

**Schlussbericht**

**Empirie zu personalisierten Preisen im  
E-Commerce**



ibi research an der Universität Regensburg GmbH  
Galgenbergstraße 25  
93053 Regensburg

trinnoactive GmbH  
Franz-Mayer-Straße 1  
93053 Regensburg

## Inhalt

<b>1</b>	<b>Ausgangslage und Motivation .....</b>	<b>4</b>
<b>2</b>	<b>Theoretische Einordnung und Fundierung .....</b>	<b>7</b>
2.1	Bestimmung und Abgrenzung .....	7
2.2	Bewertung und bestimmende Einflüsse .....	9
2.3	Auswirkungen auf den Verbraucher .....	11
<b>3</b>	<b>Projektvorgehen .....</b>	<b>13</b>
3.1	Zielsetzung .....	13
3.2	Methodik .....	13
3.3	Vorbereitung der Datenerhebung .....	16
3.3.1	Auswahl der Plattformen und Produkte .....	16
3.3.2	Design und Entwicklung der Preisabfragesoftware .....	18
3.3.3	Bereitstellung der Infrastruktur für die automatisierte Datensammlung .....	21
3.3.4	Konfiguration der Preisabfragesoftware .....	21
3.3.5	Durchführung von Qualitätssicherungsmaßnahmen ...	23
3.4	Durchführung der Datenerhebung .....	28
<b>4</b>	<b>Empirische Untersuchung .....</b>	<b>31</b>
4.1	Auswahl der Auswertungsmethodik .....	31
4.2	Deskriptive Analyse .....	35
4.2.1	Verteilung der erhobenen Datensätze .....	35
4.2.2	Deskriptive Statistik der Preisdifferenzen .....	37
4.2.2.1	Preisunterschiede pro Produkt .....	39
4.2.2.2	Preisunterschiede pro Zugriffsart .....	48
4.2.2.3	Preisunterschiede pro Proxy-Standort .....	49
4.2.2.4	Preisunterschiede und Tracker .....	50
4.2.2.5	Preisunterschiede und Cookies .....	52
4.2.2.6	Preisunterschiede und Nutzerkonto .....	53
4.2.2.7	Preisunterschiede und Social-Media-Login .....	55
4.3	Regressionsanalyse .....	56
4.3.1	Einführung .....	56
4.3.2	Zur Interpretation der Regressionsergebnisse .....	58
4.3.3	Regression mit allen Datensätzen .....	58
4.3.4	Regression nach Entfernung von Produkten ohne Preisdifferenzen .....	60
4.3.5	Regression des Produktes „IntercityHotel Hamburg“ ..	62
4.3.6	Regression des Datensatzes ohne „IntercityHotel Hamburg“ .....	65
4.3.7	Regression des manuellen Erfassungsstrangs .....	66

<b>5 Zusammenfassung und Diskussion der Ergebnisse</b> .....	<b>68</b>
5.1 Deskriptive Analyse.....	68
5.2 Regressionsanalyse.....	70
5.3 Fazit und Ausblick.....	71
<b>Abbildungsverzeichnis</b> .....	<b>73</b>
<b>Tabellenverzeichnis</b> .....	<b>75</b>
<b>Literaturverzeichnis</b> .....	<b>76</b>
<b>Anhang</b> .....	<b>78</b>
<b>Über ibi research</b> .....	<b>79</b>
<b>Über trinnovative</b> .....	<b>80</b>
<b>Impressum</b> .....	<b>81</b>

# 1 Ausgangslage und Motivation

Der Ökonom Paul Krugman stellte bereits 2000 in seinem Artikel „Reckonings; What Price Fairness?“ in der *New York Times* fest, dass im E-Commerce offenbar mindestens einmal unterschiedlichen Kunden zum selben Zeitpunkt unterschiedliche Preise angeboten wurden. Diese Preisgestaltung bezeichnet er als eine „new version of an old practice: price discrimination“.<sup>1</sup> Seither erscheint diese Beobachtung als erhärtet: Waren und Dienstleistungen unterliegen immer schneller werdenden Preisveränderungen. Ob allerdings letztere von einzelnen Kundenmerkmalen, z. B. von Kaufhistorien und Endgeräten beeinflusst werden, ist meist unklar.

Klar ist hingegen, dass sich Preise für Verbraucher auf Grundlage von Schwankungen in der Angebotsmenge, aber auch nach den erwarteten Nachfragemustern über die Tageszeit hinweg ändern können. Aufgrund der in manchen Branchen nur temporär verfügbaren Produkte, beispielsweise im Bereich der Hotels und Luftfahrt, werden dort teilweise sehr kurzfristige Maßnahmen zur Kapazitätssteuerung ergriffen. Eine dieser Maßnahmen ist die Preisgestaltung. Eine Studie zur Sicht der Verbraucher auf eine dynamische und 1:1-Preisgestaltung kommt zum Ergebnis, dass unter den 1.000 befragten Personen über 18 Jahren 55 Prozent eine flexible Preisgestaltung bei Kraftstoffen an Tankstellen, 39 Prozent bei Flugtickets und 36 Prozent bei Elektrogeräten im Online-Handel bewusst wahrnehmen, wobei branchenspezifisch eine dynamische Preisbildung insbesondere bei Flugtickets am stärksten akzeptiert wird.<sup>2</sup>

Ganz allgemein lässt sich feststellen, dass im E-Commerce die Änderung von Preisen von besonderer Relevanz ist, wie es auch schon Krugman am Beispiel von *Amazon* attestierte.<sup>3</sup> Die technologische Entwicklung ermöglicht es, dass dynamische Preisgestaltungen inzwischen wirtschaftlich realisierbar und Preise auch unmittelbar in Echtzeit veränderbar sind.<sup>4</sup> Im Vergleich zum stationären Einzelhandel profitiert von diesen Entwicklungen v. a. der Online-Handel. Preise können dort kurzfristiger verändert und beispielsweise auf Basis höherer Nachfrage oder hinsichtlich der Tageszeit angepasst werden als im stationären Handel. Im Extremfall erscheint es mit entsprechendem technischem Aufwand sogar möglich, die Preise kunden- und nutzungsspezifisch zu individualisieren.

Es lassen sich ganz grundsätzlich zwei Arten der Preisdifferenzierung unterscheiden: die dynamische Preisänderung als Anpassung auf Basis objektiver,

---

<sup>1</sup> Krugman, 2000, S. 35.

<sup>2</sup> Vgl. Krämer, Kalka & Ziehe, 2016.

<sup>3</sup> Vgl. Krugman, 2000.

<sup>4</sup> Vgl. Krämer et al., 2016.

kundenunabhängiger Faktoren auf der einen Seite und die personalisierte Preisänderung als Anpassung auf Basis individueller, kundenabhängiger Faktoren auf der anderen Seite. Auf diese beiden Arten, insbesondere die personalisierte Preisanpassung, soll in der Folge noch näher eingegangen werden. Dass die personalisierte Preisänderung aus Kundensicht sehr problematisch erscheint, lassen bereits die mahnenden Worte von Krugman erahnen, der 2000 dieses Phänomen noch als dynamische Preisgestaltung erfasste.<sup>5</sup> Mit personalisierten Preisanpassungen besteht zum Beispiel die Gefahr der Diskriminierung aufgrund der Zugehörigkeit zu einer bestimmten Nutzungsgruppe. Anbieter könnten so z. B. bei Personen mit einer höheren Kaufkraft, etwa aufgrund des Wohnorts oder des benutzten Endgeräts, auch höhere Preise setzen. Kritisch zu sehen wären solche Anpassungen dann, wenn die individuelle Preissetzung für den Kunden nicht erkennbar ist bzw. deren Gründe nicht transparent sind.

Die Fachpresse, aber auch einschlägige Studien legen die Vermutung nahe, dass personalisierte Preisdifferenzierung durchaus Einzug in die Praxis gefunden hat.<sup>6</sup> So kommen beispielsweise Schleusener und Hosell (2016) in ihren Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen zum Ergebnis, dass Preisdifferenzierungen im Bereich Tourismus und Pauschalreise vorgenommen werden und sich bei höherpreisigen Pauschalreisen eine Differenzierung nach Nutzermerkmalen sowie nach Betriebssystemen ergibt. Dies sei ein klares Anzeichen für eine personalisierte Preisgestaltung.<sup>7</sup>

Dass eine Personalisierung von Preisen einen logischen Schritt innerhalb der Entwicklung des E-Commerce darstellen kann, liegt mit Blick auf andere Trends des Internets, beispielsweise im Online-Marketing – Online-Werbung oder Anzeige von Suchmaschinenergebnissen – auf der Hand. Meist beruhen diese Vermutungen auf anekdotischen Berichten oder allenfalls stichprobenartigen Untersuchungen, die einer wissenschaftlichen Überprüfung nicht standhalten. Eine umfangreiche empirische Studie ist nötig, um den Sachverhalt belastbar zu belegen, ist aber – soweit öffentlich bekannt – noch nicht durchgeführt worden.

Im Rahmen dieser Untersuchung sollen zunächst die Ansätze zu personalisierten Preisen strukturiert nach wissenschaftlich theoretischen wie empirischen Aspekten untersucht werden. Hierfür soll zunächst die Begrifflichkeit theoretisch fundiert bestimmt und abgegrenzt werden. Auch Einflüsse auf die Preisgestaltung und Auswirkungen auf den Verbraucher werden fokussiert. Zunächst wird das Ziel der Studie beschrieben und es werden Rahmenbedingungen geklärt. Die Entwicklung eines Forschungsmodells ermöglicht schließlich die empirische Auswertung. Hierfür wurden über einen Zeitraum von 18 Wochen automatisiert und 14 Wochen manuell Daten erhoben. Vor- und Hauptanalysen

---

<sup>5</sup> Vgl. Krugman, 2000.

<sup>6</sup> Vgl. auch Abschnitt .

<sup>7</sup> Vgl. Schleusener & Hosell, 2016.

stellen die Auswertung der empirischen Untersuchung sicher, die auf der Operationalisierung dynamischer Preise und deren Einflussfaktoren fußt. Eine kritische Würdigung der empirischen Untersuchung sowie eine Zusammenfassung der wesentlichen Erkenntnisse beschließen den Bericht.

Eine derartige Untersuchung in diesem Umfang wurde nach der vorliegenden Literatur bisher noch nicht durchgeführt. Dies lässt sich mit der hohen Komplexität der Aufgabenstellung erklären. Die Komplexität erwächst u. a. aus den nachfolgenden Faktoren, die sich teilweise auch gegenseitig verstärken:

- (1) Will man die hohe Kombinatorik von Einflussfaktoren im Spektrum zwischen dynamischen und personalisierten Elementen durchgängig erfassen und einzelne Faktoren hinreichend isolieren, so erfordert dies einen sehr umfangreichen und systematischen Versuchsaufbau, der eine hohe Anzahl von Beobachtungspunkten generiert. Hierzu wurde neben dem automatisierten Strang auch ein manueller Versuchsaufbau konzipiert. Auf dieser Grundlage und unter Generierung zahlreicher Daten konnten signifikante Ergebnisse gewonnen werden, um die zentrale Forschungsfrage nach der Existenz einer personalisierten Preisdifferenzierung zu klären.
- (2) Eine weitere Komplexität erschließt sich aus der Datensammlung selbst. Die Gewinnung der Datenpunkte sollte ohne Wissen und Einflussnahme der gewählten Anbieter (Online-Shops, Marktplätze und Vergleichsportale) erfolgen. Wenn die Anbieter die Datenerhebung erkennen würden, könnten sie solche Preisabfragen entweder blocken oder sich daran anpassen, etwa durch das Einspielen nicht-personalisierter Preise. Um solche Entdeckungen zu verhindern, waren (zum Teil) aufwändige Maskierungen erforderlich. Insbesondere im automatisierten Strang wurden deshalb verschiedenste Maßnahmen gewählt. Der Versuchsaufbau wird in Abschnitt 3 im Detail beschrieben.
- (3) Ein weiteres Problem im Kontext der Preisabfragen ergibt sich durch die Einbindung von Benutzerprofilen. Für einige Webseiten wie Amazon müssen die Surf- und Kaufhistorie sowie die Profile in sozialen Netzwerken übereinstimmen und auf einem ausreichend langen – das heißt in der Regel über Jahre reichenden – Betrachtungszeitraum basieren. Diese Authentizität mit ausgeprägten Eigenschaften konnte durch den manuellen Strang simuliert und eingehend untersucht werden. Hierfür wurden teilweise auch Social-Media-Profile und Kundenkonten bei den Anbietern in diesen Versuchsaufbau miteingebunden (vgl. auch hierzu Abschnitt 3).

Diesen und weiteren Komplexitäten konnten im Rahmen dieser großangelegten Untersuchung begegnet werden. Die Kombination aus einem automatisierten Versuch, der eine große Anzahl an Datenpunkten liefert, in Verbindung mit dem manuellen Test, der auch „echte“ Kunden mit Kaufhistorien und Aktivitäten in den sozialen Medien miteinbindet, stellte sich als zielführend heraus. Im folgenden Kapitel erfolgt zunächst eine theoretische Fundierung.

## 2 Theoretische Einordnung und Fundierung

### 2.1 Bestimmung und Abgrenzung

Eine einheitliche Definition des Begriffspaares „Personalisierte Preisgestaltung“ und „Dynamische Preisgestaltung“ existiert in der einschlägigen Literatur nicht. Allerdings ist dieses Phänomen nicht unbekannt und wird bereits seit den 1990er Jahren diskutiert. Artikel wie der eingangs zitierte von Paul Krugman zeigen allerdings, dass die personalisierte Preisgestaltung wohl aus der dynamischen Preisgestaltung hervorging. In der Vergangenheit wurden unter dynamischen Preisen häufig auch personalisierte Preise verstanden. So schreibt Krugman: „Recently it came to light that Amazon has been charging different customers different prices [...] the outrage of those who paid a few dollars extra suggests that “dynamic pricing” is about to become a major consumer issue, maybe even a political issue.”<sup>8</sup> Heute werden diese beiden Begriffspaare klarer differenziert und definiert.

„Dynamische Preisgestaltung“ beschreibt hierbei häufig das Vorgehen der Unternehmen, Preise auf der Zeitachse zu ändern. „Beim dynamischen Pricing wird der Verkaufspreis an den aktuellen Markt angepasst. Der Produktpreis wird damit zu einer zeitlich schnell veränderbaren Größe“, so Krämer, Kalka und Ziehe (2016) in ihrem Artikel. Charakteristisch ist, dass Preise für alle Kunden zu einem bestimmten Zeitpunkt gleich sind. Ändern sich beispielsweise Kosten für Unternehmen durch die Änderung von Bezugspreisen oder Löhnen, aber auch durch ein sich unterjährig veränderndes Nachfrageverhalten, werden auf diese Weise Preise zur Förderung des Absatzes einheitlich für alle Kunden angepasst.<sup>9</sup> Häufig kann diese Art der Preissetzung bei saisonalen Waren, z. B. bei Winterbekleidung oder Weihnachtsdekoration, aber auch bei Kraftstoffbezug an der Tankstelle beobachtet werden. Auch die Haltbarkeit von Produkten wirkt sich auf die Preissetzung im dynamischen Kontext aus.<sup>10</sup> Dass die Relevanz der dynamischen aber zunehmend auch der personalisierten Preissetzung in der Praxis und Forschung zunimmt, stellen viele Autoren übereinstimmend fest.<sup>11</sup>

Um die personalisierte Preisgestaltung definieren zu können, erscheint zunächst eine Bestimmung des Begriffs der „Personalisierung“ sinnvoll. Personalisierung erfolgt nach Müller „auf Basis von Informationen über den Kunden“<sup>12</sup>. Die Personalisierung erfordert zunächst das Sammeln von Kundendaten auf Basis einer

---

<sup>8</sup> Krugman, 2000, S. 35.

<sup>9</sup> Vgl. Rott, 2019.

<sup>10</sup> Vgl. Zander-Hayat et al., 2016.

<sup>11</sup> Vgl. u. a. Krämer et al., 2016; Rott, 2019; Schleusener & Hosell, 2016.

<sup>12</sup> Müller, 2005, S. 20.

Interaktion zwischen dem Anbieter und Kunden.<sup>13</sup> Holland (2016) erkennt in der Personalisierung das Ziel, eine Attraktivitätssteigerung für den Kunden herzustellen. Müller (2005) sieht außerdem den Nutzen im Aufbau einer Kundenbindung.

„Unter personenbezogener, individueller Preisdifferenzierung wird verstanden, dass ein Kunde auf Basis von persönlichen Daten einen individuellen Preis angezeigt bekommt, der keinem anderen Kunden angeboten wird“,<sup>14</sup> definieren Experten in einer Untersuchung und Ausarbeitung für den Sachverständigenrat für Verbraucherfragen beim Bundesministerium für Justiz und Verbraucherschutz die personalisierte Preisgestaltung. Zu den persönlichen Daten zählen sie u. a. Informationen über IP-Adressen, die Hard- und Software, den Standort, bereits besuchte Seiten (durch Cookies und Browserhistorien übermittelt), ggf. auch in Bezug auf gewisse zeitliche Faktoren.<sup>15</sup> Rott (2019) spricht bei der Personalisierung von Preisen von einer „punktgenauere[n] Abschöpfung der Kaufkraft“<sup>16</sup>.

Dass hierbei der Einfluss der Digitalisierung und die hiermit einhergehende Möglichkeit der Datensammlung, aber auch des Datenhandels, diese Form der Preisdifferenzierung erst möglich machen, liegt auf der Hand. Hierbei sollte insbesondere die Automatisierung der Reaktion, verbunden mit immer schnelleren Reaktionszeiten genannt werden, die im ersten Schritt eine häufigere Anpassung, in einem nächsten Schritt zudem eine personalisierte Anpassung unter Zugrundelegung verschiedenster Einflussfaktoren auf die Preisbildung umfasst. Diese Art der Preissetzungsstrategie wird eingesetzt, um die Zahlungsbereitschaft der Kunden in optimaler Weise auszunutzen.<sup>17</sup> Dautzenberg et al. (2018) liefern einen Überblick über aktuelle empirische Untersuchungen zur individualisierten Preisdifferenzierung und stellen fest, dass bisher vor allem geographische Faktoren (IP-Adresse), nutzerbasierte Faktoren (Nutzerprofil, Surfhistorie, Cookies, Plugins, Pixel) und teilweise auch technische Faktoren wie Betriebssystem/Browser und/oder Endgerät/Browser untersucht wurden.<sup>18</sup> Eine Einbindung von sozialen Medien sowie eine ganzheitliche Betrachtung unterschiedlichster Einflussfaktoren auf die Preisbildung über einen längeren Zeitraum in vielen Branchen fehlt aber noch.

---

<sup>13</sup> Vgl. Müller, 2005.

<sup>14</sup> Schleusener & Hosell, 2015, S. 2.

<sup>15</sup> Vgl. Schleusener & Hosell, 2015, S. 2.

<sup>16</sup> Rott, 2019, S. 287.

<sup>17</sup> Vgl. Rott, 2019.

<sup>18</sup> Vgl. Dautzenberg, Gaßmann, Groß, Müller, Neukamp, Schmidtke & Bodenstein, 2018.

## 2.2 Bewertung und bestimmende Einflüsse

Die größte Herausforderung stellen aktuell noch die klassische Preissetzung und Preisfindung dar. Eine Anwendung von Preisdifferenzierung oder dynamischer Preissetzung wird derzeit von einer Mehrheit der Händler oder Unternehmen nicht priorisiert betrachtet. Die Applikation von dynamischen Preissetzungs-Methoden wird in Zukunft mit steigender Komplexität bewertet, da Herausforderungen durch größere zur Verfügung stehende Datenmengen (Big Data) entstehen. Zudem wurden durch die Schwerpunktstudie über die Preisdifferenzierung im Handel der Universität Köln aus dem Jahr 2017 festgestellt, dass gemeinhin wenig Preisveränderungen vorgenommen werden, jedoch eine hohe Preisdispersion besteht. Daraus ergab sich die Erkenntnis, dass Preisänderungen vor allem online aber kaum im stationären Handel stattfinden. Des Weiteren konnte vor allem bei jüngeren Menschen eine Gefährdung der Vertrauensbasis zu Unternehmen durch Preisdifferenzierung beobachtet werden. Preisänderungen werden hierbei vom Konsumenten als unfair eingestuft.<sup>19</sup>

Dabei haben sich vier Themen im Bezug zur Preisgestaltung bei Unternehmen und im Handel als relevant herausgestellt: Allgemeine Preissetzung, Preisharmonisierung/-differenzierung, Preisreferenz im Wettbewerbsumfeld und Preisdifferenzierung nach Kundengruppen/Individualisierung.<sup>20</sup>

- (1) Allgemeine Preissetzung: optimale Preisgestaltungs-Strategien bestimmen, unabhängig von Differenzierungs- und Individualisierungsmaßnahmen
- (2) Preisharmonisierung/-differenzierung: Preisgestaltungs-Entscheidungen von Multi-Channel Händlern zwischen unterschiedlichen Verkaufskanälen
- (3) Preisreferenz im Wettbewerbsumfeld: relevanten Markt und (preis-)relevante Wettbewerber bestimmen
- (4) Preisdifferenzierung nach Kundengruppen/Individualisierung: Preisveränderungen auf Basis von verhaltens-, merkmalsbezogenen und zeitabhängigen Veränderungen<sup>21</sup>

Ein Nachweis für dynamische Preisgestaltung gelingt bei angewandter zeitlicher Preisvariation. Handelt es sich jedoch um personalisierte Preise, ist dies nur schwer möglich.<sup>22</sup>

---

<sup>19</sup> Vgl. Reinartz, Haucap, Wiegand & Hunold, 2017.

<sup>20</sup> Vgl. *ibid.*

<sup>21</sup> Vgl. *ibid.*

<sup>22</sup> Vgl. Rott, 2019.

Die Wiener Arbeitskammer erhob im Jahr 2017 im Tourismus- und Reisesektor, welche Preisstrategien angewandt werden. In diesem Bereich ist eine Preisgestaltung oft sehr dynamisch umgesetzt und Preise können sich innerhalb weniger Minuten ändern. Die Ausprägung der jeweiligen Dynamik hängt dabei von der Branche ab, bei Hotelportalen oder Fluganbietern sind solche Änderungen rechtlich zulässig und können auch ökonomisch sinnvoll sein. Zu einer Einordnung sollte deshalb die Branche einbezogen werden.<sup>23</sup>

Ein wichtiges digitales Thema ist die Koordinierung der Preise zwischen stationärem und elektronischem Handel. Dabei hat auch die zunehmende Digitalisierung Einfluss auf die Dynamisierung der Preisgestaltung und das vor allem im Online-Handel. Durch die Digitalisierung kann eine bessere Vorhersagbarkeit des Nachfrageverhaltens erzielt werden, sowie schneller auf sich verändernde Wettbewerbsbedingungen reagiert werden. Daneben erlauben Preisvergleichsportale eine bessere Übersicht für Kunden. Diese neu entstandenen Möglichkeiten im Online-Bereich erlauben eine schnellere Reaktionszeit auf Ereignisse und führen somit zu häufigeren Preisanpassungen.<sup>24</sup>

Ein Einflussfaktor ist das Ziel einer attraktiven Online-Preisgestaltung, um Kaufanreiz und -bereitschaft beim Kunden zu erzeugen. Ein Grund dafür ist die Informationseinholung der Kunden. Diese informieren sich häufiger online über Preisentwicklungen als offline für den stationären Handel.<sup>25</sup>

Als Einflussfaktoren der Preisgestaltung werden zudem die unterschiedliche Dynamik, die Standortdifferenzierung, die Darstellung auf Endgeräten, die verwendeten Browser, die Versandkosten als Preisschraube und die Attribute einer Person angesehen. Auch die DSGVO-Konformität, die Branche und das Umsatzziel spielen eine Rolle bei der Preissetzung.

Echtzeitumsetzbarkeit sowie wirtschaftliche Realisierbarkeit müssen gewährleistet sein. So stellten auch diese Argumente Einflussfaktoren dar. Des Weiteren werden die Standortdifferenzierung und die individuellen Kundendaten als Basis genutzt und durch technische Aufbereitung zur Preisgestaltung angewandt. Eine Verknüpfung aller Unternehmensbereiche über alle Vertriebskanäle hinweg und eine Verbindung von extrahierten Kundenkriterien führt zu einer IT-gestützten Preisgestaltung.<sup>26</sup>

---

<sup>23</sup> Vgl. Österreichisches Institut für angewandte Telekommunikation, AIT Austrian Institute of Technology GmbH, 2019.

<sup>24</sup> Vgl. Reinartz et al., 2017; Rott, 2019.

<sup>25</sup> Vgl. Reinartz et al., 2017.

<sup>26</sup> Vgl. Krämer et al., 2016.

## 2.3 Auswirkungen auf den Verbraucher

Eine dynamische Preisgestaltung (Dynamisierung) von Preisen kann sich negativ und positiv auf den einzelnen Verbraucher auswirken. Bei einer Dynamisierung verändern Waren und Dienstleistungen ihren Preis im Zeitverlauf, was dazu führen kann, dass diese auch unter dem Durchschnittspreis angeboten werden. Im Gegensatz dazu erhalten bei einer personalisierten Preisgestaltung (Personalisierung) einzelne Verbraucher gezielt aufgrund extrahierter Einflussfaktoren einen günstigeren oder teureren Preis angezeigt als der Durchschnitt.

Die Anwendung einer Preispersonalisierung als Verkaufsstrategie wird jedoch von Verbraucherschützern sowie dem Bundeskartellamt als sehr kritisch angesehen, da sie eine negative Auswirkung auf Kundenbeziehung ausüben kann. Eine erhöhte Anzahl an Verbrauchbeschwerden im Bereich von Preisänderungen zeigt, dass Verbraucher starke Preisunterschiede bei einem Anbieter als großes Ärgernis betrachten. Häufig genannt wird hier das Beispiel von Tankstellen, die eine dynamische Preisgestaltung anwenden. Dadurch ändern sich die Preise mehrmals über den Tagesverlauf hinweg und manche Verbraucher erhalten einen besseren Preis als andere. Diese Schwankungen erzeugen somit ein Gefühl der ungleichen Behandlung und können als unfair angesehen werden. Studien ergaben, dass Verbraucher, auch wenn sie von solchen Schwankungen profitieren, diese immer noch negativ assoziieren.<sup>27</sup> Das gleiche gilt für Preisschwankungen im stationären und Online-Handel. Zwischen Händler und Verbraucher entsteht so eine Intransparenz über den Preis der angebotenen Waren oder Dienstleistungen, was zu Unzufriedenheit bei Verbrauchern und möglicherweise zu vermehrten Einkaufsabbrüchen führen kann. Haben bisher Internethändler solche Strategien angewandt, erfolgte regelmäßig eine negative Berichterstattung in den Medien.<sup>28</sup> Aus diesem Grund stellten Internetanbieter immer häufiger die Anwendung solcher Praktiken ein und Unternehmen wie Supermärkte wenden die Strategie der Preispersonalisierung aus Angst vor einer negativen Auswirkung auf die Kundenbeziehung wohl nicht an.<sup>29</sup>

Dem Ansatz der Personalisierung von Preisen wird ein distributiver Effekt zugunsten einkommensschwacher Verbraucher zugeordnet. Obwohl hier weniger Kaufkraft zugrunde liegt als bei einkommensstarken Verbrauchern, sollen durch personalisierte Preise nicht die Erwerbskosten gerecht verteilt, sondern lediglich eine Gewinnsteigerung aus Sicht der Anbieter erzielt werden.<sup>30</sup>

---

<sup>27</sup> Vgl. Reinartz et al., 2017.

<sup>28</sup> Vgl. Hannak et al., 2014; Reinartz et al., 2017.

<sup>29</sup> Vgl. Hannak et al., 2014; Reinartz et al., 2017; Schwaiger & Hufnagel, 2018.

<sup>30</sup> Vgl. Obergfell, 2017; Tietjen & Flöter, 2017; Genth, 2016.

Umgekehrt werden ggf. anderen Verbrauchern ungünstigere Konditionen angeboten. Geschieht eine solche Preispersonalisierung auf Basis persönlicher Merkmale von Verbrauchern, liegt eine Diskriminierung vor.<sup>31</sup>

Insbesondere sog. personalisierte Preise, bei denen personenabhängig unterschiedliche Preise für gleiche Produkte zur selben Zeit verlangt werden, stehen im Fokus einiger Kritiker.

Darüber hinaus können erhebliche Unterschiede zwischen den verschiedenen Formen der Preisdifferenzierung identifiziert werden. Am fairsten wird eine Bevorteilung von Mitgliedern in einem Loyalitätsprogramm (Besitzer sog. Kundenkarten) wahrgenommen, sowohl von bevorteilten Personen mit Mitgliedschaft als auch von benachteiligten Personen ohne Mitgliedschaft. Dieses Ergebnis entspricht den oben diskutierten Mechanismen: Die Mitgliedschaft in einem Loyalitätsprogramm wird über eine Selbstselektion („self selection“) herbeigeführt, d. h. Kunden entscheiden eigenmächtig, ob sie am Programm teilnehmen möchten. Somit unterliegt diese Statusänderung einer hohen Kontrolle durch den Kunden. Auch ist es allgemein anerkannt, dass Mitglieder bestimmte Privilegien (insb. Preisnachlässe) erfahren und hierfür im Gegenzug dem Unternehmen bspw. ihre persönlichen Daten und ihr Einkaufsverhalten preisgeben. Darüber hinaus existieren Kundenkarten bereits seit einiger Zeit, sodass ein gewisser Gewöhnungseffekt vermutet werden kann. Es gilt jedoch zu bedenken, dass selbst bei dieser als relativ fair wahrgenommenen Maßnahme immer noch jeder zweite bevorteilte Kunde Vorbehalte bzgl. ihrer Fairness äußert und knapp 75 Prozent der benachteiligten Kunden auch solche Rabatte nicht als fair erachten.

In der Tat kommen die meisten Beispiele für negative Wohlfahrtseffekte der Preisdifferenzierung bislang aus der Finanzbranche, von Versicherungen oder aus anderen Branchen mit länger laufenden Verträgen (etwa im Mobilfunk). Hier sind personalisierte Preise jedoch schon lange Standard (etwa bei Kreditverträgen, Schadensfreiheitsrabatten bei Kfz-Versicherungen).

Wenn Kunden im Internet mit Preissuchmaschinen Produktpreise vor dem Kauf gut vergleichen können, so sind der Personalisierung der Preissetzung allerdings auch hier Grenzen gesetzt, zumindest bei Wettbewerb.

Der Sachverständigenrat für Verbraucherfragen stellt auf seiner Website fest, dass personalisierte Preise in der Praxis bisher nur vereinzelt nachweisbar seien.<sup>32</sup>

---

<sup>31</sup> Vgl. Zeitschrift für Datenschutz, 2014.

