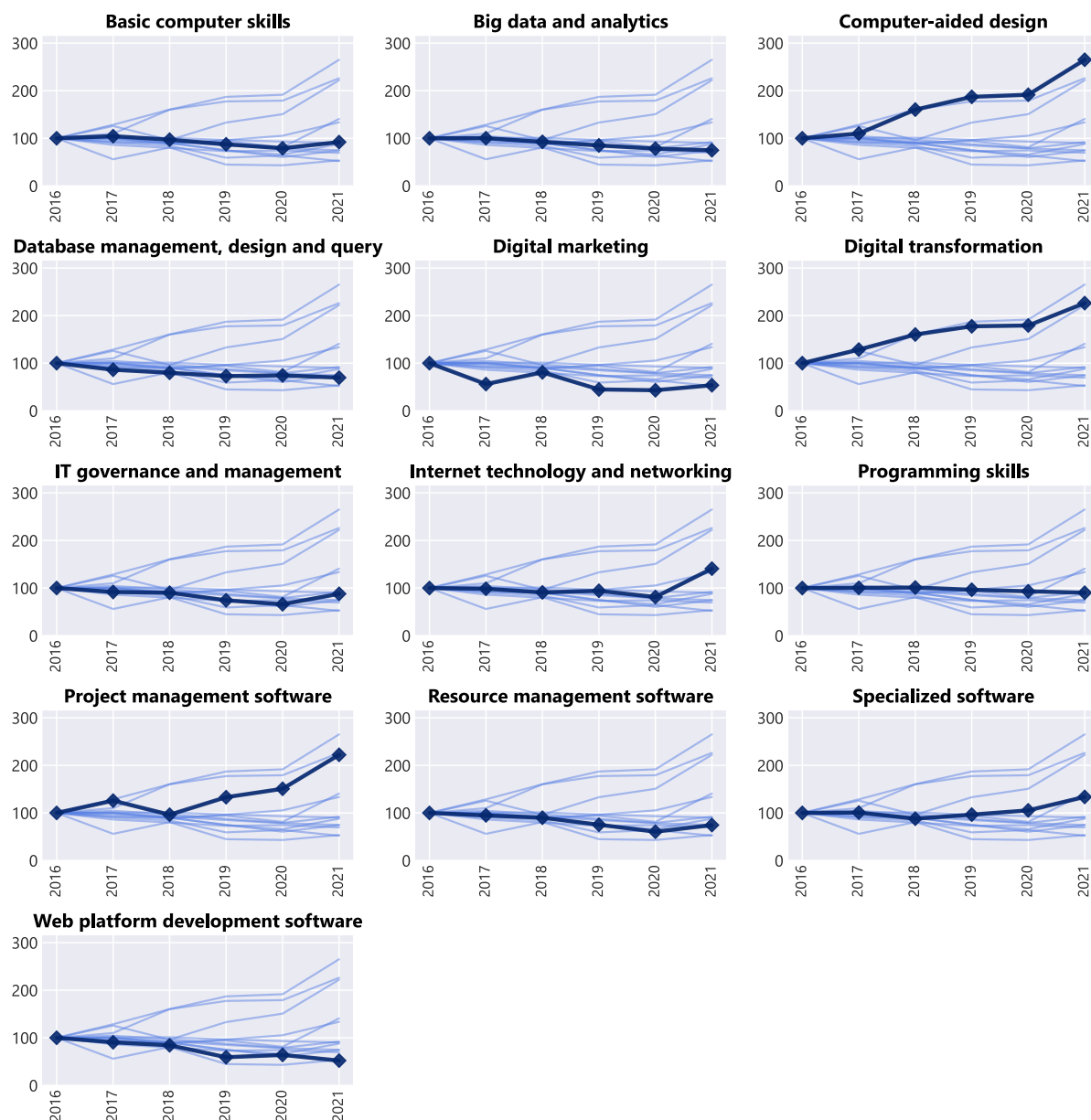
Tabel 4: Gemiddeld aantal keer gevraagde vaardigheden per 100 vacatures voor data scientist per categorie digitale vaardigheden in 2021

