



## **Können Online-Stellenanzeigen helfen, zukünftig relevante Kompetenzen zu identifizieren?**

Machbarkeitsstudie zu Vorhersagen von  
Kompetenzentwicklungen

© Bertelsmann Stiftung, Gütersloh

Oktober 2023

#### **Herausgeber**

Bertelsmann Stiftung

Carl-Bertelsmann-Straße 256, 33311 Gütersloh

[www.bertelsmann-stiftung.de](http://www.bertelsmann-stiftung.de)

#### **Verantwortlich**

Larissa Klemme

Project Manager

Nachhaltige Soziale Marktwirtschaft

Gunvald Herdin

Senior Project Manager

Nachhaltige Soziale Marktwirtschaft

#### **Autor:innen**

Dr. Fabian Stephany

Eric Sobolewski

#### **Datenmitarbeit**

Johannes Müller und Rahakavee Baskaran, &effect data solutions GmbH

#### **Redaktion**

Dr. Thomas Orthmann

#### **Zitationshinweis**

Stephany, F., Sobolewski, E. (2023). Können Online-Stellenanzeigen helfen, zukünftig relevante Kompetenzen zu identifizieren? Machbarkeitsstudie zu Vorhersagen von Kompetenzentwicklungen. Bertelsmann Stiftung (Hrsg.). Gütersloh

#### **Layout**

Markus Diekmann

#### **Bildnachweis**

© REDPIXEL – stock.adobe.com

**DOI 10.11586/2023060**

ID\_1991

# Können Online-Stellenanzeigen helfen, zukünftig relevante Kompetenzen zu identifizieren?

Machbarkeitsstudie zu Vorhersagen von  
Kompetenzentwicklungen

Fabian Stephany,  
DWG Datenwissenschaftliche Gesellschaft Berlin mbH  
und Eric Sobolewski

# Inhalt

Abstract	5
<b>1. Einleitung</b>	<b>6</b>
<b>2. Datengrundlage und Datenstruktur</b>	<b>8</b>
2.1. Online Job Advertisements als Datenquelle	8
2.2. Stichprobe aus neun Berufsgruppen	8
<b>3. Methodik</b>	<b>11</b>
3.1. Metriken	11
3.2. Zeitreihenanalyse	12
3.3. Beispiele für die Anwendung der Metriken	14
<b>4. Fokus – Die Twin Transition</b>	<b>18</b>
4.1. Was sind Green Skills?	18
4.2. Was sind Coding Skills?	21
<b>5. Vorhersagen durch Zeitreihenanalysen</b>	<b>25</b>
5.1. Zeitreihenanalyse ausgewählter Kompetenzen: Popularität und Relevanz	25
5.2. Informationsgehalt von Zeitreihenanalysen	27
5.3. Zeitreihenanalyse ausgewählter Kompetenzen: Prämie	28
5.4. Zeitreihenanalyse des Kompetenznetzwerks	29
<b>6. Fazit</b>	<b>31</b>
6.1. Vorhersagen sind möglich	31
6.2. Die Datenverfügbarkeit entscheidet über die Vorhersagequalität	31
<b>7. Referenzen</b>	<b>33</b>
Anhang	34
Abbildungsverzeichnis	40

# Abstract

---

Künstliche Intelligenz (KI) verändert die Arbeitswelt radikal. Zeitgleich steht die Wirtschaft vor einer nachhaltigen Erneuerung im Zuge der klimafreundlichen Transformation, die sich auch auf dem Arbeitsmarkt widerspiegelt. Beide Phänomene, oftmals als Zwillings-Transformation ("Twin Transition") bezeichnet, werfen die Frage auf, welche Kompetenzen in der Arbeitswelt von Morgen gefragt sein werden. Diese Frage steht auch im Fokus der vorliegenden Studie und wird durch die Analyse von Online-Stellenanzeigen („Online Job Advertisements“, kurz OJA) beleuchtet. Insbesondere wird erläutert, inwiefern sich aus der zurückliegenden Bewertung von Kompetenzen mittels Zeitreihenanalysen Aussagen über die zukünftigen Entwicklungen treffen lassen. Dabei werden OJA-Daten für mehrere Berufsgruppen aus unterschiedlichen Branchen anhand der Aspekte von Popularität, Relevanz und Preis ausgewertet. Die Ergebnisse werden anhand von einzelnen "green" (klimabezogenen) und "coding" (IT-)

Kompetenzen erläutert. Am Beispiel der Programmiersprache „Python“ lässt sich dies eindeutig illustrieren: Wir beobachten für alle drei Metriken unserer Analyse – Popularität, Relevanz und Preis – eine klare Aufwärtsbewegung. Dieser klar erkennbare Trend lässt sich entsprechend sicher in der Zukunft fortschreiben, wie unsere Zeitreihenanalyse und Cross-Validation bestätigen. Weitergehend wird der wirtschaftliche Wandel hin zu „grünen“ Kompetenzen in unserer Analyse durch Aspekte wie „Bausanierung“ und „Solarthermie“ beleuchtet. Die Ergebnisse dieser Studie zeigen, dass kurzfristige Vorhersagen (von bis zu sechs Monaten) über die zukünftige Entwicklung von Kompetenzen möglich sind. Der Zeitraum und die Qualität dieser Vorhersagen werden mitunter stark durch mangelnde Datenverfügbarkeit limitiert.

**Stichworte:** Zeitreihenanalyse, Kompetenzen, Arbeitsmarktforschung

# 1. Einleitung

---

Digitale Technologien, vor allem KI und intelligente Algorithmen, verändern unsere Arbeitswelt und die Nachfrage nach Kompetenzen nachhaltig und tiefgreifend. In Horrorszenarien des technologischen Wandels sehen manche Studien (Fuei, 2017; Arntz et al, 2016) kaum noch Platz für menschliche Arbeit und jüngste Entwicklungen im Bereich des maschinellen Lernens scheinen diese Perspektive zu untermauern.

Ein aktuelles Beispiel zu dieser Entwicklung liefert ChatGPT. Das inzwischen als globales Phänomen zu bezeichnende textbasierte Dialogsystem versetzt ganze Wirtschaftssektoren in Aufruhr. Ende November 2022 als öffentlich zugänglicher Prototyp gestartet, hat der Dienst im Januar 2023 bereits 100 Millionen<sup>1</sup> Nutzer:innen. Neben Prosa und Gedichten schreibt ChatGPT funktionierende Computercodes. Es sieht auf den ersten Blick so aus, als ob Software wie diese ganze Wertschöpfungsketten über Nacht auf den Kopf stellen und Heerscharen von Menschen in die Arbeitslosigkeit treiben könnte.

Doch wie immer lohnt sich bei komplexen Themen ein zweiter, kritischer Blick. Neuere Erkenntnisse deuten darauf hin, dass nicht die Verdrängung des Menschen durch die Maschine im Vordergrund steht, sondern dass Mensch und Maschine in Zukunft stärker zusammenarbeiten werden. Allerdings nicht mehr so, wie wir es gewohnt sind. Am Beispiel von ChatGPT wird diese Tatsache deutlich. Einer der maßgeblichen Gründe für die enorme Leistungsfähigkeit des Dienstes liegt in der Art und Weise, wie die Software „lernt“. Angewendet wird eine spezielle Form des „bestärkenden Lernens“, die auf menschlicher Interaktion aufbaut (Reinforcement Learning from Human Feedback). Das System erarbeitet auf Grundlage der Trainingsdaten selbständig Inhalte, aber Menschen beurteilen, inwiefern diese tauglich sind für den vorbestimmten Zweck. Die Maschine lernt also, während der Mensch kontrolliert und Feedback gibt, bis die gewünschten Ergebnisse erreicht werden. Ergebnis dieser Mensch-Maschine-Kooperation ist eines der

aktuell fortschrittlichsten, dialogfähigen KI-Systeme der Welt (Tian et al., 2023).

Fähigkeiten, die Menschen in der heutigen Arbeitswelt haben, werden vielleicht bald nicht mehr gebraucht, z. B. das Schreiben generischer Anzeigentexte. Andere Fähigkeiten aber gewinnen an Bedeutung oder entstehen neu, z. B. die Fähigkeit, eine KI zu trainieren (Ha-leem et. al, 2022). Genau aus diesem Ungleichgewicht entsteht der „Skill Gap“; einige Kompetenzen verlieren auf dem Arbeitsmarkt ihre Bedeutung, während sich andere Fähigkeiten zunehmender Nachfrage erfreuen (Bessen, 2014). Durch dieses Ungleichgewicht entfalten sich negative ökonomische und soziale Folgen, die den Startpunkt und maßgeblichen Betrachtungsgegenstand unseres Reports darstellen.

Die Entwicklungen im Bereich der KI und des maschinellen Lernens gehen Hand in Hand mit der dringenden Notwendigkeit, die Wirtschaft angesichts des voranschreitenden Klimawandels auf klimaneutrale Wege zu lenken. Die Herausforderungen, die diesen Transformationsprozess begleiten, sind vielfältig, wie der Fachkräftemangel aufgrund der demografischen Entwicklung, Störungen in den Lieferketten oder mögliche Energiekrisen. Besonders im Fokus stehen der strukturelle Wandel, die verstärkte Integration erneuerbarer Energien und die Umstellung auf eine CO<sub>2</sub>-neutrale Industrie. Dies führt zu einem wachsenden Bedarf an hochqualifizierten Fachkräften und entsprechenden Kompetenzen.

Bei eben diesem tiefgreifenden doppelten Strukturwandel sind drei hauptsächliche Effekte zu erwarten, die sich in den von uns untersuchten Daten nachweisen lassen:

1. Alte Kompetenzen werden verdrängt
2. Neue Kompetenzen werden verlangt
3. Kompetenzen verändern sich in ihrer Beschaffenheit über die Zeit

<sup>1</sup> <https://www.demandsage.com/chatgpt-statistics/>

Wenn man sich diese Effekte anschaut, wird schnell deutlich, dass die bisher verbreitete Methode, nämlich Berufe an sich als Analyseobjekt zu betrachten, nicht granular genug ist, um die sich verändernde Nachfrage nach beruflichen Kompetenzen zu verstehen. Durch die gestiegene Komplexität in der Arbeitswelt bedarf es einer höheren Auflösung, wenn man Veränderungen zuverlässig identifizieren will. Aus diesem Grund ist es sinnvoll, Berufe auf Fähigkeiten-Ebene zu betrachten, also ein Level tiefer zu gehen, als es in den meisten bisherigen Untersuchungen der Fall ist. Für eine quantitative Analyse in größerem Stil war dies bisher deshalb schwierig, weil Daten fehlten. Mithilfe von online generierten Massendaten, nämlich der Extraktion von Kompetenzen aus Online-Stellenanzeigen, ist dies jetzt möglich, wie die hier vorgestellte Analyse illustriert. Während zwar niemand die Zukunft vorhersagen kann, scheint eine Sache klar: Die breite Anwendung künstlicher Intelligenz wird unsere gesamte wirtschaftliche und gesellschaftliche Ordnung mindestens so tiefgreifend verändern wie die großen Erfindungen Dampfmaschine, Auto oder Computer. Unabhängig davon, ob die Mensch-Maschine-Kooperation oder der Wettbewerb zukünftig im Vordergrund stehen, ist eine tiefgreifende strukturelle Veränderung der Arbeitswelt durch den Aufstieg künstlicher Intelligenz sicher. Diese Veränderungen schlagen sich vor allem im Bereich der nachgefragten Fähigkeiten in Jobs nieder.

Daher stellt sich für den folgenden Report die Frage, inwiefern die zurückliegenden Informationen über

Kompetenzen in Online-Stellenanzeigen Aussagen über deren zukünftige Entwicklung erlauben. Dabei gehen wir in der Analyse auf die zentralen Aspekte anhand der Entwicklungen im Bereich IT-Kompetenzen am Beispiel der Programmiersprache „Python“ ein. In Zeiten, die von technologischem Umbruch und sprunghaften Entwicklungen im Bereich der künstlichen Intelligenz bestimmt sind, gilt das Erlernen von Programmiersprachen als „sicherer Hafen“ im Sinne der resilienten Weiterbildung. Die Programmiersprache „Python“ wird aufgrund ihrer vielseitigen Anwendbarkeit oftmals als Paradebeispiel für die Zukunftsfähigkeit des Programmierens bezeichnet. Unsere Analyse bestätigt diese Empfehlung. Wir beobachten für alle drei Metriken unserer Analyse – Popularität, Zentralität und Preis – eine klare Aufwärtsbewegung. Dieser klar erkennbare Trend lässt sich entsprechend sicher in der Zukunft fortschreiben, wie unsere Zeitreihenanalyse und Cross-Validation bestätigen. „Python“ wird auch in Zukunft gefragt, relevant und wertvoll sein. Ebenso weisen wir aber auch auf die Bedeutung von Datenverfügbarkeit hin. Während unsere Analyse aufzeigt, dass sich Kompetenzbewertungen grundsätzlich und in die Zukunft fortschreiben lassen, steht und fällt die Qualität dieser Vorhersagen mit der Dichte an Daten. Diesen Aspekt gilt es bei weiteren Analysen dringend zu berücksichtigen. Denn die vorliegende Analyse von Kompetenzbewertungen mit Online-Stellenanzeigen veranschaulicht deutlich: Mit ausreichenden und qualitativ hochwertigen Daten sind verlässliche Aussagen über die zukünftige Entwicklung möglich.

## 2. Datengrundlage und Datenstruktur

Der analysierte Datensatz besteht aus etwa zwei Millionen Online Job Advertisements (OJAs), die zwischen Januar 2016 und März 2022 gesammelt wurden. Zusätzlich betrachten wir in der abschließenden Zeitreihenanalyse des Reports ein Set an ausgewählten Kompetenzen in einer erweiterten Zeitreihe bis März 2023. Diese wird in den allgemeinen Statistiken im Folgenden nicht abgebildet. Zudem enthält dieser Datensatz ausschließlich OJAs, die aus neun bestimmten Berufsgruppen stammen (siehe Tabelle 1). Zur Betrachtung von Kompetenzen wird die VERBIS-Kompetenztaxonomie der Bundesagentur für Arbeit verwendet, welche 8.256 Kompetenzen umfasst<sup>2</sup>. Im vorliegenden Datensatz sind 7.050 VERBIS-Kompetenzen enthalten.

### 2.1. Online Job Advertisements als Datenquelle

Wie bereits erwähnt, stellt die Untersuchung der Änderung von Jobprofilen durch gewachsene Komplexität in der Arbeitswelt eine Herausforderung dar. Will

man aber verstehen, welche Fähigkeiten Menschen in Zukunft brauchen, um auf dem Arbeitsmarkt weiter wettbewerbsfähig zu bleiben, ist eine granulare Ansicht nötig. Vor diesem Hintergrund eignen sich die OJA-Daten sehr gut als Proxy. Zwar wurden sie nicht explizit erzeugt bzw. gesammelt, um den Forschungsgegenstand „Änderung von Skills über Zeit“ zu ergründen, doch stellen sie durch ihre Herkunft aus der „echten Arbeitswelt“ trotzdem einen hohen Mehrwert dar, denn sie beschreiben die Arbeitsnachfrage von Unternehmen in Echtzeit.

Die Grundlage für die vorliegende Studie bildet ein umfangreicher Datensatz von TextKernel BV<sup>3</sup>. Dieser zugrundeliegende Datensatz enthält mehr als 275 Millionen Stellenausschreibungen, die aus über 50.000 verschiedenen Quellen stammen (siehe Noack et al., 2022). Täglich werden neue Stellenausschreibungen in diesen Datensatz integriert. Um die Daten für spätere Analysen nutzbar zu machen, werden sie gefiltert und ergänzt. Die Erstellung dieses Datensatzes erfolgt in mehreren Schritten. Zunächst werden Duplikate innerhalb der Stellenausschreibungen entfernt. Hier-

TABELLE 1 Einteilung der Berufsgruppen

Berufsgruppe	Klassifikationsnummern (KIdB)	Anzahl der Beobachtungen (N)
Anlagenmechanik	34212, 34213, 34214	151.849
Hochbau	32122, 32123	7.501
IT	43412, 43413, 43414	167.878
Handel	71402	318.471
Logistik	51311, 51312	740.668
Industrie	26112, 26113, 26114	98.248
Elektrotechnik	26242, 26243, 26244	1.053
Gastronomie	29301, 29302	116.339
Gesundheit	82101, 82102, 82103	381.817

Quelle: Online Job Advertisements (OJA).

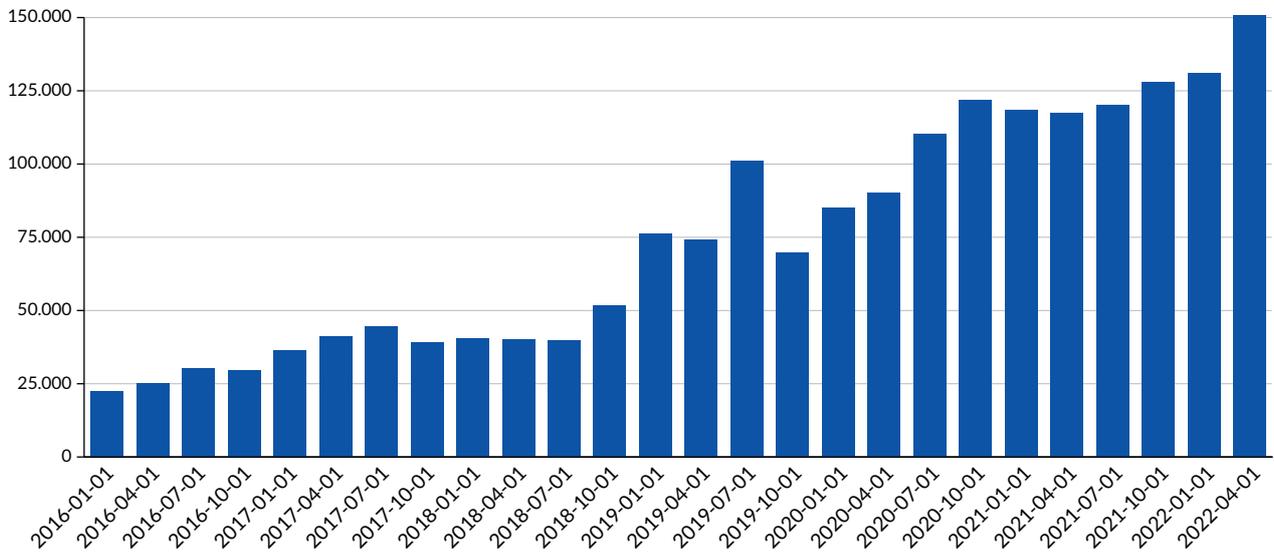
| BertelsmannStiftung

2 <https://web.arbeitsagentur.de/berufenet/taetigkeitsfelder>

3 <https://www.textkernel.com/>

ABBILDUNG 1 Zeitreihenanalyse, Anzahl an Postings 2016–2022

Anzahl in Postings



Quelle: Online Job Advertisements (OJA).

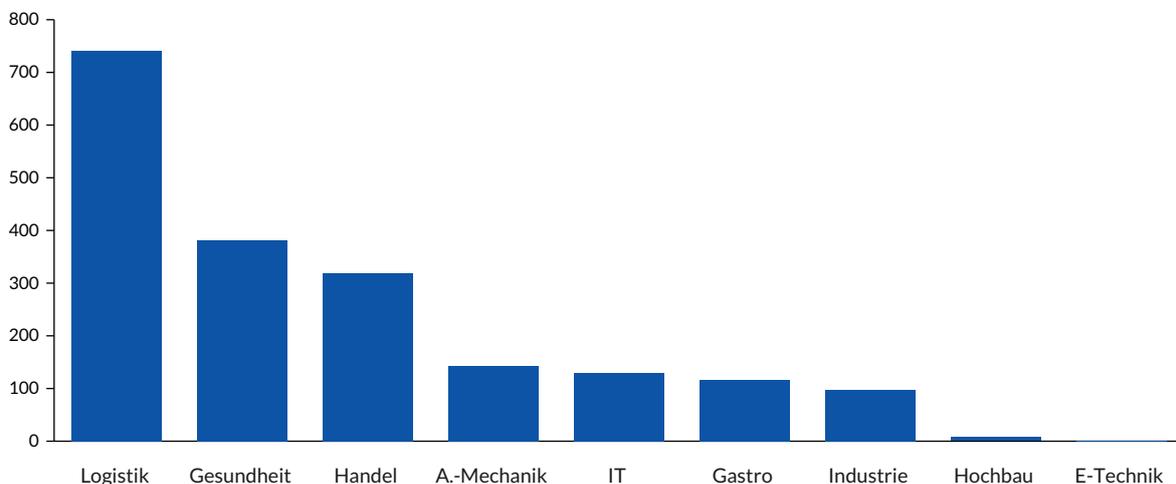
| BertelsmannStiftung

bei kommt der Duplikat-Erkennungsalgorithmus von TextKernel zum Einsatz. Nur diejenige Stellenausschreibung, die als erste veröffentlicht wurde, wird in den Datensatz aufgenommen. Es ist wichtig anzumerken, dass aufgrund der kontinuierlichen Weiterentwicklung der Methodik von TextKernel, der Zunahme von Datenquellen und dem allgemeinen Anstieg von

Online-Stellenausschreibungen absolute Zahlen – wie beispielsweise die Anzahl der Stellen, in denen eine bestimmte Kompetenz gefunden wurde – nur begrenzt über die Zeit vergleichbar sind. Daher basiert die Analyse in dieser Studie grundsätzlich auf dem relativen Anteil der Stellenausschreibungen, in denen eine bestimmte Kompetenz vorkommt.