<sup>32</sup> Vgl. Reinartz et al., 2017.

## 3 Projektvorgehen

### 3.1 Zielsetzung

Gemäß der Leistungsbeschreibung des Auftraggebers vom 25. Januar 2019 sollten mittels einer empirischen Untersuchung Erkenntnisse über das Vorhandensein sowie ggf. den Umfang und die Art und Weise der gegenwärtigen Verwendung von personalisierten Preisen bei E-Commerce-Plattformen, Online-Händlern, Web-Shops und Preisvergleichsportalen gewonnen werden. Im Rahmen der Untersuchung sollten systematisch über einen längeren Zeitraum und unter verschiedenen Nutzerszenarien Preisdaten von den wichtigsten Shops und Portalen erhoben und gespeichert werden.

Diese Daten sollten anschließend mit Methoden der deskriptiven Statistik sowie der multivariaten Datenanalyse ausgewertet werden.

Die interessierenden übergeordneten Forschungsfragen der Untersuchung waren damit:

- Wird im Markt personalisierte Preisgestaltung verwendet?
- Von welchen Kriterien hängt – für den Fall der Verwendung – die personalisierte Preisgestaltung ab und welches Ausmaß haben ihre Auswirkungen?

Hervorzuheben ist dabei die in Kapitel 2 beschriebene Differenzierung zwischen üblichen, auf Marktvorgängen oder auf objektiven Gegebenheiten basierenden (auch dynamischen) Preisänderungen einerseits und solchen, die auf subjektiven personenbezogenen Daten des Kunden beruhen, andererseits. Erstere sind aus Sicht des Verbraucherschutzes nicht zu beanstanden, während letztere für den Fall ihres Auftretens in dieser Hinsicht problematisch wären.

### 3.2 Methodik

Das Design der Untersuchung orientierte sich an der oben formulierten Zielsetzung und dem vom Auftraggeber geforderten Umfang. Es konkretisiert sich wie folgt:

Über einen Zeitraum von drei Monaten wurden die Preise erhoben, die auf unterschiedlichen Anbieter- und Preisvergleichsseiten für jeweils ein und das gleiche Produkt zum jeweils gleichen Zeitpunkt verlangt werden.

Dabei wurden insbesondere die umsatzstärksten Online-Shops und -Plattformen im deutschen E-Commerce sowie die beliebtesten Preisvergleichsportale berücksichtigt. Zur Bestimmung der zu untersuchenden Shops bzw. Plattformen wurden öffentlich verfügbare Umsatzinformationen herangezogen.

Es wurde darauf geachtet, eine möglichst breite Palette an Produktkategorien einzubeziehen. Zur Begrenzung des Datenvolumens wurde nicht jede Kategorie in jedem Shop abgefragt, sondern es fand eine Zuordnung entsprechend des typischen Angebots der Shops statt.

Ebenfalls untersucht wurden Faktoren wie der Zeitpunkt des Kaufs (Wochentage, Uhrzeiten). Dabei sollten entsprechend des typischen Kaufverhaltens mindestens vier Zeitpunkte pro Tag abgefragt werden.

Die Preise sollten daraufhin untersucht werden, ob die Preisgestaltung von nutzerindividuellen Merkmalen beeinflusst wird. Dazu zählen besonders, aber nicht ausschließlich, „technische“ Faktoren wie die Nutzung unterschiedlicher Endgeräte, Betriebssysteme, Browser oder das Setzen unterschiedlicher Datenschutzeinstellungen. In der Summe ergaben sich hier mindestens 100 relevante Kombinationen. Die Preise wurden zudem abhängig vom Standort des Nutzers untersucht. Hierfür wurden IP-Adressen aus unterschiedlichen Regionen Deutschlands genutzt. Die Preise wurden auch in Abhängigkeit der Surfhistorie und/oder Social-Media-Aktivitäten der Nutzer untersucht.

Dabei war jeweils besonders darauf zu achten, dass dynamische, aber nicht personalisierte Faktoren und personalisierte Faktoren separiert werden können.

Um die oben genannte konkretisierte Zielsetzung erfüllen zu können, ist in der verwendeten Methodik der inhärente Konflikt zwischen Rigorosität und Relevanz bestmöglich zu adressieren:

- Rigorosität verlangt nach einer auch bei hoher Kombinatorik von Parametern jederzeit reproduzierbaren, auf hohen Durchsatz ausgelegten und daher möglichst automatisierten Versuchsdurchführung. Dabei sind der Authentizität und ausgeprägten Individualität des Preissuchenden aber Grenzen gesetzt, denn bei dem vorgegebenen Ressourcen- und Zeitrahmen kann man die notwendige große Zahl virtueller Nutzer mit einer jahrelange Surf-, Präsenz- und Kaufhistorie in Shops und Sozialen Medien nicht generieren. Automatisierte Abfragen verlangen also relativ einfache Profile.
- Hohe Relevanz könnte sich aber gerade dadurch ergeben, dass Preissucher unter Umständen am ehesten dann mit personalisierten Preisen konfrontiert werden, wenn sie über eine ausgeprägte Identität mit einer Vielzahl von Eigenschaften verfügen, das heißt über eine jahrelange Surf-, Präsenz- und vor allem Kaufhistorie. Eine solche Historie kann aber mit vertretbarem Aufwand nur natürlichen Personen zugeordnet werden. Dies limitiert die Zahl der Abfragen und damit die Zahl der prüf-baren Kombinationen.

In der Methodik kann eine problemadäquate Mischung aus Rigorosität und Relevanz erreicht werden, wenn zwei Stränge parallel verfolgt werden:

- In einem zahlenmäßig breiten Strang erfolgte die Datenerhebung automatisiert über eine selbst programmierte Software-Lösung, die für eine große Zahl virtueller Nutzer regelmäßig bei den untersuchten Seiten und Produkten die Preisinformationen erhob und zusammen mit den im jeweiligen Erhebungsszenario unterstellten Nutzungsfaktoren ablegte. Die oben dargestellten Eckwerte führten zu einer Zahl von rund 75.000 Datenpunkten pro Tag, das heißt rund 7 Mio. Datenpunkten im Erhebungszeitraum der drei Monate. Durch den entsprechend großen Stichprobenumfang wird gewährleistet, dass die in der anschließenden Auswertung gewonnen Aussagen statistisch belastbar sind. Außerdem kann so sichergestellt werden, dass die personalisierte und die rein dynamische nutzerunabhängige Preissetzung mit einem akzeptablen statistischen Konfidenzniveau voneinander unterschieden werden können.
- In einem zahlenmäßig schmäleren Strang tätigten natürliche Personen manuelle Abfragen. Diese decken das Spektrum von „wenig kaufkräftig“ bis „sehr kaufkräftig“ ab und bringen sehr unterschiedliche Surf- und Kaufhistorien mit sich. Es konnte dabei aber nicht die volle Kombinatorik z. B. aller Endgeräte, Betriebssysteme, Browser und Standorte getestet werden.

Für die Auswertung wurde zunächst eine deskriptive Analyse der erhobenen Preisdaten vorgenommen. Hier erfolgte die Bestimmung der Mittelwerte der erhobenen Preise für die einzelnen Produkte bzw. Produktkategorien sowie anderer relevanter Verteilungsmaße. Diese Auswertung konnte bereits erste Hinweise auf das Vorhandensein oder Nichtvorhandensein nutzerindividueller Preissetzung liefern.

Dabei ist zu beachten, dass es nicht auf die Veränderung des Preises an sich ankommt. Entscheidend sind vielmehr Preisabweichungen. Daher wurde ein „Standardfall“ definiert, der als Grundlage für Preisvergleiche dient. Dieser entsprach einem Nutzer ohne Surf- und Kaufhistorie, der an einem Windows-Desktop-PC mit Google Chrome zum ersten Mal die entsprechende Seite besuchte.

Im Zentrum der Analyse steht aber die multivariate Untersuchung der erhobenen Daten. Hierfür wurde die multiple lineare Regression herangezogen. Diese ermöglicht die Erklärung eines abhängigen Faktors (nämlich der relativen Preisdifferenz) durch mehrere unabhängige Faktoren (nämlich der zu untersuchenden Kriterien und deren Ausprägungen). Für das hier zu untersuchende Laborexperiment – mit angesichts der Erhebungsmethodik nahezu vollständig besetzten Datenfeldern – ist die Regressionsanalyse in Form eines Fixed-Effects-Modells die geeignetste Methode.

Damit lassen sich die Fragen nach dem Einfluss der einzelnen Kriterien sowie deren Ausmaß erklären. Das Vorgehen ermöglicht ebenfalls eine Bestimmung der Güte der Auswertung. Die Analyse wurde mittels Python durchgeführt.

### 3.3 Vorbereitung der Datenerhebung

#### 3.3.1 Auswahl der Plattformen und Produkte

Es wurden zunächst 15 E-Commerce-Anbieter und fünf Vergleichsportale, zum einen anhand der Umsatzzahlen/Nutzerzahlen und zum anderen mit Beachtung einer sinnvollen Verteilung der Produktkategorien, festgelegt und mit dem Auftraggeber abgestimmt. Tabelle 1 gibt hierüber einen Überblick.

Händler	URL	Hauptproduktsegment
Amazon	www.amazon.de	Generalist
Otto	www.otto.de	Generalist
Zalando	www.zalando.de	Bekleidung
MediaMarkt	www.mediamarkt.de	Unterhaltungselektronik
Notebooksbilliger	www.notebooksbilliger.de	Unterhaltungselektronik
Tchibo	www.tchibo.de	Generalist
H&M	www.hm.com/de	Bekleidung
About You	www.aboutyou.de	Bekleidung
IKEA	www.ikea.com/de/de	Möbel & Haushaltswaren
Thomann	www.thomann.de	Hobby & Schreibwaren
Hornbach	www.hornbach.de	DIY & Garten
dm	www.dm.de	Drogerie & Gesundheit
TUI	www.tui.de	Reisen
AVIS	www.avis.de	Mietwagen
Ryanair	www.ryanair.com/de	Flüge
Booking.com	www.booking.com	Vergleichsportal, Reisen
Idealo	www.ideal.de	Vergleichsportal, Generalist
Geizhals	www.geizhals.de	Vergleichsportal, Generalist
Günstiger.de	www.guenstiger.de	Vergleichsportal, Generalist
Check24	preisvergleich.check24.de	Vergleichsportal, Generalist

**Tabelle 1: Einbezogene Online-Shops und -Portale**

Darauffolgend wurde eine Auswahl der zu beobachtenden Produkte vorgenommen. Zur Begrenzung des Datenvolumens während der Datenerhebung wurde jeweils ein Produkt je Shop abgefragt. Dieses orientierte sich am typischen Angebot der jeweiligen Anbieter. Wenn im Laufe der Datenerhebung ein gewähltes Produkt nicht mehr verfügbar war, wurde auf ein vergleichbares Produkt (Kategorie/Preis) gewechselt. Dies war insbesondere bei der Waschmaschine und bei den Stiefeletten der Fall.

Es wurde insbesondere bei den Dienstleistungen (Hotel, Flüge, Mietwagen) darauf geachtet, dass immer das exakt selbe Produkt mit denselben Optionen zugrunde gelegt wurde. Bei Hotelzimmern beispielsweise lagen immer identischen Reisedaten, identische Zimmer (Einzelzimmer, Frühstück) und identische Stornierungsoptionen zugrunde. Tabelle 2 zeigt die Produktauswahl.

AmazonBasics Staubsauger	IntercityHotel Hamburg	Ryanair-Flug Berlin - Dublin
Apple iPhone 11	LG Monitor	Seltmann-Weiden Kaffeesevice
Avis Golf	Levis Stiefelette	Specialized Fahrrad
BH Cosmetics Lidschatten	Miele Waschmaschine WCA 030	Tchibo Koffer
Diesel Jeans	Picard Schultertasche	Tesoro Bürostuhl
HM Socken	Regal Kallax	Tui Sheraton Miramar Resort
Harley Benton Gitarre	Rotband Vorschlaghammer	UGG Stiefelette

**Tabelle 2: Einbezogene Produkte**

Ebenfalls untersucht wurden der Zeitpunkt des Kaufs (Wochentage, Uhrzeiten) und der Standort des Käufers. Hierbei wurden entsprechend des typischen Kaufverhaltens vier Zeitpunkte pro Tag festgelegt. Mittels eines Proxy-Dienstes wurden im automatisierten Strang unterschiedliche Standorte in Deutschland simuliert. Im manuellen Abfragestrang wurden ebenfalls vier unterschiedliche Zeitpunkte abgefragt. Diese orientierten sich am automatisierten Testverfahren. Entsprechend standen den Probanden im Untersuchungszeitraum täglich jeweils vier Zeitfenster ab 8:00 Uhr, 12:00 Uhr, 17:00 Uhr und 20:00 Uhr zur Verfügung.

Um mit der Testphase (erste Preisabfragen der gewählten Produkte) beginnen zu können, wurden für den automatisierten Strang die prüfbaren Einflussfaktoren sinnvoll verknüpft und hieraus 48 Gerätekombinationen erstellt. In diesem Zuge wurde auch der „Standardfall“ definiert (Desktop, Windows 10, Google Chrome mit Cookie-Akzeptanz und ohne DNT-Einstellung). Außerdem wurde für den Standardfall festgelegt, dass sich hierbei weder in ein Nutzerkonto in den einschlägigen Online-Shops und Vergleichsportalen noch in ein Social-Media-Konto eingeloggt wird. Der Standardfall wurde in jeder einzelnen Preisabfrage (vier Mal täglich) für alle Produkte über den kompletten Erhebungszeitraum abgefragt und festgehalten. Der Standardfall dient in der Auswertung als Bezugsmaßstab für die Berechnung der Preisdifferenzen; denn es kommt dort nicht auf die Veränderung der Preise an, sondern ob – zum Beispiel aufgrund individueller Preissetzung – eine Abweichung besteht.

Die einzelnen Kombinationen werden nochmals in Abbildung 1 zusammengefasst; der Standardfall ist in dieser Darstellung markiert.

Endgeräte	Betriebssysteme	Browser	Cookie Akzeptanz	Do-Not-Track [DNT]-Software
1 Desktop-PC	Mac OS X	Apple Safari	ja	ja
2 Desktop-PC	Mac OS X	Apple Safari	ja	nein
3 Desktop-PC	Mac OS X	Apple Safari	nein	ja
4 Desktop-PC	Mac OS X	Apple Safari	nein	nein
5 Desktop-PC	Windows 10	Google Chrome	ja	ja
6 Desktop-PC	Windows 10	Google Chrome	ja	nein
7 Desktop-PC	Windows 10	Google Chrome	nein	ja
8 Desktop-PC	Windows 10	Google Chrome	nein	nein
9 Desktop-PC	Windows 10	Internet Explorer	ja	ja
10 Desktop-PC	Windows 10	Internet Explorer	ja	nein
11 Desktop-PC	Windows 10	Internet Explorer	nein	ja
12 Desktop-PC	Windows 10	Internet Explorer	nein	nein
13 Desktop-PC	Windows 10	Microsoft Edge	ja	ja
14 Desktop-PC	Windows 10	Microsoft Edge	ja	nein
15 Desktop-PC	Windows 10	Microsoft Edge	nein	ja
16 Desktop-PC	Windows 10	Microsoft Edge	nein	nein
17 Desktop-PC	Windows 10	Mozilla Firefox	ja	ja
18 Desktop-PC	Windows 10	Mozilla Firefox	ja	nein
19 Desktop-PC	Windows 10	Mozilla Firefox	nein	ja
20 Desktop-PC	Windows 10	Mozilla Firefox	nein	nein
21 Desktop-PC	Linux	Mozilla Firefox	ja	ja
22 Desktop-PC	Linux	Mozilla Firefox	ja	nein
23 Desktop-PC	Linux	Mozilla Firefox	nein	ja
24 Desktop-PC	Linux	Mozilla Firefox	nein	nein
25 Desktop-PC	Linux	Google Chrome	ja	ja
26 Desktop-PC	Linux	Google Chrome	ja	nein
27 Desktop-PC	Linux	Google Chrome	nein	ja
28 Desktop-PC	Linux	Google Chrome	nein	nein
29 iPhone	iOS	Apple Safari	ja	ja
30 iPhone	iOS	Apple Safari	ja	nein
31 iPhone	iOS	Apple Safari	nein	ja
32 iPhone	iOS	Apple Safari	nein	nein
33 Smartphone Android	Android	Google Chrome	ja	ja
34 Smartphone Android	Android	Google Chrome	ja	nein
35 Smartphone Android	Android	Google Chrome	nein	ja
36 Smartphone Android	Android	Google Chrome	nein	nein
37 Android-Tablet 1	Android	Google Chrome	ja	ja
38 Android-Tablet 1	Android	Google Chrome	ja	nein
39 Android-Tablet 1	Android	Google Chrome	nein	ja
40 Android-Tablet 1	Android	Google Chrome	nein	nein
41 Android-Tablet 2	Android	Google Chrome	ja	ja
42 Android-Tablet 2	Android	Google Chrome	ja	nein
43 Android-Tablet 2	Android	Google Chrome	nein	ja
44 Android-Tablet 2	Android	Google Chrome	nein	nein
45 iPad	iOS	Apple Safari	ja	ja
46 iPad	iOS	Apple Safari	ja	nein
47 iPad	iOS	Apple Safari	nein	ja
48 iPad	iOS	Apple Safari	nein	nein

**Abbildung 1: Kombinatorik technischer Einflussfaktoren**

### 3.3.2 Design und Entwicklung der Preisabfragesoftware

#### Design der Preisabfragesoftware

Ziel der automatisierten Preisabfrage war es, die für die statistische Auswertung notwendigen Daten in hinreichenden Umfang und unterer Berücksichtigung der definierten Einflussfaktoren aufzuzeichnen. Hierzu wurde eine spezielle Software zur automatisierten Aufzeichnung von Preisdaten konzeptioniert und entwickelt. Das Design der Abfragesoftware war dabei bestimmt durch insgesamt vier Haupttreiber.

Zunächst musste die Softwarearchitektur eine hinreichende zeitliche Parallelität der Abfragen erlauben. Diese Anforderung erfüllte die Architektur der Preisabfragesoftware durch eine (theoretisch) unbegrenzte horizontale Skalierung<sup>33</sup>.

Zielwebseiten können sich im Laufe der Aufzeichnung verändern. Ebenso können während der Abfrage Fehler auftreten. Dies impliziert hohe Anforderungen

<sup>33</sup> Horizontale Skalierung: Die Skalierung erfolgt durch das Hinzufügen von weiteren Recheninstanzen, die zu einer logischen Einheit verbunden werden.

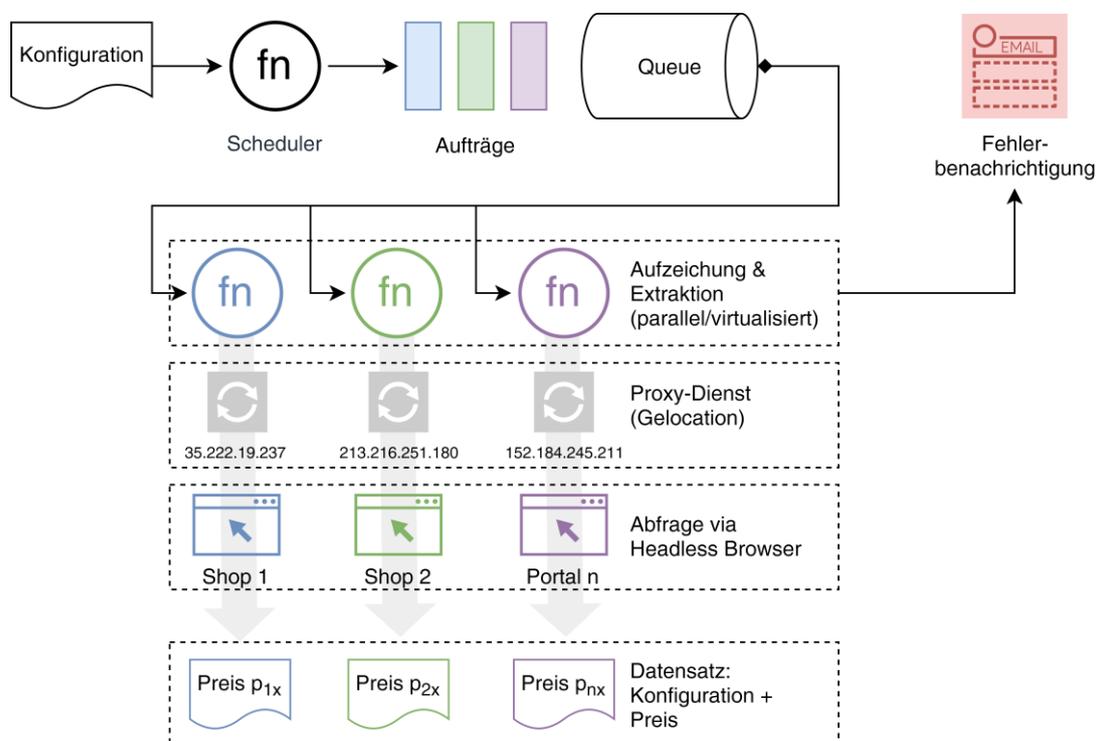
an die Robustheit der Software. In die Preisabfragesoftware wurde ein mehrschichtiges Fehler-Handling-Konzept implementiert, um diese Anforderungen umzusetzen. So waren zunächst unterschiedlich parametrisierte Abfragen untereinander unabhängig. Ein Fehler in einer Abfrage beeinflusste weitere Abfragen nicht. Trat ein Fehler in einer Abfrage auf, so wurde diese für eine festgelegte Anzahl an Versuchen automatisch wiederholt. Konnte die Abfrage dennoch nicht erfolgreich abgeschlossen werden, so wurde der fehlerhafte Versuch mit seiner Parametrierung protokolliert und ein Supportmitarbeiter benachrichtigt (vgl. hierzu auch Abschnitt 3.3.5).

Die Vermeidung der Erkennung der automatisierten Abfragen durch die Seitenbetreiber stellte einen weiteren Architekturtreiber dar. Die Preisabfrage kombinierte hierzu mehrere Gegenmaßnahmen. Durch die Verwendung von Proxy-Diensten wurde die tatsächliche (statische) IP-Adresse der Abfragesoftware verschleiert. Zudem erlaubte dies die Simulation von unterschiedlichen Standorten.

Schlussendlich musste die Preisabfragesoftware die ermittelten Daten in geeigneter Form für die nachgelagerte statistische Auswertung bereitstellen.

### **Umsetzung und Implementierung der Preisabfragesoftware**

Konkret manifestierte sich die Umsetzung der vorab beschriebenen nicht-funktionalen Anforderungen an die Preisabfragesoftware in der in Abbildung 2 dargestellten Grobarchitektur.

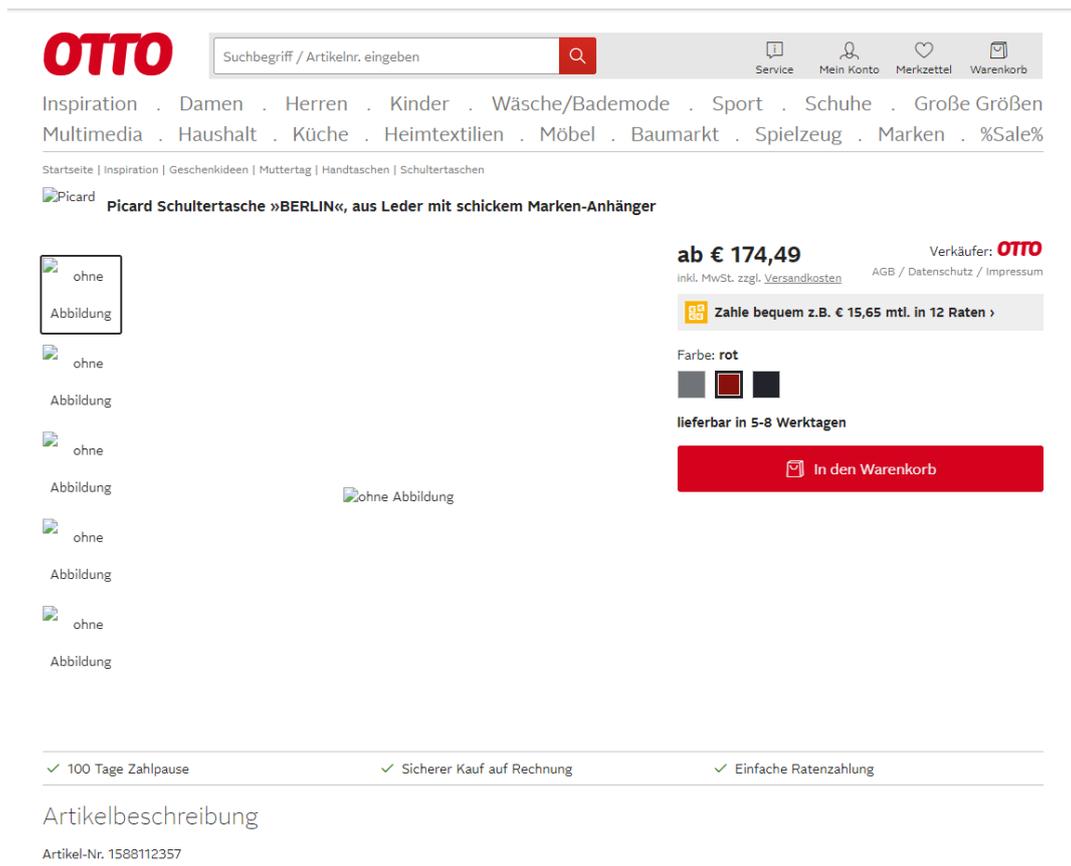


**Abbildung 2: Grobarchitektur der Preisabfragesoftware**

Basierend auf einer hinterlegten Konfiguration erzeugte ein Scheduler zeitgesteuert Aufträge, die bis zur Abarbeitung in einer Warteschlange (Queue) gepuffert wurden. Die so generierten Aufträge wurden von Arbeitern parallel und virtualisiert durchgeführt. Durch Verwendung von Docker-Containern für den Betrieb der Abfragesoftware wurde eine gleichbleibende und „saubere“ Ausführungsumgebung gewährleistet.

Für eine Abfrage wurden jeweils die Schritte „Aufzeichnung“ und „Extraktion“ ausgeführt. Für jede der 20 Zielwebseiten wurden Adapter für die Abfrage und Extraktion der Preisdaten entwickelt und parametrisiert.

Die Abfragen erfolgten dabei mittels des Headless-Browsers „Puppeteer<sup>34</sup>“, der durch eine NodeJS- bzw. JavaScript/TypeScript-Anwendung gesteuert wurde, sowie unter Verwendung eines Proxy-Dienstes. Hierbei wurde der Headless-Browser für jede Abfrage neu gestartet und konfiguriert. Dies erlaubte es wiederum, dass jede Abfrage mit einer definierten Konfiguration durchgeführt wurde und keine Einflüsse aus vorherigen Abfragen entstanden.



**Abbildung 3: Screenshot einer Beispielseite**

<sup>34</sup> <https://pptr.dev/>

Um eine möglichst schnelle Abfrage der Zielseite zu ermöglichen, wurden keine Bilder geladen (vgl. Abbildung 3). Skripte, CSS und weitere Quellen wurden geladen. Die so abgefragte Webseite wurde zunächst abgespeichert (Rohdaten) und anschließend daraus die Preisdaten extrahiert. Die extrahierten Preisdaten wurden zusammen mit ihrer Konfiguration ebenfalls abgespeichert.

Sowohl Rohdaten als auch Preisdaten wurden in einem sogenannten Objektspeicher abgelegt. Rohdaten (Konfiguration und HTML-Code) wurden als JSON-Dateien<sup>35</sup>, Preisdaten (Konfiguration und Preis) als CSV-Dateien<sup>36</sup> abgelegt. Die so gewonnenen CSV-Dateien wurden abschließend in eine einzige zusammengeführt und für die statistische Analyse aufbereitet. Weiterhin wurden zur Fehleranalyse und -behebung Screenshots der Zielseiten abgespeichert.

### 3.3.3 Bereitstellung der Infrastruktur für die automatisierte Datensammlung

Die Preisabfragesoftware wurde in einer Cloud-Umgebung in einem Cluster von virtuellen Maschinen betrieben. Dies gewährleistete eine hinreichende Parallelität der Abfragen. Um Tests und Qualitätssicherungsmaßnahmen zu erlauben, wurde die Cloud-Infrastruktur in zwei Systeme, einem Test- sowie einem Produktionssystem implementiert. Da die Erstellung der Infrastruktur automatisiert erfolgte, waren beide Systeme nahezu identisch. Dies erlaubte es, dass Testergebnisse von einem System auf das andere übertragbar waren.

### 3.3.4 Konfiguration der Preisabfragesoftware

Für die finale Konfiguration der einzelnen Zielseiten wurden für jede dieser Seiten manuell die notwendigen Parameter für die Abfrage bestimmt. Dies schloss zum einen die URL mit allen zugehörigen URL-Parametern ein. Zum anderen wurde zur Extraktion der Preisdaten ein möglichst spezifischer CSS-Selektor ermittelt, um die Preisinformation aus den Rohdaten (HTML-Code) zu extrahieren.

### **Simulation relevanter Einflussfaktoren**

Zur Modellierung der Einflussfaktoren Betriebssystem, Browser und Endgerät wurden für die möglichen Kombinationen User-Agent-Strings ermittelt. Dieser wurde bei der Anfrage mittels des zugehörigen Header-Feldes an die Zielwebseite übermittelt. Weiterhin wurde mittels JavaScript ein dazu passendes „Navigator“-Objekt<sup>37</sup> simuliert. Für die Simulation des Standortes wurden sowohl Proxy-IPs als auch die JavaScript-Geolocation-API verwendet. Um verschiedene

---

<sup>35</sup> [https://de.wikipedia.org/wiki/JavaScript\\_Object\\_Notation](https://de.wikipedia.org/wiki/JavaScript_Object_Notation)

<sup>36</sup> [https://de.wikipedia.org/wiki/CSV\\_\(Dateiformat\)](https://de.wikipedia.org/wiki/CSV_(Dateiformat))

<sup>37</sup> <https://developer.mozilla.org/de/docs/Web/API/Navigator>

Datenschutzeinstellungen zu untersuchen, wurden Abfragen mit und ohne Cookie-Akzeptanz durchgeführt. Zudem wurde die Do-Not-Track-Einstellung mittels Browsersetting bzw. DNT-Header variiert.

Nutzerkonten in den jeweiligen Online-Shops und auf den Vergleichsportalen wurden im automatisierten Strang nicht näher berücksichtigt. Diese Parameter wurden allerdings im manuellen Strang mit aufgenommen, um auch hierbei mögliche Einflüsse untersuchen zu können.