Categorie	Data Scientist	ICT	Totaal
Big data & analytics	374	40	8
Programmeervaardigheden	235	185	22
Digitale transformatie	140	96	16
Database management, design & query	108	57	7
Basiscomputervaardigheden	98	131	37
Web platform development software	21	89	10
Gespecialiseerde software	19	24	6
IT governance & management	17	14	3
Resource management software	14	36	10
Internettechnologie & netwerken	10	46	6
Project management software	7	20	3
Digitale marketing	4	2	2
Computer-aided design	2	5	6



In Figuur 4 wordt de ontwikkeling in de vraag naar digitale vaardigheden getoond. Er is vooral een grote stijging in de vraag naar vaardigheden in digitale transformatie (van 62 naar 140 op de 100 vacatures), project management software (van 3 naar 7 op de 100 vacatures) en computer aided design, hoewel nog steeds erg klein (van minder dan 1 naar 2 op de 100 vacatures).

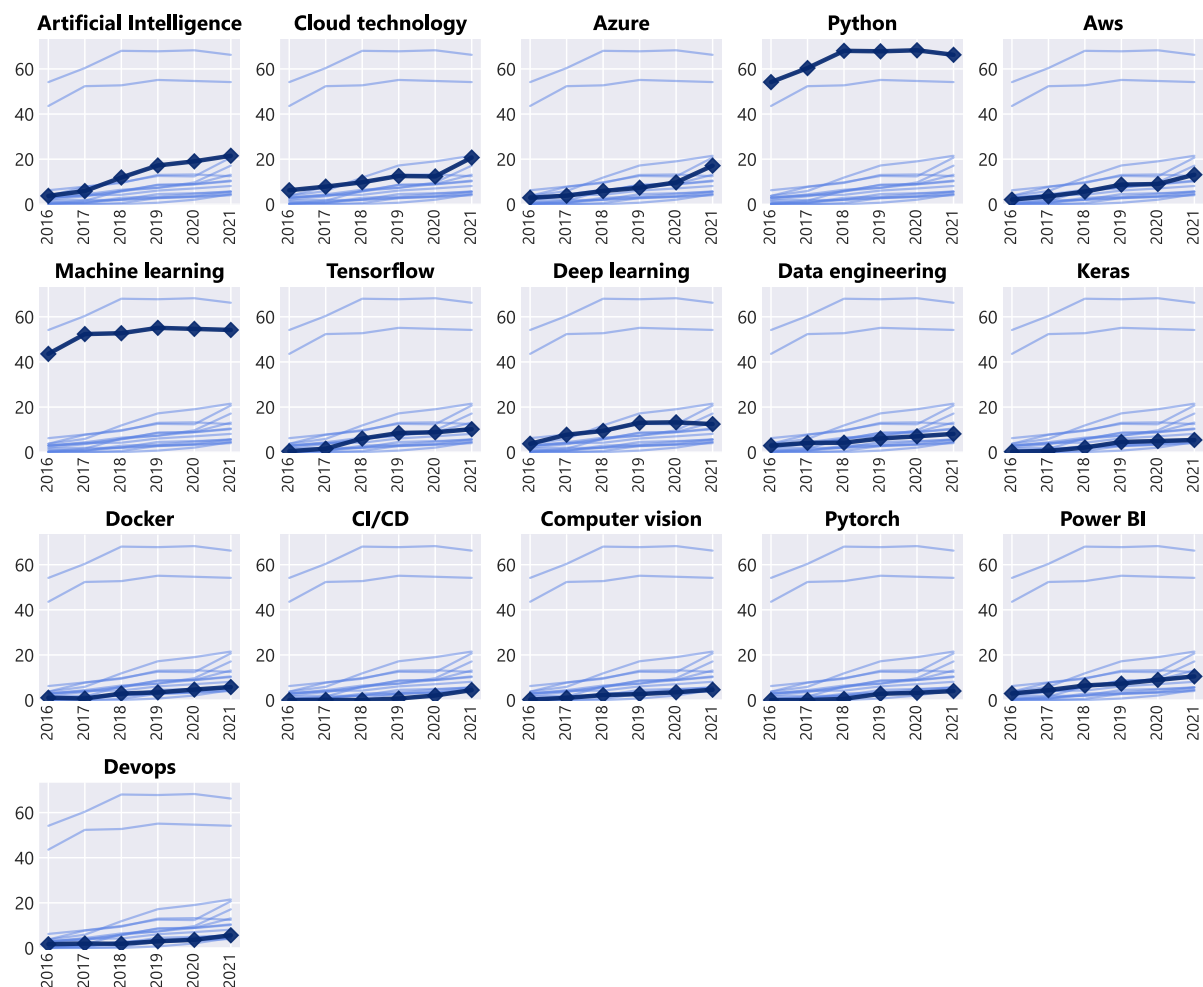
Figuur 4 Ontwikkeling in vraag naar digitale vaardigheden voor data scientist tussen 2012 en 2021 (basisjaar 2012=100), per categorie