ABBILDUNG 2 Verteilung der Postings auf Berufsgruppen

Anzahl an Postings (in Tausend)



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

In Abbildung 1 ist die Verteilung der OJA-Postings über die Zeit abgetragen. Der analysierte Datensatz besteht aus fast zwei Millionen OJAs, die zwischen Januar 2016 und April 2022 gesammelt wurden. Wie hier ersichtlich wird, steigt die Anzahl der Postings spätestens ab Oktober 2018 zunächst sprunghaft und dann relativ kontinuierlich an. Aus der Aufschlüsselung der Daten anhand der einzelnen Bundesländer wird deutlich, dass der regionale Anteil an den OJAs über die Zeit größtenteils konstant bleibt (siehe [Abbildung A1](#) im Anhang). Große Bundesländer wie Nordrhein-Westfalen, Bayern oder Baden-Württemberg sind entsprechend stark vertreten.

## 2.2 Stichprobe aus neun Berufsgruppen

In Abbildung 2 wird die Verteilung der Postings auf die verschiedenen Berufsgruppen abgetragen. Hinsichtlich der vertretenen Berufe gibt es ein klares Ungleichgewicht. Logistik-Berufe sind mit Abstand am häufigsten vertreten (doppelt so stark wie die zweithäufigste Berufsgruppe Gesundheit). Während IT, Anlagenmechanik (A.-Mechanik), Gastronomie (Gastro) und Industrieberufe in noch etwa über 100.000 OJAs vertreten sind, kommt Hochbau nur marginal zum Vorschein (weniger als 5.000 OJAs) und Elektrotechnik (E-Technik) ist kaum vertreten (weniger als 500 OJAs). Dies ist bei der Beurteilung der vorliegenden Daten zu berücksichtigen. Unternehmen, die in der Zeitarbeit tätig sind, schreiben deutlich mehr Stellen aus als Unternehmen

anderer Branchen. Der genaue Anteil an Zeitarbeitsstellen konnte hier nicht erfasst werden<sup>4</sup>. [Abbildung A2](#) im Anhang macht deutlich, dass sich für granulare Analysen, welche die Daten zusätzlich nach Zeitintervallen segmentieren, insbesondere große Berufsgruppen wie Logistik, Gesundheit oder Handel in Frage kommen, da bei den anderen möglicherweise zu geringe Datenmengen vorliegen, um Trends über die Zeit zu unterscheiden. Das Verhältnis zwischen den Berufsgruppen bleibt relativ konstant, allerdings gewinnen Logistik-OJAs weiter an Bedeutung – zu Ungunsten von Handel und Gastro, was mit der Verschiebung der Arbeitsnachfrage während der Corona-Pandemie in Verbindung stehen dürfte.

Die Bundesagentur für Arbeit gliedert die Anforderungsniveaus der Klassifikation der Berufe (KldB) in vier verschiedene Stufen der Vertikalität: Auf der ersten Stufe befinden sich Tätigkeiten, die als Hilfs- oder Anlern Tätigkeiten gelten. Diese sind in der Regel wenig komplex und erfordern oft keinen formalen beruflichen Bildungsabschluss. Die zweite Stufe umfasst fachlich ausgerichtete Tätigkeiten, die normalerweise im Rahmen einer zwei- bis dreijährigen Berufsausbildung erlernt werden. Auf der dritten Stufe finden sich komplexe Spezialist:innen-Tätigkeiten, die häufig eine Ausbildung als Meister:in, Techniker:in oder einen Bachelorabschluss voraussetzen. Die vierte Stufe (Expert:innen) beinhaltet hochkomplexe Tätigkeiten, die in der Regel von Personen mit einem Masterabschluss, Diplom oder einem ähnlichen Abschluss ausgeübt werden (Bundesagentur für Arbeit, 2011).

4 <https://jobmonitor.de/projekt>

## 3. Methodik

Zur Vermessung von Kompetenzen nutzen wir drei Metriken – Häufigkeit, Zentralität und Prämie – sowie zwei unterschiedliche zeitliche Perspektiven: eine Querschnittsanalyse aller Daten aus den Jahren 2018–2022 ohne Berücksichtigung eines Zeitverlaufs sowie eine Zeitreihenanalyse, die sich von März 2016 bis März 2023 erstreckt. Die drei verwendeten Metriken basieren auf aktueller Forschung im Bereich der quantitativen Bewertung von Kompetenzen. Durch sie werden drei unterschiedliche Konzepte abgedeckt, die bei der Beschreibung und Vorhersage im Bereich Kompetenzbewertung relevant sind, wie Stephany und Teutloff (2023) zeigen. Sie legen in ihrer aktuellen Forschung nahe, dass alle drei Metriken – gegeben der Datenbasis von OJA – eine vielseitige Betrachtung von Kompetenzbewertungen erlauben: Zum einen beschreibt der Aspekt der relativen Häufigkeit die Nachfrage nach einer bestimmten Kompetenz – wie oft wird die Programmiersprache „Python“ im Verhältnis zur allgemeinen Nachfrage im Bereich IT verlangt. Die Metrik der Zentralität beschreibt das Verhältnis von Kompetenzen untereinander. In der Fachwissenschaft spricht man hierbei von Komplementarität. Die Metrik gibt an, wie vielseitig eine Kompetenz mit anderen Fähigkeiten kombiniert wird – mit wie vielen anderen Programmiersprachen kann „Python“ gemeinsam verwendet werden. Die Metrik der Prämie letztlich beschreibt den ökonomischen „Bonus“, den der Markt für eine bestimmte Kompetenz bietet – bei ähnlichen Anforderungen, wie viel mehr kann ich durch die Anwendung von „Python“ verdienen. Zur erfolgreichen Erstellung und Auswertung dieser Metriken sind Datenverfügbarkeit und Datenqualität von höchster Bedeutung. Daher konzentriert sich die nachstehende Analyse auf ausgewählte Kompetenzen, für die über den gesamten Beobachtungszeitraum hinweg ausreichende Informationen – wie beispielsweise das im OJA-Posting aufgeführte Gehalt – vorliegen.

Die drei Metriken betrachten unterschiedliche Aspekte der Kompetenzbewertung, stehen aber dennoch in Relation zueinander. Stephany und Teutloff (2023)

zeigen, dass der Wert (Prämie) einer Kompetenz positiv sowohl mit der Nachfrage (relativ zum Angebot) als auch der Komplementarität einer Kompetenz gekoppelt ist. Da Kompetenzen nahezu ausschließlich in Kombination miteinander verwendet werden, erklärt der Grad der Komplementarität einen großen Teil des Preises. Sie zeigen, dass der Wert einzelner Kompetenzen stark von der Komplementarität abhängt – also davon, mit wie vielen verschiedenen, idealerweise hochwertigen Fähigkeiten eine Kompetenz kombiniert werden kann. Anlehnend an diese Forschung werden die hier verwendeten Metriken erstellt, wie die folgende Übersicht zeigt:

### 3.1. Metriken

- 1) **Relative Häufigkeit (Popularität)** einer Kompetenz innerhalb der OJAs einer bestimmten Berufsgruppe. Sie gibt an, wie oft eine Kompetenz innerhalb einer relevanten Vergleichsgruppe auftaucht. Das bedeutet konkret, wie viel Prozent der OJAs ( $q$ ) einer bestimmten Berufsgruppe ( $j$ ) eine bestimmte Kompetenz ( $s$ ) verlangten, wie Formel (1) zeigt:

$$(1) \quad f_{s,j} = \frac{\sum_j^n q_{j,i(S=s)}}{\sum_j^n q_{j,i}}$$

- 2) **Zentralität (Relevanz)** einer Kompetenz im Netzwerk an Kompetenzen einer Berufsgruppe. Hierbei sind zwei Kompetenzen miteinander verknüpft, wenn sie gemeinsam in einem OJA-Posting erwähnt wurden. Wird eine Kompetenz besonders oft mit anderen Kompetenzen nachgefragt, so steigt ihre Zentralität. Als Metrik für Zentralität wird für jede Kompetenz die sogenannte Pagerank-Zentralität berechnet. Die Pagerank-Zentralität misst die systemische Relevanz eines Knotenpunktes, indem sie bei der Berechnung der Zentralität auch die Zentralität der nächsten Knotenpunkte in Betracht zieht.<sup>5</sup> Formal lässt sich dies in Formel (2) ausdrücken. Hier wird die Pagerank-Zentralität ( $c$ )

5 Weitere Informationen zur Metrik: Zhang et al. (2022)

von Kompetenz (i) durch die Pagerank-Zentralität der angrenzenden Kompetenzen (j-n) und deren „einfache“ (degree<sup>6</sup>) Zentralität (d) ermittelt.

$$(2) \quad c_i = \frac{1}{n} + \sum_{j=1}^n \frac{c_j}{d_j}$$

**3) Prämie (ökonomischer Aufschlag) für das Vorhandensein einer bestimmten Kompetenz bei OJAs innerhalb einer Berufsgruppe.** Sie zeigt an, wieviel höher der Aufschlag für einen Job mit einem bestimmten Skill gegenüber einem Job ohne einen bestimmten Skill ist. Hierzu wird in den meisten Fällen eine „Prämie“ berechnet. Diese vergleicht das Durchschnittsgehalt der OJA-Postings, die eine bestimmte Kompetenz erfordern, mit jenen Postings, die diese Kompetenz nicht verlangen, wie Formel (3) zeigt. Wir berechnen hierbei den Durchschnitt der Gehälter (P) aller OJA-Postings (q) aus einer Berufsgruppe (j), in denen die Kompetenz (s) nachgefragt wird und teilen diesen Wert durch das

Durchschnittsgehalt aller Postings der gleichen Berufsgruppe, die (s) nicht nachfragen.

$$(3) \quad p_{s,j} = \frac{\sum_i^n P_i[q_{j,i}(S=s)]/n}{\sum_i^n P_i[q_{j,i}(S \neq s)]/n}$$

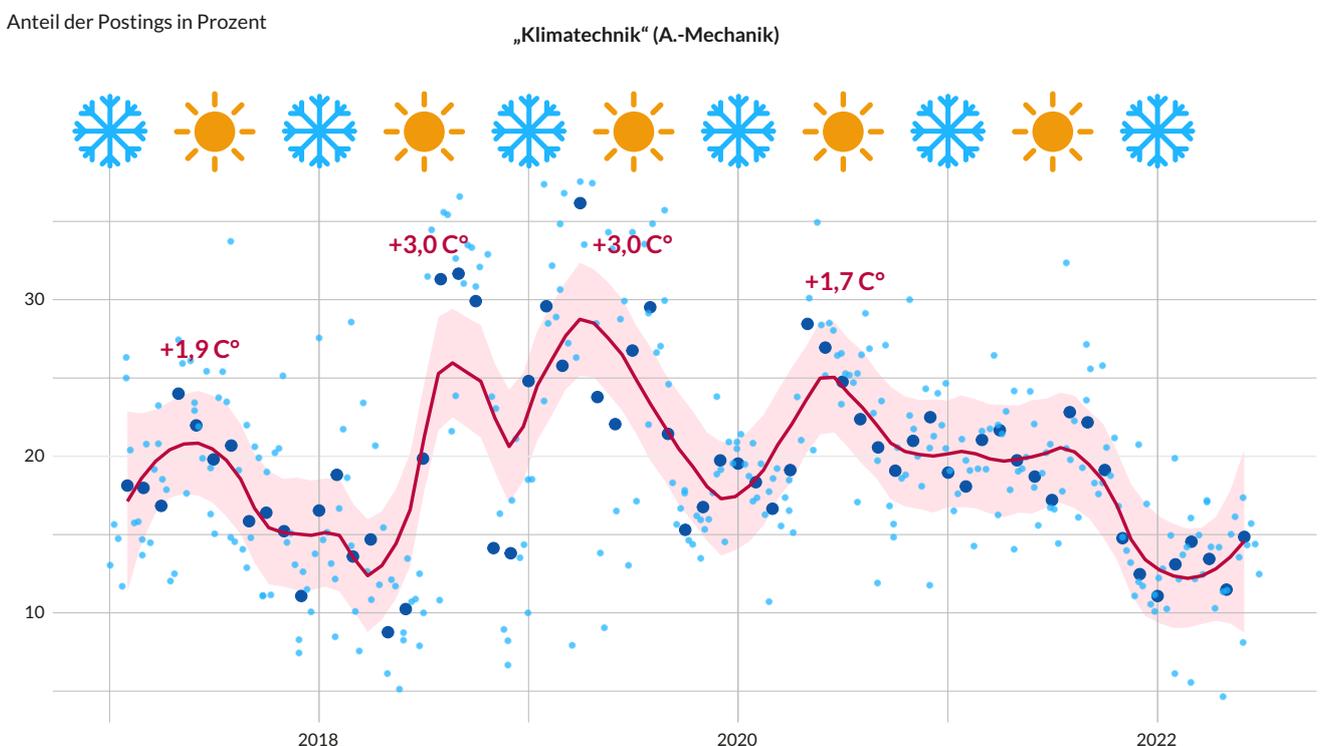
### 3.2. Zeitreihenanalyse

Im zweiten Teil der Auswertung dieser Studie betrachten wir die Entwicklung der drei Metriken über die Zeit und evaluieren deren Vorhersagbarkeit. Hierbei sind verschiedene Konzepte von Relevanz, die im Folgenden kurz erläutert werden.

#### Saisonalität, Zyklus und Trend

In Zeitreihenanalysen spielen die drei Konzepte Saisonalität, Zyklus und Trend eine zentrale Rolle bei der

ABBILDUNG 3 Was sind Saisonalität, Zyklus und Trend – Beispiel „Klimatechnik“



Quelle: Online Job Advertisements (OJA).

6 Ebenfalls Zhang et al. (2022)

Untersuchung und Prognose von Datenverläufen. Dies soll durch Abbildung 3 - die Häufigkeit der Kompetenz „Klimatechnik“ in der Berufsgruppe der Anlagenmechanik – veranschaulicht werden.

**Saisonalität:** Die Saisonalität bezieht sich auf wiederkehrende Muster in den Daten, die normalerweise mit bestimmten Zeiträumen oder Jahreszeiten in Verbindung stehen, wie das Beispiel der „Klimatechnik“ zeigt. Die Nachfrage steigt saisonal bedingt in den Sommermonaten an. Die Identifizierung von saisonalen Mustern ist entscheidend, um saisonale Anpassungen vorzunehmen und genaue Prognosen zu erstellen.

**Zyklus:** Der Zyklus beschreibt langfristige Schwankungen in Zeitreihendaten, die nicht so regelmäßig wie saisonale Muster auftreten. Diese Schwankungen können auf wirtschaftliche, demografische oder andere strukturelle Veränderungen hinweisen. Im Beispiel der Kompetenz „Klimatechnik“ werden diese zyklischen Schwankungen durch eine besonders hohe Nachfrage in auffallend heißen Sommerperioden abgebildet. Die Analyse des Zyklus kann dazu beitragen, langfristige Trends und potenzielle zyklische Entwicklungen in den Daten zu erkennen.

**Trend:** Der Trend ist die langfristige, allgemeine Richtung, in die sich eine Zeitreihe bewegt. Er repräsentiert die grundlegende Entwicklung oder das Wachstum über einen längeren Zeitraum hinweg. Beim Beispiel der Kompetenz „Klimatechnik“ ist langfristig zu erkennen, dass sich die Nachfrage zwischen 2016 und 2022 auf einem stabilen Niveau von etwa 20 Prozent hält. Das Identifizieren des Trends ist wichtig, um langfristige Vorhersagen zu treffen und potenzielle strukturelle Veränderungen oder Entwicklungen zu erkennen.

Für eine zuverlässige Vorhersage ist es entscheidend, die Trendkomponente zu isolieren. Es geht darum, die Daten zu „stationarisieren“ und die Auswirkungen anderer Komponenten jenseits des Trends auf die Vorhersagen zu reduzieren. Wenn der Trend nicht isoliert wird, könnten Vorhersagen durch saisonale oder zyklische Muster verzerrt werden, und es wäre schwierig, langfristige Entwicklung präzise zu identifizieren (Ho und Xie, 1998). Dies wird durch ein ausgewähltes Zeitreihenmodell gewährleistet.

## Zeitreihenmodelle

Für die vorliegende Analyse wird das ARIMA-Modell (Auto-Regressive Integrated Moving Average) genutzt, welches ein leistungsstarkes Werkzeug für die Analyse und Modellierung von Zeitreihen darstellt. Dieses Modell baut auf dem ARMA-Modell (Auto-Regressive Moving Average) auf, erweitert es jedoch um Differenzierung und Integration (I), um Trends zu entfernen und Stationarität zu erreichen. Konkret wird hierbei ein Differenzial der Zeitreihe – meistens des ersten Grades – berechnet (Ho und Xie, 1998).

In seiner grundlegenden Struktur kombiniert ARIMA zwei wichtige Komponenten: den autoregressiven Teil (AR) und den gleitenden Durchschnittsanteil (MA). Der AR-Teil ermöglicht die Vorhersage von Zeitreihenwerten auf der Grundlage ihrer eigenen vorherigen Werte, während der MA-Teil die Vorhersagen von den Fehlern oder Residuen vorheriger Vorhersagen ableitet. Diese Kombination macht ARIMA äußerst vielseitig und anpassbar an eine breite Palette von Zeitreihendaten. Eine bemerkenswerte Stärke des ARIMA-Modells liegt in seiner Anwendbarkeit auf verschiedene Arten von Zeitreihen. Es eignet sich sowohl für deterministische Zeitreihen, die klare Muster oder Trends aufweisen, als auch für stochastische Zeitreihen, die von zufälligen Schwankungen geprägt sind. Formal werden mit jedem ARIMA Modell drei Kennzahlen angegeben. Bei einem ARIMA (6,1,12) beispielsweise handelt es sich um ein Modell, das im AR-Teil auf die Beobachtungen der letzten sechs (6) Zeitintervalle zurückgreift, dabei mit dem ersten (1) Differenzial der Zeitreihe rechnet und einen Moving Average (MA) über zwölf (12) Zeitintervalle berechnet.

Darüber hinaus ermöglicht ARIMA die Analyse von Zeitreihen mit saisonalen Verläufen, bei denen regelmäßige Muster im Zeitverlauf auftreten. Die Fähigkeit des ARIMA-Modells zur Differenzierung und Integration ist ein entscheidender Aspekt seiner Leistungsfähigkeit. Dies ermöglicht es, Zeitreihen mit Trends zu analysieren und gleichzeitig die erforderliche Stationarität sicherzustellen. Stationarität ist ein wesentliches Konzept in der Zeitreihenanalyse, da sie sicherstellt, dass statistische Eigenschaften wie Mittelwert und Varianz über die Zeit hinweg konstant bleiben. Die Anwendungen von ARIMA sind vielfältig und erstrecken sich auf zahlreiche Bereiche in Wirtschaft und Wissenschaft. Mit ARIMA können kurzfristige Vorhersagen auf der

Grundlage vorhandener Zeitreihen getroffen werden. Darüber hinaus ist ARIMA ein wichtiges Werkzeug für die Schätzung und Validierung von Modellen in der Zeitreihenanalyse.

### Zeitreihenvalidierung

Letztlich ist bei der Bewertung der Aussagekraft von Zeitreihen und deren Fortschreibung Validierung entscheidend. Hierzu wird in der vorliegenden Analyse der Box-Ljung-Test verwendet – eine Schlüsselkomponente des ARIMA-Modells. Der Box-Ljung-Test basiert auf einer systematischen Herangehensweise an die Analyse von Zeitreihen. Die wichtigsten Schritte umfassen die Spezifikation des Modells und die Validierung des Modells. Diese Schritte werden in einem iterativen Prozess durchgeführt, um das am besten geeignete Modell für die gegebene Zeitreihe zu finden.

**Spezifikation des Modells:** Der erste Schritt des Box-Ljung-Tests besteht darin, die Struktur des Zeitreihenmodells zu spezifizieren. Dies umfasst die Bestimmung der Ordnung des autoregressiven Teils (AR), des gleitenden Durchschnittsteils (MA) und des Integrationsanteils (I) des ARIMA-Modells. Die Ordnung des Modells gibt an, wie viele vergangene Werte in die Vorhersage einbezogen werden sollen und ob eine Integration zur Herstellung der Stationarität erforderlich ist ([Abbildung A7](#) im Anhang).

**Validierung des Modells:** Sobald das Modell bestimmt wurde, erfolgt die Validierung, um sicherzustellen, dass es die Daten angemessen beschreibt. Dies beinhaltet die Überprüfung, ob die Modellresiduen (Fehler) zufällig und unkorreliert sind. Eine wichtige Komponente der Validierung ist die Verwendung von Diagnosetests, um sicherzustellen, dass das Modell den Annahmen gerecht wird und die Vorhersagen akzeptabel sind. Die Ergebnisse dieser Validierung werden später ebenfalls in [Abbildung A7](#) im Anhang dargestellt.