Interessanterweise funktionieren nicht alle Zielwebseiten ohne Cookies. So lässt sich die Seite von „mediamarkt.de“ nicht ohne Cookies verwenden. Daher wurde diese Parameterkonfiguration für „mediamarkt.de“ ausgeschlossen.

### **Verwendung von Proxy-Diensten**

Zur Simulation unterschiedlicher Standorte wurden insgesamt zwei Proxy-Dienste eingesetzt. Zum einen fand der Dienst „luminati.io“<sup>38</sup> zum anderen der Dienst „proxybonanza.com“<sup>39</sup> Verwendung.

Dienst	Standort	Bundesland	Anzahl IPs
<b>luminati.io</b>	Berlin	Berlin	19
<b>proxybonanza.com</b>	Düsseldorf	Nordrhein-Westfalen	35
<b>luminati.io</b>	Frankfurt	Hessen	73
<b>proxybonanza.com</b>	Frankfurt	Hessen	35
<b>luminati.io</b>	Hamburg	Hamburg	1
<b>proxybonanza.com</b>	Nürnberg	Bayern	30
<b>luminati.io</b>	Saarbrücken	Saarland	5
<b>luminati.io</b>	Steinheim am Albuch	Baden-Württemberg	2

**Tabelle 3: Übersicht über die Proxy-Standorte**

Tabelle 3 zeigt eine Übersicht über die vorhandenen IP-Adressen und deren Standorte. Bei allen IP-Adressen handelt es sich um sogenannte Rechenzentrum-IP-Adressen. Bei „luminati.io“ ist der Standort des Servers nicht explizit zu jeder IP-Adresse bekannt, daher wurde diese Information mittels der Geo-IP-Dienste ipgeolocation.io, ipdata.co sowie ipinfo.io ermittelt.

<sup>38</sup> <https://luminati.io/>

<sup>39</sup> <https://proxybonanza.com/>

Zu beachten ist hierbei, dass nicht alle Standorte für alle Zielwebseiten nutzbar sind. Dies ist genau dann der Fall, wenn alle IP-Adressen für diesen Standort von der Zielwebseite blockiert werden. Dennoch waren pro Zielwebseite mindestens drei bis vier Standorte für Abfragen nutzbar. Im Verlauf der Sammlung zeigte sich, dass Seitenbetreiber unterschiedlich streng bei der Sperrung verdächtiger IP-Adressen sind. Einige Seiten sperren einen nicht zu vernachlässigenden Anteil der IP-Adressen. Andere Seiten hingegen sperren nahezu keine IP-Adressen.

### 3.3.5 Durchführung von Qualitätssicherungsmaßnahmen

Sowohl im Laufe der Entwicklung als auch während des Betriebs der Preisabfragesoftware wurde eine Reihe an begleitenden Qualitätssicherungsmaßnahmen durchgeführt.

#### **Qualitätssicherungen während der Entwicklung**

Durch automatisierte Softwaretests wurden die einzelnen Bausteine und Komponenten der Software auf ihre korrekte Funktionsweise geprüft (sogenannte Unit-Tests<sup>40</sup>). Der Vorteil der Automatisierung ist, dass diese Tests bei jeder Änderung der Software ausgeführt werden können. Weiter wurde durch Integrationstests die korrekte Integration der einzelnen Komponenten verifiziert. Zudem prüften diese Tests, ob eine erfolgreiche Extraktion der Preisdaten möglich ist.

Als weitere Maßnahme wurden von den Entwicklern wechselseitige Code-Reviews durchgeführt. Das bedeutet, dass der Quellcode von einem Entwickler geschrieben und von einem anderen Entwickler im Sinne des Vier-Augen-Prinzips geprüft und ggf. Änderungen und Verbesserungen vorgeschlagen wurden. Zudem wurden Teile dieses Prozesses automatisiert, indem eine statische Code-Prüfung gegen etablierte Style-Guides und Best Practices in den Build-Prozess integriert wurden.

Mithilfe des Test- bzw. Entwicklungssystems wurden mehrere manuelle Tests über einen längeren Zeitraum durchgeführt. Hierbei wurden sowohl die korrekte Parametrierung der Abfragen, die richtige Konfiguration der Einflussfaktoren als auch die Korrektheit der gespeicherten Roh- und Preisdaten geprüft.

#### **Qualitätssicherungen während des Betriebs**

Durch die Anwendung selbst wurden bei jeder Abfrage die extrahierten Werte auf ihre Plausibilität geprüft. Dies beinhaltet zum einen die Validierung des abgefragten Produktes. So konnte sichergestellt werden, dass der Preis des gewünschten Produktes abgefragt wird. Weiterhin wurde geprüft, ob der Preis des Produktes plausibel erschien. Ergaben sich bei den Prüfungen Unstimmigkeiten

---

<sup>40</sup> <https://de.wikipedia.org/wiki/Modultest>

oder war die Abfrage des Preises aus einem anderen Grund nicht möglich, so kam es zu einem Fehler. Zunächst versuchte das System dann, durch „Retries“ (Wiederholungen) die Abfrage erneut durchzuführen. War dies ebenfalls erfolglos, so wurde die Abfrage als fehlerhaft vermerkt. Überstieg die Fehleranzahl zu einem Abfragezeitpunkt einen vorgegebenen Schwellwert, so wurde durch das System eine Benachrichtigung per E-Mail an den Support-Mitarbeiter versendet.

### **Vorbeugung der Detektion durch die Zielwebseiten**

Um eine Detektion der Preisabfragesoftware zu vermeiden, wurde eine Reihe an Vorsichtsmaßnahmen konzipiert und implementiert.

Eine dieser Maßnahmen spiegelt sich in der bereits beschriebenen Verwendung von Proxy-Diensten wider. So versuchte die Preisabfrage für einen festen Standort möglichst unterschiedliche IP-Adressen für die Abfrage zu verwenden. Zu diesem Zweck wurde für jede Abfrage zufällig eine der möglichen IP-Adressen für den Standort ausgewählt.

Weiterhin wurden die IP-Adressen statisch und dynamisch intelligent vorgefiltert. Statisch bedeutet hierbei, dass vor dem Start der Preisabfragesoftware evaluiert wurde, welche IP-Adressen für welche Zielwebseiten nutzbar sind (bzw. nicht geblockt werden). Dynamisch bedeutet dabei hingegen, dass während des Betriebs IP-Adressen in der Preisabfragesoftware pro Zielwebseite gesperrt oder freigeschaltet werden können.

Um zu vermeiden, dass trotz der Verwendung eines Proxy-Dienstes dennoch die reale IP der Preisabfragesoftware ermittelt werden kann, wurde zudem im Headless-Browser eine spezielle Erweiterung implementiert. Diese erlaubte es, das sogenannte WebRTC-IP-Leak<sup>41</sup> zu schließen.

Um unterschiedliche Browser, Betriebssysteme und Geräte möglichst glaubhaft zu simulieren, wurden neben dem Setzen des User-Agents weitere Schritte unternommen. So wurden Geräteeigenschaften wie Größe des sichtbaren Ansichtsfensters (Viewport<sup>42</sup>) spezifisch, je nach Gerät und Betriebssystem, angepasst.

Weiterhin können Details zu Browser, Betriebssystem und Gerät von den Zielwebseiten auch mittels JavaScript ausgelesen werden. Daher wurde das sogenannte Navigator-Objekt spezifisch für die jeweilige Kombination aus Browser, Betriebssystem und Gerät angepasst.

---

<sup>41</sup> <https://de.wikipedia.org/wiki/WebRTC#IP-Leak>

<sup>42</sup> <https://de.wikipedia.org/wiki/Viewport>

Hinsichtlich des zu simulierenden Standorts wurden neben der Verwendung von Proxy-IPs auch die Antworten der Geolocation-API der Browser passend verändert. Dies bedeutet, dass für eine IP aus Berlin auch die Geolocation-API einen Standort in Berlin (genauer einem Umkreis von 1km von Berlin) zurückliefert.

Eine weitere sehr wichtige Maßnahme war die Begrenzung der Anzahl der Abfragen pro Zeitpunkt und Zielwebseite. Zunächst könnte man davon ausgehen, dass man zu jedem Zeitpunkt, jede Kombination der Einflussfaktoren für jede der Zielwebseiten abfragen kann. Jedoch ergibt sich hierbei aufgrund der Kombinatorik schnell eine sehr große Anzahl an Abfragen. Aus Sicht der Preisabfrage-Software wäre dies prinzipiell kein Problem. Durch mehr Parallelisierung könnten die Abfragen im gleichen Zeitintervall durchgeführt werden. Allerdings ist ein derartiges Vorgehen für die Zielwebseiten äußerst verdächtig und könnte den Betrieb der Seite beeinträchtigen. Es ist wohl eher untypisch, dass ein spezifisches Produkt zu einer definierten Uhrzeit immer und immer wieder von sehr vielen Anfragen abgefragt wird. Reduziert man hingegen die Anzahl der Abfragen, so verschwinden diese im normalen Grundrauschen der Zielseiten. Für eine statistisch signifikante Auswertung der Preisdaten ist die volle Kombinatorik ohnehin nicht zwingend notwendig. Es reicht, an dieser Stelle ein sogenanntes Sampling durchzuführen. Anstatt die volle Kombinatorik der Abfragen tatsächlich durchzuführen, wurden für jede der Zielseiten nur 10 Prozent der Abfragen tatsächlich ausgeführt. Die Auswahl der Abfrage erfolgte dabei zufällig. Dabei ist jedoch zu beachten, dass der vorab definierte Standardfall für jede Zielwebseite immer mit aufgezeichnet wurde. Weiterhin wurden zu jedem Zeitpunkt sowohl alle Zielwebseiten als auch alle möglichen Standorte abgefragt.

### **Vorbereitung der manuellen Abfragen**

Zeitgleich zum automatisierten wurde auch der manuelle Abfragestrang vorbereitet. Durch die Anwendung des manuellen Strangs sollen Einflussfaktoren erhoben werden, die durch den automatisierten Strang nur schwierig oder gar nicht ermittelbar sind. Zu solchen zählen vor allem „weiche“ Faktoren, wie die Nutzung (Login) von sozialen Medien (Facebook und/oder Instagram) oder eine zu Grunde liegende Einkaufshistorie (Kundenkonto).

Um weiche Faktoren zu extrahieren, wurden für den manuellen Abfragestrang Probanden mit unterschiedlichen Merkmalen (Kundenkonto, Konto bei sozialen Medien, ...) zusammengesetzt. Diese wurden von Mitarbeitern von ibi research und trinno innovative repräsentiert. Die Probanden boten unterschiedliche Surf- und Kaufhistorien und deckten das Spektrum von „sehr kaufkräftig“ bis „wenig kaufkräftig“ ab.

Zunächst wurden alle Probanden erfasst, um einen möglichst passgenauen Testplan erstellen zu können. Zum einen wurde dazu die technische Ausstattung der Probanden abgefragt, dazu gehören z. B. vorhandene Endgeräte und Betriebssysteme. Des Weiteren wurde erfasst, welche der vorab ausgewählten

Shops und Portale die Probanden nutzen und wo Nutzerkonten vorhanden sind. Ebenso wurde abgefragt, welche sozialen Medien mit Nutzerkonten genutzt werden. Da ein wesentlicher Teil des Erhebungszeitraum in die Ferienmonate August und September fiel, wurden zusätzlich die zeitlichen Verfügbarkeiten abgefragt. Probanden, die aufgrund der Verwendung z. B. eines MacBooks den definierten Standardfall nicht abfragen konnten, wurden im Testplan nicht weiter berücksichtigt.

Anschließend erfolgte die Zuteilung unterschiedlicher Personas an die Probanden. Eine Persona setzt sich zusammen aus der Kombination von Einflussfaktoren, wie Endgeräten, Betriebssystemen, sozialen Konten und Kundenkonten zu Shops. Auf Basis dieser ihm zugeteilten Persona prüfte jeder Proband zu jedem festgelegten Abfragezeitpunkt den Standardfall, sowie (in der Regel bis zu drei) weitere Kombinationen (vgl. Abbildung 4 und Abbildung 5).

Link unter Verwendung der Voreinstellungen in Browser	Endgerät	Betriebssystem	Browser*	Cookie Akzeptanz	Do-Not-Track [DNT]-Software	In Nutzerkonto Shop/Portal eingeloggt	In Social Media Profil (FB, IG) eingeloggt?	
Vanilla	Desktop	Windows 10	Google Chrome	ja	nein	nein	nein	* Zur Abfrage "Vanilla" Google Chrome installieren! Anleitungen zum Import der URLs in die verschiedenen Browser sowie zu den Browsereinstellungen (Cookie Akzeptanz, Do-Not-Track-Software) liegen in Teams unter: <a href="#">ibi-Allgemein (Dateien) &gt; Manueller Test (BMV-PP)</a>
1. Kombi	Smartphone	Android	Google Chrome	ja	nein	ja	ja	
2. Kombi	Tablet	Android	Google Chrome	ja	nein	ja	nein	
Abfragezeiträume täglich zwischen		08.00 Uhr und 12.00 Uhr	10.00 Uhr und 14.00 Uhr	17.00 Uhr und 19.00 Uhr	19.00 Uhr und 22.00 Uhr			
Manueller Test von 01.07. - 30.09.2020**								
<p><b>Hinweis an alle Nutzer mobiler Endgeräte:</b> Gelegentlich werden bei Abfragen über Smartphones und Tablets nicht die richtigen Produkte angezeigt &gt;&gt; Bitte darauf achten, dass die Produkte mit dem Vanilla-Case übereinstimmen, z. B. Koffer, groß, Handtasche, schwarz etc.</p>								

↑ Falls "ja", vor der Abfrage einloggen

↑ Falls "ja", vor der Abfrage einloggen

\*\* wenn möglich, bitte ggf. Erinnerungen in GroupMail/ Outlook oder auf Smartphone/ Smartwatch einstellen, um Abfragezeiträume wann immer möglich wahrnehmen zu können. VIELEN DANK!

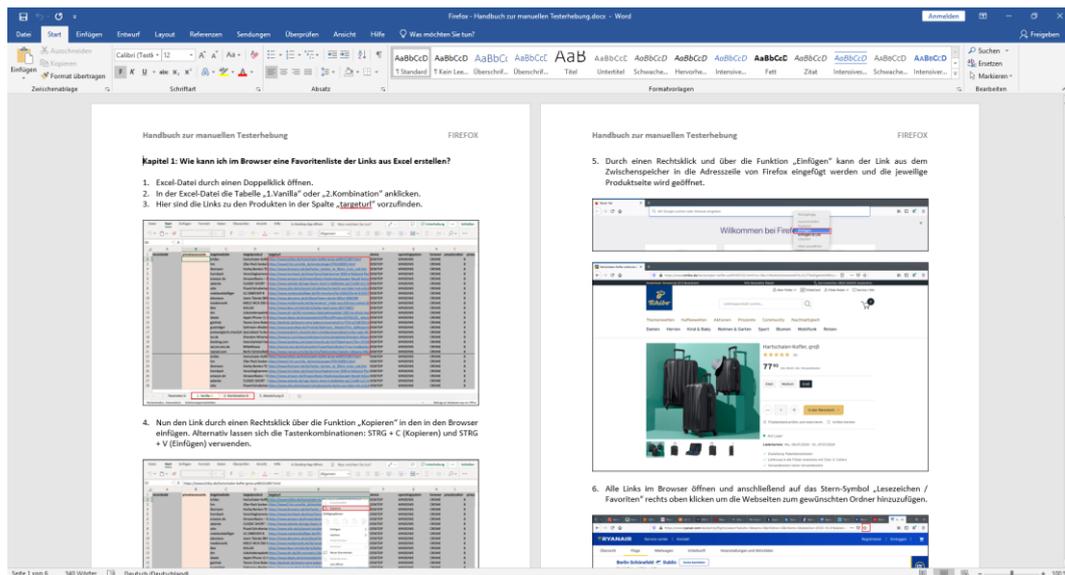
Abbildung 4: Beispielhafte Darstellung von Abfragekombinationen

Verwendete Kombination (Endgerät - Betriebssystem - Browser)	Mitarbeiter	Nummer der Kombination	Cookie-Akzeptanz	Don-Not-Track [DNT]-Software	In Shop/Portal eingeloggt	In Social-Media-Profil (FB,IG) eingeloggt?
Desktop - WIN10 - Internet Explorer		1	1	1	0	0
		1	0	1	0	1
		1	1	0	1	1
Desktop - WIN10 - Firefox		2	1	1	1	1
		1	0	0	0	0
Desktop - WIN10 - Google Chrome		1	1	1	1	1
		1	1	0	1	0
		1	0	0	0	0
Desktop - WIN 10 - Edge		1	1	0	0	0
		1	0	1	0	0
		1	1	0	1	1
Desktop - MacOS - Safari		1	1	0	1	0
		1	0	1	0	1
		1	0	0	0	0
Smartphone - Android - Google Chrome		1	1	0	1	1
		2	1	1	0	0
		2	1	1	1	1
		2	1	0	0	1
		2	1	1	0	1
Tablet - Android - Google Chrome		2	1	0	1	0
		1	1	0	0	0
		1	1	1	1	1
iPhone - iOS - Safari		1	1	0	1	0
		1	1	0	1	0
		2	1	0	1	1
		1	1	0	0	1
iPad - iOS - Safari		2	1	0	0	0
		2	1	1	1	1
Smartphone - Android - Firefox		3	1	0	1	1
Desktop - Linux - Google Chrome		2	1	0	1	0
Desktop - Linux - Firefox		3	0	0	0	0

Abbildung 5: Verteilungstabelle der Kombinationen



Vor Beginn der Erhebungsphase fand eine zentrale Einweisung bzw. Schulung der Probanden über eine Zoom-Videokonferenz statt. Dabei wurden das Vorgehen, der Ablauf und die Organisation der Preiserhebung detailliert erläutert. Ab diesem Zeitpunkt standen den Probanden mehrere Schritt-für-Schritt-Handbücher zur jeweiligen Browserkonfiguration (Favoritenlisten anlegen, Makros aktivieren/deaktivieren, Do-Not-Track-Software aktivieren/deaktivieren) zur Verfügung. Die Aufzeichnung der Videokonferenz zur Einweisung war unterstützend verfügbar (vgl. Abbildung 8).



**Abbildung 8: Handbuch zur manuellen Testerhebung (hier Firefox)**

Um möglicherweise auftretende Fragen schnell und unmittelbar klären zu können, standen den einzelnen Probanden alle Mitglieder des Projektteams als Ansprechpartner zur Verfügung.

### 3.4 Durchführung der Datenerhebung

#### **Abfragen durch die Preisabfragesoftware**

Vom 9. Juni bis zum 4. Oktober 2020, also über einen Zeitraum von fast vier Monaten, wurden durch die Preisabfragesoftware automatisiert Preise abgefragt und aufgezeichnet. Täglich wurden zeitgesteuert um 8:00, 12:00, 17:00 und 20:00 Uhr (MESZ) Abfragen an alle Zielwebseiten durchgeführt. Pro Zeitpunkt wurden mehr als 550 Abfragen durchgeführt (vgl. Sampling). Eine vollständige Abfrage aller Preisdaten dauerte pro Zeitpunkt im Mittel etwa 10 Minuten.

Die Qualität der Preisabfragen wurde täglich überwacht und geprüft. Hierzu wurden sowohl die Anzahl der Abfragen pro Tag als auch nach Bedarf die extrahierten Preise geprüft.

Im Laufe der Aufzeichnung war es notwendig, auf nicht mehr verfügbare Produkte, geänderte Links zu den Produkten sowie geänderte Seitenstrukturen zeitnah zu reagieren. Weiterhin kam es im Laufe der Aufzeichnung mehrmals dazu, dass sich eine Webseite im Wartungsmodus befand und temporär nicht verfügbar war. Eine besondere Herausforderung stellten die verwendeten Proxy-Dienste dar. Die betrachteten Webseiten wiesen ein dynamisches Blockverhalten auf, d.h. unterschiedliche Webseiten blockieren unterschiedliche IPs zu unterschiedlichen Zeiten. Um diesem Problem beizukommen, wurden die verwendeten IP-Adressen durch Support-Mitarbeiter regelmäßig nachgeschärft.

Insgesamt waren von 269.461 Abfragen 266.488 erfolgreich und 2.973, also rund 1,1 Prozent, fehlerhaft.

### **Abfragen durch natürliche Personen**

Der manuelle Abfragestrang wurde zum 6. Juli 2020 gestartet und bis zum 4. Oktober 2020 durchgeführt. Insgesamt waren 21 natürliche Personen mit der manuellen Abfrage beschäftigt. Täglich wurden durch die Probanden die Preise von 20 Produkten zu je vier Abfragezeiträumen erfasst. Dabei wurde der Standardfall inklusive bis zu drei Personas erhoben. Jeder Proband erhob somit pro Tag bis zu 320 Preise.

Die hohe Anzahl an Probanden war erforderlich, denn sie ermöglichte eine Verteilung der Abfragen, die weniger zeitaufwändig pro Person war. So wurden die Abfragen nach den Endgeräten oder Browsern der Probanden zugeordnet. Für einen Mitarbeiter wäre die Aufgabe, für drei Monate an sieben Wochentagen täglich zu vier Zeitpunkten Daten mit unterschiedlichen Grund-Kombinationen abzufragen, unmöglich machbar gewesen. Gegenüber der ursprünglichen Planung mit einer Anzahl von zehn bis zwölf Personen, die mit der Preiserhebung betraut werden, wurde der Rahmen deutlich ausgeweitet. Die Personenanzahl wurde erhöht, um ein umfangreiches, statistisch belastbares Sample zu erhalten.

Zunächst stellte sich die Bereitstellung eines Linux-Systems als problematisch dar, da keiner der Probanden über ein solches verfügte. Hier spiegelt sich der Linux-Marktanteil in Deutschland von unter drei Prozent wider.<sup>43</sup> Als Problemlösung wurde das Linux-System (Linux Mint Debian Edition, LMDE)<sup>44</sup> mit Hilfe einer virtuellen Maschine (Oracle VM VirtualBox)<sup>45</sup> simuliert. Die Preisabfragen wurden hier über Firefox und Chrome getätigt. Somit konnte auch Linux in den manuellen Abfragestrang integriert werden, allerdings mit einem im Vergleich zu den anderen Betriebssystemen kleinerem Datenumfang.

---

<sup>43</sup> Vgl. <https://de.statista.com/statistik/daten/studie/158102/umfrage/marktanteile-von-betriebssystemen-in-deutschland-seit-2009>.

<sup>44</sup> [https://de.wikipedia.org/wiki/Linux\\_Mint](https://de.wikipedia.org/wiki/Linux_Mint)

<sup>45</sup> <https://de.wikipedia.org/wiki/VirtualBox>

Wöchentlich wurden die Excel-basierten Erfassungsdateien neu erstellt und den Probanden zentralisiert zur Verfügung gestellt. Die befüllten Dateien wurden dann im wöchentlichen Turnus zentral eingereicht. Dieses Vorgehen wurde gewählt, um etwa notwendige Änderungen für inkonsistente Produktlinks anpassen zu können. Ebenso kann so eine kursorische Prüfung auf Qualität und Konsistenz der eingereichten Daten stattfinden.

Die gesammelten Datenblätter wurden wöchentlich zentralisiert abgelegt. Daraufgehend wurden diese in die CSV-Datenstruktur des automatisierten Anfragestrangs überführt. So wurde als Basis eine vereinheitlichte Datenstruktur über den automatisierten und manuellen Strang hinweg geschaffen.

## 4 Empirische Untersuchung

### 4.1 Auswahl der Auswertungsmethodik

In diesem Abschnitt sollen die nachfolgenden quantitativen Analysen zur Überprüfung dynamischer und insbesondere personalisierter Preissetzungen zunächst methodisch dargestellt werden. Zur statistischen Auswertung der gewonnenen Daten wurde die Programmiersprache Python eingesetzt.

Einzelne Schritte der Datenauswertung wurden zunächst an einer Teilstichprobe im automatisierten Strang beprobt (Pretest). Dem folgte die Hauptanalyse der Datenbasis, die auf Grundlage der automatisierten und manuellen Abfragen gewonnen werden konnten. Unter Verwendung relevanter Analysemethoden konnte eine empirische Bestimmung der forschungsleitenden Fragen vorgenommen werden. Die Auswertungsmethodik wird nachfolgend nach den Vor- und Hauptanalysen gegliedert, anschließend wird das Vorgehen geschildert:

#### **Voranalysen (Deskriptive Statistik)**

- Data Cleaning/Pre-Processing
- Detaillierte Beschreibung der Stichprobe

#### **Hauptanalyse**

- Überprüfung der Vorhersagekraft der gewählten Regressoren auf die Preisdifferenz durch eine multiple, lineare Regressionsanalyse.

Beschrieben wird die Auswahl sowie Vorgehensweise der Analysemethoden zur Feststellung von personalisierten Preisen in den zuvor erhobenen Datensätzen. Unterschieden wird zwischen der zunächst notwendigen Voranalyse mit Hilfe deskriptiver Statistik und einer anschließenden Regressionsanalyse zur weiteren Betrachtung der Einflussfaktoren.

#### **Vorgehensweise**



#### **Abbildung 9: Vorgehensweise der Auswertung**

Der Datenerhebung in manuellem und automatisiertem Strang folgte die Aufbereitung und Analyse der beiden Rohdatensätze des manuellen und automatisierten Erhebungsstrangs. Die Auswertung wurde in drei konsekutive Schritte unterteilt. Zuerst wurden die erhobenen Daten im Preprocessing für die weitere Analyse vorbereitet. Anschließend ermöglichte die deskriptive Analyse einen explorativen Einblick in die erhobenen Daten. Abschließend wurden die potenziellen Einflussfaktoren mit Hilfe einer Regressionsanalyse validiert.

Als erster Schritt der Datenanalyse hat das **Preprocessing** das Ziel, beide Datensätze in eine zu den folgenden Analysen kompatiblen Struktur zu bringen und zudem die Datenqualität zu sichern. Dazu wurden die aus dem automatisierten und manuellen Strang erhobenen Rohdaten in vier Schritten aufbereitet: Import, Cleaning, Berechnung des Preisunterschieds und Kodierung der Variablen.

Die Daten des automatisierten Erhebungsstrangs lagen als Rohdaten in Form einer CSV-Datei vor, während die Daten des manuellen Strangs pro Woche und Erfasser als einzelne Excel-Datei vorlagen. Erster Teilprozess des Preprocessing dieser beiden Rohdaten war der **Import**.

Zur Sicherung der Datenqualität war anschließend ein **Cleaning** notwendig. Der Fokus lag insbesondere auf der Preisvariablen, da diese als betrachtete Zielvariable eine möglichst geringe Fehlerrate aufweisen sollte. Dazu wurden in beiden Erhebungssträngen die Preiseinträge, die außerhalb der 25- und 75-Prozent-Quartile lagen, manuell sowie automatisiert betrachtet und korrigiert. Weiterhin wurden vor allem im manuellen Strang vorkommende Tippfehler im Cleaning bereinigt. Fehlerhafte Datensätze wie beispielsweise Testdatensätze und Duplikate wurden ebenso im Preprocessing gesucht und entweder gelöscht oder korrigiert. Als letzter Teil des Cleanings wurden die Bezeichnungen von Einflussvariablen und Produkten für die folgende Analyse uniformiert.

Die Preisdaten in beiden Rohdatensätzen lagen pro Produkt, Zeitpunkt und Einflussvariablenkombination vor. Jedoch war für die fortführende Analyse die Berechnung der relativen Preisdifferenz zum Standardfall notwendig. Deshalb wurde der in den Rohdaten erfasste Preis im Teilschritt **Berechnung Preisunterschied** in eine prozentuale Preisdifferenz umgerechnet. Zur Ermittlung von auftretenden Preisunterschieden war wegen der Datenstruktur pro Erhebungsstrang ein eigenes Vorgehen notwendig.

Grundsätzlich sind die Preise produktspezifisch und müssen normalisiert werden, um Vergleiche zwischen unterschiedlichen Produkten und damit eine Regressionsanalyse über den kompletten Datensatz hinweg zu ermöglichen. Prinzipiell wurden die Preisunterschiede zum jeweiligen Bezugspunkt, dem Standardfall, berechnet. Der Standardfall bezeichnet die im Datensatz explizit definierte Konfiguration in Bezug auf die Einflussvariablen. Als Standardfall definiert sind sämtliche Datensätze, die an einem Windows-Desktoprechner mit einem Google-Chrome-Browser ohne Cookie- oder Trackingblocker erhoben wurden und die nicht in Social Media o. ä. eingeloggt sind. Die Daten des automatisiert erhobenen Datenstrangs waren dabei so strukturiert, dass sämtlichen Datensätzen desselben Produkts, die zum selben Zeitpunkt and demselben Ort erhoben wurden, ein eindeutiger Standardfall zugeordnet werden konnte. Die manuell erfassten Datensätze hingegen enthielten zusätzlich die Information des

jeweiligen Erfassers. Hier ließ sich jeder Datengruppe, die dasselbe Produkt beschreibt und vom selben Erfasser zum gleichen Zeitpunkt erhoben wurde, ein eindeutiger Standardfall zuweisen.

Die folgend verwendete relative Preisdifferenz berechnet sich aus dem prozentualen Preisunterschied des jeweiligen Preises zum jeweiligen Standardfall. Folglich weisen Standardfälle als Bezugspunkt per Definition keinen Preisunterschied auf.

$$\text{Relative Preisdifferenz} = \frac{\text{Preis}_{\text{Produkt}} - \text{Preis}_{\text{Standardfall}}}{\text{Preis}_{\text{Standardfall}}}$$

Durch die Verwendung von relativen Differenzen ist zusätzlich die Vergleichbarkeit der berechneten Preisunterschiede über alle Produkte und möglichen Konfigurationen hinweg gewährleistet. Als zusätzliche Sicherung der Datenqualität wurden sämtliche Preisdifferenzen bei einer Preisabweichung von mehr als 18 Prozent manuell auf Plausibilität und eventuelle Erhebungsfehler überprüft. Der Wert von 18 Prozent ergab sich empirisch aus den Daten, da dies die höchste im Datensatz vorkommende Preisabweichung darstellte.

### **Kodierung der Variablen**

Erfasst wurden für die weitergehenden Analysen die Variablen Preis, Produkt, Gerät, Betriebssystem, Browser, Proxy-Standort, Tracker, Cookies, Login in ein Nutzerkonto bzw. soziale Medien und der Erfasser. Zusätzlich wurde aus der Preisvariable im Teilschritt „Berechnung Preisunterschied“ die prozentuale Preisdifferenz als Zielvariable der Regression ermittelt. Die Variablen Gerät, Betriebssystem und Browser wurden in die Variable „Zugriff“ zusammengeführt, um potentielle Multikollinearitätsprobleme der Regression zu beheben. Diese treten durch die Abhängigkeiten der Kombinationsmöglichkeiten der Variablen Gerät, Betriebssystem und Browser auf. Beispielsweise kann der Browser „Safari“ nur auf Apple-Geräten betrieben werden.