In de losse vaardigheden is vooral opvallende dat de term artificial intelligence is gestegen van 3 naar 21 op de 100 vacatures. Ook de vraag naar cloud technologieën (AWS, Microsoft Azure) is flink gestegen. De vraag naar programmeertalen, bijvoorbeeld Python, is gestegen van 55 naar 67 op de 100 vacatures. Ook de vraag naar machine learning en deep learning is gestegen. Dit is ook te zien in de vraag naar deep learning frameworks als Tensorflow, Keras en Pytorch.



Figuur 5 Stijging in vraag naar digitale vaardigheden voor data scientist





5 Transitie

In dit hoofdstuk brengen we mogelijke transitie naar data scientist in kaart, zie het hoofdrapport hoe deze analyse is uitgevoerd. Er zijn 76 beroepen die een haalbare overstap naar dit beroep hebben. Dit zijn de beroepen die een gelijkenisscore van 0,65 of meer hebben in vergelijking met data scientist. Niet alle beroepen komen in aanmerking als startberoep. In de onderstaande paragrafen wordt besproken welke beroepen ook wenselijk als startberoep.

5.1 Startberoepen

Van de 76 haalbare beroepen zijn er 36 beroepen ook wenselijk als startberoep. Een wenselijk startberoep betekent in dit geval een gelijk of lager gemiddeld bruto uurloon en geen (grote) tekorten verwacht in 2026.

Er zijn een aantal beroepsgroepen waar de haalbare en wenselijke startberoepen toe behoren, te vinden in Tabel 5. Van de haalbare startberoepen zijn beroepen als verzekeringsagenten en inkopers niet wenselijk als startberoep omdat ze een hoger bruto uurloon hebben. De meest haalbare én wenselijke startberoepen vallen in de beroepsgroep boekhouders.

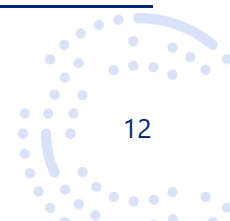
Tabel 5 Aantal haalbare en wenselijke startberoepen per beroepsgroep naar data scientist

Beroepsgroep	Aantal startberoepen	Maximale gelijkenisscore	Minimale gelijkenisscore
Boekhouders	11	0,72	0,65
Secretaresses	5	0,69	0,66
Gebruikersondersteuning ICT	6	0,68	0,65
Directiesecretaresses	5	0,68	0,65
Boekhoudkundig medewerkers	4	0,68	0,65
Productcontroleurs	1	0,67	0,67
Sociaal werkers, groeps- en woonbegeleiders	4	0,67	0,65

Tabel 6 toont de top 15 meest haalbare en wenselijke startberoepen naar data scientist.