Zusätzlich zur Validierung der vergangenen Zeitreihe mittels Box-Ljung-Test sind wir ebenso an der Verlässlichkeit der Vorhersagekraft des Modells interessiert. Hierzu nutzen wir eine Kreuzvalidierung (Cross-Validation) der vorhergesagten Zeitreihenwerte. Hierbei wird ein Segment der Zeitreihe zufällig ausgewählt und ein Zeitwert – beispielsweise die Popularität von „Python“ im Monat März 2023 – durch Vergangenheitswerte wie die vergangenen 24 Monate erklärt. Gemessen

wird die Vorhersagekraft durch einen Fehlerterm, also wie sehr der vorhergesagte Wert vom tatsächlichen Wert abweicht. Dieses Verfahren wird für jeden Wert der bekannten Zahlenreihe wiederholt, sofern ein genügend langes vorhergehendes Zeitintervall vorliegt. Am Ende wird ein allgemeiner Fehlerterm für die Vorhersage berechnet – in unserem Fall ist dies der Root Mean Square Error (RMSE), also die Wurzel des Durchschnitts aller quadrierten Einzelfehler (Bergmeir und Benítez, 2012). Je höher dieser Fehlerterm, umso weniger zuverlässig ist die Vorhersagekraft. Die einzelnen Parameter der Vorhersage-Validierung (wie das zu vorhersagende Zeitfenster oder die Länge des berücksichtigten vorhergehenden Zeitintervalls) können variiert werden (siehe [Abbildung A8](#) im Anhang).

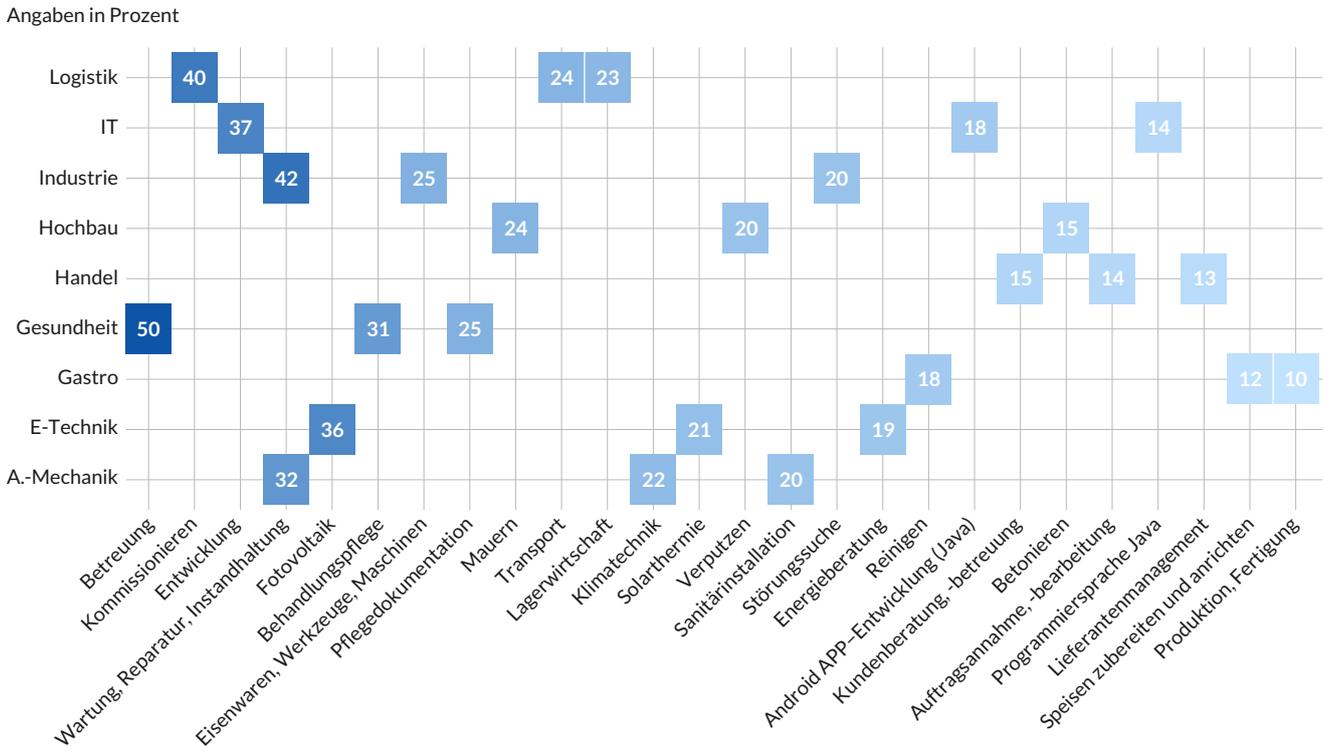
### 3.3. Beispiele für die Anwendung der Metriken

Im folgenden Abschnitt werden die verwendeten Metriken anhand von jeweils einem Beispiel erläutert.

Abbildung 4 zeigt die drei am häufigsten nachgefragten Kompetenzen pro Berufsgruppe. Auf der X-Achse sind Skills abgetragen, auf der Y-Achse Berufsgruppen. Der Farbverlauf signalisiert die Stärke der Werte. Je intensiver die Farbe ausgeprägt ist, desto höher ist der Wert. So können wir beispielsweise in dieser Grafik im Bereich Gesundheit klar erkennen, dass in etwa der Hälfte aller Job-Postings die Fähigkeit „Betreuung“ nachgefragt wird.

In [Abbildung 5](#) sind die 200 zentralen Kompetenzen im Berufsfeld Gesundheit dargestellt. Diese Netzwerkdarstellung entsteht durch die Verknüpfung von Kompetenzen (Knotenpunkte) durch Linien, wenn zwei Kompetenzen gemeinsam in einem OJA-Posting erwähnt wurden. Kompetenzen, die sich im Zentrum des Netzwerks befinden, werden gemeinsam mit vielen anderen Kompetenzen genannt. Sie erhalten entsprechend ihrer hohen Zentralität (Pagerank Centrality) einen großen Knotenpunkt und einen dunklen Farbwert. Am Beispiel der Kompetenzen „Pflegebegutachtung“ und „Pflegedokumentation“ ist gut zu erkennen, dass die zentralen Kompetenzen nicht zwangsläufig auch die am häufigsten vertretenen Kompetenzen sind. Der am häufigsten genannte Skill „Betreuung“ ist im Vergleich dazu nicht die einzige Kompetenz von relevanter Zentralität.

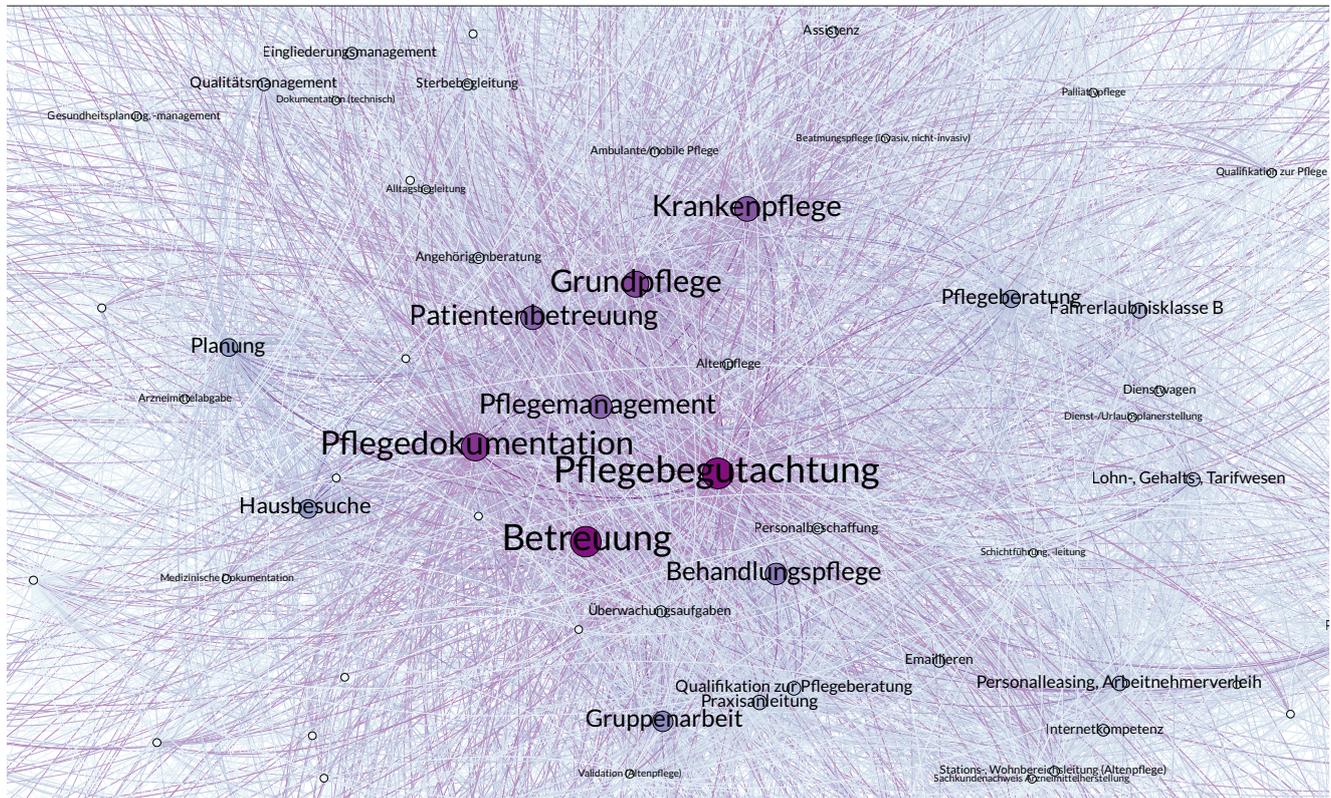
ABBILDUNG 4 Beispiel der Metrik Häufigkeit



Quelle: Online Job Advertisements (2020-2022).

BertelsmannStiftung

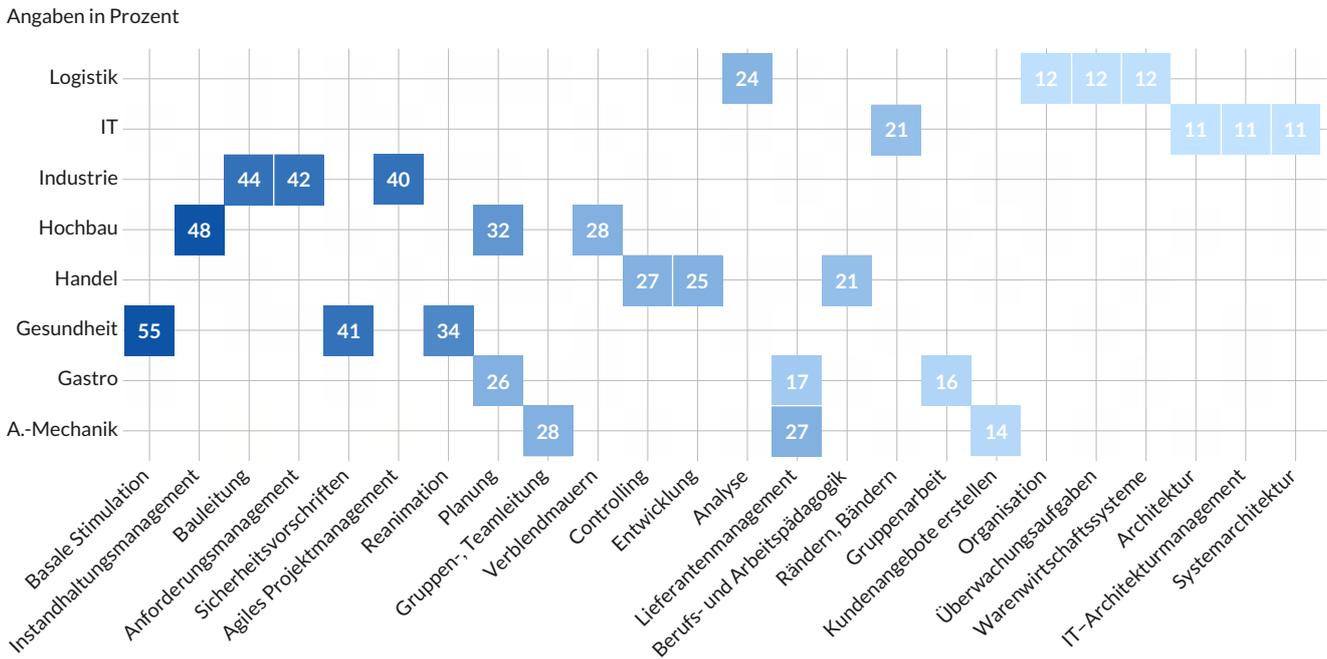
ABBILDUNG 5 Beispiel der Metrik Zentralität



Quelle: Online Job Advertisements (OJA).

BertelsmannStiftung

ABBILDUNG 6 Beispiel der Metrik Prämie



Quelle: Online Job Advertisements (2020-2022).

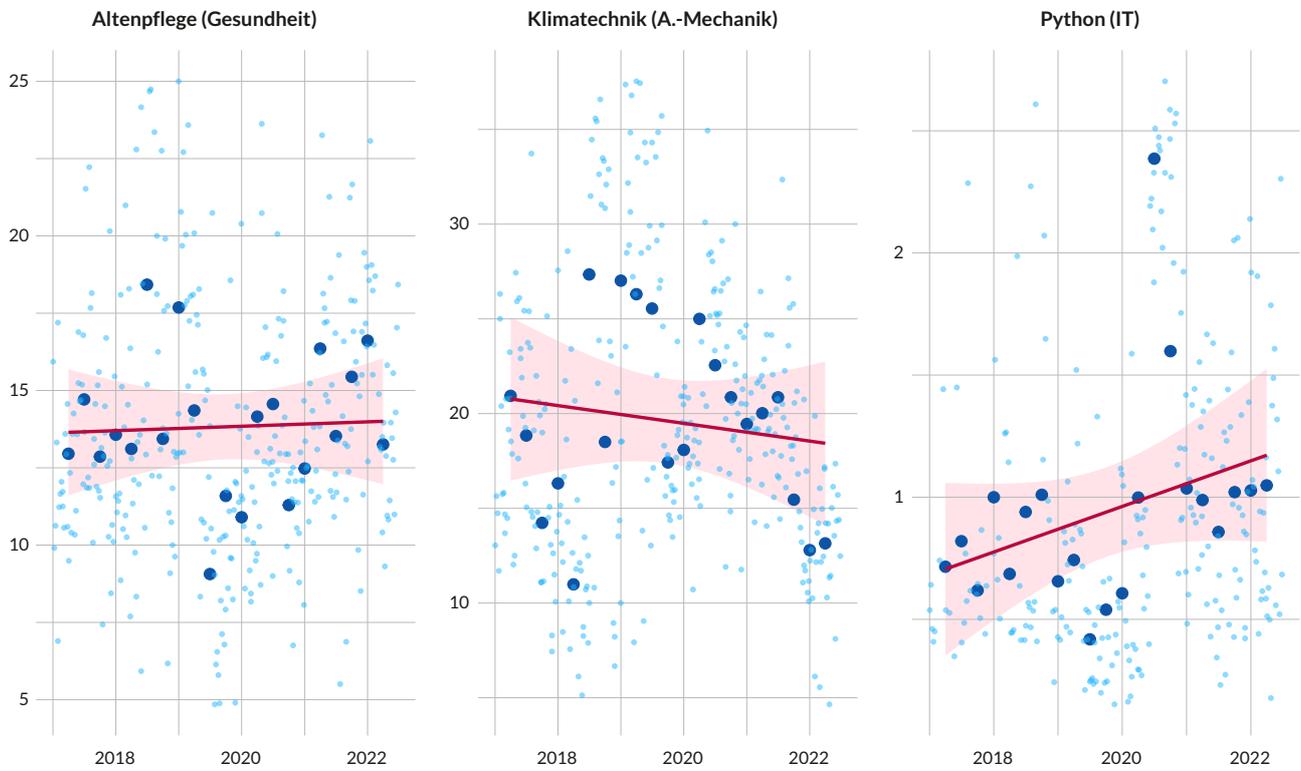
| BertelsmannStiftung

In Abbildung 6 werden die drei „wertvollsten“ Kompetenzen je Berufsfeld abgetragen. Auf der X-Achse sind Skills abgetragen, auf der Y-Achse Berufsgruppen. Der Farbverlauf signalisiert die Stärke von Werten. Je intensiver die Farbe ausgeprägt ist, desto höher ist der Wert. Gut sichtbar ist zum Beispiel, dass die Kompetenz „Reanimation“ im Bereich Gesundheit eine Prämie von 34 Prozent bedeutet. Die Prämie erlaubt es zwar nicht, direkt einen geldwerten Vorteil in Euro zu beziffern, jedoch macht sie einen relativen Vergleich unabhängig von Gehaltsniveaus möglich. In den später folgenden Zeitreihenanalysen werden verschiedene Gehaltsklassen miteinander verglichen und machen so auch eine Betrachtung mit Hinblick auf Geldwerte möglich. Eine hohe Prämie wie diese ist Anlass, um im späteren Verlauf der Analyse den Einfluss von Formalqualifikation und ggf. Zugangsbeschränkungen für bestimmte Berufe unter die Lupe zu nehmen, welche voraussichtlich einen Einfluss auf die Prämienhöhe haben.

In Abbildung 7 sehen wir ein Beispiel dafür, welche Potenziale uns eine Zeitreihenanalyse der vorliegenden Daten bietet. Abgetragen ist hier die Metrik Häufigkeit für drei Skills aus drei unterschiedlichen Bereichen. Auf der X-Achse ist der Zeitverlauf in Jahren abgetragen, auf der Y-Achse die Anzahl der Job-Postings in Prozent. Die kleinen hellblauen Punkte signalisieren Wochenwerte. Dunkelblaue, größere Punkte repräsentieren Quartalswerte. So sehen wir z. B. für den Skill „Altenpflege“ kaum eine Änderung in der Häufigkeit der Job-Postings, wohingegen der Skill „Klimatechnik“ über die Zeit an Ausprägung verliert. Der Skill „Python“ hingegen aus dem Bereich IT gewinnt über die Zeit Nennungen in Job-Postings. Hier zeigt der Trend klar nach oben. Diese generellen Tendenzen sind nicht überraschend, da beispielsweise ein Zuwachs in der Popularität von „Python“ auf die zunehmende Bedeutung dieser Programmiersprache für verschiedenste Datenanalysen in unterschiedlichen Wirtschaftszweigen zurückgeführt werden kann.

ABBILDUNG 7 Zeitreihenanalyse, Beispiel der Metrik *Häufigkeit*

Anteil der Postings in Prozent



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

## 4. Fokus– Die Twin Transition

Der Begriff **“Twin Transition”** bezieht sich auf den Übergang zu einer kohlenstoffneutralen Wirtschaft, wie sie die Europäische Union bis zum Jahr 2050 als Ziel ausgegeben hat. Eine zentrale Rolle in diesem Prozess spielen nachhaltige digitale Technologien. So hat das Joint Research Center der EU (JRC) im Juli 2022 einen Bericht über die wichtigsten Voraussetzungen für eine erfolgreiche grüne und digitale Transformation in der EU veröffentlicht (Muench et al., 2022). Der Bericht nimmt den europäischen Green Deal als Ausgangspunkt und betrachtet die Chancen und Herausforderungen bei der Erreichung dieses Ziels. Die grüne und die digitale Transformation sind miteinander verknüpft und können sich entweder gegenseitig verstärken oder aber auch miteinander in Konflikt geraten. In diesem Teil der Analyse stellen wir die Möglichkeit vor, mithilfe von Online-Stellenanzeigen die Häufigkeit und Bedeutung von Green und Coding Skills über die Zeit zu quantifizieren.

### 4.1. Was sind Green Skills?

Green Skills beschreiben Fähigkeiten, die erforderlich sind, um Produkte, Dienstleistungen und Prozesse umweltfreundlicher zu gestalten und ultimativ dem Klimawandel durch CO<sub>2</sub>-Vermeidung und CO<sub>2</sub>-Entfernung entgegenzutreten. Laut dem European Centre for the Development of Vocational Training (Cedefop) umfassen Green Skills „die notwendigen Kenntnisse, Fähigkeiten, Werte und Einstellungen, um eine nachhaltige und ressourceneffiziente Gesellschaft zu leben, zu entwickeln und zu erhalten“ (Cox et al., 2012). Diese Fähigkeiten werden entsprechend den politischen Zielen zur Klimaneutralität in allen Branchen und auf allen Beschäftigungsebenen stark nachgefragt.

