

Einführung

Im Zeitalter der datengestützten Entscheidungsfindung hat sich der Bereich Datenwissenschaft (Data Science) zu einem wichtigen Katalysator für Innovation und Fortschritt sowohl in der Industrie als auch in der Wissenschaft entwickelt. Die Rollen und Aufgaben von Data Scientists haben sich jedoch erheblich weiterentwickelt und umfassen ein breites Spektrum an Fähigkeiten, Fachwissen und Anwendungen. Um die Vielschichtigkeit dieser Rollen zu erfassen und unser kollektives Verständnis zu vermitteln, traf sich eine Gruppe von sechs Fachleuten bei Silicon Saxony und nutzte “Personas” als Methode, um unsere derzeitigen Ansichten über die Rolle von Datenwissenschaftler:innen zu formulieren.

Personas, wie sie in diesem Whitepaper verwendet werden, stellen fiktive Charaktere dar, die verschiedene Rollen und Perspektiven im Bereich der Datenwissenschaft verkörpern. Jede Persona wurde ausgearbeitet sowie im Diskurs validiert, um das Wesen einer bestimmten Facette der Datenwissenschaft zu erfassen und die erforderlichen technischen Fähigkeiten, das Fachwissen und die realen Herausforderungen zu vereinen, denen Datenwissenschaftler:innen bei ihrer täglichen Arbeit begegnen. Durch die Verwendung von Personas wollen wir ein ganzheitliches Bild der verschiedenen Anwendungen und Auswirkungen von Data Scientists vermitteln und die Lücke zwischen theoretischem Wissen und praktischer Umsetzung schließen.

Durch den Einsatz von Personas wollten wir die komplizierten Nuancen von Data-Science-Rollen in Gegenwart und Zukunft ergründen und die Breite sowie Tiefe, der in diesem dynamischen Bereich erforderlichen Fähigkeiten aufzeigen.

Dieses Papier fasst die Erkenntnisse dieser Aktivität zusammen. Es ist in zwei Teile gegliedert. Zuerst werden die Personas aus der Perspektive unseres heutigen Verständnisses analysiert. Danach konzentrieren wir uns auf einen Blick in die Zukunft dieser Personas. Der Appendix beinhaltet die Personas als Rohmaterial. Es muss an dieser Stelle unterstrichen werden, dass unsere Methode keinen Anspruch auf eine empirisch vollständige Beschreibung der Datenwissenschaftler:innen in Dresden oder Sachsen geben kann.

Personas Heute

Im Folgenden werden Unterschiede und Gemeinsamkeiten der Personas herausgearbeitet, die unsere Sicht auf die Rollen einer **Datenwissenschaftler:in im heutigen Berufsbild** widerspiegeln. Die Informationen sind aus sechs Personas (wie im Appendix wiedergegeben) extrahiert und um Erinnerungen aus den Gesprächen erweitert.

Im Workshop wurde schnell deutlich, dass eine klare Abgrenzung der Rolle einer Datenwissenschaftler:in von der Rolle einer Software-Entwickler:in oder -Ingenieur:in noch sehr schwer ist. Häufig werden Software-Ingenieur:innen mit neuen Aufgaben betraut, die sehr stark auf Datenanalyse ausgerichtet sind. Umgekehrt gibt es immer noch viele Data Scientists mit einem hohen Anteil an Aufgaben, die mit Softwareentwicklung zu tun haben. Diesen

Sachverhalt führen wir darauf zurück, dass vor allem stark digitalisierte Organisationen ihr Aufgabenfeld natürlich auf Daten erweitern und dies mit einer Veränderung der Rollenstruktur in Richtung Datenwissenschaft einhergeht.

Gemeinsamkeiten

In der Analyse der sieben erarbeiteten Personas konnten folgende Gemeinsamkeiten erkannt werden. Oftmals sind die beschriebenen Personen **jung** (Studienabsolventen oder Berufseinsteiger, circa 25 bis 35 Jahre). Wir führen dies darauf zurück, dass die Datenwissenschaft ein noch junges Aufgabenfeld ist. Häufig wurde bezüglich des Ausbildungshintergrundes ein **MINT-Studium** angeführt. Dies überrascht wenig, da es sich hier um Fachgebiete mit einem traditionellen Hang zur Digitalisierung handelt, welche in den vergangenen Jahren in der wissenschaftlichen Ausrichtung immer datengetriebener geworden sind.

