

ETL Prozess Plan

Preisstruktur für gebrauchte iPhones

Gruppe 55

Annalena Kreischer (annalena.kreischer@stud.hslu.ch)

Mélanie Bigler (melanie.bigler.01@stud.hslu.ch)

Michelle Peter (michelle.peter@stud.hslu.ch)

Modul: CIP (Data Collection, Integration und Preprocessing), HS23

Dozent: Erwin Mathis

17. November 2023

Inhaltsverzeichnis

1 Einleitung & Motivation.....	3
2 Fragestellungen.....	4
3 Quellen	5
3.1 tutti.ch.....	5
3.2 revendo.com.....	6
3.3 preisvergleich.ch	7
4 Kontextdiagramm	8
5 Risiken und Herausforderungen	9
6 Referenzlist.....	11
7 Tabellen- & Abbildungsverzeichnis	11
Anhang – Codebuch.....	12

1 Einleitung & Motivation

Ein Leben ohne Smartphone ist heutzutage für die meisten Menschen im Alltag und Berufsleben undenkbar geworden. So besaßen 93% der Schweizer*innen 2022 ein Smartphone (Statista, 2023a). Ganz besonders beliebt: Apple (47,5% der Schweizer*innen; Statista, 2023b). Neben der Funktionalität spielt bei dieser Marke auch der Status als Premiumprodukt eine Rolle bei der Wahl. Der Gebrauchtmärkte ist hierbei interessant, da ein gebrauchtes iPhone eine nachhaltige sowie kostengünstigere Alternative zu einem neuen Handy darstellt. Insbesondere iPhones sind aufgrund ihrer Langlebigkeit und des anhaltenden Markenwerts beliebt. Das Interesse spiegelt sich auch im wachsenden Gebrauchtmärkte wider, beispielsweise bei der Firma „Reverdo“, für die iPhones das wichtigste Produkt darstellt (Schurter, 2021).

Auf dem Gebrauchtmärkte besteht jedoch eine gewisse Unsicherheit bezüglich des Wertes der iPhones. Die Preisgestaltung ist oft variabel und von verschiedenen Faktoren beziehungsweise Faktorkonstellationen abhängig. Hierzu zählen neben dem Modell und dem optischen Zustand auch die Akkuleistung, der Speicherplatz, Verfügbarkeit des Originalzubehörs oder auch die Farbe des iPhones. Dies macht es für Käufer*innen schwierig, faire Preise zu erkennen und miteinander zu vergleichen. Dies wird bei privaten Gebrauchtmärkte durch unvollständige Beschreibungen verstärkt. Zudem erfordert es Zeit, sich als Käufer*in zu informieren, was ein fairer Preis ist bei welchen Faktorkonstellationen und es gibt eine Vielzahl an Plattformen für gebrauchte iPhones.

Es ist aus wirtschaftlichen und ökologischen Gründen von Bedeutung, Käufer*innen in ihrer Kaufentscheidung bei einem gebrauchten iPhone zu unterstützen. So fördert der Wiederverkauf eine nachhaltigere Konsumkultur, indem er die Lebensdauer von Produkten verlängert. Dies reduziert den Bedarf an Neuproduktion, was wiederum Ressourcen schont und Abfall und damit auch Umweltbelastungen verringert sowie Recyclingkosten spart. Darüber hinaus Menschen mit geringerem Einkommen ermöglicht, iPhones zu erschwinglichen Preisen zu erwerben. Auch aus Sicht von Apple ist der Gebrauchtmärkte förderungswert, denn ein starker Wiederverkaufsmärkte kann als Indikator für die Qualität und Langlebigkeit dienen und so die Kundenbindung fördern oder Neukunden generieren.

Unsere Motivation bestand daher darin, Käufer*innen in ihrer Kaufentscheidung für ein gebrauchtes iPhone zu unterstützen, indem wir die komplexe und zeitaufwendige Aufgabe der Preisermittlung übernehmen. Unser Ziel ist es, durch die Bereitstellung einer klaren und umfassenden Übersicht der verschiedenen Preisfaktoren und -konstellationen den Prozess der Kaufentscheidung zu vereinfachen und transparenter zu gestalten.

2 Fragestellungen

Wie bereits dargelegt, zielt unser Projekt darauf ab, potenziellen Käuferinnen und Verkäuferinnen von gebrauchten iPhones hilfreiche Anhaltspunkte zu bieten, um zu beurteilen, ob der Preis eines gebrauchten iPhones dem aktuellen Marktwert entspricht.