Die Nachfrage nach grünen Kompetenzen wird von drei Haupttrends beeinflusst: (1) der Verbesserung der Kompetenzen und Qualifikationsanforderungen in verschiedenen Berufen und Branchen, um sie umweltfreundlicher zu machen, (2) der Schaffung neuer oder erneuerter Berufe und entsprechender Qualifikations- und Kompetenzprofile im Zusammenhang mit neuen

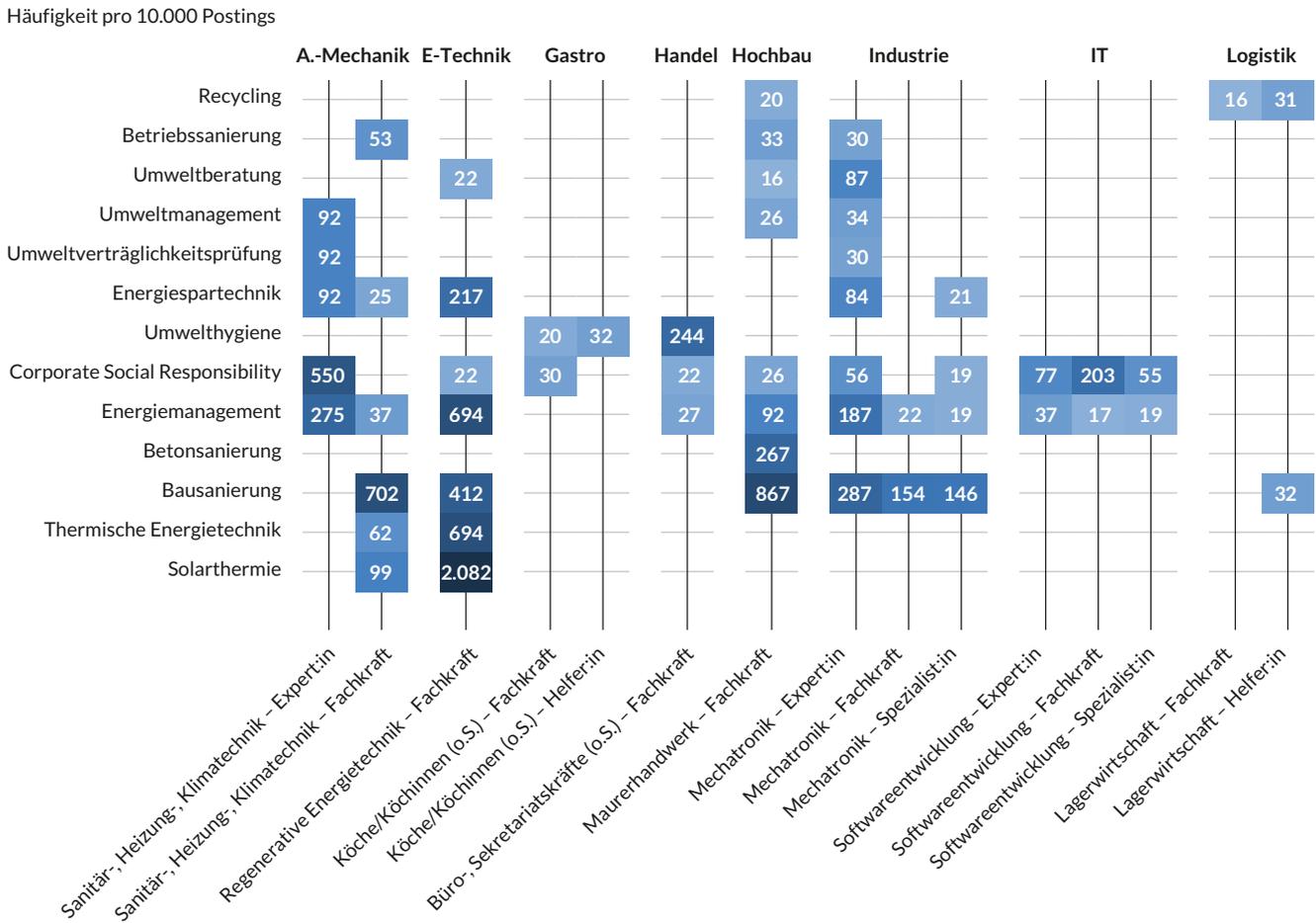
oder aufkommenden Wirtschaftstätigkeiten und (3) der Notwendigkeit für Arbeitnehmer:innen, sich an strukturelle Veränderungen anzupassen, die durch den Übergang zu einer umweltbewussteren Wirtschaft hervorgerufen werden, was zu einem Rückgang einiger traditioneller Sektoren führen und Umschulungen erfordern wird.

Die Analyse von grünen Fähigkeiten im Bereich von Online Job Advertisements bietet die Möglichkeit, eine detaillierte Untersuchung der „Ökologisierung“ der Wirtschaft, die in Zukunft wahrscheinlich immer komplexer werden wird. Vor diesem Hintergrund sind auch unsere Analyseergebnisse zu verstehen, die der oben beschriebenen Systematik folgen.

Abbildung 8 gibt an, wie häufig bestimmte Kompetenzen in den OJA-Postings einzelner Berufsgruppen verlangt werden. Die Zahlenwerte (pro 10.000 Stellenanzeigen) und Einfärbungen von hellblau (selten) zu dunkelblau (häufig) zeigen, dass grüne Kompetenzen zwar in Postings aller Berufsgruppen auftauchen, dort aber unterschiedlich stark verlangt werden. Dies kann am Beispiel der Kompetenz „Energiemanagement“ anschaulich gemacht werden, welche für den Beruf der Fachkraft Regenerative Energietechnik in 7 Prozent aller Postings (694/10.000) vertreten ist. Für eine Fachkraft im Maurerhandwerk wird diese Kompetenz allerdings nur in knapp 1 Prozent aller Stellenanzeigen (92/10.000) verlangt. Die Kompetenz „Bausanierung“ zeigt eine starke Ausprägung über Berufsgruppen hinweg und empfiehlt sich damit für weiterführende Analysen. Die Häufigkeit von grünen Kompetenzen ist sehr ungleich über Berufssparten und Berufe verteilt. Am häufigsten werden grüne Kompetenzen in den Feldern Anlagenmechanik, Energietechnik und Hochbau gefragt. Hier stechen besonders „Solarthermie“, „Bausanierung“ sowie „Energiemanagement“ hervor.

In Abbildung 9 sieht man die Verteilung unserer drei zentralen Metriken. Der betrachtete Zeitraum erstreckt sich von 2022 bis 2023. In der ersten Grafik sehen wir die Metrik Häufigkeit pro 10.000 Postings. Auf der X-Achse ist die Anzahl dargestellt, auf der Y-Achse

ABBILDUNG 8 Häufigkeit von „Green Skills“ pro 10.000 Postings



Quelle: Online Job Advertisements (OJA).

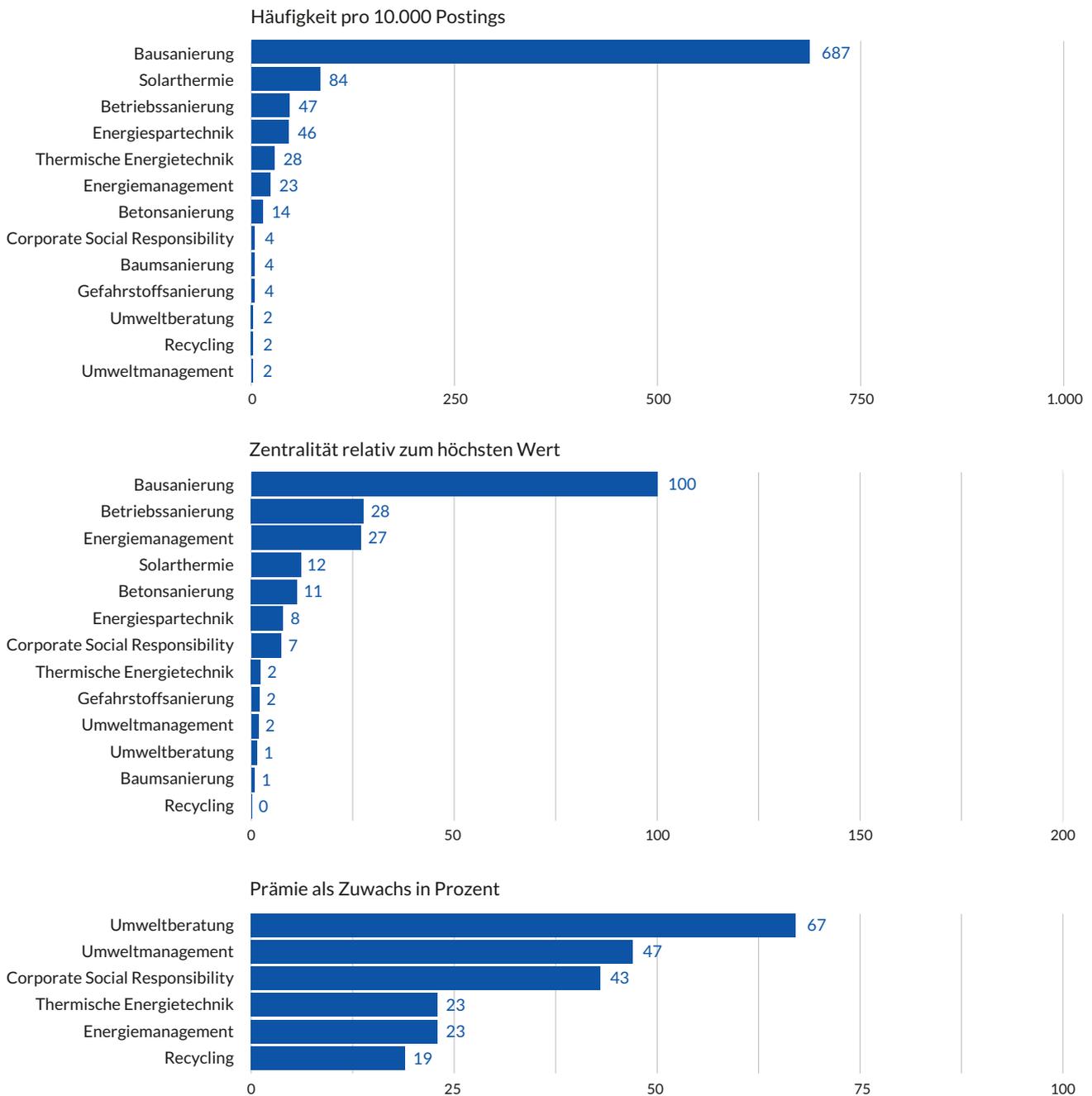
| BertelsmannStiftung

die Kompetenzen. Hier ist klar zu sehen, dass der Skill „Bausanierung“ bei weitem die höchste Anzahl an Nennungen aufweist. In der zweiten Grafik sehen wir die Metrik Zentralität. Auf der X-Achse ist die Ausprägung der Zentralität notiert, auf der Y-Achse wieder die Kompetenzen. Auch hier ist der Skill „Bausanierung“ der am stärksten vertretene Wert. Bemerkenswert ist aber auch, dass die Kompetenz „Energiemanagement“ eine deutlich zentralere Bedeutung hat, als man ihr nach reiner Häufigkeit zuschreiben würde. In der letzten Grafik sehen wir Informationen zur Metrik Prämie. Auf der X-Achse ist der Zuwachs in Prozent angegeben. Interessant ist hier, dass die populärste Kompetenz („Bausanierung“) nicht zwangsläufig die mit der höchsten Prämie ist. Insbesondere „Umweltberatung“, „Corporate Social Responsibility (CSR)“ und „Umweltmanagement“ werden höher entlohnt. Popularität ist also nicht zwangsläufig mit dem höchsten ökonomischen

Mehrwert für Arbeitnehmer:innen verknüpft. Fraglich ist, inwiefern Makro-Effekte hier eine Rolle spielen wie beispielsweise gesetzliche Regelungen oder der Trend zur Corporate Social Responsibility. Das Thema energieeffizientes Bauen drängt sich hier besonders auf.

In Abbildung 10 sehen wir die Häufigkeit pro 100 Postings von ausgewählten Green Skills im Bereich Anlagenmechanik und Hochbau. Auf der X-Achse ist der Zeitverlauf in Jahren beschrieben, auf der Y-Achse die Anzahl an Nennungen der jeweiligen Kompetenz pro 100 Postings. Die Punkte stellen Ausprägungen pro Zeitpunkt dar, die Linien zeichnen den Trend nach. Auf Grundlage dessen können wir gegenläufige Trends für die Kompetenz „Bausanierung“ erkennen. Während sie im Maurerhandwerk immer weniger gefragt wird, nimmt ihre Popularität im Beruf der Fachkraft für Sanitär-, Heizungs- und Klimatechnik zu.

ABBILDUNG 9 Verteilung der Metriken – Häufigkeit, Zentralität und Prämie auf Green Skills



Quelle: Online Job Advertisements (OJA).

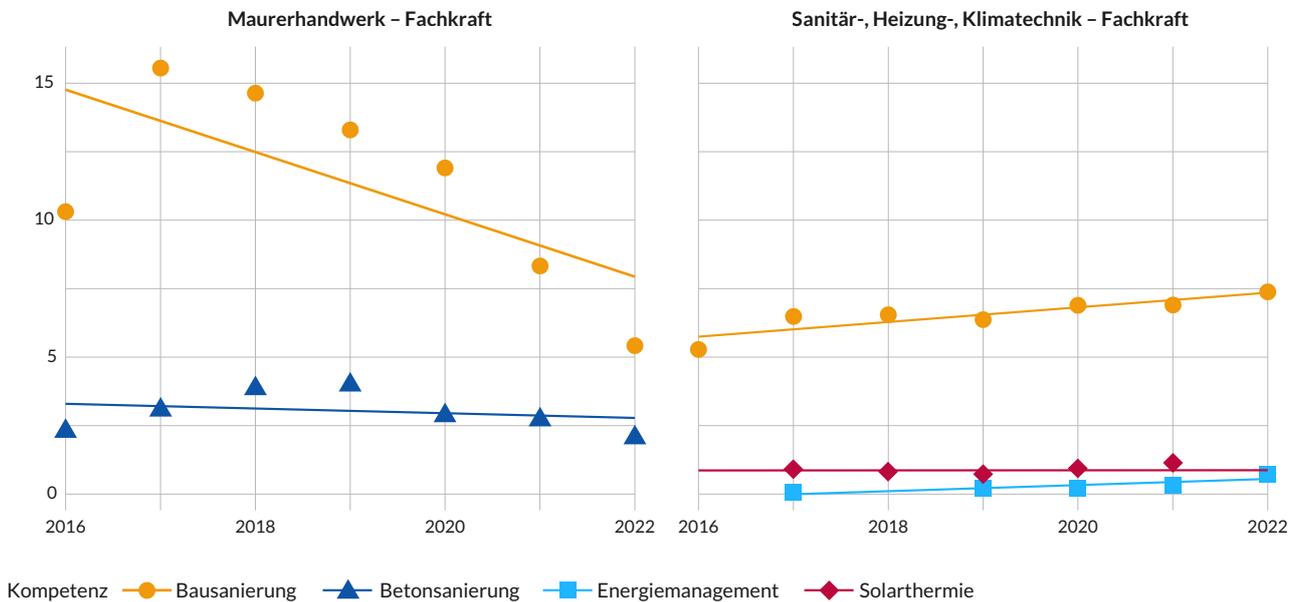
| BertelsmannStiftung

Insgesamt ist zu bemerken, dass wir in den letzten Jahren einen generellen Aufwärtstrend in allen Berufen in puncto Gehalt beobachten können (siehe [Abbildung A4](#) im Anhang). Postings, die entsprechende grüne Kompetenzen verlangen, liegen nicht zwangsläufig über den Gehaltsangaben von Postings ohne grüne Kompetenz. Am Beispiel von „Bausanierung“ im Maurerhandwerk allerdings ist ein klarer Trend hin zu

einer grünen Kompetenz-Prämie zu erkennen. Diese Ergebnisse sind zwar signifikant, heben sich aber nicht sehr von der allgemeinen Situation auf dem Arbeitsmarkt ab. Ausgebildete Fachkräfte haben in den letzten Jahren relevante Gehaltszuwächse gesehen. Auch hier kommen für die Ursache wieder Makro-Trends in Betracht wie die zunehmende Überalterung der Gesellschaft, der sektorspezifische Fachkräftemangel sowie

ABBILDUNG 10 Zeitreihenanalyse, Entwicklung der Metrik Häufigkeit pro 100 Postings im Vergleich

Häufigkeit pro 100 Postings



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

die Tatsache, dass die Abitur- und Studienquoten in der letzten Dekade stetig angestiegen sind. [Abbildung A4](#) im Anhang zeigt (wie auch schon die vorherigen) die Stärke von Big Data, welche in der Identifikation von Trends auf granularer Ebene liegt. Allerdings kommt mit der Möglichkeit zur Identifikation diverser Trends auch die Notwendigkeit hinzu, die einzelnen Interpretationsspielräume abzustechen und mögliche überschneidende Trendursachen, wie wir sie hier andeuten, zu unterscheiden.

## 4.2. Was sind Coding Skills?

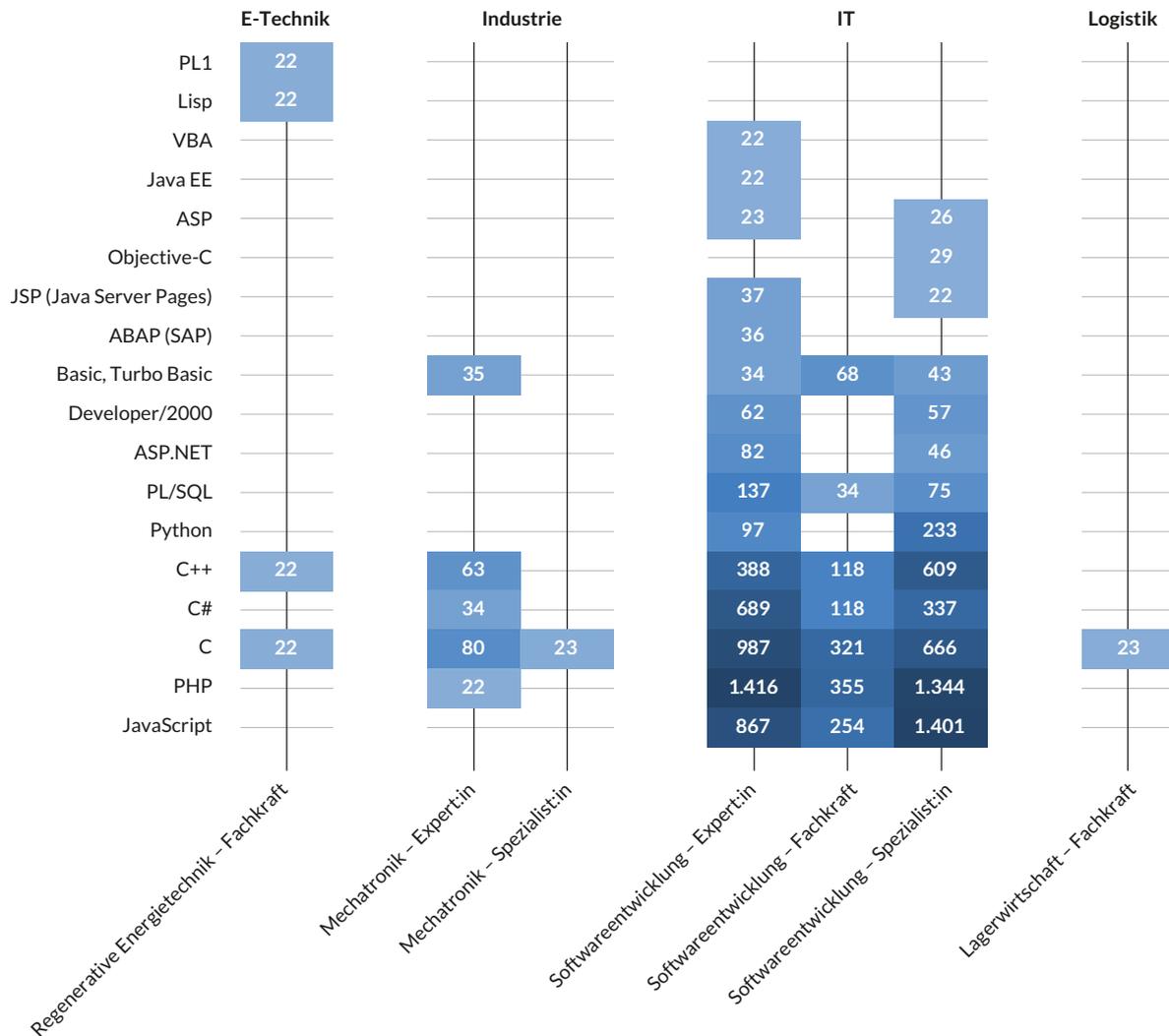
Alle Arten von modernen Informationssystemen und Anwendungen erfordern Programmier- bzw. Skriptsprachen. Unterschiedliche Sprachen eignen sich besser oder schlechter für verschiedene Anwendungsfälle, was sich in divergierenden Prämien abzeichnen sollte. Während z. B. „JavaScript“ und „PHP“ Plattformen sind, die besonders in der Web-Entwicklung eingesetzt werden, eignen sich Sprachen wie „Python“ oder „Java“ eher für sogenannte Backend-Anwendungen. In ihrer Gesamtheit sind Coding Skills entsprechend dem Konzept der Twin Transition besonders interessant für einen Blick in die Zukunft, vor allem aufgrund

ihrer wechselseitigen Beziehung zu den vorher beschriebenen Green Skills. Die hier vorliegende Analyse beschränkt sich auf Programmiersprachen an sich als „Coding Skills“ – also ob ein OJA-Posting eine Programmiersprache wie „Python“ oder „Java“ als Anforderung enthält. Dabei gilt zu beachten, dass auch Metawissen über Programmiersprachen oder Anwendungswissen über die Funktionsweise von Programmiersprachen im Allgemeinen zur erweiterten Definition von „Coding Skills“ gehören können (Popat und Starkey, 2019).

Abbildung 11 zeigt die Häufigkeit der Nennung verschiedener Programmiersprachen pro 10.000 Postings. Auf der X-Achse sind Berufsbezeichnungen lokalisiert, auf der Y-Achse diverse Programmiersprachen. Die Zahlen im Diagramm stellen die Anzahl der Nennungen dar. Die Farbcodierung gibt an, wie stark die Ausprägungen sind. Je dunkler die Farbe, desto höher die Zahl. Festhalten lässt sich hier, dass Programmiersprachen am häufigsten im Bereich der Softwareentwicklung gefragt sind, was wenig verwundert. Stellenweise finden wir noch relevante Nachfrage in Mechatronik und Energietechnik. Ein Umstand, der sich durch die zunehmende Ausbreitung von „(Industrial) Internet of Things“-Anwendungen erklären ließe.