Des Weiteren werden die Personas als **teamfähig, agil und kommunikationsstark** beschrieben. Wir führen dies darauf zurück, dass Datenwissenschaft oft eine fachübergreifende Aktivität ist und daher nur durch gute Kommunikation die Bedeutung der Daten effektiv herausgearbeitet werden kann. Zusätzlich spiegelt vor allem die Beschreibung der Teamfähigkeit die häufig anzutreffende Spaltung zwischen Datengeber:in und Datenanalyst:in wider. Es sind also immer mindestens zwei Personen involviert, häufig mehr. Ein weiteres Motiv ist eine inhärente **Neugier**, welche den Personas zugeschrieben wird. Da die Methoden der Datenwissenschaft als horizontales Werkzeug in vielen Bereichen einsetzbar sind, erschließt sich diese Eigenschaft schnell. Ein:e Datenwissenschaftler:in ohne Scheu vor neuen Fachgebieten wird die anwendungsspezifischen Anforderungen, Randbedingungen und Kundenerwartungen leichter erschließen und ihre Lösungen darauf ausrichten können. Im Gegensatz dazu sehen wir bei Lösungen von Data Scientists, deren Interesse vorrangig der Datenwissenschaft gilt, weniger Kundenorientierung und folglich ein hohes Risiko, dass die Lösungen später von den Anwendern nicht akzeptiert werden.

Data Scientists wird zudem die Fähigkeit zugeschrieben, Software gut **automatisieren sowie parallelisieren** zu können. Wir führen dies auf den Sachverhalt zurück, dass es in der Datenwissenschaft häufig Aufgaben gibt, welche über die verschiedenen Datensätze hinweg unabhängig voneinander bearbeitet werden können. Bzgl. der Software-Umgebung wurden die Sprache "**Python**" sowie die Python-Bibliotheken **Pytorch, Pandas usw.** genannt, die zur Zeit des Workshops sehr beliebt waren. Fachlich wurden weiterhin häufig ein **gutes Verständnis von Mathematik oder Statistik** angegeben. Da gegenwärtig das Paradigma des probabilistischen Lernens vorherrscht, ist dies nur folgerichtig. Schließlich wurde aber auch die Fähigkeit des **wissenschaftlichen Arbeitens** benannt. Dies rührt aus dem Fakt, dass es sehr viele verschiedene Analysemethoden gibt, welche für neue Datensätze häufig neu evaluiert und verglichen werden müssen.

Unterschiede

Im Folgenden sollen Unterschiede in den Personas hervorgehoben werden. Die Einordnung dieser Unterschiede bleibt jedoch schwer, da wir nur sieben Personas erarbeiten konnten.

Häufig wurden Data Scientists als **Quereinsteigende** beschrieben. Durch die hohe Bandbreite von vertretenen Organisationen beim Workshop, insbesondere bzgl. der Organisationsgröße, kann dies möglicherweise auf die Arbeitsmarktsituation zurückgeführt werden. Abschlüsse im Bereich Datenwissenschaft sind vermutlich noch nicht sehr weit verbreitet. In der Diskussion ist auch eine Aufspaltung zwischen Personen mit **Methodenfokus und/oder Generalist:innen** benannt worden. Dies kann wieder auf den Sachverhalt zurückgeführt werden, dass Daten überall sind und es den einzelnen Persönlichkeiten sowie Umgebungsbedingungen obliegt, diese Daten entweder eher tiefgehend oder allgemein zu bearbeiten. In diesem Zusammenhang wurde die Unterscheidung **Domain Expert und/oder Method Specialists** beleuchtet. Diese Differenzierung spiegelt erneut die Einteilung in Datengeber:in und Datenanalyst:in wider und zeigt, dass auch Personen ohne direkten Kontakt zu Algorithmen das Vokabular der Analyse beherrschen oder diesem gegenüber zumindest offen eingestellt sein sollten. In diesem Zusammenhang wurde darauf hingewiesen, dass eine Analyse bzw. **Einsicht der Daten auch ohne Quellcode** wichtig ist, um vor allem im Dialog mit Data Scientists die Bedeutung der Daten gemeinsam zu erfahren. Dies wurde im Kontrast zu vielen Quereinsteigenden angesprochen, die zu fokussiert auf die Datenanalyse mit Software sind und dadurch möglicherweise (unbewusst) kommunikative Barrieren aufbauen. Mitunter wurde die berufliche **Erfahrung mit Datenbanken** positiv beleuchtet, wobei dies im Gegensatz zu großvolumigen Daten gesehen wurde, welche selten in Datenbanken verfügbar sind. Schließlich wurde zum Teil auch die **Fähigkeit, KI-Methoden zu hinterfragen** und deren Ergebnisse kritisch zu untersuchen, angesprochen. Hierin spiegelt sich zwar erneut die Fähigkeit der Personas zum wissenschaftlichen Arbeiten wider. Gleichwohl wurde damit auch eine generelle Skepsis gegenüber KI-Methoden teilweise zum Ausdruck gebracht.