Unterschieden wurden die erfassten und generierten Variablen in kontinuierliche, kategoriale und binäre Variablen. Kontinuierliche Variablen sind numerische Variablen, die zwischen zwei beliebigen Werten sämtliche Werten annehmen können. Kontinuierlich ist im Datensatz die Zielvariable Preis und die daraus abgeleitete Preisdifferenz. Kategoriale Variablen weisen eine abzählbare Anzahl an Ausprägungen auf. Produkt, Gerät, Betriebssystem, Browser, Proxy-Standort, Erfasser und die kombinierte Variable Zugriff sind die im Datensatz auftretenden kategorialen Variablen. Binäre Variablen sind ein Spezialfall der kategorialen Variablen mit nur zwei Ausprägungen: falsch (0) und richtig (1). Die Aktivierung von Cookies, Trackern, der Login in ein Nutzerkonto und die Social-Media-Nutzung werden auf diese Weise im Datensatz erfasst. Tabelle 4 fasst die Variablen zusammen.

<b>Kontinuierliche Variable</b>	<b>Kategoriale Variable</b>	<b>Binäre Variable</b>
Preis	Produkt	Cookies
Preisdifferenz	Gerät	Tracker
	Betriebssystem	Nutzerkonto
	Browser	Social Media
	Proxy-Standort	
	Erfasser	
	Zugriff	

**Tabelle 4: Variablen**

**Deskriptive Analyse**

Mit dem Ziel, die Personalisierung von Preisen anhand der manuell und automatisiert erhobenen Datensätze zu ermitteln, war vor tiefergehenden Analysen zunächst eine deskriptive Analyse notwendig. Ermöglicht wurde mit dieser explorativen Vorgehensweise die unvoreingenommene Betrachtung der vorhandenen Daten. Nach einer Analyse der erhobenen Variablen wurde die Zielvariable Preisdifferenz entlang der erhobenen unabhängigen Variablen betrachtet.

**Regressionsanalyse**

Die Regressionsanalyse dient der Findung und Messung der unabhängigen Einflussvariablen auf die abhängige Zielvariable „relative Preisabweichung“. Die Auswertung der enthaltenen p-Values ermöglicht zudem die Evaluierung der Übertragbarkeit von in dem Datensatz gefundenen Zusammenhängen auf die Grundgesamtheit.

**Technologie**

Der gesamte Schritt der Datenanalyse vom Preprocessing über die deskriptive Analyse bis hin zur Regression wurde in Python durchgeführt, da Python spezielle, zur Umsetzung von Regressionsanalysen notwendige und hierbei viele Anpassung- und Individualisierungsmöglichkeiten anbietet. Ebenso kann die Ausgabe von Graphen zur Auswertung einfach realisiert werden. Verwendet wurden die Open-Source-Bibliotheken numpy, pandas, matplotlib, statsmodels, patsy, matplotlib und seaborn.

**Vorbereitung der Auswertung**

Bereits während der Erhebungsphase wurde das geplante Auswertungskonzept überprüft. Dies diente zur Klärung, ob die vorgesehene statistische Auswertung in Kombination mit der vorliegenden Datenstruktur durchführbar sein würde.

Mögliche Anpassungen wären so vor Beginn des Erhebungszeitraums noch ohne größeren Aufwand machbar gewesen.

Das geplante Vorgehen wurde intensiv getestet, um spätere Anpassungen, die mit hohem Aufwand verbunden sind, zu vermeiden und eine fundierte Basis zu erhalten. Grundsätzlich wurden die Preise durch die beiden Abfragestränge (manuell, automatisiert) erhoben. Die statistische Auswertung beruht jedoch auf einem Vergleich der Preisabweichungen. Diese entstehen durch einen Abgleich des Standardfalls mit den erhobenen Daten mit abweichenden Einflussfaktoren. Als Ergebnis der Pretests stellte sich heraus, dass eine Zuordnung der erfassten Preise zum zugehörigen Standardfall möglich ist.

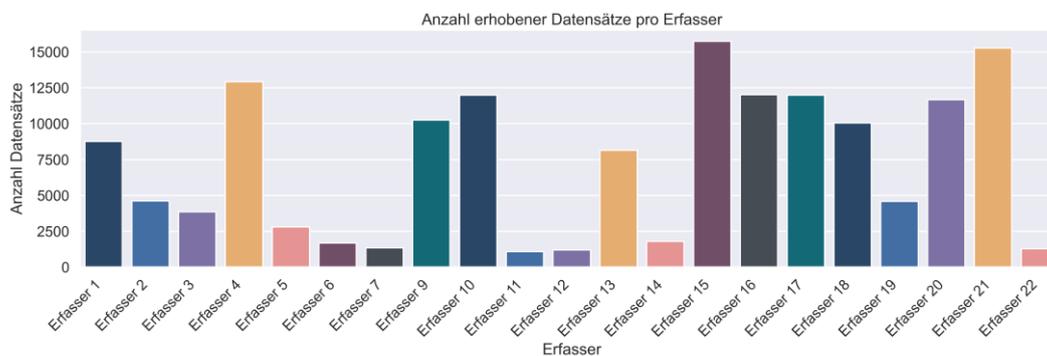
Zur Untermauerung des Vorhabens und der Umsetzbarkeit wurden Vorabtests mit R und Python durchgeführt. Letztendlich wurde Python zur Umsetzung der statistischen Auswertung ausgewählt.

## 4.2 Deskriptive Analyse

Im Folgenden wird die im manuellen und automatisierten Strang erhobene Stichprobe genauer betrachtet. Nach einer Darstellung der Verteilung der erhobenen Daten auf die Einflussvariablen Produkt und verwendetes Gerät pro Erhebungsstrang wird schließlich in der deskriptiven Analyse die Zielvariable Preis genauer untersucht.

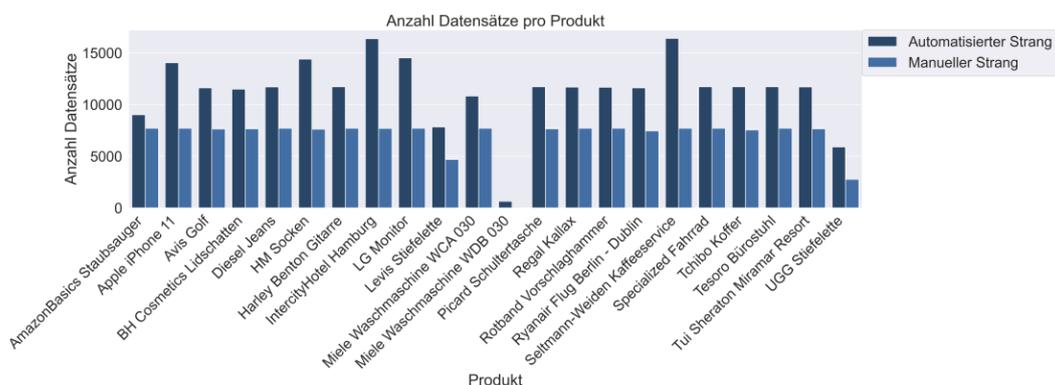
### 4.2.1 Verteilung der erhobenen Datensätze

Im Zeitraum vom 9. Juni bis zum 4. Oktober 2020 wurden insgesamt 403.204 verwertbare Datensätze mit jeweils 17 Datenpunkten (Spalten) erhoben. Diese Zahl exkludiert bereits die im Preprocessing entfernten Duplikate und fehlerhaften Datensätze. Die Gesamtzahl der Datensätze verteilt sich auf 250.091 im automatisierten Strang und 153.113 im manuellen Strang erhobene Datensätze. Die Anzahl der pro Erfasser erhobenen Datensätze fasst Abbildung 10 zusammen. Hier zeigen sich stark unterschiedliche Erhebungsstärken.



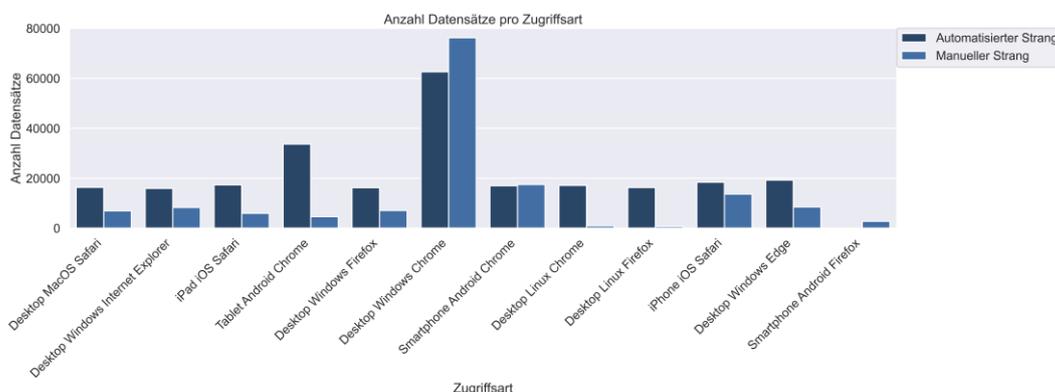
**Abbildung 10: Anzahl erhobener Datensätze nach Erfasser**

Abbildung 11 stellt die pro Produkt erhobenen Datensätze dar. Auffällig ist die geringere Anzahl an Datensätzen der Produkte „Miele Waschmaschine WBD030“ sowie der beiden Stiefeletten „UGG Stiefelette“ und „Levis Stiefelette“. Dies ist auf die Änderung der Verfügbarkeit der beiden Produkte zurückzuführen. Sobald das zuerst verfügbare Produkt aus dem Sortiment genommen wurde, war ein vergleichbares Produkt als Ersatz aufgenommen worden. Die restliche Variation der Anzahl an Datensätzen ist beim automatisierten Strang auf von den Online-Shops geblockte Anfragen, beim manuellen Strang auf nicht erfasste Preisdaten auf Grund von Abwesenheit des Erfassers zurückzuführen. Zur Datenanalyse sind letztendlich sämtliche verfügbaren Datensätze verwendet worden.



**Abbildung 11: Anzahl Datensätze nach Produkt**

Abbildung 12 stellt die Verteilung der erhobenen Preisdatensätze auf die dazu verwendeten Geräte dar. Auffällig ist die hohe Anzahl an Datensätzen, die mit einem Windows-Rechner und Chrome abgerufen wurden. Da diese Zugriffsart Teil der Konfiguration des Standardfalls ist, muss zu jeder anderen Konfiguration ein entsprechender Standardfall existieren, um die Preisdifferenzen zu berechnen. Die Kombination „Smartphone Android Firefox“ wurde nur im manuellen Strang erfasst. Um die Kombinationen „Desktop Linux Chrome“ und „Desktop Linux Firefox“ im manuellen Strang vollständig zu erheben, standen personalbedingt zu wenig Ressourcen zur Verfügung (vgl. dazu Abschnitt 3.4).



**Abbildung 12: Anzahl Datensätze nach Zugriffsart**

## 4.2.2 Deskriptive Statistik der Preisdifferenzen

Weiterhin betrachtet wird der gesamte Datensatz bestehend aus automatisiert und manuell erhobenen Daten. Im Fokus stehen dabei die Daten mit Preisdifferenzen. Berechnet werden die Preisdifferenzen, wie bereits in Kapitel 4.1 beschrieben, prozentual zum korrespondierenden Standardfall.

Insgesamt lagen 403.204 Datensätze vor. 115.169 dieser Datensätze stellten Standardfälle dar, die als Referenzpunkt für die Berechnung von Preisdifferenzen für 286.011 Datensätze dienten. Daneben gab es noch 2.024 Datensätze, für die in der Auswertung kein Standardfall definiert werden konnte. Das lag in der Regel daran, dass der entsprechende Standardfall in diesem Abfragezeitpunkt nicht erhoben worden war, etwa weil die Website zum Zeitpunkt der automatisierten Abfrage nicht erreichbar war.

Von den 286.011 Datensätzen wiesen 9.478 Datensätze Preisdifferenzen (ungleich null) zum Standardfall auf, das entspricht einem relativen Anteil von 3,31 Prozent.

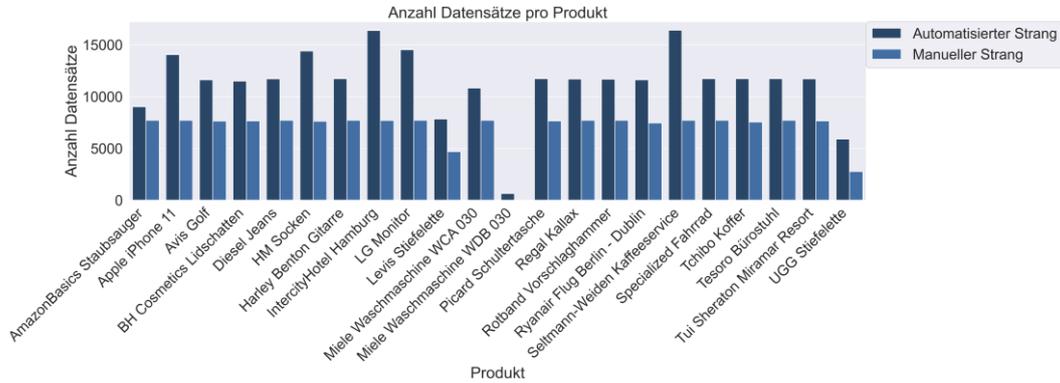
Im Folgenden werden diese Datensätze mit Preisdifferenz entlang der beim Erhebungsdesign vordefinierten Parameter genauer betrachtet. Diese vordefinierten Analysedimensionen sind Produkt, Gerät, Betriebssystem, Browser, Zugriff, Proxy-Standort, Tracker, Cookies, Login in ein Nutzerkonto bzw. Soziale Medien und der Erfasser.

	Erhobene Datensätze	Erhobene Datensätze mit potenziellen Preisdifferenzen (ohne Standardfälle)	Datensätze mit Preisdifferenzen	Anteil Datensätze mit Preisdifferenz an erhobenen Datensätzen (ohne Standardfälle) in Prozent
AmazonBasics Staubsauger	16.727	11.406	8	0,0702 %
Apple iPhone 11	21.758	15.615	300	1,9213 %
Avis Golf	19.247	13.623	27	0,1982 %
BH Cosmetics Lidschatten	19.150	13.339	2	0,0150 %
Diesel Jeans	19.422	13.752	16	0,1164 %
HM Socken	22.011	15.636	2	0,0128 %
Harley Benton Gitarre	19.425	13.753	1	0,0073 %
IntercityHotel Hamburg	24.069	17.454	8.571	49,1063 %
LG Monitor	22.230	15.812	13	0,0823 %
Levis Stiefelette	12.520	8.914	3	0,0337 %
Miele Waschmaschine WCA 030	18.536	13.038	51	0,3912 %
Picard Schultertasche	19.378	13.736	132	0,9610 %
Regal Kallax	19.403	13.738	5	0,0364 %
Rotband Vorschlaghammer	19.378	13.704	8	0,0584 %
Ryanair Flug Berlin - Dublin	19.059	13.475	15	0,1114 %
Seltmann-Weiden Kaffeeservice	24.108	17.484	3	0,0172 %
Specialized Fahrrad	19.428	13.760	9	0,0655 %
Tchibo Koffer	19.262	13.661	31	0,2270 %
Tesoro Bürostuhl	19.431	13.762	99	0,7194 %
Tui Sheraton Miramar Resort	19.349	13.681	176	1,2865 %
UGG Stiefelette	8.677	6.188	6	0,0970 %
Summe	403.204	286.011	9.478	3,3139 %

**Tabelle 5: Anzahl Datensätze mit Preisdifferenz**

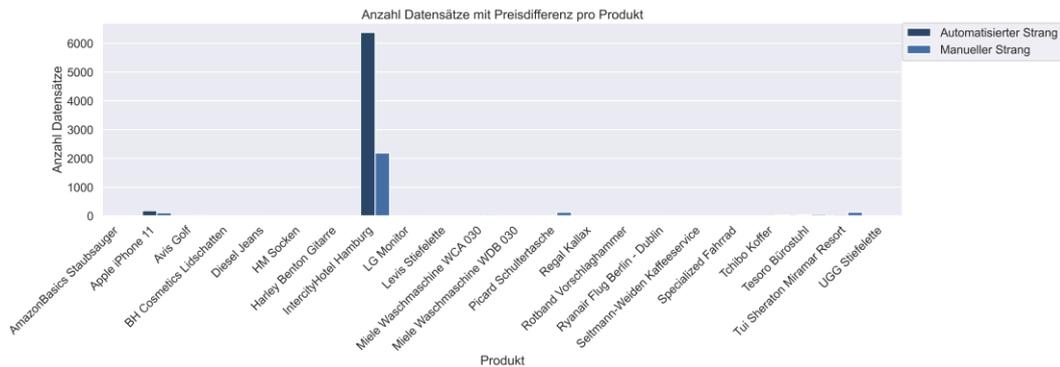
#### 4.2.2.1 Preisunterschiede pro Produkt

Insgesamt ist die Anzahl der Datensätze pro Produkt relativ gleichmäßig verteilt (siehe Abbildung 13). Ausnahmen sind die bereits beschriebenen Änderungen der Produkte „Miele Waschmaschine WBD 030“ und der beiden Stiefeletten.



**Abbildung 13: Anzahl Datensätze nach Produkt**

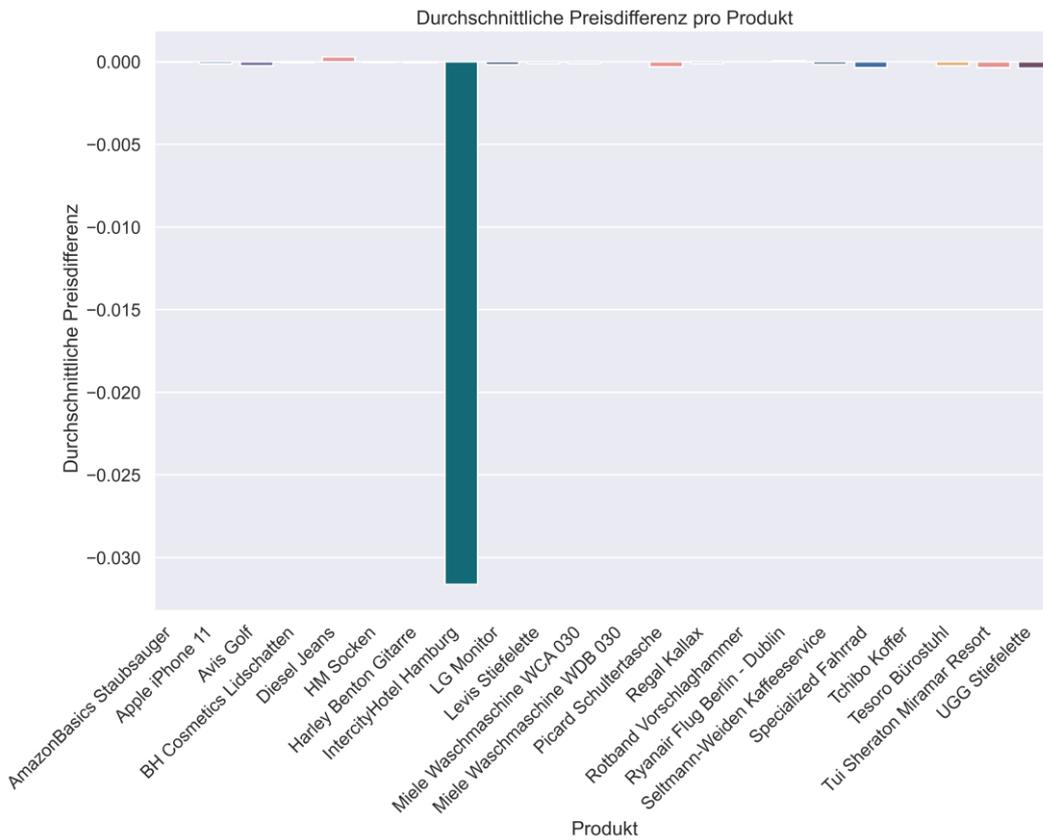
Abbildung 14 stellt die auftretenden Preisdifferenzen, aufgeteilt nach den erhobenen Produkten dar. Auffällig ist die hohe Häufigkeit von Preisdifferenzen des „IntercityHotel Hamburg“ mit 8.571 Preisabweichungen, das entspricht ca. 49,11 Prozent der Datensätze des „IntercityHotel Hamburg“ mit potentiellen Preisabweichungen. Ebenso eine nennenswerte Anzahl an Preisabweichungen weisen das „Apple iPhone 11“ mit 300 Abweichungen (1,92 Prozent), das „Tui Sheraton Miramar Resort“ mit 176 Abweichungen (1,29 Prozent), die „Picard Schultertasche“ mit 132 Abweichungen (0,96 Prozent) und der „Tesoro Bürostuhl“ mit 99 Abweichungen (0,72 Prozent) auf. Weniger als 0,4 Prozent der Datensätze der restlichen Produkte weisen Preisabweichungen auf.



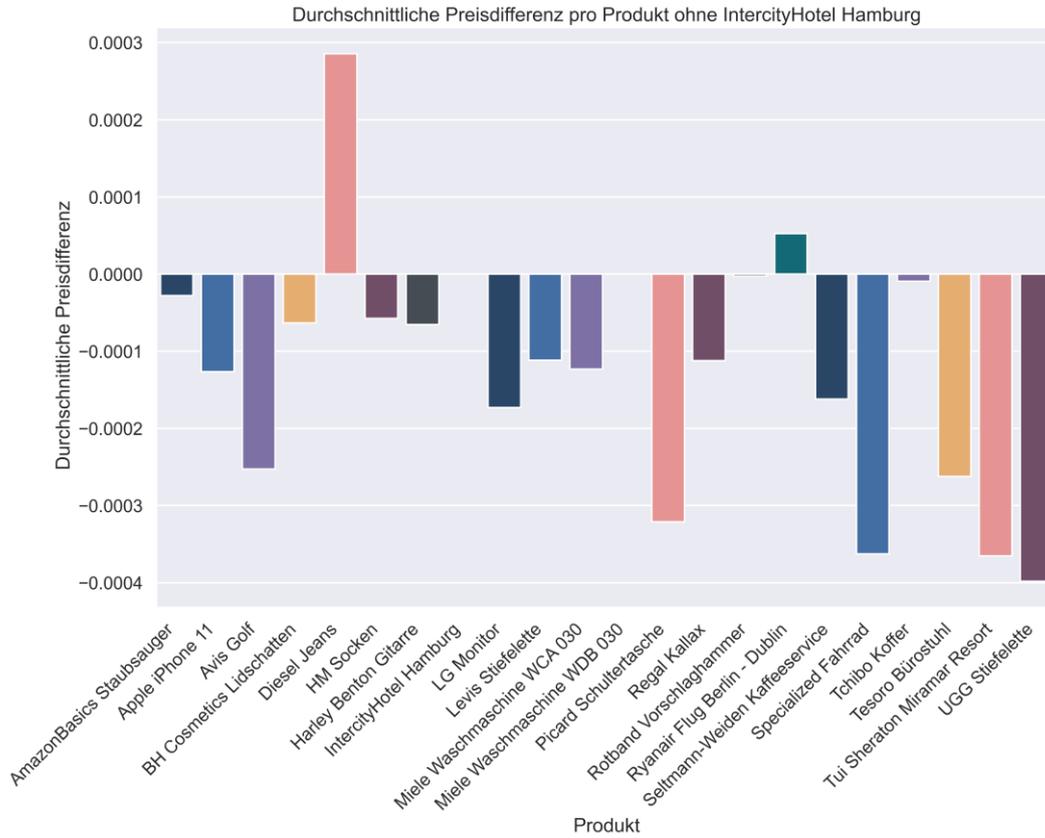
**Abbildung 14: Anzahl Datensätze mit Preisdifferenz nach Produkt**

Da die Personalisierung von Preisen produktspezifisch ist, liegt der Fokus weiterführender Analysen auf Produkten mit mehr als 40 Preisabweichungen. Abbildung 15 stellt schließlich die durchschnittlichen Preisdifferenzen pro Produkt dar und gibt damit auch einen Eindruck für die Stärke der zuvor festgestellten Preisdifferenzen. Hier ist besonders der hohe durchschnittliche Preisunterschied des „IntercityHotel Hamburg“ auffällig, während die restlichen Produkte einen

verschwindend geringen durchschnittlichen Preisunterschied aufweisen. Um einen besseren Einblick der Preisunterscheide der übrigen Produkte zu erhalten, exkludiert Abbildung 16 das IntercityHotel Hamburg. Bemerkenswert ist, dass die auftretenden Preisdifferenzen im Schnitt negativ sind und sich in einem sehr geringen Spektrum von -0,04 Prozent und +0,03 Prozent bewegen. Einzig die Produkte „Diesel Jeans“ und „Ryanair Flug Berlin - Dublin“ weisen eine im Durchschnitt positive Preisdifferenz auf.

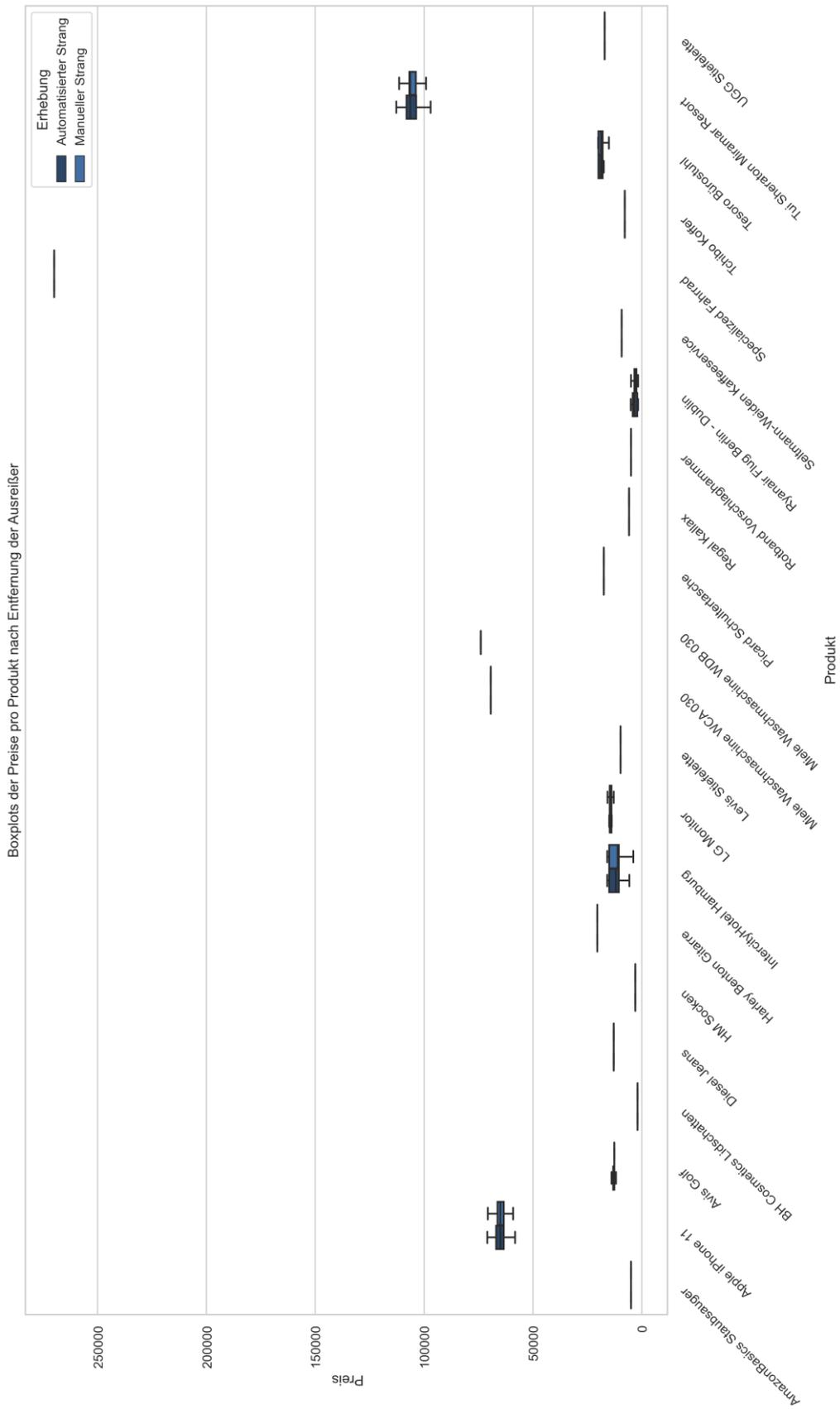


**Abbildung 15: Durchschnittliche Preisdifferenz nach Produkt**



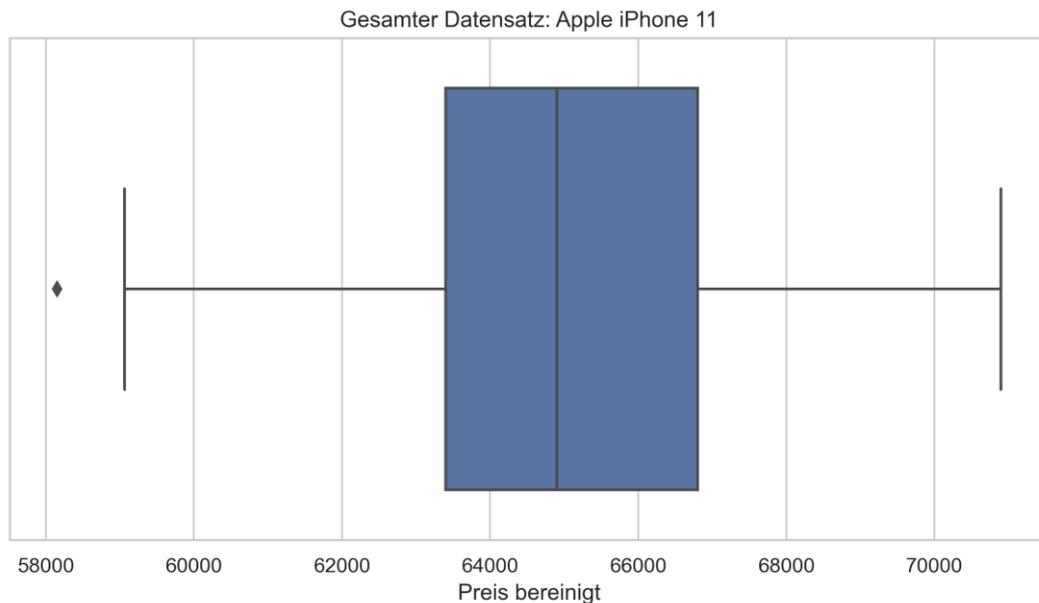
**Abbildung 16: Durchschnittliche Preisdifferenz nach Produkt ohne „IntercityHotel Hamburg“**

Zusammenfassen lassen sich die gewonnen Erkenntnisse bezüglich der Preisunterschiede pro Produkt mit Hilfe von Boxplots. Boxplots ermöglichen die Darstellung des Zusammenhangs einer kategorialen Variablen (Produkt) mit einer kontinuierlichen Variablen (Preis/Preisdifferenz). Zusätzlich zu dem bereits analysierten Durchschnittswert der Preisdifferenz enthalten Boxplots Information zu Variabilität und Verteilung der Zielvariable. Dazu werden fünf charakteristische Datenpunkte dargestellt: Minimum, erstes Quartil, Median, drittes Quartil und Maximum. Sämtliche außerhalb des Maximums und Minimums gelegenen Datenpunkte sind Ausreißer, welche maximal 0,7 Prozent der gesamten Daten darstellen. Abbildung 17 stellt die Boxplots der Preise der jeweiligen Produkte dar.