ABBILDUNG 11 Häufigkeit von Programmiersprachen

Häufigkeit pro 10.000 Postings



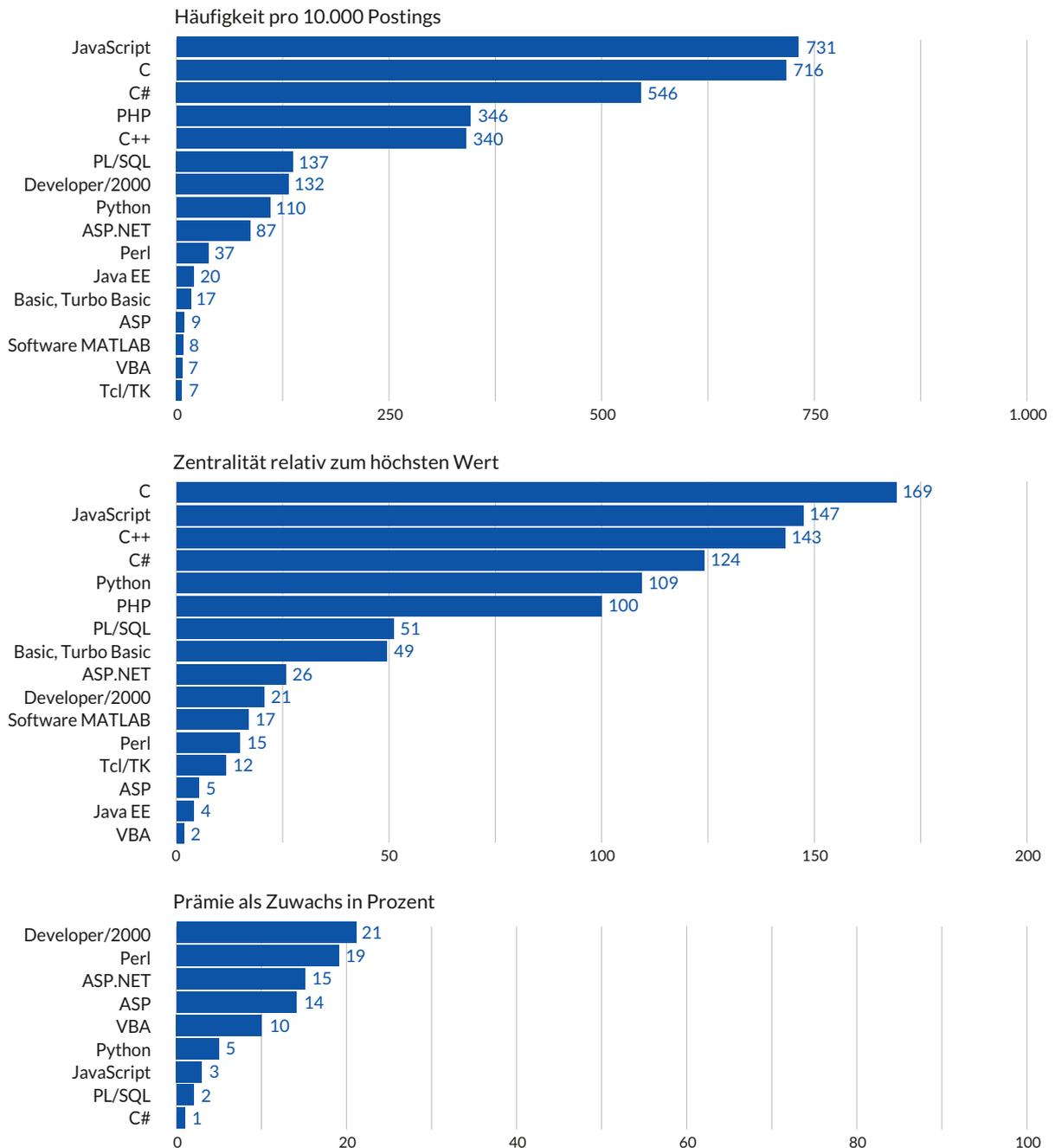
Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

In Abbildung 12 fokussieren wir auf die Verteilung unserer drei zentralen Metriken. Der betrachtete Zeitraum erstreckt sich von 2022 bis 2023. In der obersten Grafik sehen wir die Metrik Häufigkeit pro 10.000 Postings. Auf der X-Achse ist die Anzahl der Nennungen dargestellt, auf der Y-Achse die Kompetenzen. Hier sehen wir, dass die Sprachen „JavaScript“, „C#“ und „C“ die höchste Anzahl Nennungen verzeichnen. In der mittleren Grafik sehen wir die Metrik Zentralität. Auf der X-Achse ist die Ausprägung der Zentralität notiert, auf der Y-Achse wieder die Kompetenzen. Hier sind die Spitzenreiter identisch mit den häufigst genannten Kompetenzen „C“, „JavaScript“ und „C#“. An vierter Stelle liegt „Python“. In

der Abbildung 12 sehen wir Informationen zur Metrik Prämie. Auf der X-Achse ist der Zuwachs in Prozent angegeben. Wie hier deutlich wird, sind insbesondere Programmiersprachen für spezielle Anwendungen besonders „wertvoll“ wie „Developer/2000“ (Datenbank-Programmiersprache), „ASP.NET“ (eignet sich für Open Source Web Applications) bzw. selten eingesetzte Spezialsprachen wie „Perl“ (Allzweck-Programmiersprache). Die Prämie könnte sich hier durch die geringe Verfügbarkeit dieser Sprachfähigkeiten erklären sowie einen stärkeren Anstieg in der Nachfrage. Die populäre Allzweckprogrammiersprache „Python“ ist mit fünfprozentiger Prämie ebenfalls unter den wertvollen Coding-Kompetenzen vertreten.

ABBILDUNG 12 **Häufigkeit, Zentralität und Prämie** von Programmiersprachen in OJAs



Quelle: Online Job Advertisements (OJA).

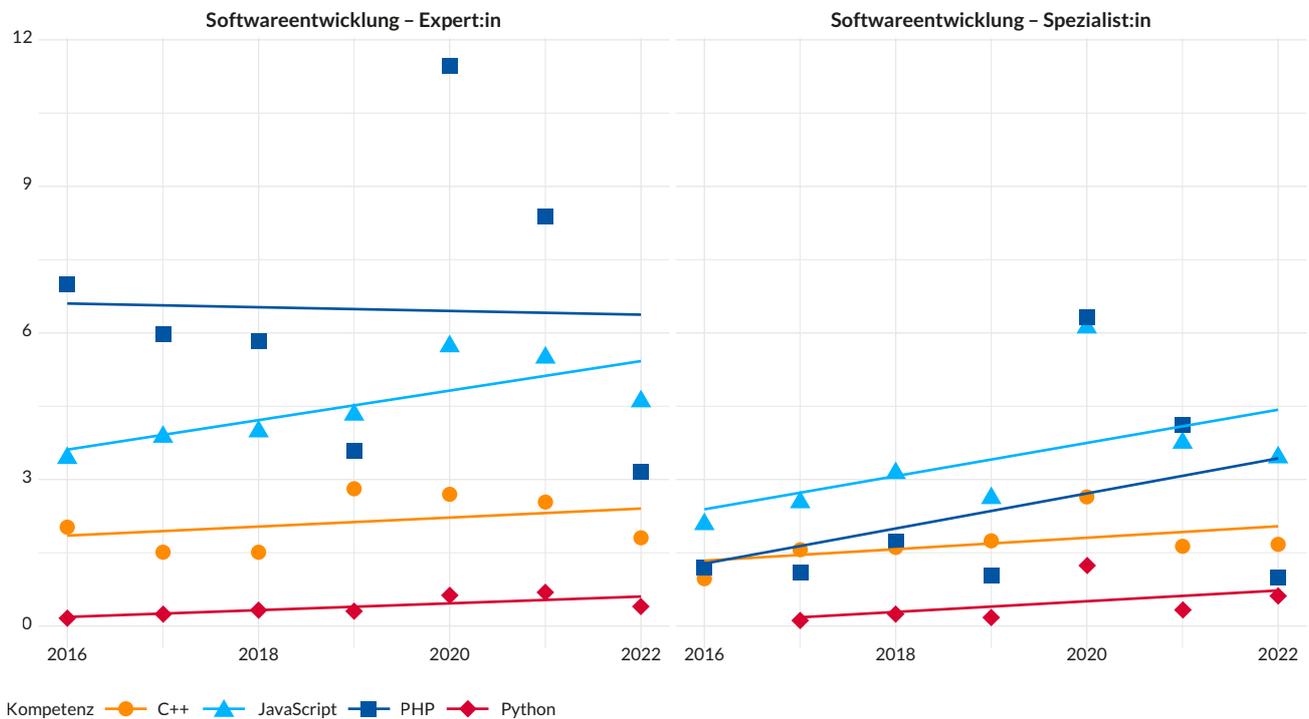
| BertelsmannStiftung

Analog zur vorherigen Betrachtung von Green Skills sehen wir in Abbildung 13 die Häufigkeit pro 100 Postings von ausgewählten Coding Skills. Auf der X-Achse ist der Zeitverlauf in Jahren beschrieben, auf der Y-Achse die Anzahl an Nennungen der jeweiligen Kompetenz pro 100 Postings. Die Punkte stellen Ausprägungen pro Zeitpunkt dar, die Linien zeichnen den Trend nach. In dieser Grafik ist gut erkennbar, dass

sich die Häufigkeit von Kompetenzen für das Level „Expert:in“ recht stabil darstellt. Für Spezialist:innen ist der Trend leicht ansteigend. Dies könnte ein Hinweis darauf sein, dass eine bestimmte Gruppe von Programmiersprachen wie „PHP“, „JavaScript“, „C++“ und „Python“ sich erst im Berufsfeld der Expert:innen etabliert haben, um sich nun auch im Bereich der Spezialist:innen auszubreiten.

ABBILDUNG 13 Zeitreihenanalyse, Entwicklung der Metrik *Häufigkeit* für ausgewählte Programmiersprachen

Häufigkeit pro 100 Postings



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

Weiterhin ist ein klarer Prämienzuwachs für „Python“ zu beobachten (siehe [Abbildung A6](#) im Anhang), welcher sich durch die enorme Nachfrage nach Kompetenzen im Umgang mit großen Datenmengen (hier findet „Python“ oftmals Anwendung) in den letzten Jahren erklären lassen könnte. „JavaScript“, welches oftmals im Bereich der Frontend-Entwicklung zum Einsatz kommt, hat seine Prämie von 2016 bis 2019 eingebüßt.

Dies könnte durch eine nachlassende Nachfrage im Bereich der kundenspezifischen Frontend-Entwicklung zu erklären sein, da sich eigene komplexe Websites oder Web-Anwendungen zunehmend mit interaktiven „Web-Baukästen“ erstellen lassen. Die Prämien für weitere Kompetenzen können den [Abbildungen A3](#) und [A5](#) im Anhang entnommen werden.

# 5. Vorhersagen durch Zeitreihenanalysen

## 5.1. Zeitreihenanalyse ausgewählter Kompetenzen: Popularität und Relevanz

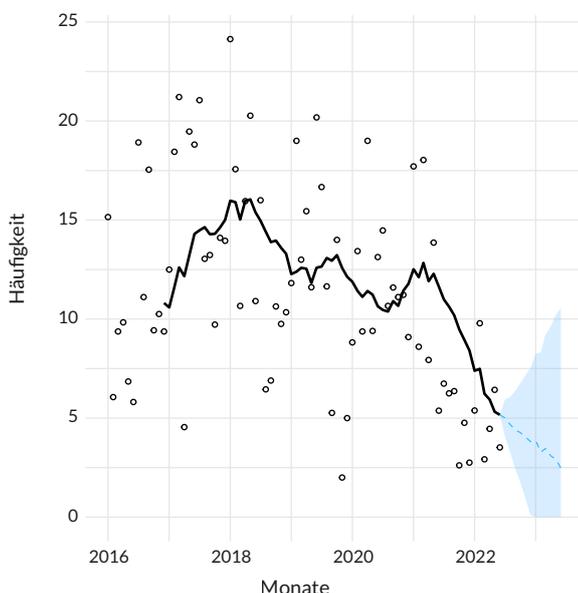
Eine granulare Erhebung der vorgestellten Metriken erlaubt uns Vorhersagen über die Bedeutung von Kompetenzen in der nahen Zukunft. Am Beispiel von drei Kompetenzen – „Bausanierung“ im Hochbau, „Python“ im Bereich IT, sowie „Solarthermie“ in der Anlagenmechanik – wird das Vorgehen bei der Vorhersage durch Zeitreihenanalysen veranschaulicht. Die gewählte Darstellung ist beispielhaft für die zunehmende Bedeutung von Green und Coding Skills. Für jede der drei Kompetenzen werden im Folgenden zwei Zeitreihen mit Monatsdaten aufgeführt. Zum einen betrachten wir die Veränderung der relativen Häufigkeit einer Kompetenz – in wie vielen OJAs einer Branche die Kompetenz erwähnt wird – zum anderen stellen wir die Veränderung der Zentralität im Netzwerk angrenzender Kompetenzen dar.

Auf Grundlage dieser Werte wird mit einem Vorhersage-Modell (Adhikari und Agrawal, 2013) eine Vorhersage über zukünftige Entwicklungen gemacht. Bei den Zeitreihen (Januar 2016 bis März 2023) handelt es sich um ein ARIMA-(6,1,12)-Modell, welches die zukünftigen Werte für die kommenden zwölf Monate extrapoliert. Das Modell wurde mittels eines Box-Ljung-Tests validiert (siehe [Abbildung A7](#) im Anhang). Hier wurde auch das sechsmonatige Intervall für den AR-Teil des Modells identifiziert.

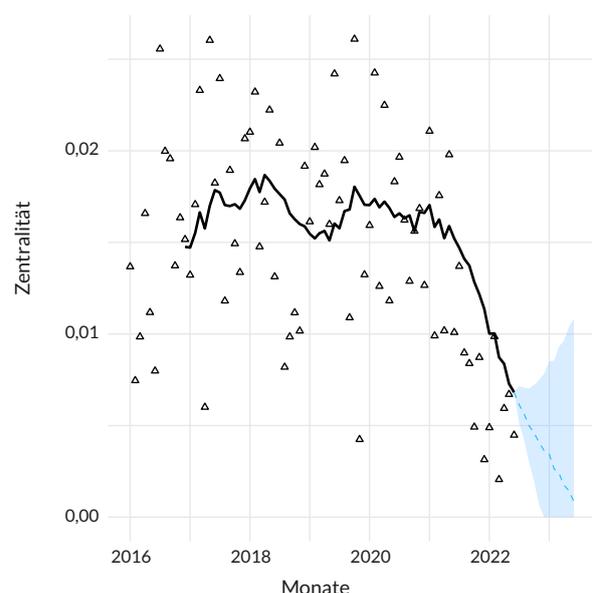
Das Beispiel der Kompetenz „Bausanierung“ lässt eine klare Trendentwicklung von Häufigkeit und Zentralität erkennen. In der linken Abbildung 14 ist ein klar abfallender Trend zu erkennen, den die Vorhersage des Modells fortschreibt. Während im Jahr 2018 noch knapp 15 Prozent aller Anzeigen im Bereich Hochbau „Bausanierung“ erwähnten, sind es Anfang 2022 nur etwa 5 Prozent. Für den Sommer 2023 wird diese Entwicklung auf 3 Prozent fortgeschrieben. Ähnlich sieht

ABBILDUNG 14 Zeitreihenanalyse „Bausanierung“ (Hochbau) sowie Vorhersage von Zeitreihenwerten

Note: Bausanierung in Hochbau, Prozent aller OJAs, ARIMA-(6,1,12)-Modell und 95%-Konfidenz



Note: Bausanierung in Hochbau, Pagerank-Zentralität, ARIMA-(6,1,12)-Modell und 95%-Konfidenz



Quelle: Online Job Advertisements (2016–2023).

dies für die Zentralität von „Bausanierung“ aus. Hier ist in der rechten Abbildung ebenfalls eine abfallende Entwicklung zu erkennen. Die Pagerank-Zentralität von „Bausanierung“ im Netzwerk der Kompetenzen in der Branche Hochbau ist in den Jahren zwischen 2018 und 2021 stabil und fällt danach deutlich ab. Diese Entwicklung wird von dem Vorhersage-Modell fortgeschrieben, indem es einen weiteren Abfall der Zentralität für 2023 prognostiziert. Als Schlussfolgerung lässt sich aus dieser Analyse ableiten, dass die „Bausanierung“ – trotz weiterer Bemühungen um umweltbewusste Energieeinsparungen – an Bedeutung verlieren wird.

Ein anderes Bild zeichnet sich bei der Auswertung der Zeitreihen der Kompetenz „Python“ in der IT-Branche ab (Abbildung 15). Hier ist zwischen Januar 2016 und März 2023 ein klarer und stetiger Zuwachs sowohl mit Hinblick auf die Häufigkeit als auch auf die Zentralität erkennbar. Diese klaren Trends werden entsprechend vom Zeitreihenmodell fortgeschrieben. Für die Häufigkeit sieht die Vorhersage beispielsweise einen Wert von 2 Prozent für das Frühjahr 2024 voraus. Diese Analyse lässt erkennen, dass „Python“ sowohl an Popularität als auch an systemischer Relevanz gewinnen

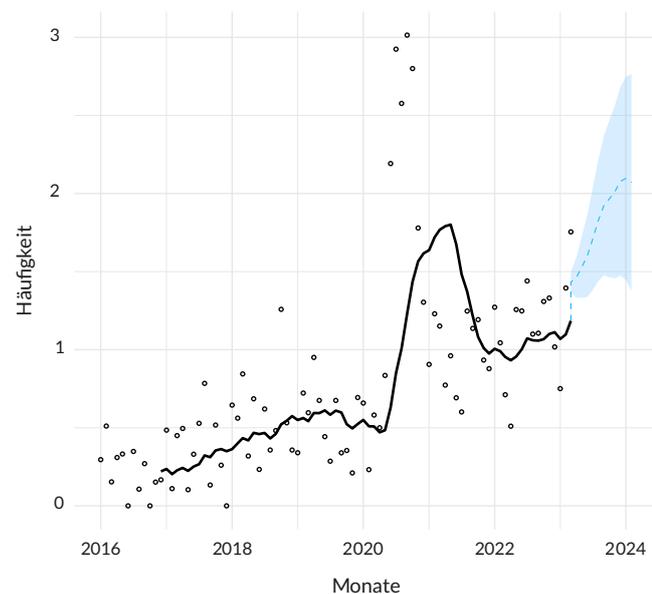
wird. Ein weiterer Einblick zum Aspekt der veränderten systemischen Relevanz von „Python“ folgt im nächsten Abschnitt der Analyse.

Letztlich betrachten wir die Entwicklung der Kompetenz „Solarthermie“ in der Branche der Anlagenmechanik (Abbildung 16). Hier ist mit Hinblick auf beide Metriken ein stabiles Niveau zu erkennen, das ebenso für die kommenden zwölf Monate fortgeschrieben werden kann. Beispielsweise wurde „Solarthermie“ in knapp ein Prozent aller OJAs der Branche Anlagenmechanik verlangt. Dieser Wert kann relativ sicher auf die kommenden zwölf Monate vorgeschrieben werden.

Allgemein zeigen die bisher vorgestellten Zeitreihenanalysen, dass bei der zeitlichen Entwicklung von „Python“ deutlich klare Trends erkennbar sind – im Gegensatz zu den beiden grünen Kompetenzen. Dies könnte darauf hindeuten, dass es sich bei „Python“ (einer sogenannten generellen Programmiersprache) um eine vielseitig einsetzbare Kompetenz handelt, was bei den Kompetenzen „Bausanierung“ und „Solarthermie“ (mit ganz speziellen Anwendungsfeldern) nicht der Fall ist.