Personas Morgen

Im Folgenden werden Unterschiede und Gemeinsamkeiten der Personas herausgearbeitet, welche unseren Blick auf die Rollen einer **Datenwissenschaftler:in im zukünftigen Berufsfeld** widerspiegeln. Die Informationen sind aus sechs Personas (wie im Appendix wiedergegeben) extrahiert und um Erinnerungen aus den Gesprächen erweitert.

Im Workshop ist klar geworden, dass eine Klassifizierung der/s (typischen) „Data Scientist“ sehr differenziert zu betrachten ist. Wir gehen davon aus, dass sich aufgrund der zukünftigen Entwicklungen am Arbeitsmarkt und den Anforderungen im Berufsfeld eine Zweiteilung herausbilden wird: Zum einen Spezialist:innen für ganz bestimmte Aufgaben im Entwicklungsprozess und zum anderen Allrounder, die ein breites Spektrum von Fähigkeiten abbilden, das jedoch mit Abstrichen in der tieferen Expertise einhergeht. Die Wahrscheinlichkeit ist groß, dass sich vielfältige Rollen von Data Scientists

herauskristallisieren werden. Es wird daher besonders darauf ankommen, dass diese unterschiedlichen Menschen gut zusammenarbeiten können.

Gemeinsamkeiten

Was die meisten Data Scientists mitbringen, ist ihre **Begeisterung und Neugier**, Lösungen und Antworten auf die Herausforderungen der zukünftigen Gesellschaft, Wirtschaft und Umwelt zu finden. Unabhängig davon, ob es sich um eine:n Spezialist:in oder Allrounder:in handelt, ist ein **grundlegendes Verständnis bzgl. Datenumgang, -qualität und -struktur** erforderlich, um brauchbare Ergebnisse für die zu entwickelnden Anwendungen zu erzielen. Ein wichtiger Punkt ist zudem die **(selbst-)kritische und verantwortungsbewusste Auseinandersetzung im Umgang mit Datenmaterial** an sich. Oft spiegeln diese relevante Geschäftsprozesse wider oder es handelt sich um sensible personenbezogene Daten.

Die zukünftigen Data Scientists sollten über ein **gutes Überblickswissen zu aktuellen und bewährten Methoden** verfügen. Oft wird der **Blick über den sprichwörtlichen Tellerrand ein entscheidender Erfolgsfaktor** sein. Beim Thema Künstliche Intelligenz stehen wir noch am Anfang einer Entwicklung, daher ist ein starrer Methodenkoffer nicht hilfreich. Letztlich geht es darum, Erfahrungen zu sammeln und eine stetige Auseinandersetzung mit dem aktuellen Stand der Erkenntnisse.

Gerade in der Umsetzung von Projekten wird es auf **Teamwork** ankommen, dies setzt Vertrauen voraus, ebenso die Bereitschaft, mit anderen in Dialog treten zu wollen und die Fähigkeit, **Feedback zu empfangen und auch zu geben**. Hier ist die Fähigkeit zu **guter Kommunikation** gefragt, die bereichsübergreifend, verständlich und respektvoll sein muss. **Selbständiges Arbeiten** wird nach wie vor notwendig sein, um gute und belastbare Ergebnisse zu produzieren.