Tabel 6 Top 15 haalbare en wenselijke startberoepen naar data scientist

Startberoep	Beroepsgroep	Gelijkenisscore	Gem. bruto uurloon
Analist/researcher	Boekhouders	0,72	€ 24,8
Medewerker planning en controle	Boekhouders	0,69	€ 24,8
Data-entry-medewerker	Secretaresses	0,69	€ 19,6
Medewerker kredietanalyse	Boekhouders	0,68	€ 24,8
IT specialist	Gebruikersondersteuning ICT	0,68	€ 22,0
Kredietanalist	Boekhouders	0,68	€ 24,8
Administratief projectmedewerker	Directiesecretaresses	0,68	€ 22,8





Medewerker pensioenen	Boekhoudkundig medewerkers	0,68	€ 20,5
Webmaster	Gebruikersondersteuning ICT	0,68	€ 22,0
Investment manager	Boekhouders	0,67	€ 24,8
Subsidiecoördinator	Boekhouders	0,67	€ 24,8
Quality Assurance Engineer	Productcontroleurs	0,67	€ 19,3
Office manager	Directiesecretaresses	0,67	€ 22,8
Actuarieel medewerker	Boekhouders	0,67	€ 24,8
Zorgadviseur	Sociaal werkers, groeps- en woonbegeleiders	0,67	€ 20,4

5.2 Skills gap

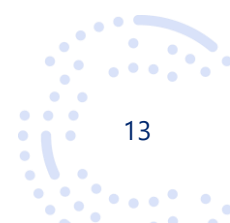
Het is ook mogelijk om van een transitie tussen beroepen de skills gap te bekijken. Als voorbeeld hiervan bekijken we de overgang van medewerker kredietanalyse naar data scientist. Het belang van een skill wordt uitgedrukt op een schaal van 0 tot 100. Ook een niveau van een skill wordt uitgedrukt op een schaal van 0 tot 100. We spreken van een skills gap als de skill belangrijk genoeg is in het doelberoep, een minimaal belang van 30 en als het gevraagde niveau van het startberoep minimaal 20 onder het gevraagde niveau van het doelberoep ligt.

In Tabel 7 wordt de grootste gap in kennis getoond, gesorteerd op hoe belangrijk dit is voor data scientist. Vooral kennis in wiskunde en wetenschap, engineering en technologie en onderwijs en training wordt bij de data scientist op een hoger niveau gevraagd dan voor de medewerker kredietanalyse.

Tabel 7 Skills gap in kennis van medewerker kredietanalyse naar data scientist

Categorie	Kennis	Belang doelberoep	Niveau startberoep	Niveau doelberoep	Gap
Engineering en technologie	Computers en elektronica	66	43	68	25
Wiskunde en wetenschap	Wiskunde	91	68	89	21
Wiskunde en wetenschap	Biologie	31	1	38	36
Onderwijs en training	Onderwijs en training	48	29	59	30

In vaardigheden is er op verschillende skills sprake van een skills gap, zie Tabel 8. Voor wetenschap, programmeren en wiskunde wordt een hoger niveau gevraagd dan voor de medewerker kredietanalyse.





Tabel 8 skills gap in vaardigheden van medewerker kredietanalyse naar data scientist

Categorie	Vaardigheden	Belang doelberoep	Niveau startberoep	Niveau doelberoep	Gap
Inhoud	Wiskunde	94	59	82	23
Inhoud	Wetenschap	63	5	57	52
Technische vaardigheden	Programmeren	50	14	52	37

In capaciteiten zoals visuele en verbale vaardigheden is er geen sprake van skills gap van medewerker kredietanalyse naar data scientist.

In de werkactiviteiten wordt voor een data scientist op een aantal punten een hoger niveau gevraagd dan voor medewerker kredietanalyse, bij het communiceren, anderen adviseren, creatief denken, interactie met computers en identificeren van informatie.