ABBILDUNG 15 Zeitreihenanalyse „Python“ (IT) sowie Vorhersage von Zeitreihenwerten

Note: Python in IT, Prozent aller OJAs, ARIMA-(6,1,12)-Modell und 95%-Konfidenz



Note: Python in IT, Pagerank-Zentralität, ARIMA-(6,1,12)-Modell und 95%-Konfidenz

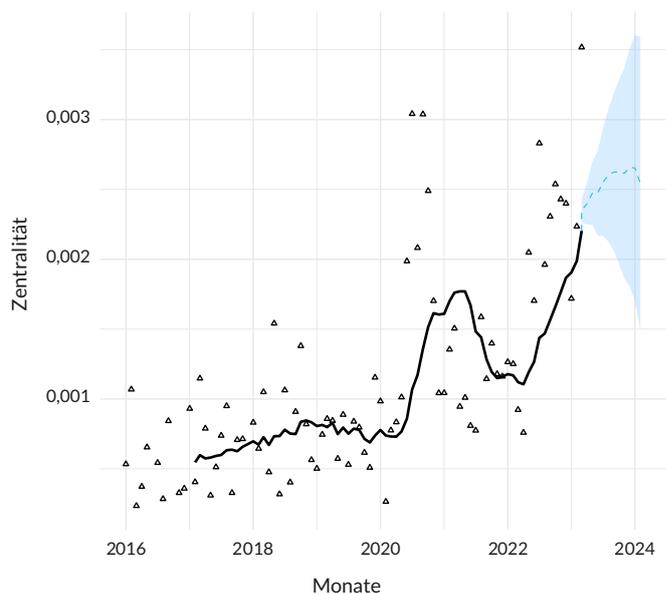
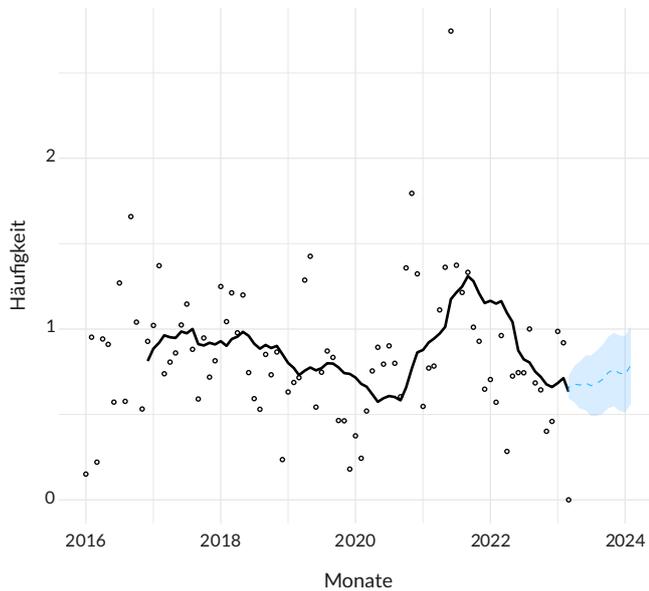
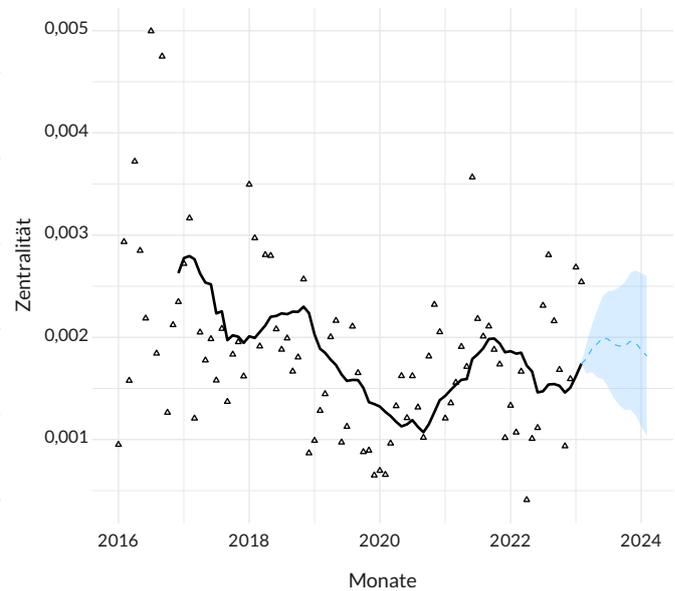


ABBILDUNG 16 Zeitreihenanalyse „Solarthermie“ (Anlagenmechanik) sowie Vorhersage von Zeitreihenwerten

Note: Solarthermie in Anlagenmechanik, Prozent aller OJVs, ARIMA-(6,1,12)-Modell und 95%-Konfidenz



Note: Solarthermie in Anlagenmechanik, Pagerank-Zentralität, ARIMA-(6,1,12)-Modell und 95%-Konfidenz



Quelle: Online Job Advertisements (2016–2023).

| BertelsmannStiftung

## 5.2. Informationsgehalt von Zeitreihenanalysen

Um Aufschluss darüber zu erlangen, ob eine Zeitreihe weitere verwertbare Informationen für eine Vorhersage enthält, kann eine Autokorrelationsanalyse durchgeführt werden. Diese beschreibt, wie stark die zufällig gewählten Werte einer Zeitreihe mit den vorherigen Werten korrelieren. Zeigt sich bei keinem der sogenannten Lags (Vergangenheitswerte) eine Korrelation zu den aktuellen Werten, so kann die Zeitreihe als zufällig interpretiert werden, was in der Fachsprache als "white noise" bezeichnet wird.

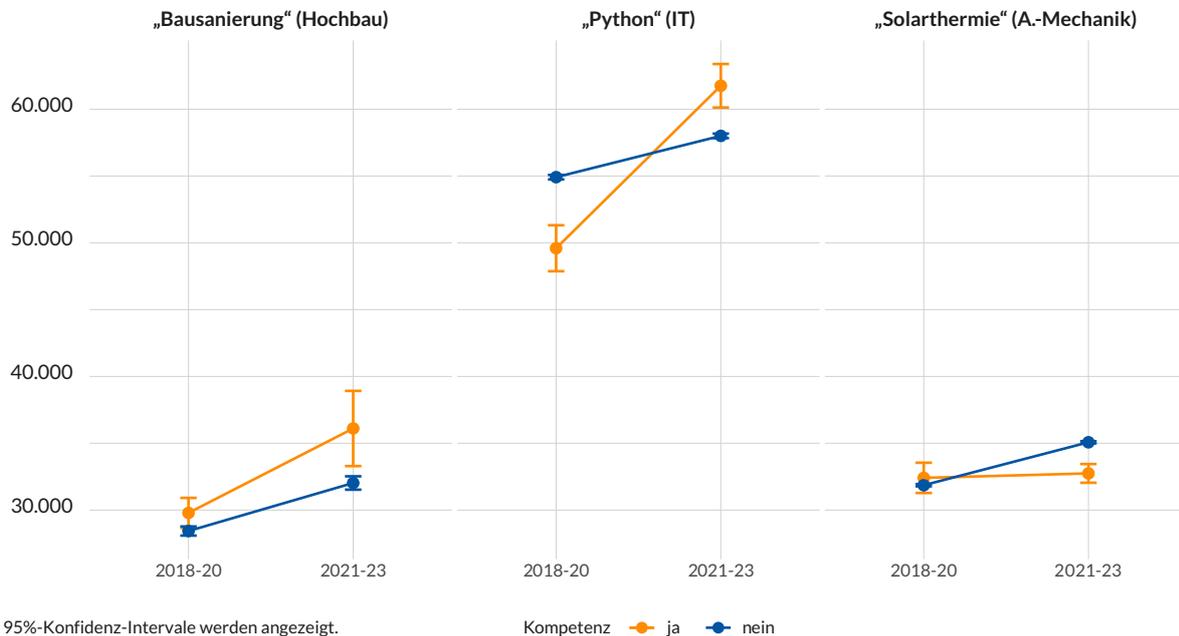
Bei den vorliegenden Zeitreihen von Häufigkeiten zeigt sich jedoch eine klare Autokorrelation, die durch den entsprechenden statistischen Test (Box-Ljung Test) bestätigt werden kann (siehe [Abbildung A7](#) im Anhang). Für alle drei Kompetenzen haben Werte aus den vergangenen 3 bis 4 Monaten (hohe Ausschläge für einen Lag von 1 bis 4 Monaten) eine starke Vorhersagekraft. Bei „Python“ ist zudem ein längerfristiger Trend er-

kennbar, der auf eine stetig steigende Popularität der Programmiersprache hinweist – die positive Autokorrelation fällt stetig über einen Zeithorizont von gut zwei Jahren (hier in 25 Monaten abgebildet) ab. Unabhängig davon, welchen Beobachtungszeitraum man wählt, waren die Werte des Vorjahres stets niedriger als in der jüngsten Vergangenheit.

Um Vorhersagen mit Vergangenheitswerten, wie wir sie hier verwenden, zu validieren, kann ein sogenanntes Cross-Validation-Verfahren herangezogen werden. Hierbei wird eine Zeitreihe zufällig ausgewählt und ein Zeitfenster (hier Monatswerte) durch Vergangenheitswerte (hier die letzten 25 Monate) erklärt. Gemessen wird die Vorhersagekraft durch einen Fehlerterm (hier der Root Mean Square Error, RMSE). Je höher dieser Fehlerterm, umso weniger zuverlässig ist die Vorhersagekraft. Um die Lesbarkeit der Grafik zu erleichtern, wird auf der Y-Achse der Wert 1-RMSE im Verhältnis zur Vorhersage mit aktuellen Werten ( $t=0$ ) abgetragen. Hohe Werte stehen somit für eine hohe Vorhersagekraft mit einem Maximum von 100. Auf der X-Achse

ABBILDUNG 17 Entwicklung der Gehälter im Jahresdurchschnitt zwischen beiden Gruppen

Gehalt im Jahresdurchschnitt



Note: 95%-Konfidenz-Intervalle werden angezeigt.

Kompetenz — ja — nein

Quelle: Online Job Advertisements (2018–2023).

| BertelsmannStiftung

wird die Vorhersagegüte in Abhängigkeit vom Vorhersagezeitraums (1–25 Monate) abgetragen. Bei der Vorhersagekraft der relevanten Zeitreihen (Häufigkeit der Fertigkeiten in Job-Postings) zeigt sich, dass die Vorhersagekraft abflacht, je weiter wir in die Zukunft blicken (siehe [Abbildung A8](#) im Anhang). Während jedoch für „Bausanierung“ die Vorhersagekraft stetig abflacht, je weiter man in die Zukunft blicken möchte, bleibt eine relativ hohe Vorhersagekraft im Falle von „Python“ und „Solarthermie“ auch noch in einem Zeitfenster von 25 Monaten gewährleistet.

### 5.3. Zeitreihenanalyse ausgewählter Kompetenzen: Prämie

Eine granulare Analyse auf Basis von Wochen und Monatswerten ist für tiefere Betrachtungen, die beispielsweise Gehaltsinformationen beinhalten oder nach bestimmten Ausbildungsniveaus unterscheiden, nur schwer möglich. Hierzu wäre eine größere Datengrundlage empfehlenswert. Auf der Basis von Jahreswerten jedoch lassen sich mit den vorliegen-

den Daten auch Aussagen über die Entwicklung von Prämien treffen. Für drei ausgewählte Kompetenzen – „Bausanierung“ (Hochbau), „Python“ (IT) sowie „Solarthermie“ (Anlagenmechanik) – können hier die Gehaltsinformationen der OJAs für die Jahre 2018–20 und 2021–2023 gegenübergestellt werden.

In [Abbildung 17](#) werden jeweils pro Jahresintervall angebotene Durchschnittsgehälter von OJAs verglichen, die die jeweilige Kompetenz erforderten, mit Gehaltswerten aus OJAs, welche die entsprechende Kompetenz nicht verlangten. An den vorliegenden drei Beispielen können zwei unterschiedliche Trends abgebildet werden. Die Gehälter von OJAs, welche „Bausanierung“ forderten, stiegen jedoch von 2018–20 auf 2021–23 etwas stärker an als die Entlohnung für jene OJAs, die die Kompetenz nicht erwähnten. Für „Python“ hingegen ist eine stark wachsende Prämie zu erkennen. OJAs, die „Python“ verlangten, zeigen einen stärkeren Anstieg der gebotenen Gehälter als OJAs, welche keine Kenntnisse in „Python“ voraussetzen. Für „Solarthermie“ sind keine signifikanten Änderungen zu erkennen.

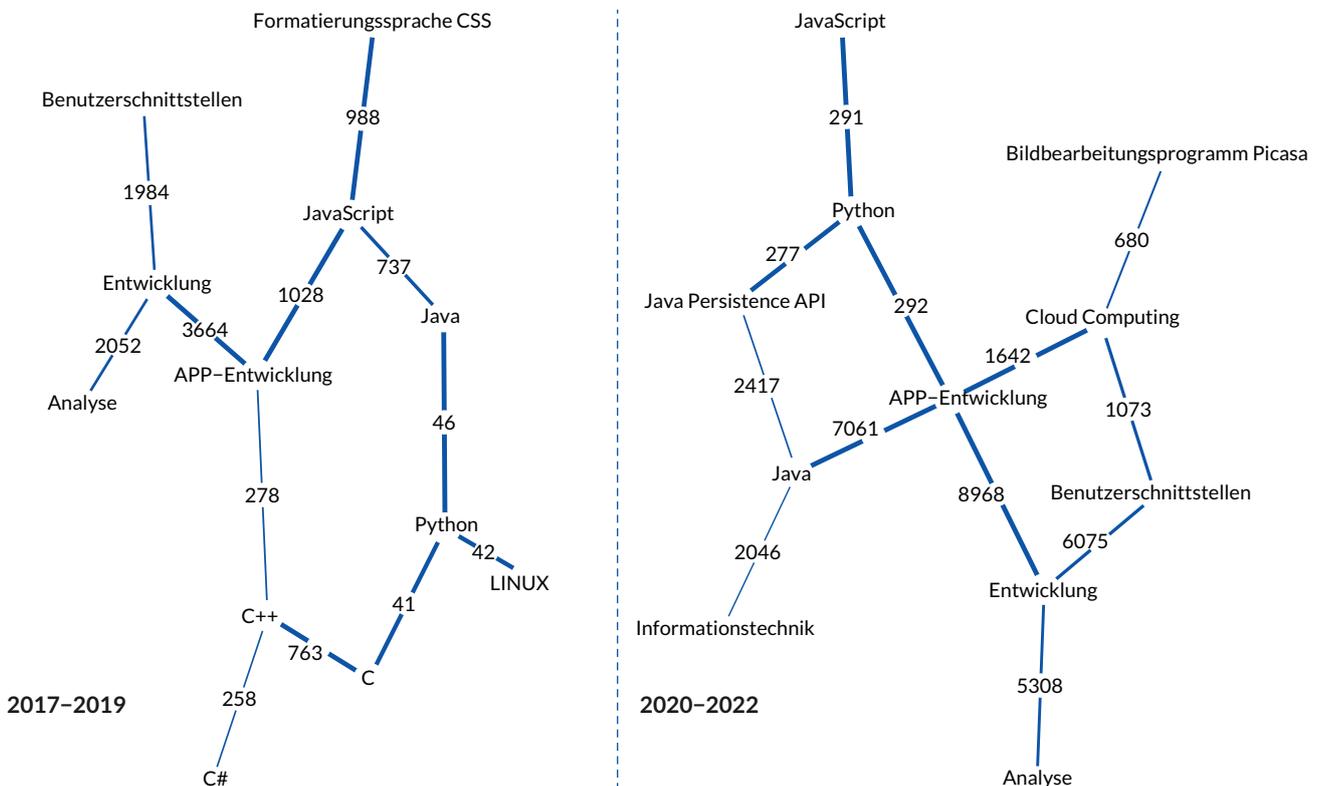
## 5.4. Zeitreihenanalyse des Kompetenznetzwerks

In den vorangegangenen Analysen ist insbesondere die Entwicklung von „Python“ – stetig wachsende Häufigkeit, steigende Zentralität und positive Prämie – herausgestochen. Daher wird im letzten Schritt der Analyse eine Betrachtung des Kompetenznetzwerks von „Python“ über die Zeit vorgenommen, um weitere Evidenz dafür vorzulegen, wie Popularität, Relevanz und Prämie im Falle der Programmiersprache zusammenhängen. Hierzu bedienen wir uns einer vereinfachten Darstellung des Kompetenznetzwerks in der direkten Nachbarschaft von „Python“. Als Grundlage dieser vereinfachten Netzwerkkabbildung dient ein sogenanntes „backboning“ (Coscia und Neffke, 2017): Dargestellt werden nur die drei am häufigsten genannten Kompetenzen im Umfeld von „Python“ sowie wiederum deren drei am häufigsten genannten Nachbarn, welche jedoch nicht zuvor bereits als Nachbar von „Python“ genannt wurden. Dieses Verfahren wird für OJAs aus den Jahren 2017–2019 sowie 2020–2022 angewendet (Abbildung

18). Die Zahlen der Verbindungselemente zwischen den Kompetenzen geben an, in wie vielen Anzeigen beide Kompetenzen jeweils gemeinsam genannt wurden.

In beiden Zeiträumen wird sichtbar, dass die Kompetenz „App-Entwicklung“ im Zentrum des reduzierten Netzwerks steht. Um diese herum paaren sich verschiedene Programmiersprachen sowie Softwareanwendungen. Mit Hinblick auf die Kompetenz „Python“, lässt die Netzwerkanalyse über die Zeit mindestens zwei klare Trends erkennen. Zum einen wird deutlich, wie „Python“ über die Zeit seine Stellung von der Peripherie des Netzwerks verschiebt. War „Python“ im ersten Zeitintervall nur über die konkurrierende Programmiersprache „Java“ mit dem Zentrum des Netzwerks verknüpft, so ist die Programmiersprache jetzt direkt mit „App-Entwicklung“ im Zentrum des Netzwerks verbunden. Ebenso wird sichtbar, dass „Python“ nun deutlich häufiger mit anderen Kompetenzen verknüpft ist. Im ersten Zeitintervall noch wurde „Python“ mit 46 Nennungen am häufigsten gemeinsam mit „Java“ erwähnt. Im Zeitraum von 2020–2022

ABBILDUNG 18 Netzwerkanalyse über die Zeit von „Python“ (IT)

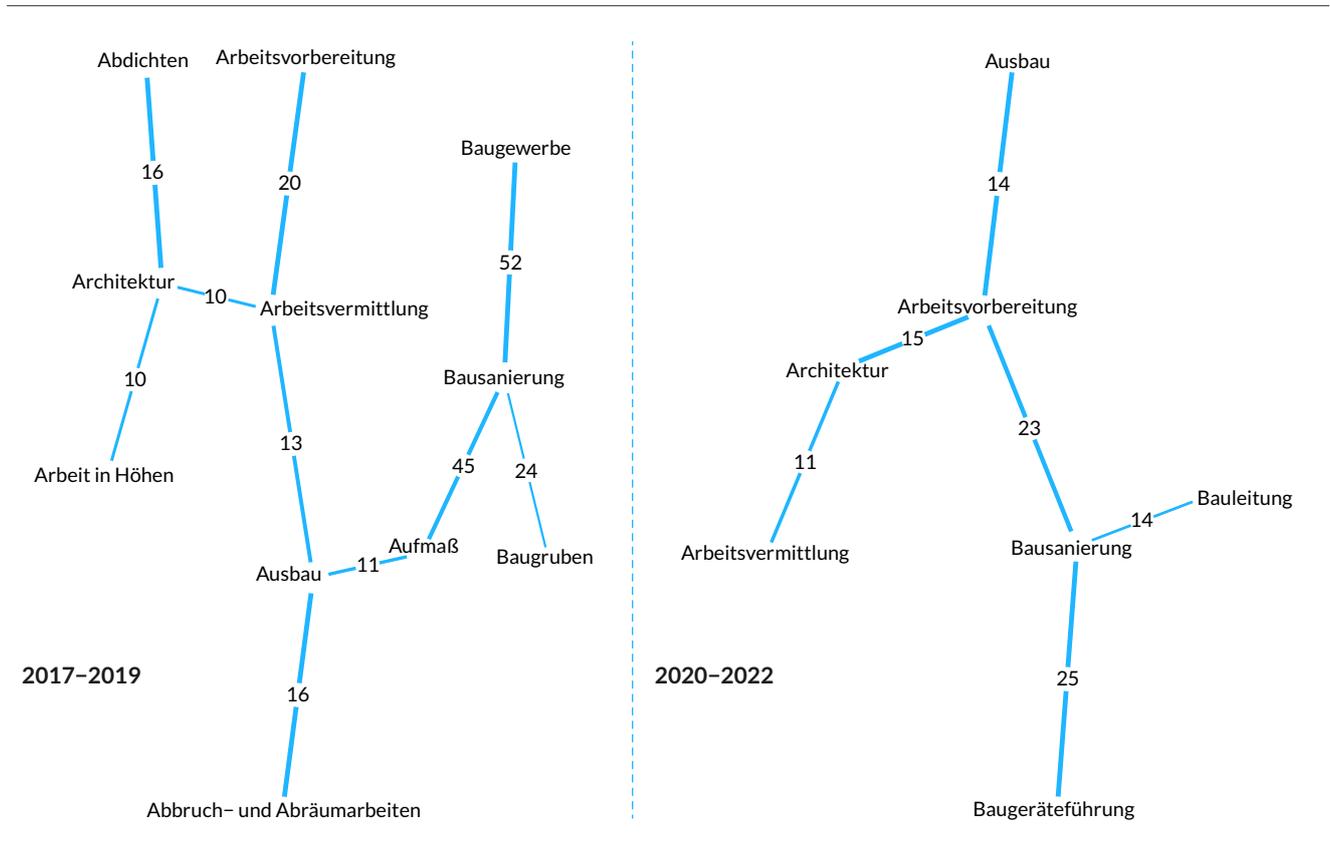


Quelle: Online Job Advertisements (OJA).

jedoch wurde „Python“ bereits 292 Mal direkt im Zusammenhang mit „App-Entwicklung“ genannt. Die zunehmende Vernetzung der Programmiersprache mit relevanten und profitablen Anwendungen wie „App-Entwicklung“ lässt vermuten, warum „Python“ in den letzten sechs Jahren an Bedeutung und ökonomischer Prämie gewonnen hat.

Im Vergleich zu „Python“ zeigt Abbildung 19 die Entwicklung des Netzwerks um die Kompetenz „Bausanierung“. Hier wird sichtbar, dass „Bausanierung“ seine Positionierung im Netzwerk nicht verbessern konnte. Ebenso ist zu erkennen, dass „Bausanierung“ im jüngsten Zeitraum (2020–22) nur noch verhältnismäßig selten in Verbindung mit anderen Kompetenzen genannt wird.