Unterschiede

Es ist heute noch nicht absehbar, in welchen Bereichen von Gesellschaft, Wirtschaft und Umwelt Datenwissenschaftler:innen eine Rolle einnehmen werden. Daher wird es wahrscheinlich vielfältige und kontextabhängige Konstellationen geben, in denen **Data Scientists mit unterschiedlichen Kompetenzprofilen** wirken werden. Trotzdem wird es in vielen Bereichen erfolgsentscheidend sein, dass eine **stärkere Spezialisierung in den Kompetenzprofilen** entsteht und diese Spezialist:innen gut zusammenarbeiten. Dies setzt aber voraus, dass unser Bildungssystem eine Antwort auf diese spezifischen Anforderungen findet. Die heutige Realität bringt (noch) viele Quereinsteigende hervor, die in die Rolle der/s Data Scientist hineinwachsen.

Außerdem erkennen wir neben den Fähigkeitsprofilen auch eine **Auffächerung der Spezialisierung bezüglich domänenspezifischen Wissens und auf der anderen Seite zu domänenübergreifendem Wissen**. Jedes Einsatzfeld erfordert von

Datenwissenschaftler:innen ein Verständnis des speziellen Hintergrundwissens in dieser Domäne. Es wird kaum möglich sein, dass zukünftige Data Scientists sich in allen Domänen auskennen. Daher wird es notwendig werden, dass diese sich auf bestimmte Einsatzfelder spezialisieren. Den/die Allrounder:in in diesem Bereich wird es trotzdem geben und diese:r hat die Herausforderung, sich vorab auf die Gegebenheiten des jeweiligen Kontextes einstellen zu müssen.

Selbst bei Softskills wird sich eine tiefere Spezialisierung herauskristallisieren. Zum einen die **Kommunikation im Projektteam** untereinander und zum anderen die **Kommunikation mit den Anwendern und Stakeholdern**. Bei aller fachlichen Expertise werden in zukünftig international zusammenwirkenden Teams gerade die überfachlichen Kompetenzen eine erfolgsentscheidende Rolle einnehmen. Einzelakteure werden durch co-kreative und -intelligente Teams ersetzt. Dies setzt im Team als Ganzes betrachtet die Fähigkeiten sowie den Willen zur Kooperation, Veränderung, Verbindlichkeit, Neugier, Empathie, Grit (Ausdauer und Leidenschaft für langfristige Ziele), Selbstführung, (Kritik)-Resilienz und Wettbewerbsorientierung voraus.

Wir dürfen uns vom Gedanken verabschieden, perfekte Data Scientists zu fordern und sollten stattdessen anfangen, Datenwissenschaftler:innen mit spezifischen fachlichen und überfachlichen Stärken und Schwächen zu sehen und diese in kontextbasiert zusammengestellten Teams kooperieren zu lassen.

Zusammenfassung

Der Workshop hat sechs Stimmen aus dem Feld der Datenwissenschaft zusammengebracht. Es waren Vertreter aus Industrie, Wirtschaft, Wissenschaft und Verwaltung anwesend.

Data Scientists in der Gegenwart wurden zum Zeitpunkt des Workshops als jung, neugierig, teamfähig, agil und kommunikationsstark beschrieben. Sie kommen häufig mit einem abgeschlossenen MINT-Studium auf den Arbeitsmarkt. Sie beherrschen das wissenschaftliche Arbeiten und verfügen über eine starke Kompetenz im Bereich Mathematik und Statistik. Des Weiteren meistern sie Automatisierung und Parallelisierung von Aufgaben. Sie benutzen häufig die Programmiersprache Python und populäre Data Science Bibliotheken wie Pytorch, Pandas, etc. darin. Generell wurde in der Diskussion klar, dass es ein hohes Potential für Quereinsteigende in diesem Bereich gibt, da immer mehr Fachdomänen einen datengetriebenen Charakter erhalten.