Tabel 9 skills gap in activiteiten van medewerker kredietanalyse naar data scientist

Categorie	Activiteiten	Belang doelberoep	Niveau startberoep	Niveau doelberoep	Gap
Informatie identificeren en evalueren	Het schatten van de kwantificeerbare kenmerken van producten, evenementen of informatie	65	34	60	26
Redenering en besluitvorming	Creatief denken	71	47	75	28
Complexe en technische activiteiten	Interactie met computers	94	45	71	26
Communiceren en interactie	De betekenis van informatie voor anderen interpreteren	86	53	86	34
Communiceren en interactie	Communiceren met personen van buiten de organisatie	68	44	76	32
Coördineren, ontwikkelen, beheren en adviseren	Geef overleg en advies aan anderen	71	44	76	32

De grootste verschillen in context- en omgevingsfactoren zijn dat de frequentie van besluitvorming bij een data scientist minder vaak is dan bij een medewerker kredietanalyse. Evenzo is het herhalen van dezelfde taken minder belangrijk bij een data scientist.





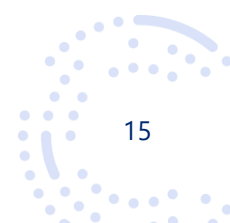
A Categorieën vaardigheden

Tabel 10 Categorieën voor algemene vaardigheden

Categorie	Voorbeelden
Critical thinking	Kritisch denken, beredeneren, detailgericht
Creativity	Innovatief, creatief, origineel
Collaboration	Samenwerken, teamspeler
Communication	Communicatie, schrijven, spreken, presenteren
Computational thinking	Wiskunde, analytisch, statistiek
Flexibility	Flexibel, aanpassen
Leadership	Coördineren, delegeren, overtuigen, onderhandelen
Self-starter	Initiatief, proactief, zelfstandig
Result-orientation	Resultaatgericht, doelgericht
Problem solving	Oplossingsgericht
Active learning	Actief leren
Planning and organization	Planning, organisatie, agile
Service-orientation	Klantgericht
Quality Management	Kwaliteit, kwaliteitsborging

Tabel 11 Categorieën voor digitale vaardigheden

Categorie	Voorbeelden
Specialized software	File versioning software, industrial control software, medical software, map creation software and compliance software like: Blackboard, Git, SVN, arcgis, SCADA
Computer-aided design	Computer-aided design and manufacturing (CAD/CAM) software, computer based training software and pattern design software, like: Catia, CAD, E-plan, Cadence, Autocad, civil 3d
Resource management software	Inventory management software, customer relationship management (CRM) software, materials requirements planning logistics and supply chain software like: SAP, ERP, Primavera
Basic computer skills	Word processing, presentation and spreadsheet software, internet browser software, electronic mail software, operating system software and backup or archival software like: Microsoft Office, Windows operating system, Solaris, Unix, TextPad, Ubuntu
Database management, design and query	Query and processing language, database user interface and query software, object oriented data base management software, metadata management software and database reporting software like: SQL, MYSQL, datawarehouse, netezza, database management, RDBMS, NoSQL
Big data and analytics	Data analytics, natural language processing, parallel computing, machine learning, artificial intelligence, business intelligence and data analysis software like: Hadoop, Spark, Hive, Pig, Tableau, Rapidminer,





	Logistic regression, Support vector machine, K means, Text analytics
Programming skills	Object or component oriented development software, development environment, program testing software, compiling software like: C++, C#, Perl, Java, Lisp, prolog++, Julia, Python
Internet technology and networking	Network monitoring software, network security and virtual private network, application server software VPN, internet protocol IP multimedia subsystem software and equipment software like: LAN, WAN, DNS, webserver, traceroute, weblogic
Web platform development software	Bv: HTML, javascript, django, angularjs, php, css, drupal, joomla, Typescript, requirejs, dhtml, Ruby on Rails, ngrx
Project management software	Bv.: Devops, Content workflow software, Microsoft Project
IT governance and management	Bv.: Information management, IT governance, IT infrastructure, data governance, IT frameworks, Information systems coordination
Digital marketing	Bv.: Adsense tracker, Digital marketing
Digital transformation skills	3D printing, Artificial intelligence, Blockchain, Cloud computing, Cybersecurity, Docker, internet of things, Robotics