ABBILDUNG 19 Netzwerkanalyse über die Zeit von „Bausanierung“ (Hochbau)



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

## 6. Fazit

### 6.1. Vorhersagen sind möglich

Kann man aus der Vergangenheit Vorhersagen über die Zukunft ableiten? Dies ist eine der grundlegenden Fragestellungen im Bereich der angewandten Zeitreihenforschung. Im vorliegenden Fall bedeutet dies ganz konkret, ob die zurückliegende Bewertung von Kompetenzen Aussagen über deren zukünftige Entwicklung erlaubt. Die zugrundeliegenden Daten für die hier vorgestellte Analyse stammen aus einer mehrjährigen Zeitreihe deutscher Online-Stellenanzeigen. Die Ergebnisse unserer Analyse beantworten die eingangs gestellte Frage klar und zeigen, dass Vergangenheitsdaten verlässliche Vorhersagen über zukünftige Entwicklungen erlauben. Weiß man, wie sich Kompetenzen hinsichtlich ihrer Popularität, Relevanz und ökonomischen Bewertung in der Vergangenheit entwickelt haben, so kann man daraus Vorhersagen über zukünftige Entwicklungen ableiten.

Am Beispiel der Programmiersprache „Python“ lässt sich dies eindeutig illustrieren. In Zeiten, die von technologischem Umbruch und sprunghaften Entwicklungen im Bereich der künstlichen Intelligenz bestimmt sind, gilt das Erlernen von Programmiersprachen als gute Grundlage für eine nachhaltige Weiterbildung. Die Programmiersprache „Python“ wird aufgrund ihrer vielseitigen Anwendbarkeit oftmals als Paradebeispiel für die Zukunftsfähigkeit des Programmierens bezeichnet. Unsere Analyse bestätigt diese Empfehlung. Wir beobachten für alle drei Metriken unserer Analyse – Popularität, Zentralität und Preis – eine klare Aufwärtsbewegung. „Python“ wird zunehmend mehr nachgefragt. Sie gewinnt ebenso an Relevanz. In der Kombination mit anderen Programmiersprachen oder Anwendungsfeldern nimmt „Python“ eine immer zentralere Rolle ein. Auch der ökonomische Wert von „Python“ steigt weiterhin. Wir beobachten, dass OJAs, die „Python“ fordern, anhaltend steigende Gehälter offerieren. Dieser klar erkennbare Trend lässt sich entsprechend sicher in der Zukunft fortschreiben, wie unsere Zeitreihenanalyse und Cross-Validation bestätigen. „Python“ wird auch in Zukunft gefragt, relevant und wertvoll sein. Parallel zu den Fortschritten im Be-

reich künstlicher Intelligenz und maschinelles Lernen wird zunehmend deutlich, dass die Wirtschaft angesichts der voranschreitenden Klimaveränderungen auf Klimaneutralität umgestellt werden muss. Der Strukturwandel, insbesondere die verstärkte Integration erneuerbarer Energien und die Umstellung auf eine CO<sub>2</sub>-neutrale Industrie, führt die dringende Nachfrage nach hochspezialisierten Fachkräften und den entsprechenden Fähigkeiten in den Mittelpunkt. Die in dieser Analyse herausgestellten grünen Kompetenzen „Bausanierung“ und „Solarthermie“ zeigen unterschiedliche Entwicklungen. Als Schlussfolgerung lässt sich aus unserer Analyse ableiten, dass die „Bausanierung“ – trotz weiterer Bemühungen um umweltbewusste Energieeinsparungen – an Bedeutung verlieren wird. Es ist ein klar abfallender Trend zu erkennen, den die Vorhersage des Modells fortschreibt. Für „Solarthermie“ ist mit Hinblick auf unsere Metriken ein stabiles Niveau zu erkennen, das ebenso für die kommenden zwölf Monate fortgeschrieben werden kann.

### 6.2. Die Datenverfügbarkeit entscheidet über die Vorhersagequalität

Bei der Ableitung zukünftiger Entwicklungen aus vergangenen Daten gibt es im Bereich der Kompetenzbewertung allerdings klare Einschränkungen, die für zukünftige, weiterführende Forschung berücksichtigt werden sollten. Zwar können wir zeigen, dass Vorhersagen über die zukünftige Entwicklung von Kompetenzen grundsätzlich möglich sind, nicht aber für einen beliebig langen Zeitraum. Unsere Analysen zeigen klar auf, dass die Vorhersagekraft mitunter stark abnimmt, je weiter wir in die Zukunft blicken. In den meisten Fällen ist dies ab einer Vorhersage der Fall, die über ein Zeitfenster von sechs Monaten hinausgeht. Grundsätzlich gilt, dass stringente Muster aus der Vergangenheit – Entwicklungen mit geringer Varianz – zuverlässiger in die Zukunft fortgeschrieben werden können. Eine Fortsetzung des klaren Aufwärtstrends in der Popularität von „Python“ kann weiter in die Zukunft prognostiziert

werden als die unstete Entwicklung der Kompetenz „Bausanierung“.

Abschließend weisen wir auf die Bedeutung von Datenverfügbarkeit hin. Während unsere Analyse aufzeigt, dass sich Kompetenzbewertungen grundsätzlich und in die Zukunft fortschreiben lassen, steht und fällt die Qualität dieser Vorhersagen mit der Dichte an Daten. Zeitreihenmodelle schreiben die Varianz vergangener Daten in die Zukunft fort. Ihre Aussagekraft leidet stark darunter, wenn für vergangene Zeitintervalle keine verlässlichen Mittelwerte (mit geringer Varianz) berechnet werden können. Dieses Kriterium wurde leider nur für einige wenige Kompetenzen in dem hier vorliegenden Datensatz erfüllt. Zwar steht am Startpunkt unserer Analyse ein Datensatz mit fast zwei Millionen OJAs, doch auf dem Weg zur Analyse und Vorhersage von Zeitreihen dünnt sich dieser Datensatz immer weiter aus. Nicht alle OJAs erfüllen Qualitätskriterien für eine Datenanalyse, da nur ein Teil von ihnen alle relevanten Informationen zu Kompetenzen oder Branchen enthält. Weiterhin

können aussagekräftige Bewertungen von Kompetenzen nur innerhalb einer Berufsgruppe erfolgen. Zudem müssen Stellenanzeigen Gehaltsinformationen ausweisen, damit eine ökonomische Bewertung von Kompetenzen überhaupt möglich wird. In dieser Hinsicht sind deutsche Stellenanzeigen eine Besonderheit. Im Jahr 2022 enthält im Schnitt nur jede zehnte OJA Gehaltsinformationen. Im Vergleich dazu geben Daten aus Großbritannien (30 %) oder den USA (mehr als 50 %) deutlich häufiger Auskunft über Gehälter. Diese Einschränkungen verringern die Anzahl an Observationen, die für eine robuste Zeitreihenanalyse unabdingbar sind, immer weiter. So wird aus „Big Data“ sehr schnell „Small Data“. In vielen Fällen zu „klein“, um verlässliche und robuste Zeitreihenanalysen zu gewährleisten. Diesen Aspekt gilt es bei weiteren Analysen dringend zu berücksichtigen, denn die vorliegende Analyse von Kompetenzbewertungen mit Online-Stellenanzeigen veranschaulicht deutlich: Mit ausreichenden und qualitativ hochwertigen Daten sind verlässliche Aussagen über die zukünftige Entwicklung möglich.

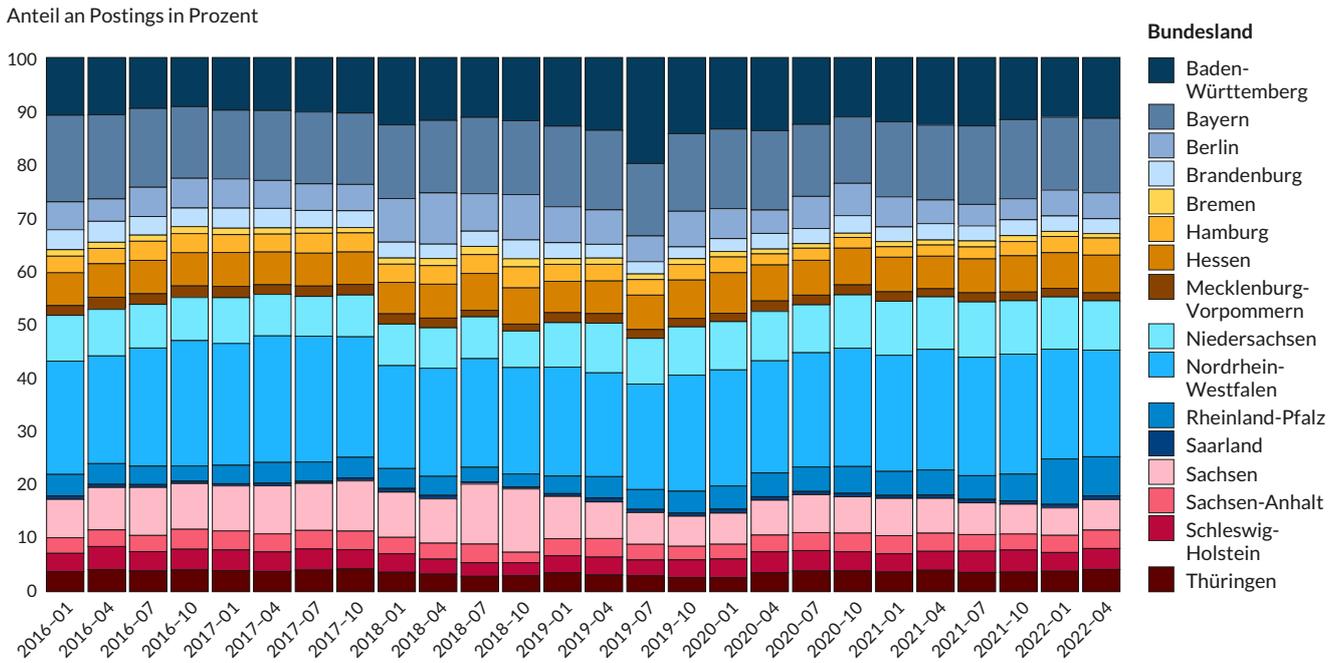
## 7. Referenzen

---

- Adhikari, R. und Agrawal, R. K. (2013).** „An introductory study on time series modeling and forecasting”. arXiv preprint arXiv:1302.6613.
- Arntz, M., Gregory, T. und Zierahn, U. (2016).** *The risk of automation for jobs in OECD countries: A comparative analysis*.
- Bergmeir, C. und Benítez, J. M. (2012).** „On the use of cross-validation for time series predictor evaluation”. *Information Sciences*. 191. 192–213.
- Bessen, J. (2014).** “Employers aren’t just whining – the “skills gap” is real”. *Harvard Business Review*. 25.
- Bundesagentur für Arbeit (2011).** *Klassifikation der Berufe 2010*. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. Nürnberg.
- Coscia, M. und Neffke, F. M. (2017).** “Network backboning with noisy data”. *IEEE 33rd international conference on data engineering*. (ICDE). (pp. 425–436). IEEE.
- Cox, A., Carta, E., Marangozov, R. und Newton, B. (2012).** *Green skills and environmental awareness in vocational education and training: Synthesis report*.
- Fuei, L. K. (2017).** “Automation, computerization and future employment in Singapore”. *Journal of Southeast Asian Economies*. 388–399.
- Muench, S., Stoermer, E., Jensen, K., Asikainen, T., Salvi, M. und Scapolo, F. (2022).** *Towards a green and digital future*. EUR 31075 EN. Publications Office of the European Union. Luxembourg. ISBN 978-92-76-52452-6, doi:10.2760/54, JRC129319.
- Noack, M., Ziegler und M., Müller, J. (2022).** „Kompetenzwandel in Krisenzeiten – Welche Soft Skills jetzt zählen“. Bertelsmann Stiftung. Gütersloh.
- Haleem, A., Javaid, M. und Singh, R. P. (2022).** „An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges”. *BenchCouncil transactions on benchmarks, standards and evaluations*. 2(4). 100089.
- Ho, S. L. und Xie, M. (1998).** “The use of ARIMA models for reliability forecasting and analysis”. *Computers & industrial engineering*. 35 (1–2), 213–216.
- Popat, S. und Starkey, L. (2019).** “Learning to code or coding to learn? A systematic review”. *Computers & Education*, 128, 365-376.
- Stephany, F. und Teutloff O. (2023).** “What is the Price of a Skill? The Value of Complementarity”. <http://dx.doi.org/10.2139/ssrn.4234078>
- Tian, H., Lu, W., Li, T. O., Tang, X., Cheung, S. C., Klein, J., und Bissyandé, T. F. (2023).** „Is ChatGPT the Ultimate Programming Assistant – How far is it?”. arXiv preprint arXiv:2304.11938.
- Zhang, P., Wang, T. und Yan, J. (2022).** PageRank centrality and algorithms for weighted, directed networks. *Physica A: Statistical Mechanics and its Applications*. 586. 126438.

# Anhang

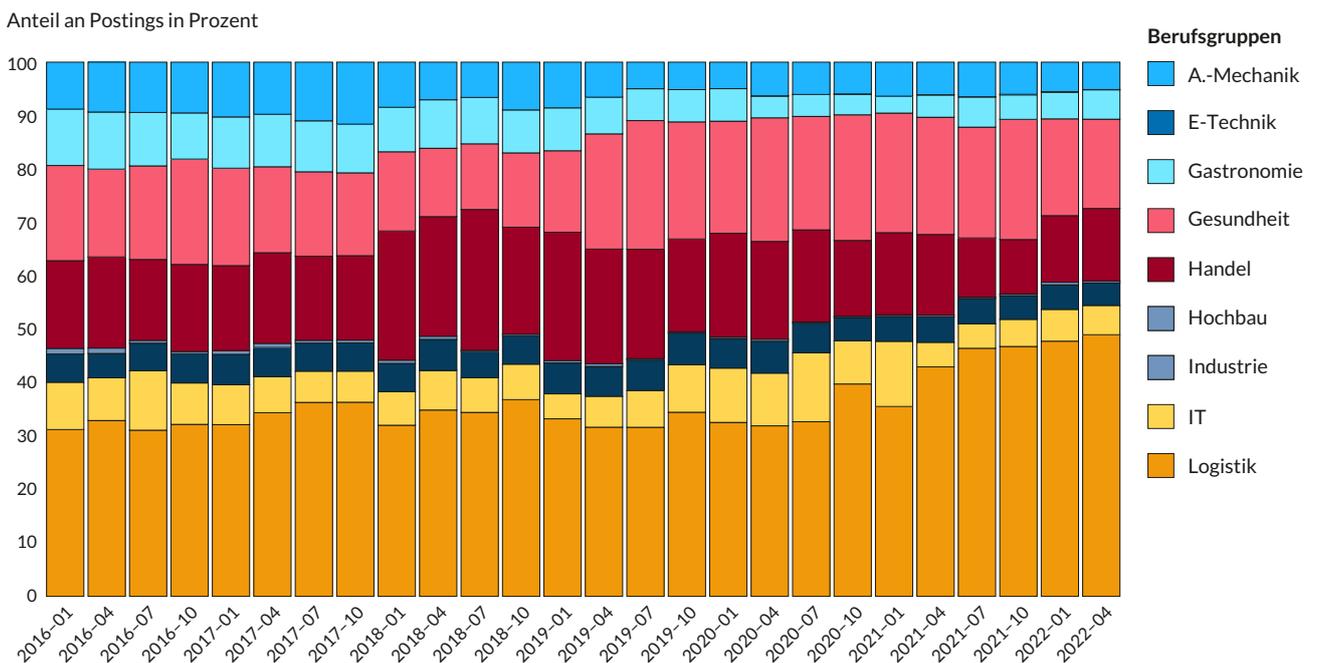
ABBILDUNG A1 Zeitreihenanalyse, Verteilung der Job-Postings über die deutschen Bundesländer



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

ABBILDUNG A2 Zeitreihenanalyse, Verhältnis zwischen den Berufsgruppen



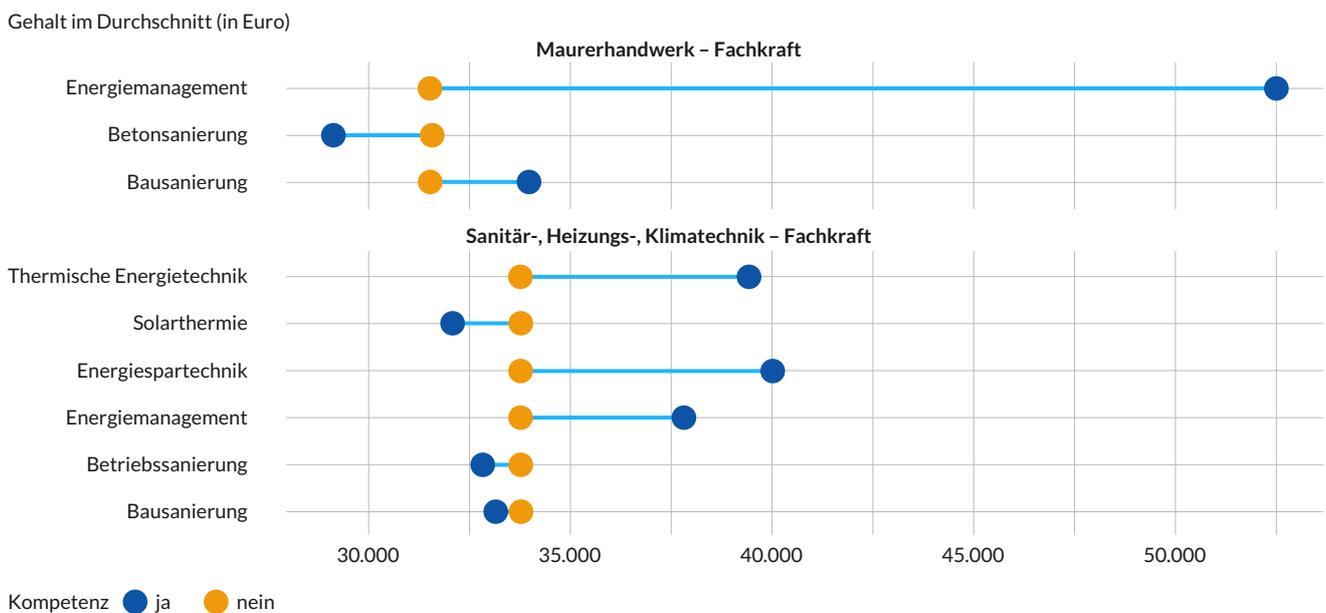
Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

Auf der X-Achse der Abbildung A1 ist der Zeitverlauf abgetragen, auf der Y-Achse die Anzahl der Postings. Jeweils eine Farbe stellt ein Bundesland dar, wie der Legende rechts der Grafik zu entnehmen ist. Da die Position der Bundesländer konstant bleibt, kann eine Änderung der Verhältnisse über die Zeit leicht erfasst werden.

Abbildung A2 hält die Entwicklung der Anzahl an OJA-Postings pro Berufsgruppe über die Zeit fest. Die Berufsgruppen sind in verschiedenen Farben voneinander abgesetzt und in einer gleichbleibenden Reihenfolge visualisiert. So kann man Veränderungen in den Mengenverhältnissen über die Zeit schnell identifizieren.

ABBILDUNG A3 Zeitreihenanalyse, Gehalt im Durchschnitt (Zeitraum 2020–2022) ausgewählter Green Skills



Quelle: Online Job Advertisements (2020–2022).