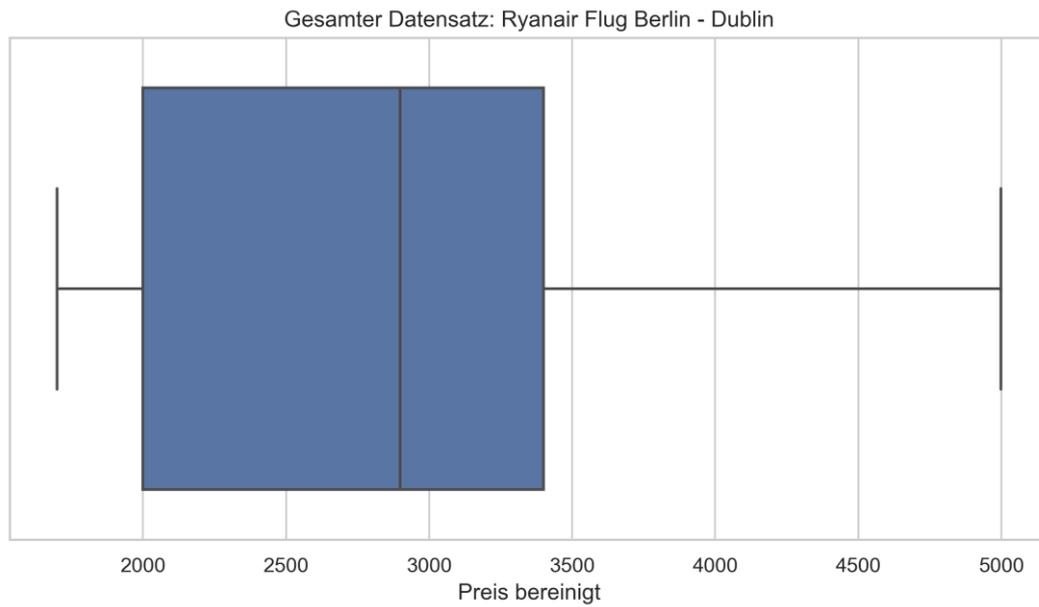


**Abbildung 17: Boxplots der Preise nach Produkt nach Entfernung der Ausreißer**

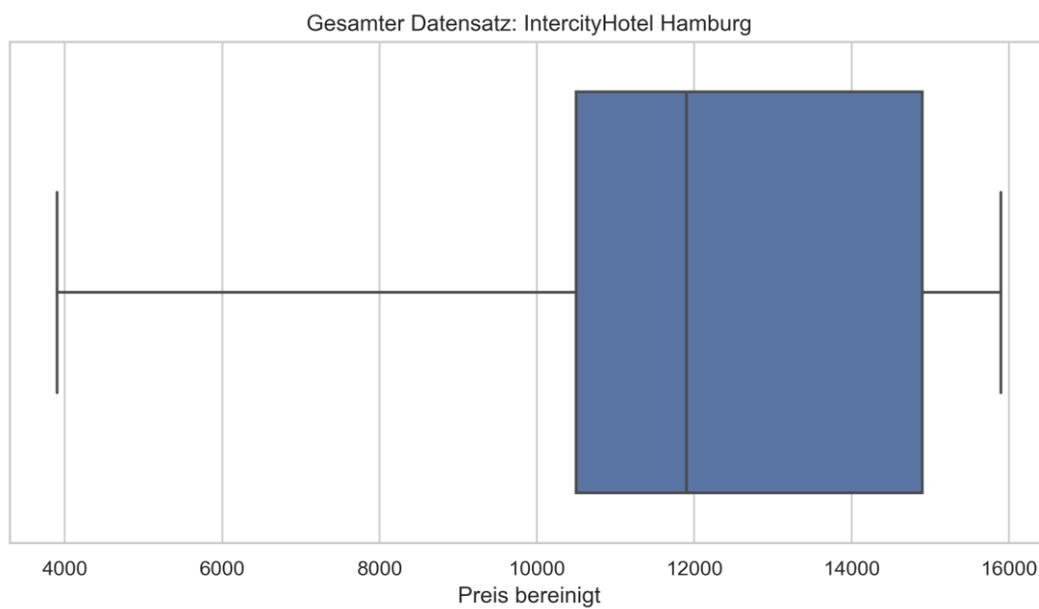
Zu beachten ist, dass die Produktpreise im Gegensatz zur Preisdifferenz auch eine Zeitkomponente enthalten und somit auch eine Streuung des Preises bei sonst konstanten Einflussvariablen anzeigen. Zusätzlich wurden Ausreißer, welche extreme Datenpunkte darstellen, zur verbesserten Darstellung entfernt. Wie schon aufgrund der Anzahl an Preisdifferenzen zu erwarten, zeigen die schon aufgrund der Anzahl an Datensätzen mit Preisdifferenz identifizierten Produkte „IntercityHotel Hamburg“, „Apple iPhone 11“, „Tui Sheraton Miramar Resort“, „Picard Schultertasche“ und „Tesoro Bürostuhl“ eine Streuung der Preisvariable und damit Preisveränderungen auf. Zur besseren Darstellung sind die um die Ausreißer bereinigten Boxplots der Produkte Apple iPhone 11, Ryanair Flug Berlin – Dublin, IntercityHotel Hamburg, Tesoro Bürostuhl und Tui Sheraton Miramar Resort in den Abbildungen Abbildung 18 bis Abbildung 22 einzeln dargestellt. Eine weitere Auffälligkeit der Boxplots aller Produkte ist eine Verteilung der Preisvariable der Produkte „Avis Golf“ und „LG Monitor“. Dies ist auf eine Preisveränderung im Laufe des Erhebungszeitraums zurückzuführen.



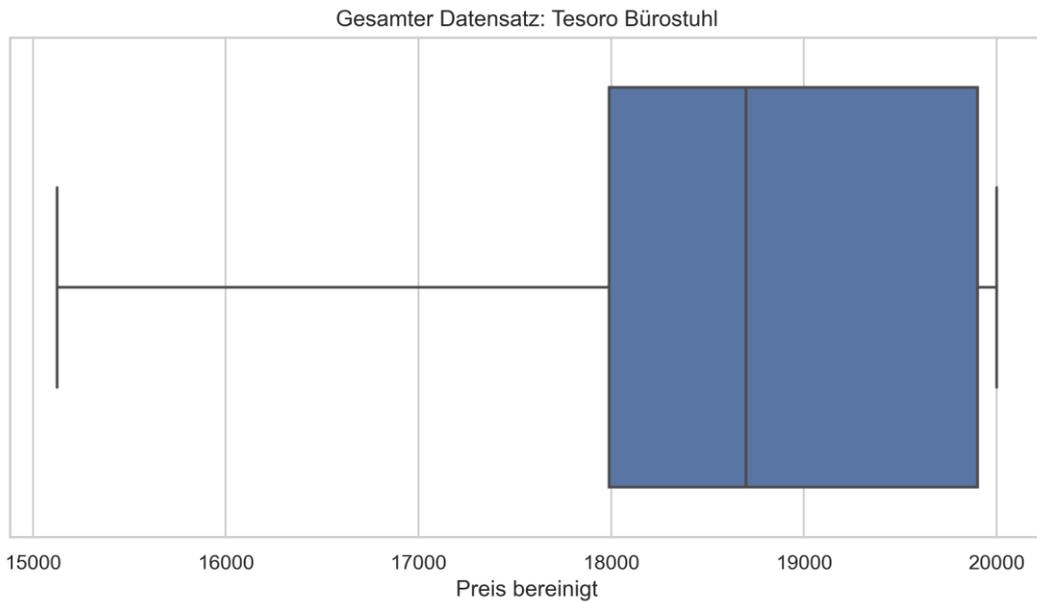
**Abbildung 18: Boxplot bereinigter Preis Apple iPhone 11**



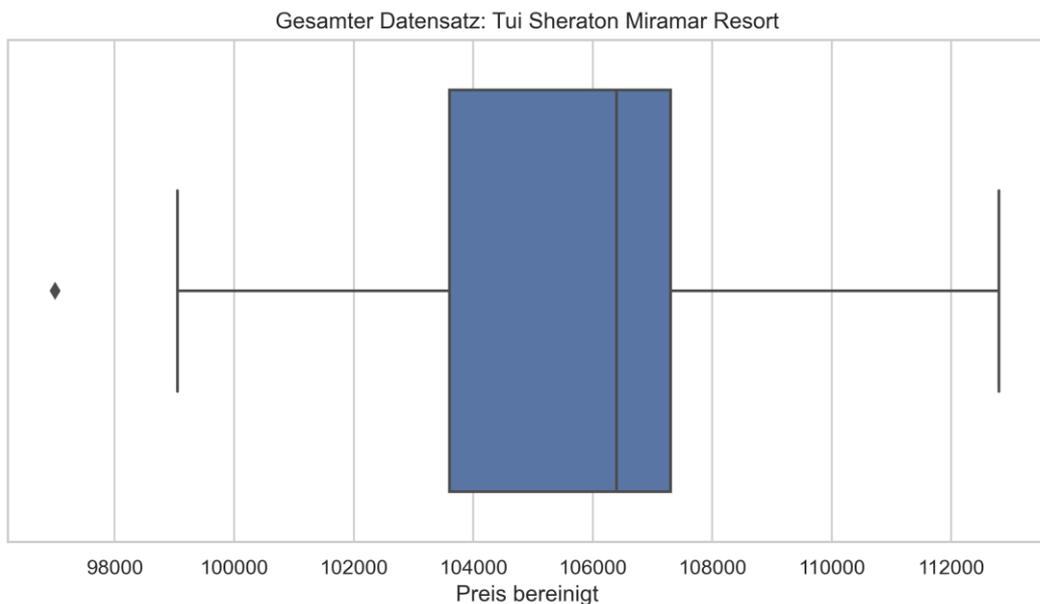
**Abbildung 19: Boxplot bereinigter Preis Ryanair Flug Berlin - Dublin**



**Abbildung 20: Boxplot bereinigter Preis IntercityHotel Hamburg**



**Abbildung 21: Boxplot bereinigter Preis Tesoro Bürostuhl**



**Abbildung 22: Boxplot bereinigter Preis Tui Sheraton Miramar Resort**

Abbildung 23 stellt die Preisdifferenzen unabhängig von zeitlichen Veränderungen dar. Die Berechnungsform der Preisdifferenzen betrachtet per Definition nur punktuelle Preisveränderungen und bereinigt damit die Boxplots der Preise um die Zeitkomponente. Pro Produkt werden die Boxplots der relativen Preisunterschiede der einzelnen Produkte gegenübergestellt. Hier zeigt sich, dass das „IntercityHotel Hamburg“ eine systematische Verteilung der Preisdifferenz zwischen 0 und -10 Prozent aufweist. Sämtliche Preisdifferenzen der übrigen Produkte werden nur als Ausreißer dargestellt, da die Mehrheit der Datensätze keine Preisdifferenzen aufweist.

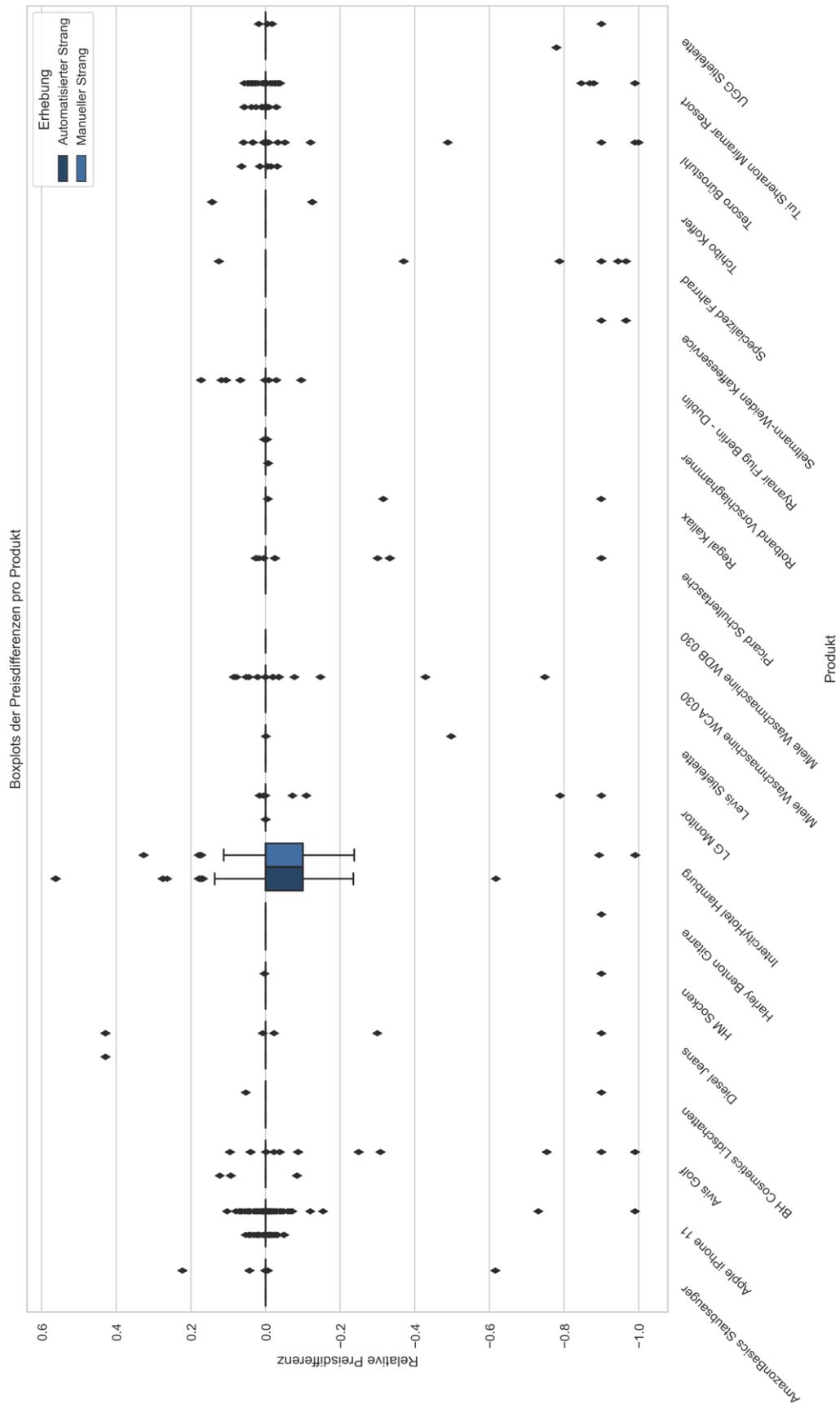
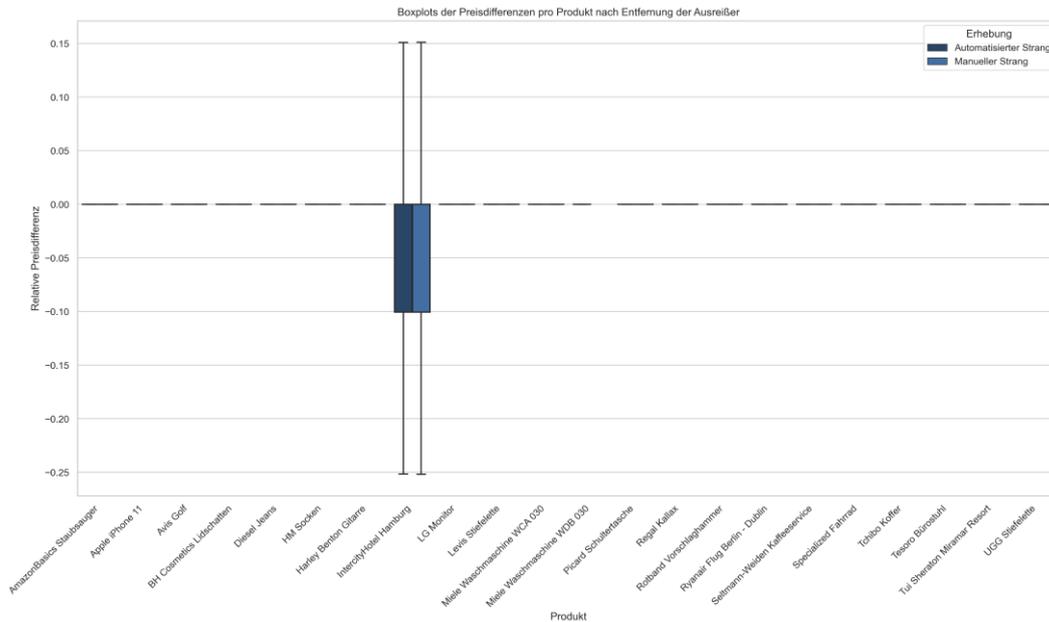
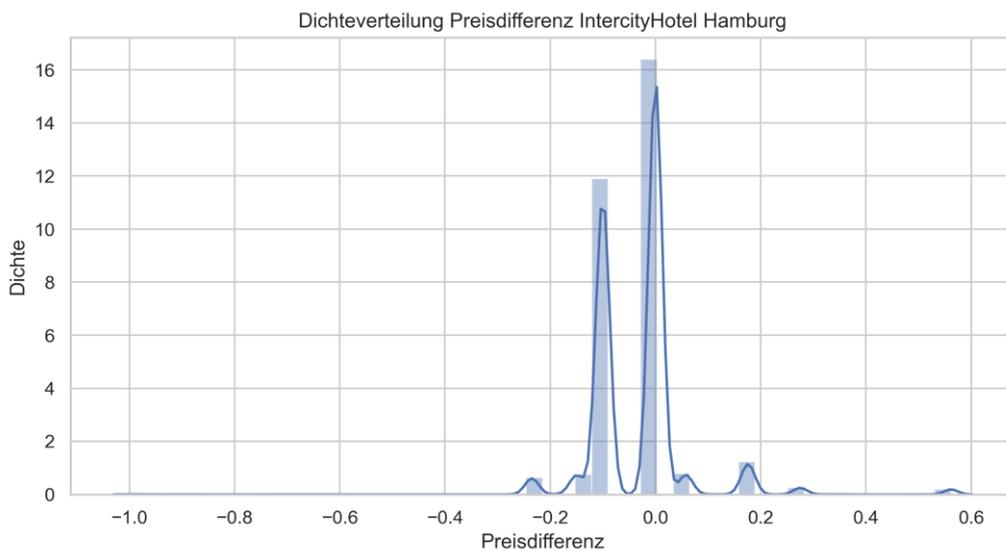


Abbildung 23: Boxplots der Preisdifferenzen nach Produkt

Abbildung 24 stellt die Preisdifferenzen der Produkte nach Entfernung der Ausreißer dar. Hier ist noch einmal die systematische Preisabweichung des „IntercityHotel Hamburg“ sichtbar. Mit Blick auf die Dichteverteilung in Abbildung 25 wird ersichtlich, dass die Mehrheit der Preisdifferenzen auf eine Preisdifferenz von -10 Prozent zurückzuführen ist.



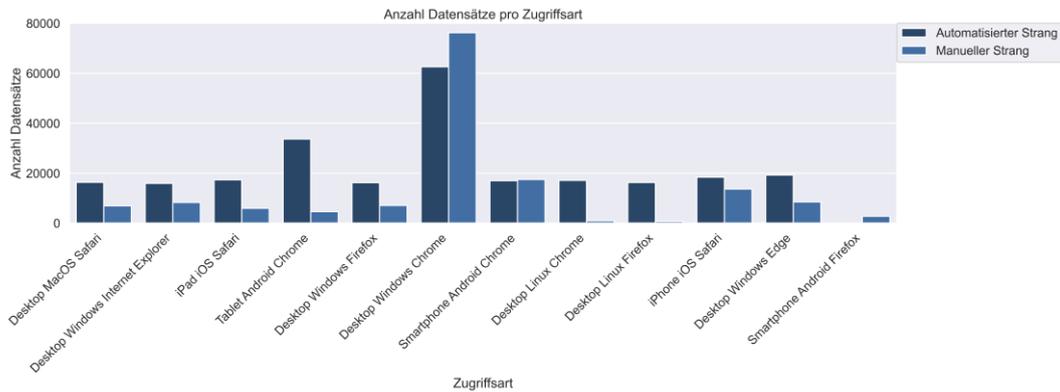
**Abbildung 24: Boxplots der Preisdifferenzen nach Entfernung der Ausreißer**



**Abbildung 25: Dichteverteilung „IntercityHotel Hamburg“**

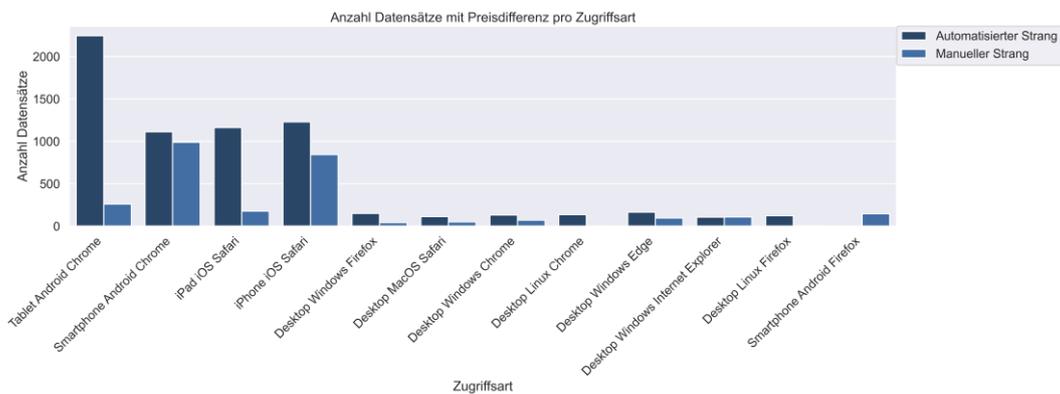
#### 4.2.2.2 Preisunterschiede pro Zugriffsart

Mit Fokus auf die pro Zugriffsart erhobenen Datensätze ist die hohe Anzahl der erhobenen Kombination „Desktop Windows Chrome“ auffällig, vgl. Abbildung 26. Notwendig ist diese hohe Anzahl von Standardfällen als Referenzwert zur Berechnung der Preisdifferenzen.

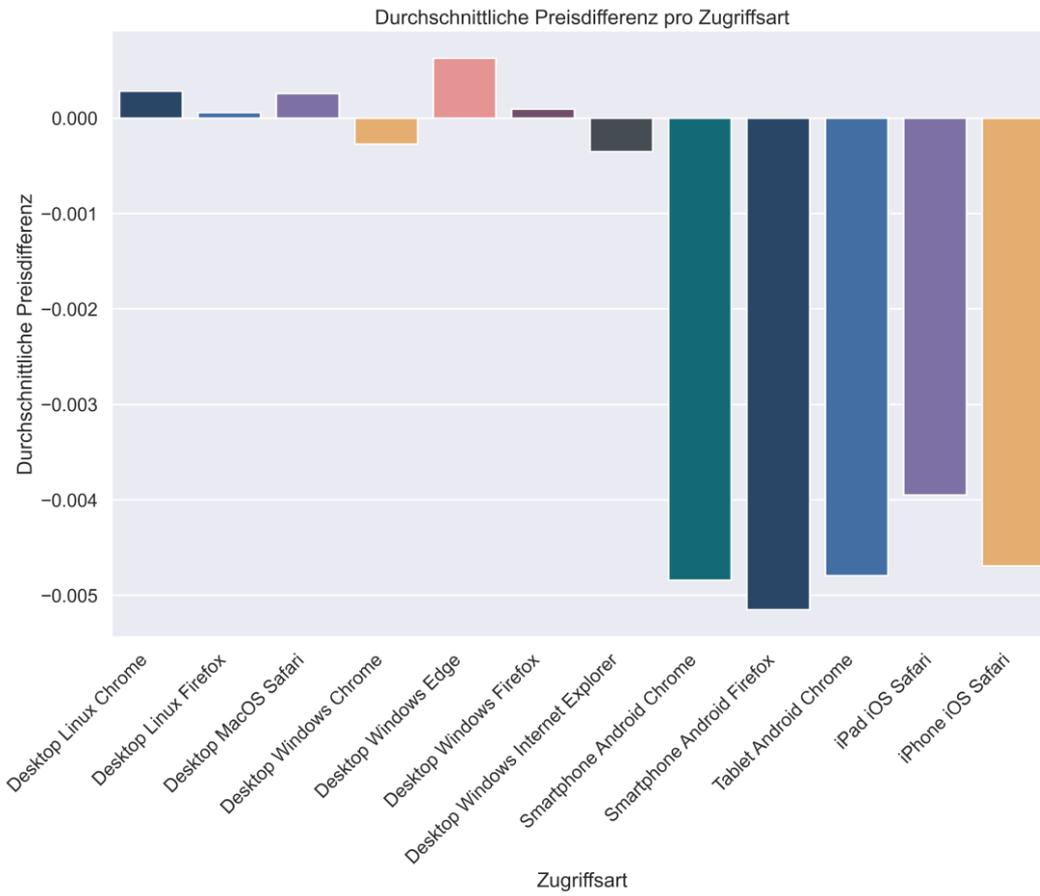


**Abbildung 26: Anzahl Datensätze nach Zugriffsart**

Nach Kalkulation der Preisdifferenzen stellt sich eine erhöhte Anzahl an Datensätzen mit Preisdifferenzen bei Verwendung eines mobilen Gerätes heraus. Abbildung 27 und Abbildung 28 stellen die Preisdifferenzen mit Hilfe der Anzahl an Datensätzen sowie der durchschnittlichen Preisdifferenz pro Zugriffsart dar. Bei Abfrage des Preises mit einem mobilen Gerät sind die Preise zwischen 0,4 Prozent und 0,5 Prozent günstiger.



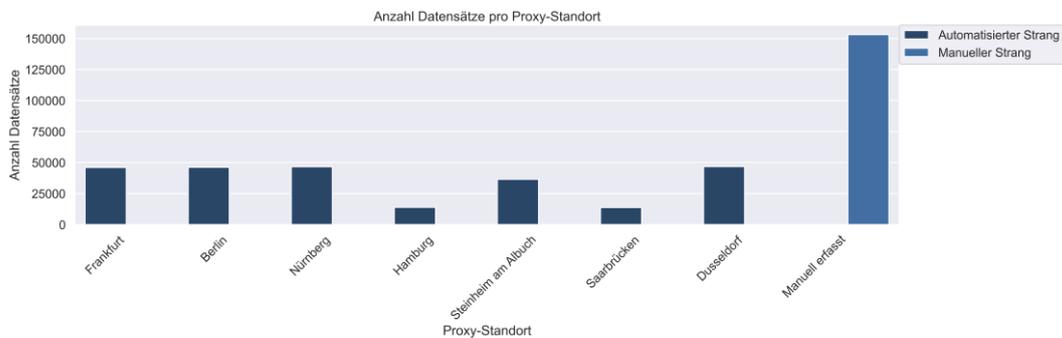
**Abbildung 27: Anzahl Datensätze mit Preisdifferenz nach Zugriffsart**



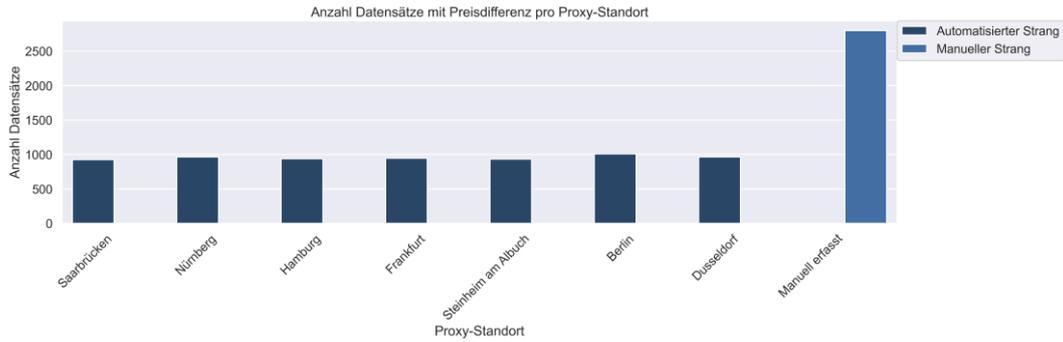
**Abbildung 28: Durchschnittliche Preisdifferenz nach Zugriffsart**

#### 4.2.2.3 Preisunterschiede pro Proxy-Standort

Die Anzahl an Datensätzen, die pro Proxy-Standort erhoben wurden, variiert mit Ausnahme von Hamburg und Saarbrücken kaum. Dort wurden nur knapp die Hälfte an Datensätzen erhoben. Die Datensätze des manuellen Erhebungsstranges werden nicht aufgrund der IP unterschieden und in der Grafik in eine Vergleichskategorie zusammengefasst. Die Anzahl der pro Proxy-Standort erhobenen Datensätze mit Preisdifferenzen sind sehr gleichmäßig verteilt.