Wir haben uns entschieden, unser Projekt auf iPhones ab dem Baujahr 2018, beginnend mit dem iPhone XS, zu fokussieren. Eine Liste der relevanten iPhone-Modelle ist im Anhang zu finden. Dies ist vor allem damit zu begründen, dass Apple Software-Updates und Support im Schnitt fünf Jahre lang anbietet (Maciej, 2020). Dies bedeutet, dass iPhones ab 2018 mindestens bis Ende dieses Jahres die neuesten iOS-Updates erhalten werden, was sie für Käufer*innen attraktiver macht. Zudem ist die Wahrscheinlichkeit bei neueren Modellen höher, dass die Akkuleistung in einem annehmbaren Rahmen liegt. Doch auch abgesehen davon tendieren Nutzer*innen aus persönlicher Erfahrung dazu, Modelle zu bevorzugen, die nicht zu alt sind, um von neueren Technologien und Designs zu profitieren. Es ist also eine strategische Entscheidung, die sicherstellen soll, dass unser Projekt technologisch relevant, wirtschaftlich sinnvoll und für die Zielgruppe attraktiv ist. Beim Vergleich der Modelle berücksichtigen wir die Aspekte Modellversion (z. B. iPhone 11), Speichergrösse (z. B. 64 GB), Farbe und den Zustand des Geräts.

Aus diesen Überlegungen ergibt sich folgende Haupt-Forschungsfrage für die Arbeit:

Was sind die durchschnittlichen Preise für iPhone-Modelle ab dem Baujahr 2018 auf dem Gebrauchtwarenmarkt im Vergleich zu den Preisen auf dem Erstmarkt?

Daraus möchten wir ableiten, welche Preise pro Modell dem Marktwert entsprechen. Ebenfalls möchten wir mit folgenden Fragen die Preisgestaltung noch genauer untersuchen:

- Wie stark beeinflusst der Zustand (optischer Zustand und Akkukapazität) eines iPhones dessen Preis auf dem Gebrauchtmakrt?
- In welchem Tempo und Umfang verringern sich die Preise von gebrauchten iPhones im Vergleich zum Verkaufspreis der jeweiligen neuen iPhones? Gibt es Unterschiede zwischen den Modellen?
- Gibt es Preisunterschiede zwischen iPhones verschiedener Farben und, falls ja, wie gross sind diese?
- Unterscheiden sich die Preise für gebrauchte iPhones zwischen den verschiedenen Gebrauchtmakrtplattformen, z. B. zwischen spezialisierten und allgemeinen Handelsplattformen?

Es ist zu beachten, dass die aufgeführten Fragestellungen nicht das gesamte Spektrum möglicher Untersuchungsaspekte abdecken. Weitere relevante Fragen könnten im Verlauf der Forschung identifiziert und integriert werden.

3 Quellen

Um die Fragestellung zu beantworten und eine geeignete Datengrundlage zu Verfügung zu stellen, planen wir, Daten von drei relevanten Webseiten zu sammeln: tutti.ch, revendo.com und preisvergleich.ch. Diese Plattformen bieten eine reiche Auswahl an iPhones und repräsentieren verschiedene Facetten des Marktes – von privaten Verkäufen über professionell aufbereitete Geräte bis hin zu neuen Geräten.

Unser Ziel ist es, eine Übersicht der durchschnittlichen Preise zu erstellen, die nach detaillierte Angaben zu Modell, Zustand, Spezifikationen und weiteren relevanten Faktoren aufgeschlüsselt ist. Durch das Sammeln und Aggregieren dieser Daten können wir tiefere Einblicke in die Preisstrukturen und -trends gewinnen und so die Entscheidungsfindung für potenzielle Käufer*innen erleichtern.

Im Rahmen dieser Arbeit legen wir Wert darauf, die Daten in Übereinstimmung mit den rechtlichen Vorgaben zu sammeln. Die zu nutzenden Datensätze enthalten keine personenbezogenen Informationen, sie können also nicht auf individuelle Personen zurückgeführt werden können. Dies erleichtert die Einhaltung der Datenschutzkonformität. Zudem prüfen wir die robots.txt-Datei jeder Webseite und stellen sicher, dass unsere Datenscraping-Prozesse nur solche Bereiche der Webseiten abdecken, die explizit erlaubt sind.