Datenwissenschaftler:innen in der Zukunft wurden zum Zeitpunkt des Workshops als Menschen mit Begeisterung und Neugier beschrieben, die Lösungen und Antworten auf die Herausforderungen der zukünftigen Gesellschaft, Wirtschaft und Umwelt finden. Sie haben ein grundlegendes Verständnis bzgl. Datenumgang, -qualität und -struktur. Sie gehen (selbst-)kritisch und verantwortungsbewusst mit Datenmaterial und Vorhersagen von KI-Vorhersagen um. Sie haben gutes Überblickswissen zu aktuellen und bewährten Methoden, scheuen sich aber nicht, den Blick über den Tellerrand zu wagen. Dies wird ein entscheidender Erfolgsfaktor

sein. Außerdem zeigen sie Teamwork und können (kritisches) Feedback geben sowie empfangen. Es werden also Fähigkeiten zur guten und effektiven Kommunikation benötigt. Schließlich wird von diesen Data Scientists ein hohes Maß an selbstständigem Arbeiten erwartet.

Appendix

Der Appendix umfasst die Rohdaten des Workshops. Im Folgenden sind also die Personas verschriftlicht worden und diesem Dokument der Vollständigkeit halber beigelegt.

Verschriftlichtung der Personas für Data Scientists heute

Persona Sahil

- in Indien Computer Science studiert
- Ausbildungsschwerpunkt: KI
- nach Berufsstart mit wenig KI-Inhalt, dann Data-Science-Stelle angenommen

- Soft skills:
 - fleißig ***
 - zielorientiert **
 - techn. begabt ***
 - kunden-/problemorientiert 0 (wenig bis gar nicht)

- Hard skills:
 - ML/DL-Methoden ***
 - Computer Science ***
 - Statistik *
 - Domain Knowledge 0 (wenig bis gar nicht)

Persona Tina

- Bio-Master, 3 Jahre Berufserfahrung nach Master in der Bioanalytik
- Aktueller Job / Task: Datenanalyse einer Werkstoffdatenbank (Bilder und Tabellen) -> anderes Fachgebiet

- Soft skills:
 - Neugier
 - Struktur
 - Kritisch
 - Wertschätzend

- Hard skills:
 - Python for data: MONAI, Pytorch, Pandas, skimage, napari
 - testing frameworks
 - visualisation
 - git, CI/CD
 - parallelisation and HPC

Persona John

- hat ein Studium im MINT-Bereich absolviert
- nach Uni: eine Tätigkeit in der IT-Beratung aufgenommen

- dort Tool-Kompetenz und Arbeitsmethodik gelernt
- Fokus im Studium war hauptsächlich Aufbau von Fachexpertise
 - Kam oft ins Straucheln bei Themen wie:
 - Führung
 - Menschenkenntnis
 - Kommunikation
 - Teamwork
- Der praktische Bezug und die Anwendung von Wissen fehlt ihm im Studium.

- Soft skills:
 - intrinsisch motiviert, oft zu ungeduldig
 - beherrscht Englisch
 - Fokus liegt auf Work-Life-Balance

- Hard skills:
 - verfügt über Kenntnisse in objektorientierter Programmierung (Java, C#, ...)
 - hat sich im Studium umfangreiche Kenntnisse zur Statistik angeeignet
 - verfügt über Wissen aus dem Bereich Operations Research

Persona Tina L.

- MINT-Studium
- 30 Jahre
- 1-2 Berufsrollen im CV

- Soft skills:
 - agile Methoden
 - Teamplayer
 - kommunikative Kompetenz
 - Resilienz
 - Neugierde
 - Offenheit

- Hard skills:
 - Statistik
 - Mathe
 - Datenmanagement

Persona Nick

- Absolvent Elektrotechnik

- Soft skills:
 - fachlich eigenmotiviert
 - guter "Early Adopter"

- Hard skills:

- python, pandas, pytorch
- automation

Persona Dörte

- Domain Data Scientist
- Quereinsteigerin aus Ing-Domäne, Statistik (Maschinenbau, Elektrotechnik,...)