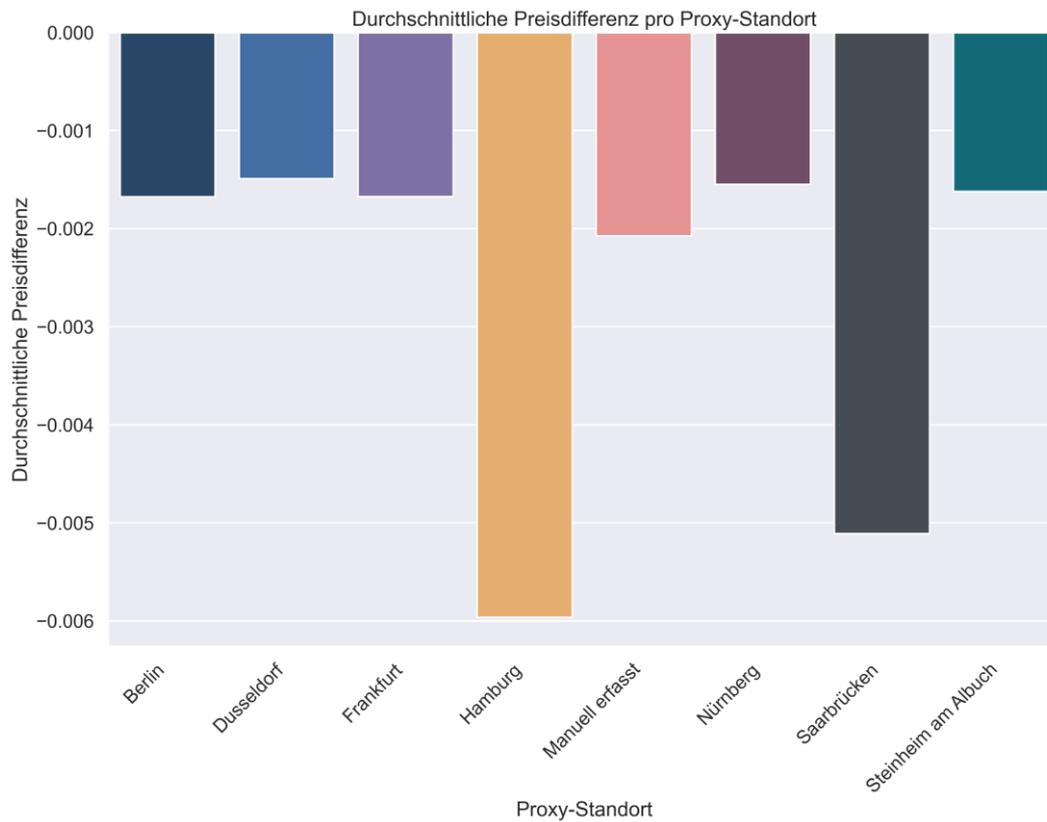


**Abbildung 29: Anzahl Datensätze nach Proxy-Standort**



**Abbildung 30: Anzahl Datensätze mit Preisdifferenz nach Proxy-Standort**

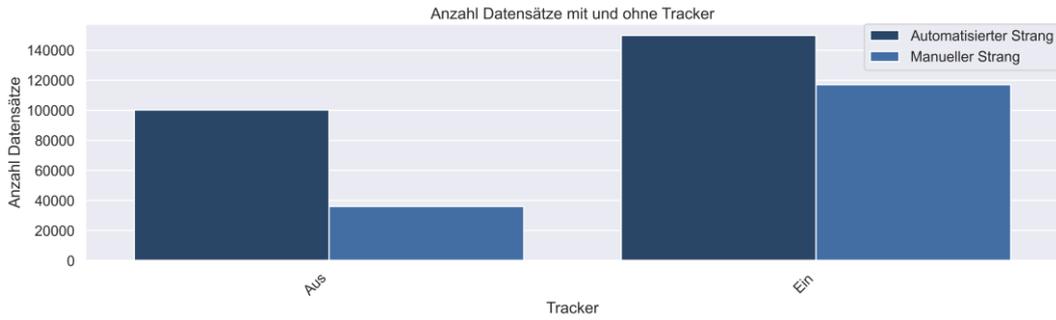
Betrachtet man die Höhe der erfassten Preisunterschiede in Abbildung 31, so weisen Hamburg und Saarbrücken eine geringfügig höhere, durchgehend negative relative Preisdifferenz auf.



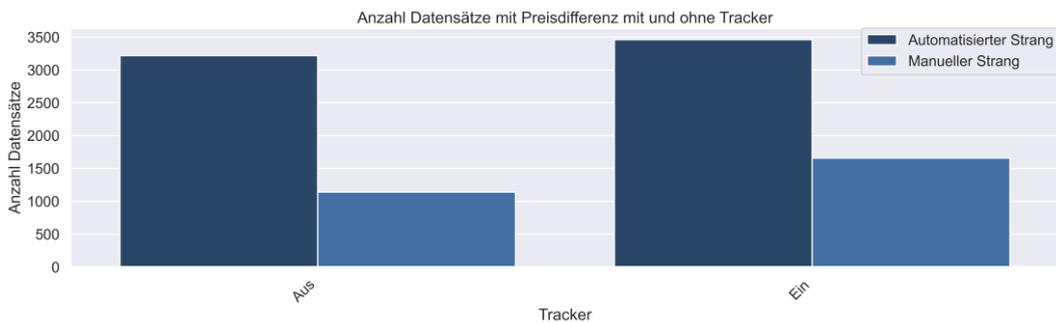
**Abbildung 31: Durchschnittliche Preisdifferenz nach Proxy-Standort**

#### 4.2.2.4 Preisunterschiede und Tracker

Abbildung 32 zeigt, dass generell mehr Datensätze mit erlaubten Trackern als ohne Tracker erhoben wurden. Die Anzahl an Datensätzen mit Preisdifferenzen verteilt sich relativ ähnlich zu der Anzahl an erhobenen Daten.

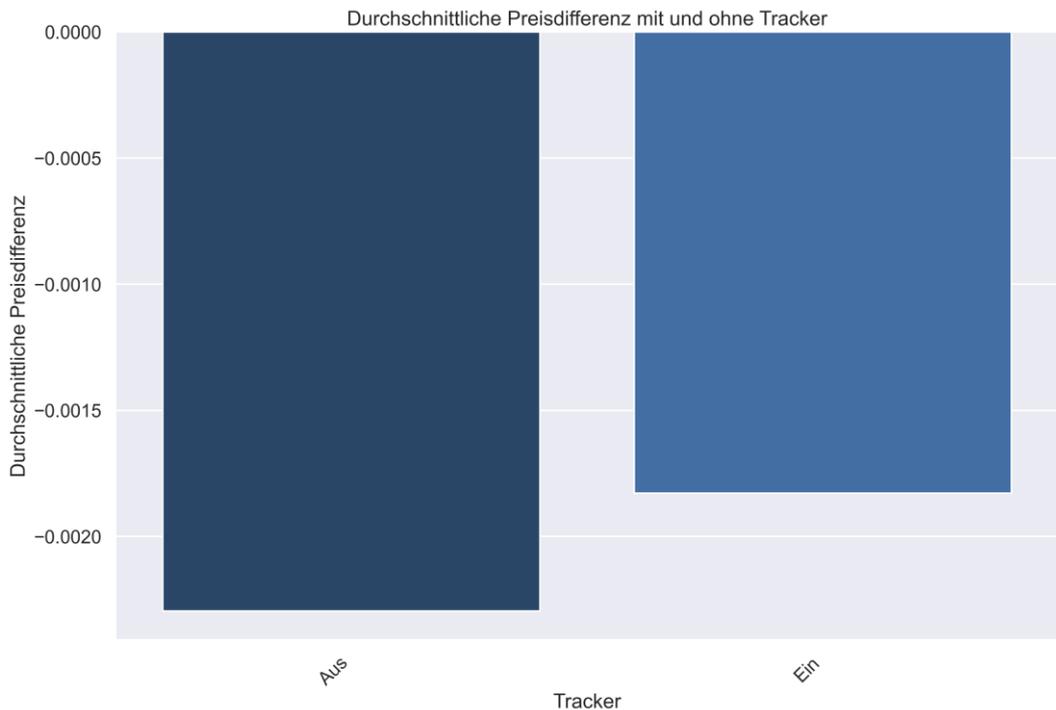


**Abbildung 32: Anzahl Datensätze mit und ohne Tracker**



**Abbildung 33: Anzahl Datensätze mit Preisdifferenz mit und ohne Tracker**

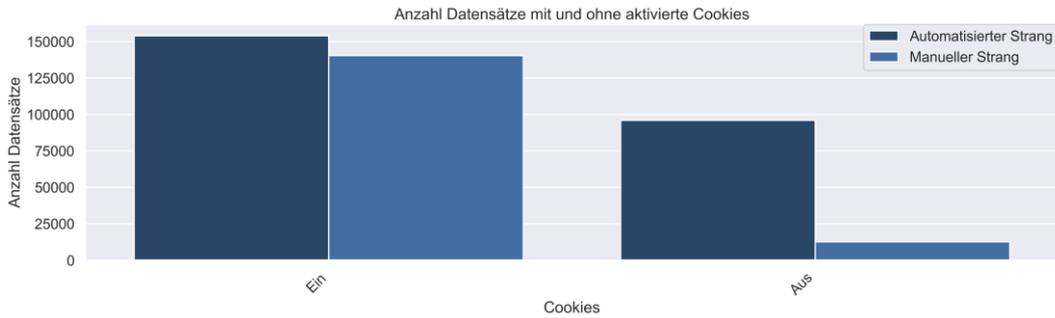
Abbildung 34 stellt die durchschnittliche Preisdifferenz mit und ohne erlaubte Tracker dar, wobei ausgeschaltete Tracker durchschnittlich eine etwas höhere negative Preisdifferenz aufweisen.



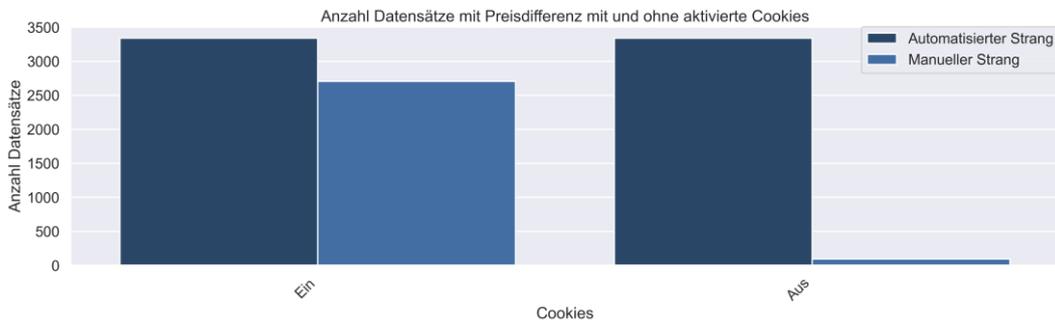
**Abbildung 34: Durchschnittliche Preisdifferenz mit und ohne Tracker**

#### 4.2.2.5 Preisunterschiede und Cookies

Abbildung 35 und Abbildung 36 stellen die Anzahl der erhobenen Datensätze sowie die Verteilung der Datensätze mit Preisdifferenz mit und ohne erlaubte Cookies dar. Hier zeigen sich keine Auffälligkeiten.

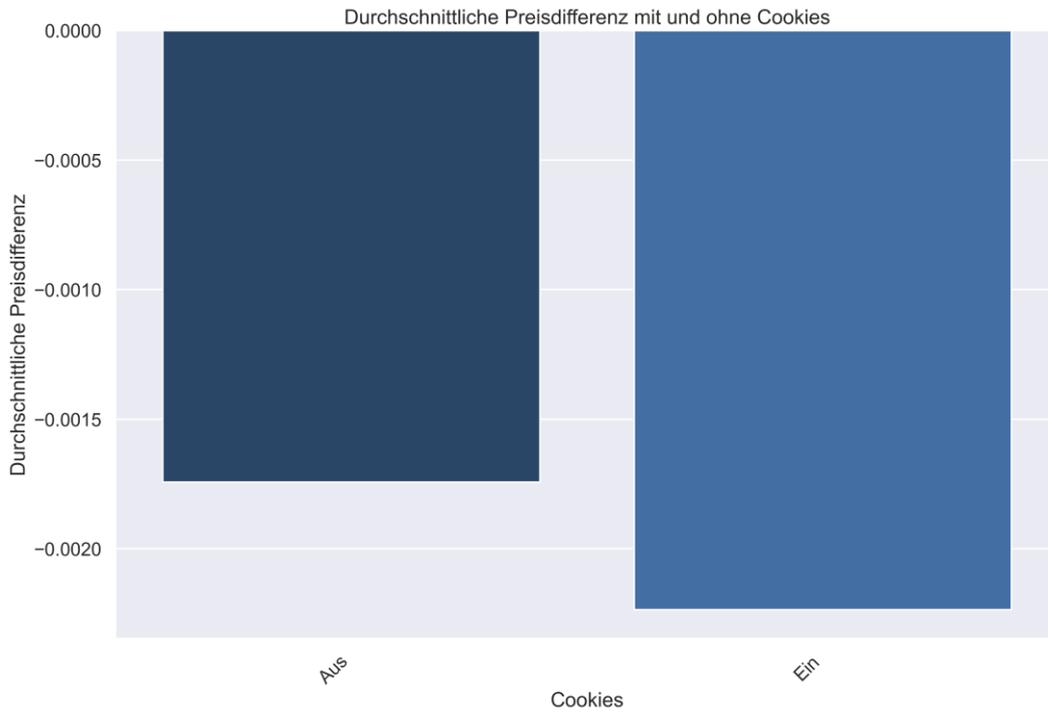


**Abbildung 35: Anzahl Datensätze mit und ohne aktivierte Cookies**



**Abbildung 36: Anzahl Datensätze mit Preisdifferenz mit und ohne aktivierte Cookies**

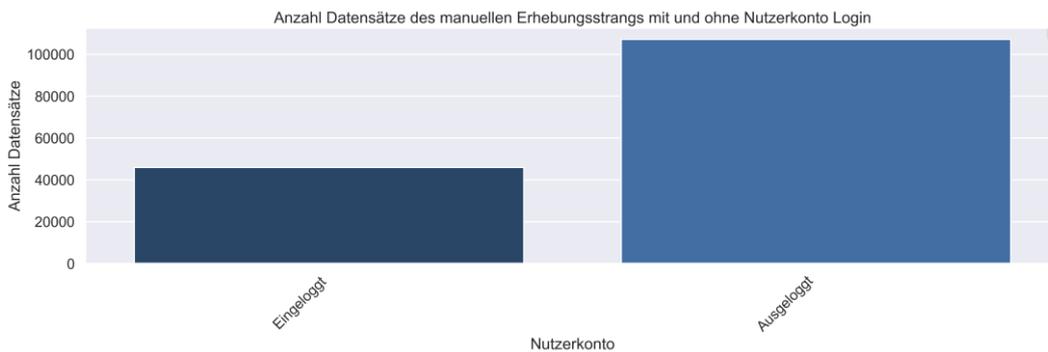
Abbildung 37 zeigt geringe Unterschiede in der Preiserhebung mit und ohne Cookies.



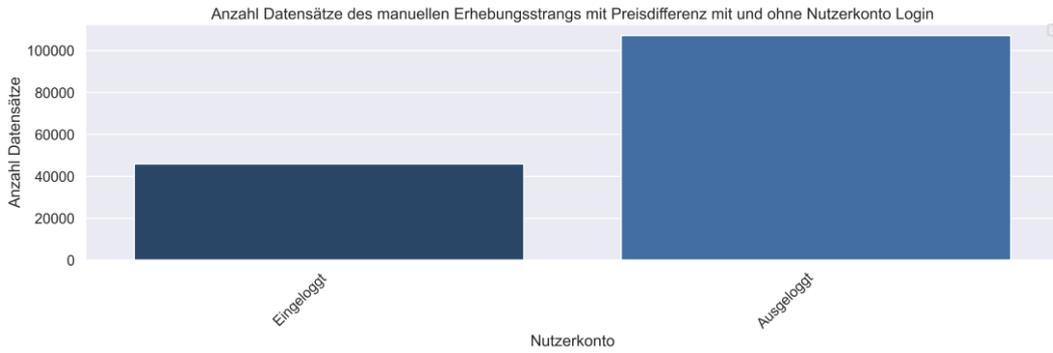
**Abbildung 37: Durchschnittliche Preisdifferenz mit und ohne aktivierte Cookies**

#### 4.2.2.6 Preisunterschiede und Nutzerkonto

Die Verteilung der Anzahl an Datensätzen gesamt und mit Preisdifferenz der Analysevariable Nutzerkonto ist in Abbildung 38 und Abbildung 39 dargestellt. Unterschieden wird, ob der Preis mit einem eingeloggten Nutzerkonto oder ohne ein eingeloggtes Nutzerkonto erhoben wurde.

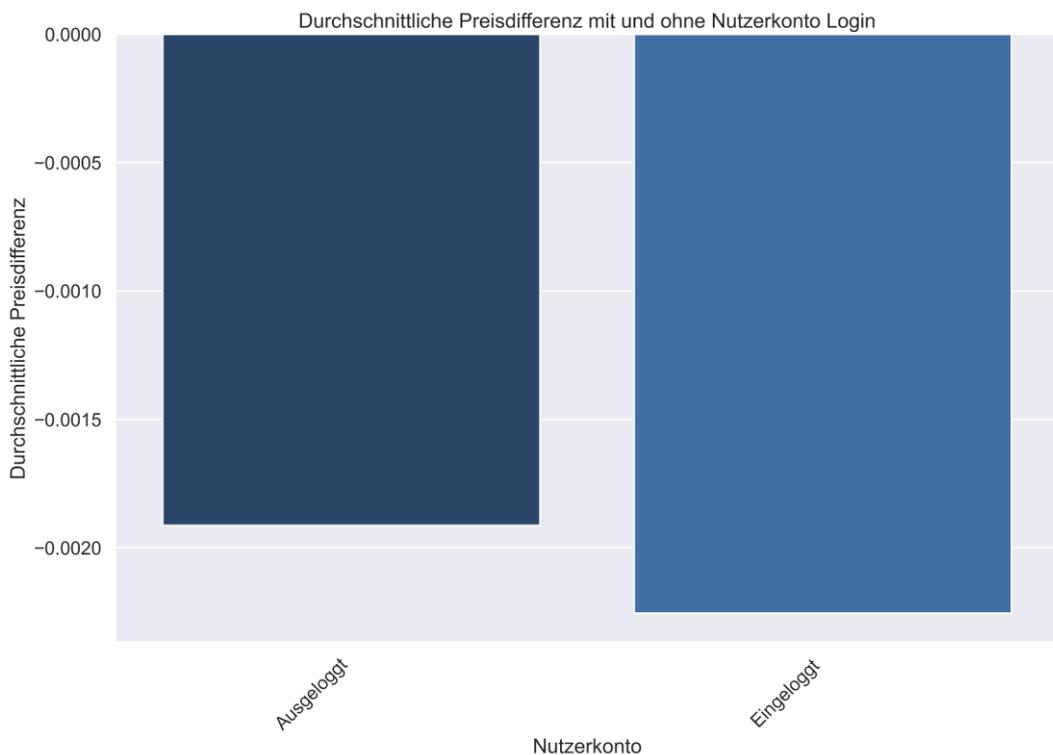


**Abbildung 38: Anzahl Datensätze des manuellen Erhebungsstrangs mit und ohne Nutzerkonto-Login**



**Abbildung 39: Anzahl Datensätze mit Preisdifferenzen des manuellen Erhebungsstrangs mit und ohne Nutzerkonto-Login**

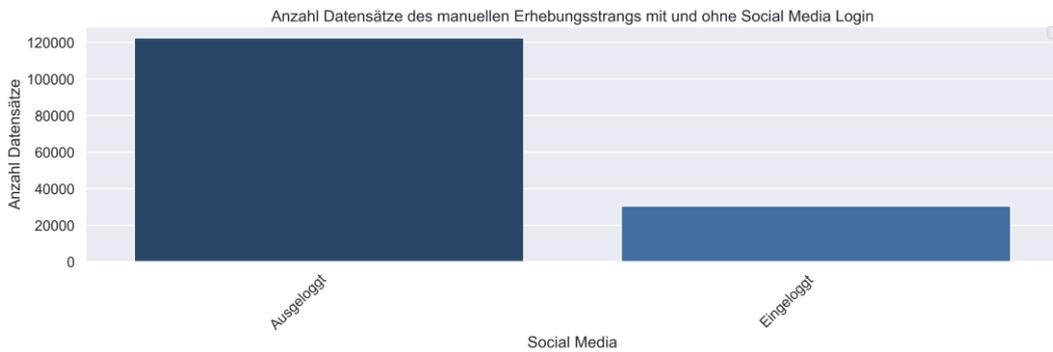
Der Vergleich der beiden Durchschnittspreise in Abbildung 40 zeigt nur geringe Differenzen auf.



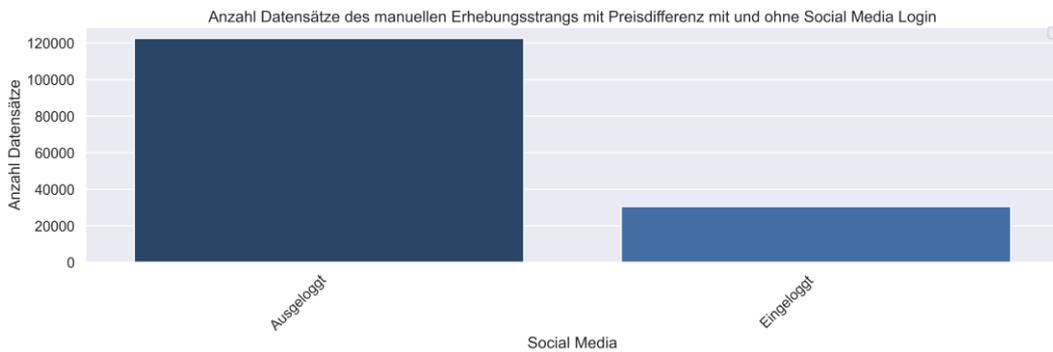
**Abbildung 40: Durchschnittliche Preisdifferenz mit und ohne Nutzerkonto-Login**

#### 4.2.2.7 Preisunterschiede und Social-Media-Login

Vergleichbar zum Nutzerkonto wird auch der Einfluss eines Social-Media-Profiles auf die erfassten Preisdifferenzen betrachtet. Abbildung 41 und Abbildung 42 stellen die erhobenen Datensätze den Datensätzen mit Preisdifferenzen gegenüber.

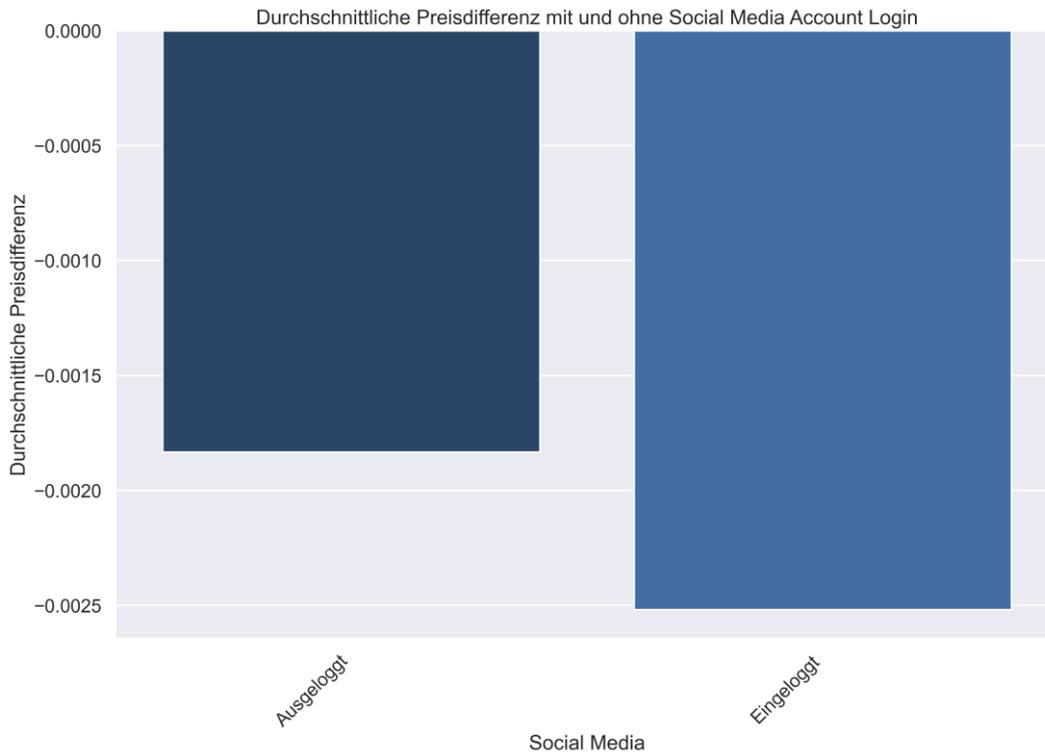


**Abbildung 41: Anzahl Datensätze des manuellen Erhebungsstrangs mit und ohne Social-Media-Login**



**Abbildung 42: Anzahl Datensätze mit Preisdifferenzen des manuellen Erhebungsstrangs mit und ohne Social-Media-Login**

Abbildung 43 vergleicht die erhobenen Durchschnittspreise mit und ohne Login, welche nur eine geringe Differenz aufweisen.



**Abbildung 43: Durchschnittliche Preisdifferenz mit und ohne Social-Media-Account-Login**

## 4.3 Regressionsanalyse

### 4.3.1 Einführung

Nach Betrachtung der deskriptiven Statistik der Preisdifferenzen entlang der gegebenen Analysedimensionen wird mittels einer multivariaten Regressionsanalyse die Stärke dieser potentiellen Einflussfaktoren ermittelt.

Ein multivariates Regressionsmodell untersucht quantitativ, wie sich eine abhängige Variable – in unserem Fall die relative Preisdifferenz – unter dem Einfluss mehrerer unabhängiger Variablen (Einflussfaktoren) ändert. Dabei erlaubt das Modell auch die quantitative Schätzung des Ausmaßes des Einflusses der einzelnen unabhängigen Variablen. Für die Auswertung wurde ein Vorgehen nach der Methode der Kleinst-Quadrat-Schätzung gewählt.

Insgesamt werden verschiedene Regressionen durchgeführt und verglichen: Regression mit allen Datensätzen, Regression nach Entfernung von Produkten ohne Preisdifferenzen, Regression des auffälligen Produktes „IntercityHotel Hamburg“ sowie Regression der auffälligen Produkte ohne „IntercityHotel Hamburg“ und zuletzt Regression der Daten des manuellen Strangs und der nur dort erfassten Variablen.

Zusätzlich zu den unterschiedlichen Größen der Datensätze variieren auch die betrachteten unabhängigen Variablen. Dies hängt damit zusammen, dass die

Variablen Erfasser, Social Media und Nutzerkonto nur für den manuellen Strang erhoben wurden. Wird der Einfluss der exklusiv im manuellen Strang erhobenen Variablen auf den gesamten Datensatz betrachtet, wäre die Variable immer ein Indikator für die Verschiedenheit der beiden erhobenen Datensätze, da die im automatisierten Strang erhobenen Daten dort keine Werte aufweisen. Über beide Erhebungsstränge erfasste Variablen sind: Produkt, Cookies, Tracker und Zugriff. Diese Variablen werden in sämtlichen Regressionen analysiert.

Tabelle 6 gibt einen Überblick der in den Regressionen verwendeten „Feature-Sets“ oder unabhängigen Variablen. Zu beachten ist, dass sich bei Verwendung des Feature-Set 2 der Datensatz auf die manuell erhobenen Daten reduziert.

Die Regressionsergebnisse werden am Ende der Abschnitte 4.3.3 bis 4.3.7 im Detail dargestellt.

Regression	Alle Datensätze (4.3.3)	Entfernung von Produkten < 40 Preisdifferenzen (4.3.4)	IntercityHotel Hamburg (4.3.5)	Ohne IntercityHotel Hamburg (4.3.6)	Manueller Strang (4.3.7)
Feature-Set 1	Produkt, Cookies, Tracker, Zugriff	Produkt, Cookies, Tracker, Zugriff	Produkt, Cookies, Tracker, Zugriff	Produkt, Cookies, Tracker, Zugriff	Produkt, Cookies, Tracker, Zugriff
Feature-Set 2			+ Erfasser, Social Media, Nutzerkonto	+ Erfasser, Social Media, Nutzerkonto	+ Erfasser, Social Media, Nutzerkonto

**Tabelle 6: Feature-Sets der Regressionen**

Zur Beurteilung und Verbesserung der Modellqualität wurde jeweils auch ein Test auf den Varianzinflationsfaktor durchgeführt, der die Relevanz der einzelnen Einflussvariablen evaluiert. Getestet wird dabei, inwiefern sich das Modell verbessert, wenn die jeweils betrachtete Variable aus dem Modell entfernt wird. Damit wird dem Multikollinearitätsproblem Rechnung getragen. Liegt der Varianzinflationsfaktor unter fünf, so sollte die zugehörige Variable im Modell behalten werden.

### 4.3.2 Zur Interpretation der Regressionsergebnisse

Zur Berechnung der Regressionen wurde Python verwendet. Im Folgenden soll kurz skizziert werden, wie der Ergebnisoutput in den Tabellen Tabelle 7 bis Tabelle 14 interpretiert werden kann und welche Größen von Bedeutung sind.

Vereinfacht kann bei einer linearen Regression, wie sie hier zugrunde gelegt wird, die abhängige Variable (hier die relative Preisdifferenz,  $y$ ) durch die unabhängigen Variablen (Einflussfaktoren,  $x_i$ ) wie folgt ausgedrückt werden:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \varepsilon$$

Dabei quantifizieren die Beta-Werte den Einfluss des betrachteten Einflussfaktors auf die relative Preisdifferenz. Die Beta-Koeffizienten geben an, in welchem Ausmaß sich die abhängige Variable in Bezug verändert, wenn sich die unabhängige Variable ändert. Im Output von Python entspricht dies der Spalte „coef“.

Die beiden nächsten Spalten „std err“ und „t“ sind zur Interpretation zunächst nicht unbedingt notwendig. „std err“ entspricht dem Standardfehler der Regression, „t“ ist einfach „coef“ geteilt durch „std err“.

Interessant ist weiterhin der p-Wert. Er dient zur Einschätzung der Signifikanz und gibt an, ob der ermittelte Koeffizient (Beta) der jeweiligen Variablenausprägung unterschiedlich zu der Referenzkategorie ist. Die Referenzkategorie stellt dabei die am häufigsten auftretende Kategorie einer Variablen dar. Ist der p-Value kleiner als das vordefinierte Signifikanzniveau von 5 Prozent, so kann der Koeffizient als der Preisunterschied zu der Referenzkategorie interpretiert werden. Die Preisdifferenzen sind indirekt über den Intercept des Regressionsmodells erfasst. Im Output von Python entspricht dies der Spalte „P>|t|“.

Der Angabe des p-Werts äquivalent ist die Angabe des Intervalls [0.025;0.975]. 95 Prozent der Werte liegen in diesem Intervall. Vereinfacht ausgedrückt ist Beta aussagekräftiger, je enger das Intervall ausfällt.