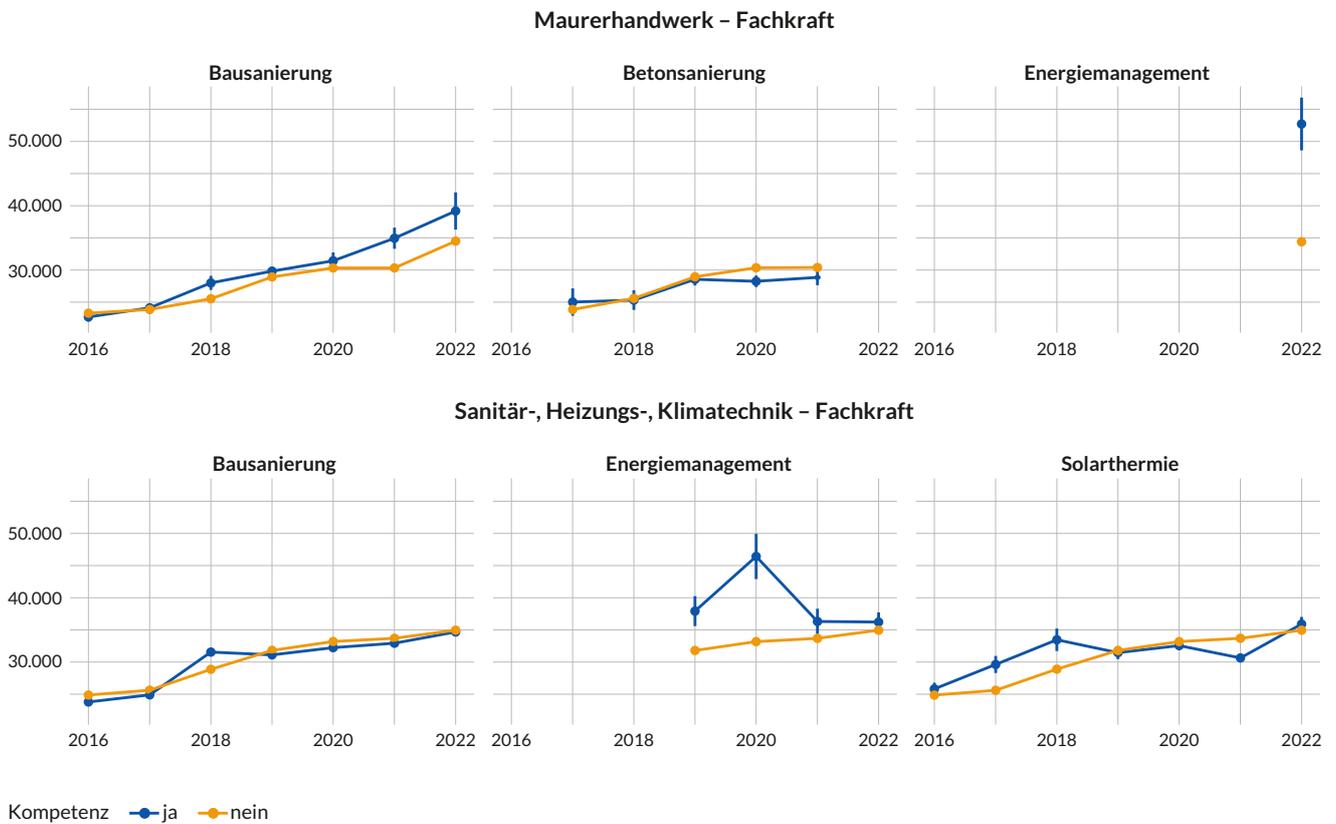
| BertelsmannStiftung

In Abbildung A3 sehen wir das Gehalt im Durchschnitt für ausgewählte Kompetenzen im Vergleich. Auf der X-Achse ist das durchschnittliche Gehalt eingezeichnet, auf der Y-Achse sehen wir ausgewählte Green Skills im Vergleich. In blau werden hier die Durchschnittsgehälter für Jobs abgetragen, die eine entsprechende Kompetenz verlangen. In orange wird das durchschnittliche Jahresgehalt jener Ausschreibungen dargestellt, die diese Kompetenz nicht voraussetzen. Es wird deutlich, dass Green Skills eine ökonomische Prämie realisieren. Diese Prämie ist insbesondere für „Thermische Energie-

technik“, „Energiespartechnik“ sowie „Energiemanagement“ ausgeprägt. Interessant sind die Unterschiede zwischen den Berufen. Während die „Bausanierung“ für Maurerfachkräfte eine Prämie bietet, ist diese im Beruf der Fachkraft für Sanitär-, Heizungs- und Klimatechnik nicht sichtbar. Gleichzeitig hat „Energiemanagement“ für das Maurerhandwerk eine deutlich höhere Prämie als im Bereich Sanitär-, Heizungs- und Klimatechnik. Dies könnte auf die wachsende Nachfrage an energieeffizienter Sanierung von Gebäuden wie im Bereich Wärmedämmung zurückzuführen sein.

ABBILDUNG A4 Entwicklung der Gehälter pro Berufsgruppe

Gehalt im Jahresdurchschnitt (in Euro)

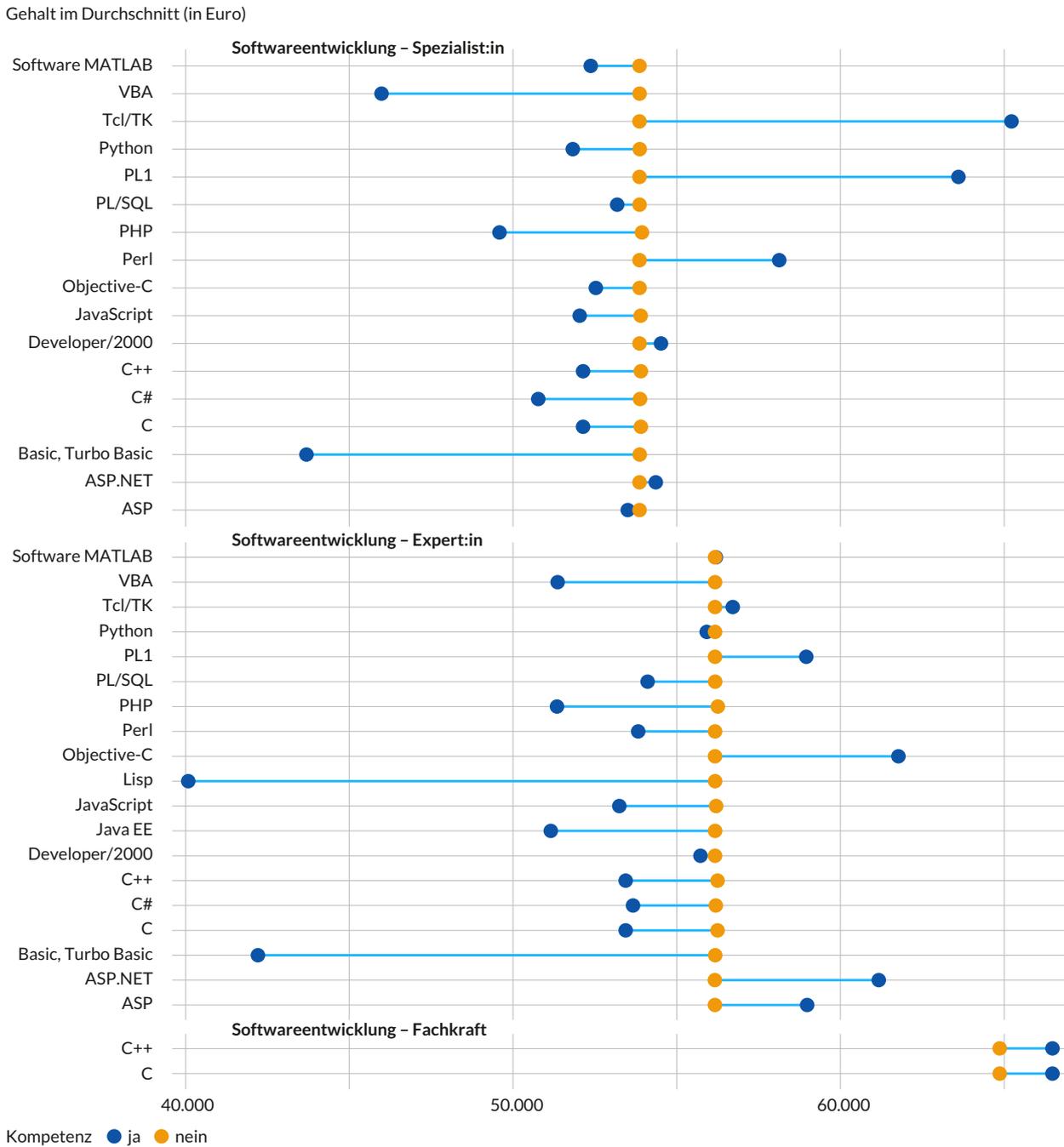


Quelle: Online Job Advertisements (2016–2022).

BertelsmannStiftung

Abbildung A4 stellt die Gehaltsentwicklung für zwei verschiedene Fachkräfte im Jahresdurchschnitt über die Zeit dar. Auf der X-Achse ist jeweils der Zeitverlauf in Jahren eingetragen, auf der Y-Achse das Gehalt im Jahresdurchschnitt in Euro.

ABBILDUNG A5 Zeitreihenanalyse, Gehalt im Durchschnitt Zeitraum 2020–2022 pro Sprache



Quelle: Online Job Advertisements (2020–2022).

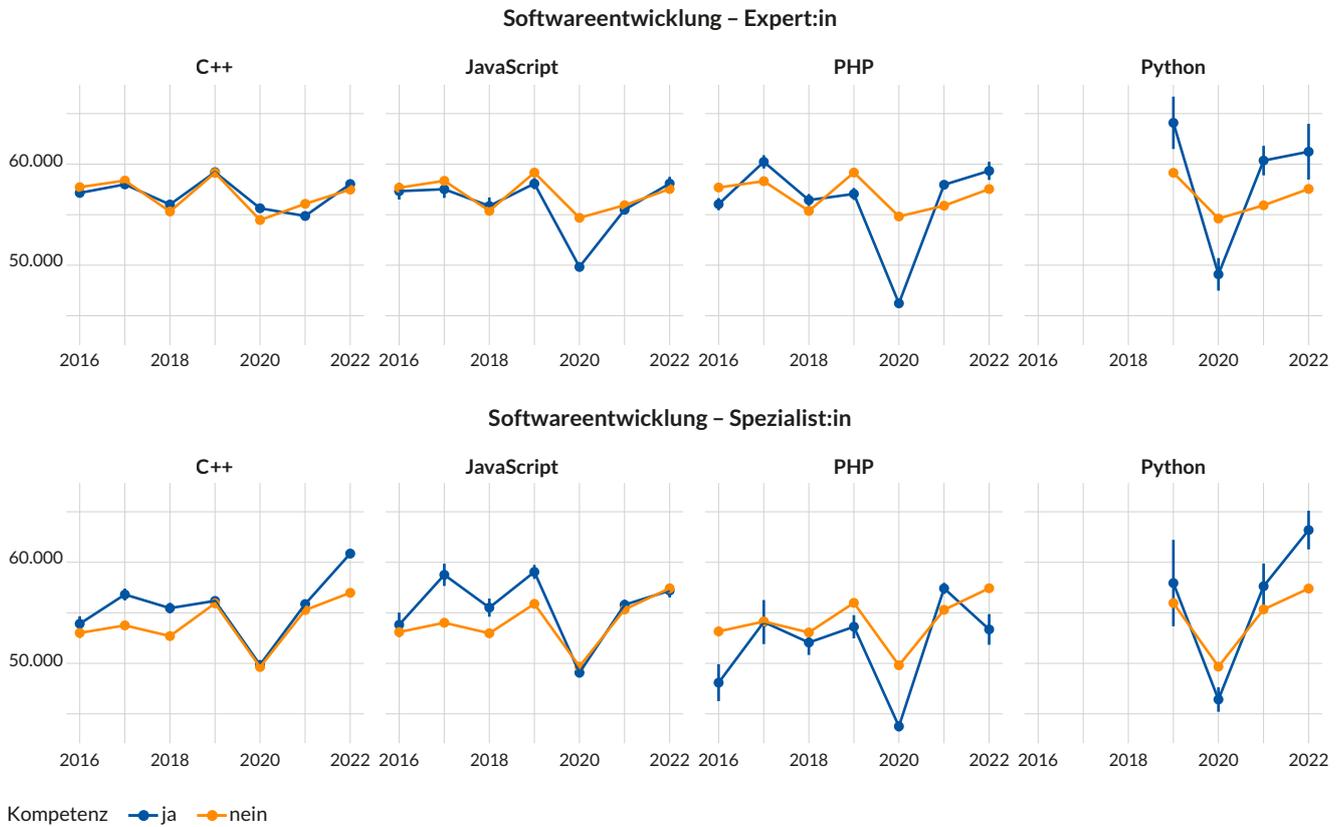
| BertelsmannStiftung

Ähnlich wie bei den Green Skills ist in Abbildung A5 auf der X-Achse das durchschnittliche Gehalt eingezeichnet, auf der Y-Achse sehen wir ausgewählte Coding Skills im Vergleich. In dunkelblau werden hier die Durchschnittsgehälter für Jobs abgetragen, die eine entsprechende Kompetenz verlangen. In orange wird das durchschnittliche Jahresgehalt jener Ausschreibungen dargestellt, die diese Kompetenz nicht voraus-

setzen. Klare Level-Unterschiede sind erkennbar in der Unterscheidung zwischen den Beschreibungen „Spezialist:in“ zu „Expert:in“ zu „Fachkraft“. Einige Programmiersprachen scheinen eine negative Prämie zu haben. Dies könnte darauf zurückzuführen sein, dass es sich um veraltete und wenig anspruchsvolle Programmiersprachen, wie „Basic“ oder „Turbo Basic“, handelt.

ABBILDUNG A6 Gehalt im Jahresdurchschnitt

Gehalt im Jahresdurchschnitt (in Euro)

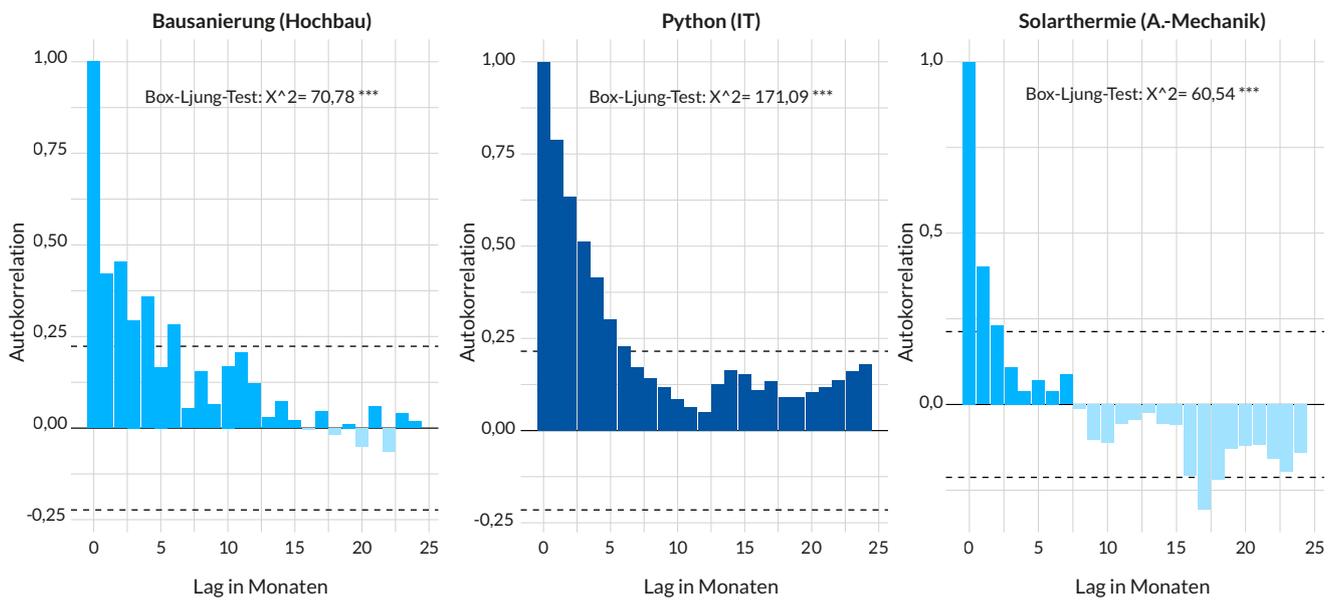


Quelle: Online Job Vacancies in Anlagenmechanik und Hochbau (2016–2022).

| BertelsmannStiftung

In Abbildung A6 sehen wir die Entwicklung von Prämien für ausgewählte Sprachen. Auf der X-Achse ist jeweils der Zeitraum abgetragen, auf der Y-Achse das Jahresgehalt in Euro.

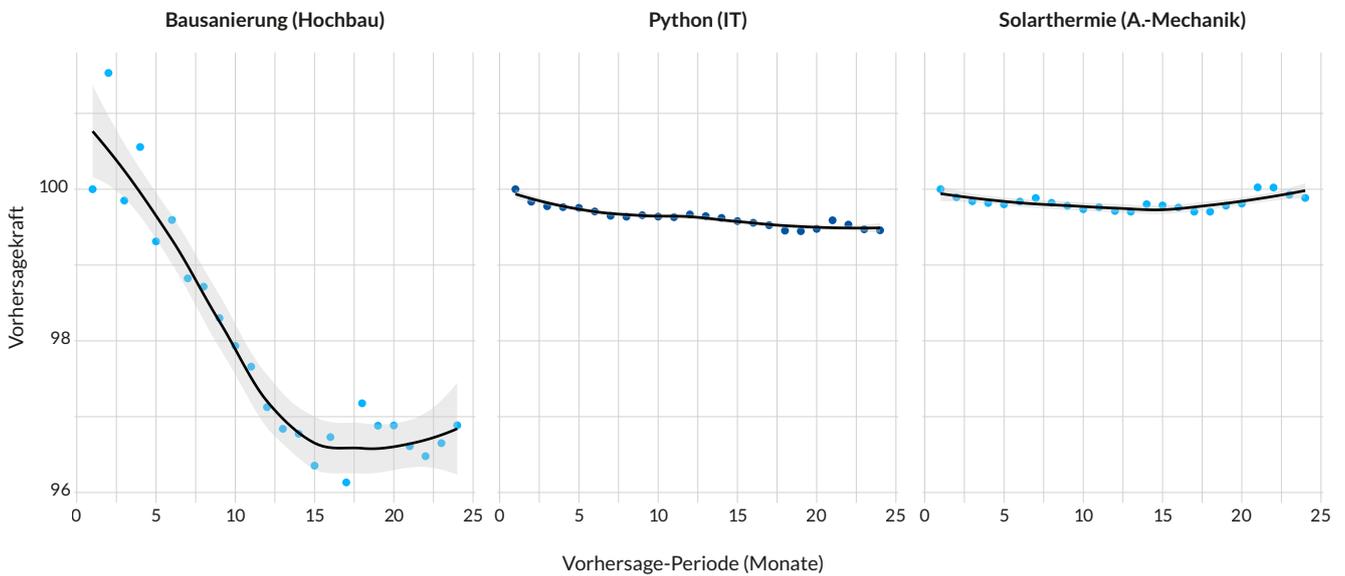
ABBILDUNG A7 Autokorrelationsanalyse, Zeitraum 25 Monate



Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

ABBILDUNG A8 Cross-Validation Analyse von Vorhersagen im 25-Monate Fenster



Anmerkung: Vorhersagekraft gemessen an 1-RSME (at t=0): 100

Quelle: Online Job Advertisements (OJA).

| BertelsmannStiftung

# Abbildungsverzeichnis

---

ABBILDUNG 1	Zeitreihenanalyse, Anzahl an Postings 2016-2022	9
ABBILDUNG 2	Verteilung der Postings auf Berufsgruppen	9
TABELLE 1	Einteilung der Berufsgruppen	10
ABBILDUNG 3	Was sind <i>Saisonalität</i> , <i>Zyklus</i> und <i>Trend</i> – Beispiel „Klimatechnik“	12
ABBILDUNG 4	Beispiel der Metrik <i>Häufigkeit</i>	15
ABBILDUNG 5	Beispiel der Metrik <i>Zentralität</i>	15
ABBILDUNG 6	Beispiel der Metrik <i>Prämie</i>	16
ABBILDUNG 7	Zeitreihenanalyse, Beispiel der Metrik <i>Häufigkeit</i>	17
ABBILDUNG 8	Häufigkeit von Green Skills pro 10.000 Postings	19
ABBILDUNG 9	Verteilung der Metriken <i>Häufigkeit</i> , <i>Zentralität</i> und <i>Prämie</i> auf Green Skills	20
ABBILDUNG 10	Zeitreihenanalyse, Entwicklung der Metrik <i>Häufigkeit</i> pro 100 Postings im Vergleich	21
ABBILDUNG 11	<i>Häufigkeit</i> von Programmiersprachen in OJAs	22
ABBILDUNG 12	<i>Häufigkeit</i> , <i>Zentralität</i> und <i>Prämie</i> von Programmiersprachen in OJAs	23
ABBILDUNG 13	Zeitreihenanalyse, Entwicklung der Metrik <i>Häufigkeit</i> für ausgewählte Programmiersprachen	24
ABBILDUNG 14	Zeitreihenanalyse „Bausanierung“ (Hochbau) sowie Vorhersage von Zeitreihenwerten	25
ABBILDUNG 15	Zeitreihenanalyse „Python“ (IT) sowie Vorhersage von Zeitreihenwerten	26
ABBILDUNG 16	Zeitreihenanalyse „Solarthermie“ (Anlagenmechanik) sowie Vorhersage von Zeitreihenwerten	27
ABBILDUNG 17	Entwicklung der Gehälter im Jahresdurchschnitt zwischen beiden Gruppen	28
ABBILDUNG 18	Netzwerkanalyse über die Zeit von „Python“ (IT)	29
ABBILDUNG 19	Netzwerkanalyse über die Zeit von „Bausanierung“ (Hochbau)	30
ABBILDUNG A1	Zeitreihenanalyse, Verteilung der Job-Postings über die deutschen Bundesländer	34
ABBILDUNG A2	Zeitreihenanalyse, Verhältnis zwischen den Berufsgruppen	34
ABBILDUNG A3	Zeitreihenanalyse, Gehalt im Durchschnitt (Zeitraum 2020-2022) ausgewählter Green Skills	35
ABBILDUNG A4	Entwicklung der Gehälter pro Berufsgruppe	36
ABBILDUNG A5	Zeitreihenanalyse, Gehalt im Durchschnitt Zeitraum 2020-2022 pro Sprache	37
ABBILDUNG A6	Zeitreihenanalyse, Entwicklung von Gehältern im Jahresdurchschnitt, Zeitraum 2016-2022	38
ABBILDUNG A7	Autokorrelationsanalyse, Zeitraum 25 Monate	39
ABBILDUNG A8	Cross-Validation Analyse von Vorhersagen im 25-Monate Fenster	39

**Bertelsmann Stiftung**

Carl-Bertelsmann-Straße 256  
33311 Gütersloh  
Deutschland  
+49 5241 810

**Larissa Klemme**

Project Manager  
Nachhaltige Soziale Marktwirtschaft  
+49 5241 8181141  
larissa.klemme@bertelsmann-stiftung.de

**Gunvald Herdin**

Senior Project Manager  
Nachhaltige Soziale Marktwirtschaft  
+49 5241 8181464  
gunvald.herdin@bertelsmann-stiftung.de

[www.bertelsmann-stiftung.de](http://www.bertelsmann-stiftung.de)