- Hard skills:
 - Domänenwissen in Theorie und Praxis -> viele Jahre
 - keine bis wenig Code-Erfahrung
 - Excel- und Textfiles

- Soft skills:
 - Prozessenerfahrung außerhalb von IT-Projekten
 - "gefährliches Halbwissen"

Persona Gerd

- General Data Scientist
- Informatik-Studium (mgl. Promotion)

- Hard skills:
 - Data Engineering + Statistik + Domäne
 - Python, R, Logik
 - RDBMS/SQL, Data Management, NoSQL

- Soft Skills:
 - Interesse, Neugier, Fähigkeit zum "Reindenken"
 - wissenschaftliches Arbeiten, agile Projektarbeit

Verschriftlichung der Personas für Data Scientists morgen

Persona Klaus, Projektleiter

- ML-Ops
 - komplette “life-cycle-expertise”
 - Technology Readiness Level 9
 - Interesse an Funktion sozialer Systeme und Prozesse

Persona Maria, Data-Scientist

- Hat Data-Science studiert
- In jungen Jahren bereits Kenntnisse im Bereich Informatik erworben
- Weiß mit Menschen umzugehen (soziale Kompetenz)
- Kann domainübergreifend denken (interdisziplinär)
- Versucht Perspektiven der Mitarbeitenden zu übernehmen (Empathie)
- **Soft skills**
 - Menschenkenntnis
 - Selbstreflektion
 - Spagat Soft- vs. Hard-Skills
 - Intrinsische Motivation
- **Hard skills**
 - Objektorientierte Programmierung (OOP)
 - Methoden

Persona Anna Maria

- 30 Jahre
- Hat Soziologie mit Nebenfach Data-Science studiert
- War während des Studiums im Ausland (Fremdsprache und interkulturelle Kompetenz)
- **Hard skills**
 - Mathematik, Statistik
 - Sozialwissenschaftliche Expertise
 - Change-Management
- **Soft skills**
 - Kommunikative Kompetenz
 - Neugierde
 - Agile Methoden
 - Kritischer Umgang mit Daten

Persona Jean, Physik-Master

- Master in Physik
- PhD oder 2 bis 3 Jahre Berufserfahrung in Industrie
- Task: Sprach- + Gestensteuerung einer Sim./Dil. Twin am Experiment
- **Soft skills**
 - Neugier
 - Struktur

- Team-Skills
- Wertschätzung
- Selbstvertrauen -> Community
- **Hard skills**
 - Code-tools in mindestens 2 Modalitäten
 - Basics in Statistik
 - Repro. Standards

Persona Tina

- Hat ein MINT-Fach studiert
- 5 Jahre Berufserfahrung (digitale Transformation)
- Eingebettet in Fachabteilung
- **Soft skills**
 - Interdisziplinär
 - Kommunikativ, kann begeistern
 - Kundenorientiert
 - Neugierig
- **Hard skills**
 - Statistik
 - ML/DC
 - MLOps
 - Business background

Verbund/Team D.S.G.O.

Dörte 2.0

- Domain Data Scientist
- Quereinsteigerin aus Ing-Domäne
- **Hard skills:**
 - Domänenwissen in Theorie und Praxis -> viele Jahre
 - Kennt Lösungen und Anwendungsfälle
 - Grundlegende IT-Kompetenz → Kann mit IT Expert:innen reden
- **Soft skills:**
 - Kann vertraut mit Stakeholder/Anwender kommunizieren → “einer von uns”

Persona Sylvia

- Integrator
- Unterstützt bzw. gestaltet Einführung der Anwendung/Lösung
- Baut Ängste und Vorurteile ab
- **Hard skills:**
 - HR, Organisation, Psychologie
 - Change Management, Partizipative Methoden
- **Soft skills:**
 - Kommunikation mit anderen Teamrollen

- Kommunikation mit verschiedenen Mitarbeitenden in der Organisation des Stakeholders

Persona Gerd 2.0

- Data Scientist
- Informatik-Hintergrund (Studium, evtl. Promotion)
- **Hard skills:**
 - Statistik, Modelle, Daten(vor)verarbeitung
 - SQL/RDBMS, Datenmanagement
- **Soft skills:**
 - Kommunikation mit anderen Teamrollen
 - Erklären und Dokumentieren der Lösung → Nachvollziehbarkeit

Persona Olaf

- MLOPS
- Überführt Modelle in Anwendung und rollt diese aus
- Informatik-Hintergrund
- **Hard skills:**
 - Data und Software Engineering
 - Docker, Microservices
 - Modellevaluation und -wartung
- **Soft skills:**
 - Kommunikation mit anderen Teamrollen