Das Maß für die Anpassungsgüte  $R^2$  („R-squared“) schließlich beschreibt, wie viel Prozent der im Datensatz auftretenden Streuung durch das Modell erklärt werden können. Das angepasste  $R^2$  („Adj. R-squared“) passt den Wert an die verwendete Anzahl von Freiheitsgraden an und bezieht somit die Anzahl an inkludierten Variablen mit ein.

### 4.3.3 Regression mit allen Datensätzen

Zunächst wurde die Regression mit allen verfügbaren 286.011 Datensätzen durchgeführt. Tabelle 7 zeigt den Ergebnisoutput.

Dep. Variable:	pricediffrel	R-squared:	0.089			
Model:	OLS	Adj. R-squared:	0.089			
Method:	Least Squares	F-statistic:	825.3			
No. Observations:	286011	AIC:	-1,29E+09			
Df Residuals:	285976	BIC:	-1,29E+09			
Df Model:	34					
Covariance Type:	nonrobust					
Referenzkategorien Seltmann-Weiden Kaffeeservice, Desktop Window s Chrome						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0018	0.000	6.887	0.000	0.001	0.002
AmazonBasics_Staubsauger	0.0002	0.000	0.706	0.480	-0.000	0.001
Apple_iPhone_11	5,19E-02	0.000	0.186	0.852	-0.000	0.001
Avis_Golf	-5,41E-02	0.000	-0.187	0.852	-0.001	0.001
BH_Cosmetics_Lidschatten	0.0002	0.000	0.555	0.579	-0.000	0.001
Diesel_Jeans	0.0005	0.000	1.671	0.095	-8.34e-05	0.001
HM_Socken	0.0001	0.000	0.466	0.641	-0.000	0.001
Harley_Benton_Gitarre	0.0001	0.000	0.464	0.643	-0.000	0.001
IntercityHotel_Hamburg	-0.0315	0.000	-116.311	0.000	-0.032	-0.031
LG_Monitor	4.23e-06	0.000	0.015	0.988	-0.001	0.001
Levis_Stiefelette	9,12E-02	0.000	0.277	0.782	-0.001	0.001
Miele_Waschmaschine_WCA_030	0.0002	0.000	0.529	0.597	-0.000	0.001
Miele_Waschmaschine_WDB_030	0.0001	0.001	0.123	0.902	-0.002	0.002
Picard_Schultertasche	-0.0001	0.000	-0.422	0.673	-0.001	0.000
Regal_Kallax	8,68E-02	0.000	0.301	0.764	-0.000	0.001
Rotband_Vorschlaghammer	0.0002	0.000	0.685	0.494	-0.000	0.001
Ryanair_Flug	0.0002	0.000	0.850	0.396	-0.000	0.001
Specialized_Fahrrad	-0.0002	0.000	-0.565	0.572	-0.001	0.000
Tchibo_Koffer	0.0002	0.000	0.635	0.525	-0.000	0.001
Tesoro_Bürostuhl	-6,31E-02	0.000	-0.219	0.827	-0.001	0.001
Tui_Sheraton_Miramar_Resort	-0.0002	0.000	-0.558	0.577	-0.001	0.000
UGG_Stiefelette	-0.0003	0.000	-0.681	0.496	-0.001	0.000
Desktop_Linux_Chrome	0.0008	0.000	3.263	0.001	0.000	0.001
Desktop_Linux_Firefox	0.0006	0.000	2.355	0.019	0.000	0.001
Desktop_MacOS_Safari	0.0007	0.000	2.952	0.003	0.000	0.001
Desktop_Window_s_Edge	0.0010	0.000	4.453	0.000	0.001	0.001
Desktop_Window_s_Firefox	0.0005	0.000	2.115	0.034	3.64e-05	0.001
Desktop_Window_s_Internet_Explorer	8,00E-02	0.000	0.344	0.731	-0.000	0.001
Smartphone_Android_Chrome	-0.0044	0.000	-20.307	0.000	-0.005	-0.004
Smartphone_Android_Firefox	-0.0052	0.001	-10.147	0.000	-0.006	-0.004
Tablet_Android_Chrome	-0.0043	0.000	-20.365	0.000	-0.005	-0.004
iPad_iOS_Safari	-0.0035	0.000	-14.698	0.000	-0.004	-0.003
iPhone_iOS_Safari	-0.0044	0.000	-20.189	0.000	-0.005	-0.004
donottrackenabled	-0.0004	9.81e-05	-4.033	0.000	-0.001	-0.000
cookiesenabled	-0.0002	0.000	-2.010	0.044	-0.000	-5.07e-06
Omnibus:	170.437.924	Durbin-Watson:	1.985			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1.893.839.586.710			
Skew :	-0.973	Prob(JB):	0.00			
Kurtosis:	401.640	Cond. No.	34.2			

**Tabelle 7: Regression mit allen Datensätzen**

Die Referenzkategorien sind in diesem Fall das Produkt „Seltmann-Weiden Kaffeeservice“, und die Zugriffsvariable „Desktop Windows Chrome“. Tracker und Cookies weisen als Binärvariablen keine Referenzkategorie auf.

Auffällig ist zunächst, dass bis auf das IntercityHotel Hamburg sämtliche Produkte nicht signifikant unterschiedlich zum „Seltmann-Weiden Kaffeeservice“ sind. Dies liegt vermutlich an der Asymmetrie des Datensatzes, die mit der hohen Anzahl an Datensätzen mit Preisdifferenz des „IntercityHotel Hamburg“ begründet ist. Die Preisdifferenzen des „IntercityHotel Hamburg“ sind im Durchschnitt um minus 3,15 Prozent höher als die Preisdifferenzen des „Seltmann-Weiden Kaffeeservice“. Diese Differenz ist auch signifikant.

Bei Betrachtung der Zugriffsvariablen zeigt sich, dass – mit Ausnahme der Zugriffskombination „Desktop Windows Internet Explorer“ –sämtliche Kombinationen der Zugriffsvariable signifikant sind. Die Koeffizienten beim Zugriff über die verschiedenen Desktopgeräte sind jedoch mit Werten zwischen +0,05 Prozent und +0,10 Prozent nur unwesentlich von null verschieden.

Beim Zugriff über mobile Endgeräte ergaben sich hingegen deutlich stärkere Ausprägungen der Koeffizienten (zwischen -0,35 Prozent und -0,52 Prozent,

also rund um einen Faktor von 10 höher). Auch wenn die Koeffizienten hier relativ gering sind, ist zu beachten, dass die mobilen Geräte allgemein eine negative Preisdifferenz verzeichnen und damit einen geringeren Preis verzeichnen als die Referenzkategorie. Diese Variablen waren auch durchgehend signifikant.

Von Interesse ist auch die Betrachtung der Einflussvariablen „Akzeptanz von Cookies“ und „Aktivierung der Do-not-track-Einstellung“. Beide Ergebnisvariablen sind in der vorliegenden Regression signifikant; allerdings ist mit Werten von -0,02 Prozent und -0,04 Prozent deren Höhe sehr überschaubar und nur wenig von null verschieden.

Das  $R^2$  der Regression lag bei 0,089. Es können also 8,9 Prozent der im Datensatz vorliegenden relativen Preisdifferenzen durch das Regressionsmodell erklärt werden.

Als Zwischenfazit ist festzustellen, dass insbesondere die Einflussvariable „IntercityHotel Hamburg“ als auch die Zugriffe über mobile Endgeräte von Bedeutung sind und in den folgenden Berechnungen tiefer analysiert werden sollten.

#### 4.3.4 Regression nach Entfernung von Produkten ohne Preisdifferenzen

Um die Analyseergebnisse zu verbessern und aussagekräftiger zu machen, wurde in einem nächsten Schritt der erhobene Datensatz weiter reduziert. Wie Tabelle 5 in Abschnitt 4.2.2 zeigte, gab es im Datensatz eine ganze Reihe von Produkten, bei denen nahezu keine relativen Preisdifferenzen auftraten. Das bedeutet nicht, dass es nicht zu Änderungen der erhobenen Preise gekommen wäre, sondern dass nahezu immer der identische Preis erhoben wurde wie im jeweiligen Standardfall. Beispielsweise ergaben sich beim Produkt „HM Socken“ von 15.636 erhobenen Fällen nur in zwei Datensätzen eine Abweichung von Preis des Standardfalls.

Auf das Regressionsmodell übertragen bedeutet dies, dass bei einer hohen Anzahl von Datensätzen die abhängige Variable „relative Preisdifferenz“ den Wert null annahm. Daher wurde im zweiten Regressionsmodell Produkte mit weniger als 40 Preisabweichungen (das entspricht einem relativen Anteil von weniger als 0,2 Prozent) entfernt. Tabelle 8 fasst die Verteilung der verbleibenden Datensätze auf die Produkte „IntercityHotel Hamburg“, Apple iPhone 11“, „Tesoro Bürostuhl“, „Picard Schultertasche“, „Tui Sheraton Miramar Resort“ und Miele Waschmaschine WCA 030“ zusammen.

Produkt	Anzahl Datensätze (ohne Standardfälle)
IntercityHotel Hamburg	17.454
Apple iPhone 11	15.615
Tesoro Bürostuhl	13.762
Picard Schultertasche	13.736
Tui Sheraton Miramar Resort	13.681
Miele Waschmaschine WCA 030	13.038
Summe	87.286

**Tabelle 8: Datensatz Regression nach Produkten**

Referenzkategorien des Datensatzes nach Entfernung von Produkten ohne Preisdifferenzen sind das „IntercityHotel Hamburg“ und die als Standardfall definierte Zugriffsvariable „Desktop Windows Chrome“. Tabelle 9 zeigt die Ergebnisse des Regressionsmodells.

Dep. Variable:	pricediffrel	R-squared:	0.109			
Model:	OLS	Adj. R-squared:	0.109			
Method:	Least Squares	F-statistic:	594.8			
No. Observations:	87286	AIC:	-3,04E+08			
Df Residuals:	87267	BIC:	-3,04E+08			
Df Model:	18					
Covariance Type:	nonrobust					
Referenzkategorien: IntercityHotel Hamburg, Desktop Windows Chrome						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.0248	0.001	-40.865	0.000	-0.026	-0.024
Apple_iPhone_11	0.0316	0.000	67.551	0.000	0.031	0.032
Miele_Waschmaschine_WCA_030	0.0318	0.001	62.996	0.000	0.031	0.033
Picard_Schultertasche	0.0314	0.000	64.954	0.000	0.030	0.032
Tesoro_Bürostuhl	0.0315	0.000	65.107	0.000	0.031	0.032
Tui_Sheraton_Miramar_Resort	0.0314	0.000	64.826	0.000	0.030	0.032
Desktop_Linux_Chrome	0.0018	0.001	2.326	0.020	0.000	0.003
Desktop_Linux_Firefox	0.0010	0.001	1.346	0.178	-0.000	0.003
Desktop_MacOS_Safari	0.0013	0.001	1.850	0.064	-7.88e-05	0.003
Desktop_Window_s_Edge	0.0025	0.001	3.651	0.000	0.001	0.004
Desktop_Window_s_Firefox	0.0008	0.001	1.119	0.263	-0.001	0.002
Desktop_Window_s_Internet_Explorer	-7,82E-02	0.001	-0.110	0.912	-0.001	0.001
Smartphone_Android_Chrome	-0.0157	0.001	-23.555	0.000	-0.017	-0.014
Smartphone_Android_Firefox	-0.0182	0.002	-11.621	0.000	-0.021	-0.015
Tablet_Android_Chrome	-0.0148	0.001	-23.058	0.000	-0.016	-0.014
iPad_iOS_Safari	-0.0114	0.001	-15.764	0.000	-0.013	-0.010
iPhone_iOS_Safari	-0.0151	0.001	-22.532	0.000	-0.016	-0.014
donotrackenabled	-0.0014	0.000	-4.702	0.000	-0.002	-0.001
cookiesenabled	-0.0003	0.000	-1.037	0.300	-0.001	0.000
Omnibus:	5,74E+07	Durbin-Watson:	1.981			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	46738677131			
Skew:	1.785	Prob(JB):	0.00			
Kurtosis:	116.307	Cond. No.	18.7			

**Tabelle 9: Regression nach Entfernung von Produkten ohne Preisdifferenzen**

Dabei ist insbesondere der Vergleich mit der ersten Regression aus Abschnitt 4.3.3 von Bedeutung.

Zunächst zeigt sich wieder der Einfluss des „IntercityHotel Hamburg“, der sich mit einem (signifikanten) Koeffizienten von -2,48 Prozent niederschlägt. Die Preisdifferenzen sämtlicher Produkte sind signifikant von dem „IntercityHotel Hamburg“ verschieden und um durchschnittlich rund drei Prozent höher.

Bei den Zugriffsvariablen ist auffällig, dass sämtliche mobile Zugriffsarten signifikanten, negativen Einfluss auf die Preisdifferenz ausüben. Die Ausprägungen reichen von -1,14 Prozent bis -1,82 Prozent. Insofern wird das Ergebnis der ersten Regression bestätigt. Bezüglich der Zugriffe über Desktopgeräte gibt es nur in zwei Fällen (Chrome/Linux und Windows/Edge) signifikante Ergebnisse, die Koeffizienten sind aber in beiden Fällen nahe null.

Das Deaktivieren von Trackern hat einen signifikanten, aber mit -0,14 Prozent extrem geringen negativen Einfluss auf die Preisdifferenz. Für den Einfluss von Cookies auf die relative Preisdifferenz konnte kein signifikantes Ergebnis belegt werden.

Mit 10,9 Prozent ist das  $R^2$  dieser zweiten Regression etwas höher als in der ersten Auswertung.

#### 4.3.5 Regression des Produktes „IntercityHotel Hamburg“

Bereits die deskriptive Analyse hat ergeben, dass das Produkt „Intercity Hotel Hamburg“ aus dem Datensatz heraussticht: In beinahe jedem zweiten Fall gab es eine Abweichung vom Preis des jeweiligen Standardfalls. Aus diesem Grund wurde im nächsten Schritt der Auswertung ausschließlich auf dieses Produkt fokussiert und der Datensatz zur genaueren Betrachtung auf die Einträge dieses Produktes reduziert. Dabei wurden zwei Regressionen berechnet: Zunächst die Kombination aus manuellem und automatisierten Erhebungsstrang mit Feature-Set 1, anschließend nur mit den Daten des manuellen Erhebungsstrangs mit den dort zusätzlich erhobenen Variablen (Feature-Set 2).

Tabelle 10 zeigt den Output des mit Feature-Set 1 berechneten Modells. Grundlage waren die 17.454 Datensätze mit den Preisen des Produkts „InterCityHotel Hamburg“, erhoben auf der Plattform Booking.com.

Dep. Variable:	pricediffrel	R-squared:	0.191			
Model:	OLS	Adj. R-squared:	0.190			
Method:	Least Squares	F-statistic:	316.0			
No. Observations:	17454	AIC:	-3,76E+07			
Df Residuals:	17440	BIC:	-3,75E+07			
Df Model:	13					
Covariance Type:	nonrobust					
Referenzkategorie	Desktop Windows Chrome					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0051	0.002	2.200	0.028	0.001	0.010
Desktop_Linux_Chrome	0.0020	0.003	0.611	0.541	-0.004	0.008
Desktop_Linux_Firefox	-0.0013	0.003	-0.398	0.690	-0.008	0.005
Desktop_MacOS_Safari	0.0025	0.003	0.811	0.418	-0.004	0.009
Desktop_Window_s_Edge	0.0076	0.003	2.532	0.011	0.002	0.013
Desktop_Window_s_Firefox	-0.0009	0.003	-0.274	0.784	-0.007	0.005
Desktop_Window_s_Internet_Explorer	-0.0009	0.003	-0.297	0.767	-0.007	0.005
Smartphone_Android_Chrome	-0.0847	0.003	-28.971	0.000	-0.090	-0.079
Smartphone_Android_Firefox	-0.1073	0.007	-14.539	0.000	-0.122	-0.093
Tablet_Android_Chrome	-0.0766	0.003	-27.580	0.000	-0.082	-0.071
iPad_iOS_Safari	-0.0589	0.003	-18.820	0.000	-0.065	-0.053
iPhone_iOS_Safari	-0.0822	0.003	-28.032	0.000	-0.088	-0.076
donotrackenabled	-0.0064	0.001	-5.020	0.000	-0.009	-0.004
cookiesenabled	0.0003	0.001	0.244	0.807	-0.002	0.003
Omnibus:	1,35E+07	Durbin-Watson:	1.486			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	556649222			
Skew:	3.358	Prob(JB):	0.00			
Kurtosis:	29838	Cond. No.	18.1			

**Tabelle 10: Regression des Produkts „IntercityHotel Hamburg“ (Feature-Set 1)**

Es zeigt sich, dass sämtliche Mobilgeräte Preisdifferenz zum Standardfall aufweisen. Diese Preisdifferenz bewegt sich in einem Rahmen von -5,89 Prozent bis -10,73 Prozent. Alle Ergebnisse in Bezug auf die mobilen Endgeräte sind auch signifikant. Die Abweichung ist von Bedeutung. Sie kann durch einen Rabatt erklärt werden, den das Hotel in manchen Fällen bei der Buchung über mobile Endgeräte gewährte. In Abschnitt 5.1 folgt eine inhaltliche Begründung. Die Do-not-track-Einstellung ist ebenfalls signifikant, deren Einfluss ist jedoch mit einer Auswirkung von -0,64 Prozent auf die Preisdifferenz sehr gering.

Cookies haben keinen signifikanten Einfluss auf Preisdifferenzen.

Wird das umfangreichere Feature-Set 2 auf den kleineren manuellen Datensatz angewendet, so ist weiterhin Desktop Windows Chrome die Referenzkategorie. Die Ergebnisse dieser Regression zeigt Tabelle 11.

Dep. Variable:	pricediffrel	R-squared:	0.462			
Model:	OLS	Adj. R-squared:	0.460			
Method:	Least Squares	F-statistic:	249.5			
No. Observations:	4381	AIC:	-1,21E+07			
Df Residuals:	4365	BIC:	-1,20E+07			
Df Model:	15					
Covariance Type:	nonrobust					
Referenzkategorie	Desktop Windows Chrome					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0095	0.003	2.859	0.004	0.003	0.016
Desktop_Linux_Chrome	-0.0059	0.011	-0.555	0.579	-0.027	0.015
Desktop_Linux_Firefox	-0.0080	0.011	-0.747	0.455	-0.029	0.013
Desktop_MacOS_Safari	0.0239	0.005	4.767	0.000	0.014	0.034
Desktop_Window_s_Edge	0.0216	0.004	4.892	0.000	0.013	0.030
Desktop_Window_s_Firefox	0.0005	0.004	0.106	0.915	-0.008	0.009
Desktop_Window_s_Internet_Explorer	0.0081	0.005	1.796	0.073	-0.001	0.017
Smartphone_Android_Chrome	-0.0877	0.004	-21.462	0.000	-0.096	-0.080
Smartphone_Android_Firefox	-0.1114	0.006	-17.561	0.000	-0.124	-0.099
Tablet_Android_Chrome	-0.1036	0.005	-19.987	0.000	-0.114	-0.093
iPad_iOS_Safari	0.0618	0.005	11.912	0.000	0.052	0.072
iPhone_iOS_Safari	-0.0902	0.004	-22.294	0.000	-0.098	-0.082
donotrackenabled	-0.0293	0.003	-10.044	0.000	-0.035	-0.024
cookiesenabled	0.0020	0.004	0.516	0.606	-0.006	0.009
Social_Media	-0.0020	0.002	-0.836	0.403	-0.007	0.003
Nutzerkonto	0.0001	0.002	0.049	0.961	-0.005	0.005
Omnibus:	2,77E+06	Durbin-Watson:	0.730			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	428769284			
Skew:	-2.022	Prob(JB):	0.00			
Kurtosis:	51296	Cond. No.	20.2			

**Tabelle 11: Regression des Produkts „IntercityHotel Hamburg“ (Feature-Set 2)**

Bezüglich der Zugriffsvariable sind Desktop Windows Edge und Desktop MacOS Safari signifikant um rund 2 Prozent von der Referenzkategorie verschieden.

Bezüglich der Preisabfragen über mobile Geräte zeigt sich dasselbe Bild wie bei Feature-Set 1: Die Preise sind signifikant in einem Spektrum von -8,77 Prozent bis -11,14 Prozent niedriger. Allerdings sticht ein Wert heraus: Beim Zugriff über den Browser Safari auf dem Endgerät iPad ergibt sich eine positive Preisabweichung von 6,18 Prozent, die ebenfalls signifikant ist. Der Grund liegt offenbar darin, dass „auf dem iPad ... Safari die Desktop-Version von Websites [anzeigt]“<sup>46</sup>. Für Booking.com sieht es also so aus, als erfolge der Zugriff über einen Desktop-PC. Ein Rabatt, den das Hotel ggf. für den Zugriff über die mobile Website eingestellt hat, wird somit nicht an den Kunden ausgespielt.

Die Aktivierung der Do-not-track-Einstellung hat ebenfalls einen signifikanten Einfluss von -2,93 Prozent auf die relative Preisdifferenz.

Die weiteren betrachteten Größen haben dagegen keinen signifikanten Einfluss. Weder das Aktivieren bzw. Vorhandensein von Cookies, noch der Login in einen Social-Media-Account oder ein Nutzerkonto beeinflussen die Preisdifferenzen signifikant.

<sup>46</sup> Vgl. <https://support.apple.com/de-de/guide/ipad/ipad999d68f9/ipados>

### 4.3.6 Regression des Datensatzes ohne „IntercityHotel Hamburg“

Da die Regression des gesamten Datensatzes wegen der asymmetrischen Verteilung der Preisdifferenzen hauptsächlich durch das „IntercityHotel Hamburg“ beeinflusst ist, wird dieses aus dem Datensatz entfernt, um Zusammenhänge der restlichen Produkte zu betrachten.

Dep. Variable:	pricediffrel	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	4.431			
No. Observations:	69832	AIC:	-3,86E+08			
Df Residuals:	69814	BIC:	-3,85E+08			
Df Model:	17					
Covariance Type:	nonrobust					
Referenzkategorien:	Apple iPhone 11, Desktop Windows Chrome					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.0005	0.000	-1.977	0.048	-0.001	-4.04e-06
Miele_Waschmaschine_WCA_030	0.0002	0.000	1.163	0.245	-0.000	0.001
Picard_Schultertasche	-0.0002	0.000	-1.029	0.303	-0.001	0.000
Tesoro_Bürostuhl	-0.0001	0.000	-0.697	0.486	-0.000	0.000
Tui_Sheraton_Miramar_Resort	-0.0002	0.000	-1.294	0.196	-0.001	0.000
Desktop_Linux_Chrome	0.0010	0.000	3.303	0.001	0.000	0.002
Desktop_Linux_Firefox	0.0011	0.000	3.479	0.001	0.000	0.002
Desktop_MacOS_Safari	0.0009	0.000	3.249	0.001	0.000	0.002
Desktop_Window_s_Edge	0.0011	0.000	4.065	0.000	0.001	0.002
Desktop_Window_s_Firefox	0.0011	0.000	3.644	0.000	0.000	0.002
Desktop_Window_s_Internet_Explorer	4.53e-06	0.000	0.016	0.987	-0.001	0.001
Smartphone_Android_Chrome	0.0008	0.000	2.887	0.004	0.000	0.001
Smartphone_Android_Firefox	0.0010	0.001	1.688	0.092	-0.000	0.002
Tablet_Android_Chrome	0.0009	0.000	3.533	0.000	0.000	0.001
iPad_iOS_Safari	0.0004	0.000	1.431	0.152	-0.000	0.001
iPhone_iOS_Safari	0.0013	0.000	4.644	0.000	0.001	0.002
donottrackenabled	-0.0002	0.000	-1.400	0.161	-0.000	6.73e-05
cookiesenabled	-0.0006	0.000	-4.680	0.000	-0.001	-0.000
Omnibus:	2.26E+08	Durbin-Watson:	1.958			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3,10726E+13			
Skew :	-54.849	Prob(JB):	0.00			
Kurtosis:	3269044	Cond. No.	18.8			

**Tabelle 12: Regression des Datensatzes ohne „IntercityHotel Hamburg“ (Feature-Set 1)**

Die Ergebnisse der Regression des Feature-Sets 1 in Tabelle 12 zeigen keine signifikanten Unterschiede der Preisdifferenzen der betrachteten Produkte „Miele Waschmaschine WCA 030“, „Picard Schultertasche“, „Tesoro Bürostuhl“ und „Tui Sheraton Miramar Resort“ zum Referenzprodukt „Apple iPhone 11“.

Bis auf die Zugriffsarten „Desktop Windows Internet Explorer“, „Smartphone Android Firefox“ und „iPad iOS Safari“, die nicht signifikant sind, weisen sämtliche Zugriffsarten zu der Referenzkategorie „Desktop Windows Chrome“ verschiedene Preisdifferenzen auf. Diese Preisdifferenzen sind im Schnitt alle um nur etwa 0,1 Prozent höher, also nahezu vernachlässigbar gering.

Die Erlaubnis, Cookies zu verwenden, ist als Einflussvariable signifikant, deren Einfluss mit -0,06 Prozent jedoch sehr klein.

Das Bestimmtheitsmaß  $R^2$  ist mit 0,1 Prozent extrem klein. Das bedeutet, dass nur 0,1 Prozent der beobachteten Preisabweichungen durch das Regressionsmodell erklärt werden können. Die Schlussfolgerung liegt nahe, dass es im vorliegenden Datensatz keine systematischen Preisabweichungen gibt.

Dep. Variable:	pricediffrel	R-squared:	0.003			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	3.297			
No. Observations:	21877	AIC:	-9,58E+07			
Df Residuals:	21857	BIC:	-9,56E+07			
Df Model:	19					
Covariance Type:	nonrobust					
Referenzkategorien:	Apple iPhone 11, Desktop Window s Chrome					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0002	0.001	0.324	0.746	-0.001	0.002
Miele_Waschmaschine_WCA_030	5,83E-04	0.001	0.101	0.920	-0.001	0.001
Picard_Schultertasche	-0.0006	0.001	-1.018	0.309	-0.002	0.001
Tesoro_Bürostuhl	-0.0005	0.001	-0.940	0.347	-0.002	0.001
Tui_Sheraton_Miramar_Resort	-0.0009	0.001	-1.553	0.120	-0.002	0.000
Desktop_Linux_Chrome	0.0030	0.002	1.435	0.151	-0.001	0.007
Desktop_Linux_Firefox	0.0002	0.002	0.071	0.944	-0.004	0.004
Desktop_MacOS_Safari	0.0030	0.001	3.010	0.003	0.001	0.005
Desktop_Window s_Edge	0.0029	0.001	3.242	0.001	0.001	0.005
Desktop_Window s_Firefox	0.0016	0.001	1.858	0.063	-8.87e-05	0.003
Desktop_Window s_Internet_Explorer	-0.0011	0.001	-1.191	0.234	-0.003	0.001
Smartphone_Android_Chrome	0.0027	0.001	3.255	0.001	0.001	0.004
Smartphone_Android_Firefox	0.0020	0.001	1.559	0.119	-0.001	0.004
Tablet_Android_Chrome	0.0021	0.001	2.003	0.045	4.38e-05	0.004
iPad_iOS_Safari	0.0005	0.001	0.473	0.636	-0.002	0.003
iPhone_iOS_Safari	0.0032	0.001	3.954	0.000	0.002	0.005
donotrackenabled	-0.0006	0.001	-1.117	0.264	-0.002	0.000
cookiesenabled	-0.0030	0.001	-3.921	0.000	-0.004	-0.001
Social_Media	0.0009	0.000	1.942	0.052	-8.48e-06	0.002
Nutzerkonto	0.0001	0.000	0.277	0.782	-0.001	0.001
Omnibus:	5,75E+07	Durbin-Watson:	1.966			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1,00643E+12			
Skew:	-31.311	Prob(JB):	0.00			
Kurtosis:	1051895	Cond. No.	20.8			

**Tabelle 13: Regression des Datensatzes ohne „IntercityHotel Hamburg“ (Feature-Set 2)**

Dieselbe Schlussfolgerung muss bei Betrachtung von Feature-Set 2 mit dem manuellen Datensatz gezogen werden. Viele der betrachteten Einflussvariablen zeigen keinen signifikanten Einfluss. Wo Variablen signifikant sind, etwa beim Setzen von Cookies, ist deren Ausmaß extrem gering, im Cookie-Beispiel etwa nur -0,3 Prozent.

Auch hier liegt mit Blick auf das Bestimmtheitsmaß von 0,3 Prozent der Schluss nahe, dass im vorliegenden Datensatz keine systematischen Preisabweichungen vorliegen.

#### 4.3.7 Regression des manuellen Erfassungsstrangs

Die Regression auf Daten des manuellen Erfassungsstrangs enthält für das komplette Feature-Set 2 erhobene Datenpunkte und ermöglicht so die konsistente Analyse der Variablen Erfasser, Social Media und Nutzerkonto. Betrachtet

werden exklusiv die Produkte mit mehr als 40 Preisdifferenzen (vgl. Abschnitt 4.3.4). Tabelle 14 zeigt die Ergebnisse der Analyse.

Dep. Variable:	pricediffrel	R-squared:	0.120			
Model:	OLS	Adj. R-squared:	0.120			
Method:	Least Squares	F-statistic:	188.6			
No. Observations:	26258	AIC:	-9,33E+07			
Df Residuals:	26238	BIC:	-9,32E+07			
Df Model:	19					
Covariance Type: nonrobust						
Referenzkategorien:		Tesoro Bürostuhl, Desktop Window s Chrome				
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0069	0.001	7.083	0.000	0.005	0.009
IntercityHotel_Hamburg	-0.0324	0.001	-42832	0.000	-0.034	-0.031
Miele_Waschmaschine_WCA_030	0.0003	0.001	0.439	0.661	-0.001	0.002
Picard_Schultertasche	-0.0003	0.001	-0.419	0.675	-0.002	0.001
Tui_Sheraton_Miramar_Resort	-0.0006	0.001	-0.769	0.442	-0.002	0.001
Desktop_Linux_Chrome	0.0015	0.003	0.522	0.602	-0.004	0.007
Desktop_Linux_Firefox	-0.0012	0.003	-0.403	0.687	-0.007	0.005
Desktop_MacOS_Safari	0.0064	0.001	4.674	0.000	0.004	0.009
Desktop_Window_s_Edge	0.0060	0.001	4.945	0.000	0.004	0.008
Desktop_Window_s_Firefox	0.0015	0.001	1.216	0.224	-0.001	0.004
Desktop_Window_s_Internet_Explorer	0.0006	0.001	0.459	0.647	-0.002	0.003
iPad_iOS_Safari	0.0107	0.001	7.518	0.000	0.008	0.013
iPhone_iOS_Safari	-0.0123	0.001	-11.116	0.000	-0.014	-0.010
Smartphone_Android_Chrome	-0.0124	0.001	-11.088	0.000	-0.015	-0.010
Smartphone_Android_Firefox	-0.0169	0.002	-9.697	0.000	-0.020	-0.013
Tablet_Android_Chrome	-0.0156	0.001	-10.930	0.000	-0.018	-0.013
donotrackenabled	-0.0054	0.001	-6.727	0.000	-0.007	-0.004
cookiesenabled	-0.0021	0.001	-2.029	0.042	-0.004	-7.27e-05
Social_Media	0.0004	0.001	0.554	0.580	-0.001	0.002
Nutzerkonto	0.0001	0.001	0.220	0.826	-0.001	0.001
Omnibus:	3,93E+07	Durbin-Watson:	2.088			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	52156804310			
Skew:	-8.802	Prob(JB):	0.00			
Kurtosis:	220628	Cond. No.	20.7			

**Tabelle 14: Regression des manuellen Erfassungsstrangs**

Im manuell erhobenen Datensatz liegt nun wieder ein mit -3,24 Prozent beobachtbarer Einfluss des Produkts „IntercityHotel Hamburg“ vor, der signifikant ist. Die restlichen Produkte weisen keine signifikanten Unterschiede auf.

Bezüglich der Zugriffsvariable sind die mobilen Geräte durchgehend signifikant zu der Referenzkategorie „Desktop Windows Chrome“ verschieden, im Ausmaß von -1,23 bis -1,69 Prozent. Damit bestätigen sich die Ergebnisse der Analyse aus den Abschnitten 4.3.4 und 4.3.5. Auffällig ist wiederum die positive signifikante Preisdifferenz beim Zugriff über den Safari-Browser beim Endgerät iPad. Der Grund hierfür wurde in Abschnitt 4.3.5 dargestellt.

Cookies haben einen signifikanten, aber mit -0,21 Prozent extrem kleinen Einfluss auf die relative Preisdifferenz.

Von besonderem Interesse ist noch der Einfluss von Logins in Nutzerkonten oder soziale Medien, da dieses Feature-Set 2 nur im manuellen Erfassungsstrang erhoben wurden. Die Ergebnisse sind allerdings nicht signifikant.

## 5 Zusammenfassung und Diskussion der Ergebnisse

### 5.1 Deskriptive Analyse

Der gewählte Versuchsaufbau ergab umfangreiches Datenmaterial für die Auswertung. Durch die beiden Erhebungsstränge (automatisiert/manuell) sowie den langen Erhebungszeitraum konnten insgesamt über 400.000 Datensätze mit rund sieben Millionen Datenpunkten erhoben werden. Rund 250.000 Datensätze entfielen auf die automatisierte Erhebung, rund 153.000 wurden durch natürliche Personen erfasst.

Die Grundlage der Auswertung bildeten nicht die erhobenen absoluten Preise, sondern die relative Preisdifferenz zu einem definierten Standardfall, bei dem die Plattformen mit hoher Sicherheit keine Preispersonalisierung betreiben. Deshalb wurde in jedem einzelnen Erfassungsvorgang dieser Standardfall erhoben und im Rahmen des Preprocessing die relativen Preisdifferenzen berechnet. Die Anzahl der ausgewerteten Datensätze betrug damit rund 286.000. Von diesen wiesen rund 9.500 eine Preisdifferenz auf (rund 3,3 Prozent).

Bereits die erste Betrachtung der Daten ergab, dass ein Großteil dieser Preisabweichungen auf nur ein einziges Produkt bzw. eine einzige Plattform zurückzuführen war, nämlich der Buchung eines Hotels über Booking.com.

Die übrigen Preisdifferenzen waren extrem gering und bewegten sich in einem Spektrum von -0,04 Prozent bis +0,03 Prozent. Überwiegend waren diese Differenzen negativ. Es wurden also – im Fall des Auftretens einer Differenz – geringere Preise veranschlagt als im Standardfall. Auch bei der Betrachtung der relativen Preisdifferenzen in Abhängigkeit von einzelnen Einflussfaktoren änderte sich dieser Eindruck nicht (vgl. ausführlich dazu Abschnitt 4.2.2).

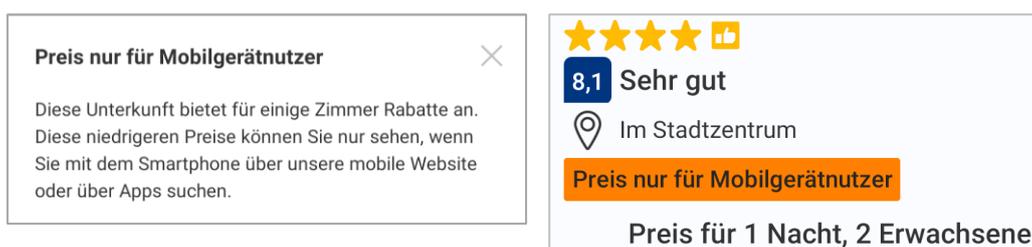
Die Abweichungen bei der Buchung des Produkts „IntercityHotel Hamburg“ über die Plattform „Booking.com“ sind erläuterungsbedürftig. Booking.com ist eine Plattform, über die Unternehmen der Hotellerie und ähnlicher Branchen Reiseunterkünfte (Hotel-/Pensionszimmer, Ferienwohnungen etc.) anbieten können. Booking.com tritt dabei als Vermittler zwischen Anbieter und Endkunden auf, betreibt also selbst keine Hotels. Die Vermittlung erfolgt über die Websites und Apps des Unternehmens. Die Leistungen von Booking.com bestehen im Wesentlichen aus der Bereitstellung der Plattform und der Kundenakquise. Dafür stellt Booking.com den Anbietern im Buchungsfall Gebühren in Rechnung.

Die Preise legt dabei grundsätzlich nicht Booking.com fest, sondern der jeweilige Anbieter bzw. Vermieter selbst.

Booking.com bietet den Vermietern die Möglichkeit, eine „Handy-Rate“ anzubieten, also einen Sondertarif für die Buchung über die App oder einen mobilen

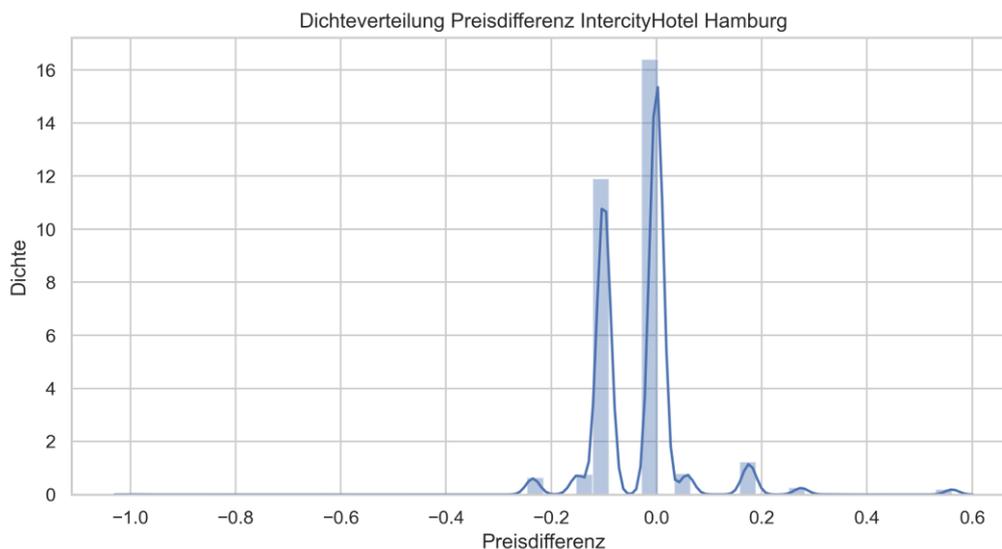
Browser. Dieser greift in der Regel dann, wenn der Kunde ein Smartphone oder Tablet benutzt. Dies erhöhe die Konversion um bis zu 28 Prozent, weil „80 Prozent der Reisenden [...] bei der Suche einer Reise eine mobile App“ nutzten und „50 Prozent der Unterkunftssuchanfragen und -buchungen [...] auf Mobilgeräten“ getätigt würden.<sup>47</sup> Booking.com empfiehlt einen Rabatt in Höhe von 15 Prozent.

Bei der Buchung über Mobilgeräte wird dem Kunden angezeigt, dass es sich um einen rabattierten Preis aufgrund des benutzten Endgeräts handelt (vgl. Abbildung 44).



**Abbildung 44: Preis nur für Mobilgerätenutzer<sup>48</sup>**

Damit lässt sich der überwiegende Anteil der Preisabweichungen des Produkts „IntercityHotel Hamburg“ in den erhobenen Daten erklären. Wie Abbildung 45 zeigt, wurde offenbar in der Mehrzahl der mobilen Buchungen ein Rabatt von 10 Prozent gewährt.



**Abbildung 45: Dichteverteilung „IntercityHotel Hamburg“<sup>49</sup>**

<sup>47</sup> <https://partner.booking.com/de/l%C3%B6sungen/handy-rate>; vgl. dazu auch <https://partner.booking.com/de/hilfe/raten-verf%C3%BCgbarkeiten/preise-f%C3%BCr-mobilger%C3%A4te>

<sup>48</sup> Bildquelle: Booking.com (Android-App).

<sup>49</sup> Abbildung 45 entspricht Abbildung 25 und dient nochmals zur Verdeutlichung.

## 5.2 Regressionsanalyse

Als Hauptbestandteil der Auswertung kam die multivariate lineare Regressionsanalyse in Form der Kleinste-Quadrat-Methode zum Einsatz. Um ein möglichst aussagekräftiges Ergebnis zu erzielen, wurde der vorliegende Datensatz mehrfach sinnvoll beschnitten und jeweils ausgewertet.

Zunächst wurde eine Regression mit allen zur Verfügung stehenden rund 286.000 Datensätzen gerechnet (vgl. Abschnitt 4.3.3). Die dabei berechneten Koeffizienten waren nahezu durchgehend nur unwesentlich von null verschieden. Das bedeutet, dass der Einfluss der betrachteten unabhängigen Variablen auf den abhängigen Faktor „relative Preisdifferenz“ extrem gering ausfiel. In zwei Feldern gab es Hinweise auf systematische Differenzen. Zum einen ergab sich für das Produkt „IntercityHotel Hamburg“ eine Abweichung von über 3 Prozent, zum anderen zeigte sich ein signifikanter (negativer) Einfluss bei Zugriffen über mobile Endgeräte.

Es zeigte sich bei der deskriptiven Analyse, dass es Produkte gab, bei denen überhaupt oder nahezu keine Preisdifferenzen auftraten. Um die Qualität der Analyse zu verbessern, wurde diese in einem nächsten Schritt aus dem Datensatz entfernt (vgl. Abschnitt 4.3.4). Es ergaben sich insbesondere bei den Zugriffen über Mobilgeräte Koeffizienten zwischen -0,0152 und -0,0114, also Abweichungen zwischen -1,52 und -1,14 Prozent. Dabei lag auf Grundlage der Ergebnisse der deskriptiven Analyse zu vermuten, dass hier das Produkt „IntercityHotel Hamburg“ einen hohen Einfluss auf die Modellergebnisse ausübte.

Um die Modellqualität weiter zu verbessern, wurde daher der Datensatz getrennt und zunächst nur das Produkt „IntercityHotel Hamburg“ betrachtet (vgl. Abschnitt 4.3.5), anschließend der restliche Datensatz ohne das Produkt (vgl. Abschnitt 4.3.6). Hier wurde auch jeweils der (Teil-)Datensatz des manuellen Erhebungsstrangs zusätzlich ausgewertet, da nur in diesem Strang die weichen qualitativen Faktoren wie der Login in ein Nutzerkonto oder Login in Social-Media-Konten abgefragt wurden („Feature-Set 2“, vgl. dazu Tabelle 6).

Bei der Betrachtung des Produkts „IntercityHotel Hamburg“ wurden die Eindrücke aus der deskriptiven Analyse bestätigt. Es ergaben sich Koeffizienten zwischen -6 und -11 Prozent; diese waren auch signifikant. Nutzerkonto- oder Social-Media-Logins hatten dagegen offenbar keinen Einfluss.

Mit Blick auf das Bestimmtheitsmaß  $R^2$  von immerhin 46,2 Prozent erklärt dieses Modell mit Feature-Set 2 die vorliegenden relativen Preisdifferenzen am besten durch die unabhängigen Variablen. Allerdings betrachtet dieses Modell nur einen Spezialfall und hat auch den kleinsten zugrunde liegenden Datensatz. Die Aussagekraft kann damit nicht verallgemeinert werden. Zur Analyse aller vorliegenden Einflussfaktoren scheint es weniger geeignet als umfassendere Modelle mit größeren Datensätzen.

Die Regressionsanalyse des nun final bereinigten Datensatzes ergab keine auffälligen Koeffizienten. Bei einigen Variablen traten Preisunterschiede auf, zum Teil waren diese auch signifikant. Allerdings waren diese extrem klein. Bei der Betrachtung von Feature-Set 2, der wie beschrieben auf den manuellen Erhebungsstrang beschränkt war, ergab sich ein nahezu identisches Ergebnis. Den größten signifikanten Einfluss hatte hier die Verwendung von Cookies, die zu einer Preisdifferenz von -0,2 Prozent führte.

Mit Bestimmtheitsmaßen ( $R^2$ ) von 0,1 bzw. 0,3 Prozent in Abschnitt 4.3.6 ergab sich, dass nur ein verschwindend geringer Anteil der vorhandenen Preisdifferenzen durch das Regressionsmodell erklärbar sind. Die Schlussfolgerung liegt nahe, dass die betrachteten unabhängigen Variablen keinen systematischen Einfluss auf die relative Preisdifferenz haben.

Von Interesse war zuletzt noch die Betrachtung des vollständigen manuell erhobenen Datensatzes (vgl. Abschnitt 4.3.7), denn nur dort wurde das komplette Feature-Set 2 erhoben, also insbesondere die genannten weichen Faktoren. Dabei hatte aber weder der Login in ein Nutzerkonto der Plattform noch der Login in soziale Medien einen merklichen Einfluss auf die Preisdifferenzen.

### 5.3 Fazit und Ausblick

Die zu Beginn der Untersuchung formulierten Forschungsfragen waren: Wird im Markt personalisierte Preisgestaltung verwendet? Von welchen Kriterien hängt – für den Fall der Verwendung – die personalisierte Preisgestaltung ab und welches Ausmaß haben ihre Auswirkungen? Die vorliegende Untersuchung hat sich dem Thema empirisch genähert und mittels einer breit angelegten Datenerhebung im deutschen E-Commerce erstmals systematisch analysiert.

Personalisierte Preise sind dabei zu unterscheiden von in der Regel auf Marktgegebenheiten zurückzuführenden Preisänderungen. Personalisierte Preisgestaltung liegt dann vor, wenn ein Kunde auf Basis von persönlichen Daten einen individuellen Preis angezeigt bekommt, der anderen Kunden nicht angeboten wird.

Die durchgeführte statistische Auswertung der gesammelten Daten konnte nicht bestätigen, dass personalisierte Preisgestaltung im Markt vorliegt und von Anbietern aktiv eingesetzt wird.

Im Datensatz ließ sich ein Einfluss des verwendeten Endgeräts auf die relative Preisdifferenz signifikant belegen. Durch eine vertiefte Regressionsanalyse ließ sich dieser Effekt auf ein Produkt bzw. eine Plattform im Datensatz eingrenzen. Die deskriptive Analyse wurde damit bestätigt. Beim fraglichen Sachverhalt handelt es sich offenbar um marktübliches Verhalten.

Nicht bestätigt werden konnte ein Einfluss der Do-not-track-Einstellung in Browsern auf den Preis, ebenso wenig der Einfluss des Setzens oder Nicht-Setzens von Cookies. Die Ergebnisse waren hier nur zum Teil signifikant und dann nur geringfügig von null verschieden.

Untersucht wurde ebenfalls, ob es Preisdifferenzen gibt, wenn der Nutzer in ein Nutzerkonto beim jeweiligen Anbieter bzw. der jeweiligen Plattform oder auf einem sozialen Netzwerk wie Facebook eingeloggt ist. Ein Einfluss dieser Faktoren konnte nicht belegt werden. Die Ergebnisse waren durchweg nicht signifikant.

Eine ganze Reihe weiterer potentieller Einflussfaktoren wurde untersucht, bei denen wiederum kein Einfluss auf die angezeigten Preise festgestellt werden konnte. Dazu zählten das verwendete Betriebssystem des für den Kauf benutzten Endgeräts, der verwendete Browser, der Wochentag oder die Uhrzeit des Kaufs und der auf Basis der IP-Adresse übermittelte Standort.

Zusammenfassend lässt sich attestieren, dass eine Differenzierung nach Endgeräten bei der Preisgestaltung zumindest im Teil auf dem Markt bereits erkennbar ist. Andererseits spielen Nutzermerkmale, wie von Schleusener und Hosell (2016) angenommen, nach den vorliegenden Erkenntnissen offenbar noch keine größere Rolle.

Doch auch wenn klare Anzeichen einer personalisierten Preisgestaltung fehlen, können marketinggetriebene Preissetzungen, wie sie teilweise zu beobachten sind, bereits eine Vorstufe der personalisierten Preissetzung darstellen. Die individuelle Preissetzung war bei der Buchung des „IntercityHotel Hamburg“ für mobile Gerätenutzer zwar erkennbar, dennoch werden hier jenseits einer dynamischen Preisgestaltung einzelne Kundengruppen durch Hotels und die angebotene Handy-Rate bevorteilt.

Waren und Dienstleistungen unterliegen wie eingangs erwähnt schneller werdenden Preisveränderungen. Dynamische Preisgestaltungen im E-Commerce, aber auch in Branchen wie Hotellerie und Reisen, stellen bereits eine gängige Praxis dar. Dennoch konnte eine Preisdiskriminierung im Sinne von Krugman (2000) nicht aufgedeckt werden. Die technologischen Möglichkeiten erscheinen gegeben. Auf die Anwendung einer personalisierten Preisgestaltung wird jedoch offenbar in der Praxis noch verzichtet.

## Abbildungsverzeichnis

Abbildung 1: Kombinatorik technischer Einflussfaktoren .....	18
Abbildung 2: Grobarchitektur der Preisabfragesoftware .....	19
Abbildung 3: Screenshot einer Beispielseite .....	20
Abbildung 4: Beispielhafte Darstellung von Abfragekombinationen .....	26
Abbildung 5: Verteilungstabelle der Kombinationen .....	26
Abbildung 6: Beispielhafte Darstellung einer Erfassungsdatei in Microsoft Excel .....	27
Abbildung 7: VBA-Editor zur Excel-Programmierung.....	27
Abbildung 8: Handbuch zur manuellen Testerhebung (hier Firefox).....	28
Abbildung 9: Vorgehensweise der Auswertung .....	31
Abbildung 10: Anzahl erhobener Datensätze nach Erfasser .....	35
Abbildung 11: Anzahl Datensätze nach Produkt .....	36
Abbildung 12: Anzahl Datensätze nach Zugriffsart .....	36
Abbildung 13: Anzahl Datensätze nach Produkt .....	39
Abbildung 14: Anzahl Datensätze mit Preisdifferenz nach Produkt.....	39
Abbildung 15: Durchschnittliche Preisdifferenz nach Produkt .....	40
Abbildung 16: Durchschnittliche Preisdifferenz nach Produkt ohne „IntercityHotel Hamburg“ .....	41
Abbildung 17: Boxplots der Preise nach Produkt nach Entfernung der Ausreißer.....	42
Abbildung 18: Boxplot bereinigter Preis Apple iPhone 11 .....	43
Abbildung 19: Boxplot bereinigter Preis Ryanair Flug Berlin - Dublin .....	44
Abbildung 20: Boxplot bereinigter Preis IntercityHotel Hamburg .....	44
Abbildung 21: Boxplot bereinigter Preis Tesoro Bürostuhl .....	45
Abbildung 22: Boxplot bereinigter Preis Tui Sheraton Miramar Resort.....	45
Abbildung 23: Boxplots der Preisdifferenzen nach Produkt .....	46
Abbildung 24: Boxplots der Preisdifferenzen nach Entfernung der Ausreißer.	47
Abbildung 25: Dichteverteilung „IntercityHotel Hamburg“ .....	47
Abbildung 26: Anzahl Datensätze nach Zugriffsart .....	48
Abbildung 27: Anzahl Datensätze mit Preisdifferenz nach Zugriffsart.....	48
Abbildung 28: Durchschnittliche Preisdifferenz nach Zugriffsart .....	49

Abbildung 29: Anzahl Datensätze nach Proxy-Standort.....	49
Abbildung 30: Anzahl Datensätze mit Preisdifferenz nach Proxy-Standort ....	50
Abbildung 31: Durchschnittliche Preisdifferenz nach Proxy-Standort.....	50
Abbildung 32: Anzahl Datensätze mit und ohne Tracker .....	51
Abbildung 33: Anzahl Datensätze mit Preisdifferenz mit und ohne Tracker ...	51
Abbildung 34: Durchschnittliche Preisdifferenz mit und ohne Tracker .....	51
Abbildung 35: Anzahl Datensätze mit und ohne aktivierte Cookies .....	52
Abbildung 36: Anzahl Datensätze mit Preisdifferenz mit und ohne aktivierte Cookies .....	52
Abbildung 37: Durchschnittliche Preisdifferenz mit und ohne aktivierte Cookies .....	53
Abbildung 38: Anzahl Datensätze des manuellen Erhebungsstrangs mit und ohne Nutzerkonto-Login .....	53
Abbildung 39: Anzahl Datensätze mit Preisdifferenzen des manuellen Erhebungsstrangs mit und ohne Nutzerkonto-Login .....	54
Abbildung 40: Durchschnittliche Preisdifferenz mit und ohne Nutzerkonto- Login .....	54
Abbildung 41: Anzahl Datensätze des manuellen Erhebungsstrangs mit und ohne Social-Media-Login.....	55
Abbildung 42: Anzahl Datensätze mit Preisdifferenzen des manuellen Erhebungsstrangs mit und ohne Social-Media-Login.....	55
Abbildung 43: Durchschnittliche Preisdifferenz mit und ohne Social-Media- Account-Login .....	56
Abbildung 44: Preis nur für Mobilgerätenutzer .....	69
Abbildung 45: Dichteverteilung „IntercityHotel Hamburg“ .....	69

## Tabellenverzeichnis

Tabelle 1: Einbezogene Online-Shops und -Portale.....	16
Tabelle 2: Einbezogene Produkte .....	17
Tabelle 3: Übersicht über die Proxy-Standorte.....	22
Tabelle 4: Variablen .....	34
Tabelle 5: Anzahl Datensätze mit Preisdifferenz .....	38
Tabelle 6: Feature-Sets der Regressionen .....	57
Tabelle 7: Regression mit allen Datensätzen .....	59
Tabelle 8: Datensatz Regression nach Produkten .....	61
Tabelle 9: Regression nach Entfernung von Produkten ohne Preisdifferenzen	61
Tabelle 10: Regression des Produkts „IntercityHotel Hamburg“ (Feature-Set 1) .....	63
Tabelle 11: Regression des Produkts „IntercityHotel Hamburg“ (Feature-Set 2) .....	64
Tabelle 12: Regression des Datensatzes ohne „IntercityHotel Hamburg“ (Feature-Set 1) .....	65
Tabelle 13: Regression des Datensatzes ohne „IntercityHotel Hamburg“ (Feature-Set 2) .....	66
Tabelle 14: Regression des manuellen Erfassungsstrangs .....	67

## Literaturverzeichnis

[DAUTZENBERG, GABMANN, GROß, MÜLLER, NEUKAMP, SCHMIDTKE, BODENSTEIN 2018]

Individualisierte Preisdifferenzierung im deutschen Online-Handel. Potsdam: Verbraucherzentrale Brandenburg e. V., 2018.

[GENTH 2016]

Dynamische Preise: ein Gewinn für Handel und Verbraucher. In: Wirtschaftsdienst, 96, 863-868.

[HANNAK, SOELLER, LAZER, MISLOVE, WILSON 2014]

Measuring Price Discrimination and Steering on E-commerce Web Sites. DOI: 10.1145/2663716.2663744.

[HOLLAND 2016]

Dialogmarketing: Offline- und Online-Marketing, Mobile- und Social-Media-Marketing. München: Verlag Franz Vahlen, 2016.

[KRAEMER, KALKA, ZIEHE 2016]

Personalisiertes und dynamisches Pricing aus Einzelhandels- und Verbrauchersicht. Marketing Review St. Gallen, Ausgabe 6, 2016.

[KRUGMAN 2000]

Reckonings; What Price Fairness? New York: New York Times, 04.10.2000, Section A, S. 35.

[MÜLLER 2005]

Kundenbindung im E-Commerce: Personalisierung als Instrument des Customer Relationship Marketing. Wiesbaden: Deutscher Universitätsverlag, 2005.

[OBERGFELL 2017]

Personalisierte Preise im Lebensmittelhandel – Vertragsfreiheit oder Kundenbetrug? In: Zeitschrift für Lebensmittelrecht, 44, S. 290-301.

[ÖSTERREICHISCHES INSTITUT FÜR ANGEWANDTE TELEKOMMUNIKATION,  
AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH 2017]

Algorithmen und Preise: Die Individualisierung von Preisen im Online-Handel.  
Wien: Österreichisches Institut für angewandte Telekommunikation, AIT Aus-  
trian Institute of Technology GmbH, 2019.

[REINARTZ, HAUCAP, STÜBER 2017]

Preisdifferenzierung und -dispersion im Handel. Ausgewählte Schriften der IFH  
Förderer I Band 5, Köln: Gesellschaft zur Förderung des Instituts für Handels-  
forschung an der Universität zu Köln e. V., 2017.

[ROTT 2019]

A Consumer Perspective on Algorithms. In: The Transformation of Economic  
Law, Lucila de Almeida (Hrsg.), Oxford: Hart Publishing, S. 43-63.

[SCHLEUSENER, HOSELL 2016]

Expertise zum Thema "Personalisierte Preisdifferenzierung im Online-Handel".  
Berlin: Sachverständigenrat für Verbraucherfragen beim Bundesministerium  
der Justiz und für Verbraucherschutz.

[SCHWAIGER, HUFNAGEL 2018]

Gutachten zum Thema „Handel und elektronische Bezahlssysteme“. München:  
Ludwig-Maximilians-Universität München, Institut für Marktorientierte Unter-  
nehmensführung.

[TIETJEN 2017]

Dynamische und personalisierte Preise: Welche lauterkeitsrechtlichen Schran-  
ken gelten für Unternehmen? In: Praxis im Immaterialgüter- und Wettbe-  
werbsrecht (GRUR-Prax), 9, S. 546-548.

[ZANDER-HAYAT, REISCH, STEFFEN 2016]

Personalisierte Preise – Eine verbraucherpolitische Einordnung. In: Verbrau-  
cher und Recht, 31, S. 403-409.

[ZEITSCHRIFT FÜR DATENSCHUTZ 2014]

Zeitschrift für Datenschutz (ZD), S. 570, 573.

## Anhang

Dem Dokument auf Datenträger beigelegt sind die folgenden Dateien und Unterlagen:

1. full\_dataset.zip: kompletter Rohdatensatz (Formate: CSV und Python Pickle)
2. final\_set.zip: für Auswertung verwendete Datensätze (Format: CSV und Python Pickle)
3. regression.zip: Ergebnisse (Format: XLSX)

## Über ibi research



research

an der Universität  
Regensburg GmbH

Seit 1993 bildet die **ibi research an der Universität Regensburg GmbH** eine Brücke zwischen Universität und Praxis. Das Beratungs- und Forschungsinstitut betreibt anwendungsorientierte Forschung und Beratung mit Schwerpunkt auf Innovationen rund um Finanzdienstleistungen und den Handel.

ibi research arbeitet auf den Themenfeldern E-Business, Retail Banking und Governance & Controlling. Zugleich bietet ibi research in Zusammenarbeit mit seinen Partnern umfassende Beratungsleistungen zur Umsetzung der Forschungs- und Projektergebnisse an.

Das ibi-Partnernetzwerk ([www.ibi.de/partnernetzwerk.html](http://www.ibi.de/partnernetzwerk.html)) stellt heute den umfangreichsten Innovationsverbund im deutschsprachigen Finanzdienstleistungsbereich dar. Im Mittelpunkt steht der Netzwerkgedanke. ibi research stellt den Partnern die Forschungsergebnisse zur Verfügung, führt auf neutraler Plattform Experten der Finanzdienstleistungs- und Handelsbranche zusammen, fördert den intensiven Dialog und unterstützt dadurch den wertvollen Austausch von Ideen, Konzepten und Erfahrungen.

## Über trinnovative



Die **trinnovative GmbH** wurde 2014 als IT-Beratungs- und Dienstleistungsunternehmen in Regensburg gegründet. Der Tätigkeitsschwerpunkt der trinnovative liegt in der projektbasierten Unterstützung ihrer Kunden bei Entwicklungsvorhaben im Bereich Software und Data Science. Die drei Gründer der trinnovative sind erfahrene IT-Berater und können langjährige Berufs- und Projekterfahrung vorweisen. Aktuell zählt das Unternehmen acht Mitarbeiter.

Die Kunden der trinnovative stammen aus verschiedensten Branchen, so z. B. aus dem E-Commerce, der Logistik oder aus dem Bereich E-Government bzw. dem öffentlichen Sektor. Neben dem Projektgeschäft hält die trinnovative GmbH Kontakte zur aktuellen Forschung an Hochschulen und Universitäten und pflegt diese u. a. durch die Betreuung von Bachelor- und Masterarbeiten. Teil der Unternehmensphilosophie ist es, stets Software zu entwickeln, die dem Stand der Technik nicht nur entspricht, sondern diesen stetig vorantreibt.

## Impressum

### **Empirie zu personalisierten Preisen im E-Commerce – Schlussbericht**

ibi research an der Universität Regensburg GmbH:

Holger Seidenschwarz

Johann Faltermeier

Quirin Wierer

Nils Deichner

Dr. Georg Wittmann

trinnovative GmbH:

Dr. Andreas Beer

Dr. Johann Schenkl

### **Kontakt**

ibi research an der Universität Regensburg GmbH

Galgenbergstraße 25

93053 Regensburg

Tel.: 0941 943 1901

Fax: 0941 943 1888

E-Mail: [info@ibi.de](mailto:info@ibi.de)

Internet: [www.ibi.de](http://www.ibi.de)

trinnovative GmbH

Franz-Mayer-Straße 1

93053 Regensburg

Tel.: 0941 604 889 450

Fax: 0941 604 889 459

E-Mail: [info@trinnovative.de](mailto:info@trinnovative.de)

Internet: [www.trinnovative.de](http://www.trinnovative.de)