3.1 tutti.ch

tutti.ch ist eine Online-Plattform in der Schweiz, die als Kleinanzeigenmarkt dient. Dort werden viele verschiedene Gebrauchsgüter zum Verkauf oder Tauschen sowohl von Privatpersonen als auch von professionellen Händlern angeboten. Gerade bei tutti.ch ist daher die Einhaltung der robots.txt-Richtlinien wichtig zum Schutz der Sicherheit der Nutzer*innendaten. Die robots.txt erlaubt grundsätzlich den Zugriff auf die Webseite, mit bestimmten Ausnahmen. Gesperrt sind persönliche Bereiche wie die Favoriten, personalisierte Werbung oder Nutzer*innendaten sowie Funktionen wie promote und payment.

Als Grundlage dient die Suche auf tutti.ch mit dem Suchbegriff «iphone». Ausgewählt ist die Kategorie Handys, um Zubehör wie Kabel auszuschliessen. Zudem ist die Option «Nur Inserate von tutti.ch» aktiviert sowie «Deutsch» ausgewählt. Die Suche ergab einen Bestand von 3255 iPhones (16.11.2023), um einen groben Rahmen für die Datenmenge zu geben. tutti.ch bietet wenig vordefinierte Felder, um Informationen zu Produkten detailliert anzugeben. Daher sind wesentliche Details wie der Zustand des Gerätes oder die Farbe nicht systematisch erfasst, sondern in freien Textfeldern und dem Titel ohne einheitliche Formatierung zu finden. Wir können somit den Titel, die Beschreibung sowie den Preis scrapen, um daraus im nachfolgenden Transformationsprozess die Zielvariablen zu extrahieren (geplant sind derzeit Modell, Speicherplatz, Farbe, Zustand, Akkukapazität, Zubehör).

Eine Herausforderung bei tutti.ch werden neben verschiedenen Schreibweisen sowie Rechtschreibfehlern vor allem fehlende Daten sein. Diese können fehlen, weil sie von den Verkäufer*innen nicht angegeben sind,

oder weil sie in Form von Bildern zu finden sind (wie Screenshots der Akkukapazität). Diese Faktoren können die Qualität und Konsistenz der gesammelten Daten beeinträchtigen und erfordern daher besondere Aufmerksamkeit während des Scraping- und Datenverarbeitungsprozesses. Wir haben beschlossen, als ersten definitiven Schritt Beobachtungen auszuschließen, bei denen wesentliche Informationen wie Modell und Speicherkapazität fehlen. Der finale Datensatz soll schließlich um den Durchschnittspreis pro Modell (aufgeschlüsselt nach Speicherkapazität, Akku- und optischem Zustand) sowie die Preisabweichung von diesem Durchschnitt ergänzt werden. Zusätzlich wird der Preis von Schweizer Franken in Euro umgerechnet, um den Vergleich für potentielle Käufer*innen von iPhones aus dem Euro-Ausland einfacher zu gestalten. Eine Spalte soll die Datenquelle (tutti.ch) angeben.

3.2 revendo.com

Zur Analyse ziehen wir als weitere Datenquelle die Website <https://revendo.com/de-ch/kaufen/k/iphone/> heran. Auf dieser Seite ist bereits ein Filter für die Marke iPhone eingestellt. Revendo zählt zu den renommiertesten Online-Shops in der Schweiz für den An- und Verkauf von Elektronik, insbesondere Mobiltelefone. Die umfassende Auswahl bei Revendo erlaubt uns, eine breite Palette an refurbished und gebrauchten Geräten zu untersuchen, was uns einen Einblick in die Dynamik des Sekundärmarktes und die Preisentwicklung ermöglicht.

Grundsätzlich erlaubt die robots.txt-Datei von revendo.com Webcrawlern den Zugriff auf die meisten Bereiche der Website, setzt jedoch spezifische Einschränkungen. Sie verbietet das Crawlen von URLs, die nach dem Slash ('/') ein Fragezeichen ('?') enthalten, und den Zugriff auf Verzeichnisse, die sensible Benutzerinformationen wie Kontodaten oder Checkout-Prozesse enthalten könnten. Da unser Fokus auf dem Kaufbereich und der Marke iPhone liegt, prüften wir, ob in diesen spezifischen URLs ein '?' vorkommt, was jedoch nicht der Fall ist. Daher steht dem Scrapen dieser URLs nichts im Wege. Zusätzlich haben wir untersucht, in welchen Bereichen der Website '?' in den URLs häufig auftreten. Es stellte sich heraus, dass dies hauptsächlich im Verkaufsbereich der Fall ist, der für unser aktuelles Projekt jedoch irrelevant ist.

Zum Zeitpunkt des 17.11.2023 bestehen 270 Einträge, die gescraped werden können. Die folgenden Informationen sollen daraus extrahiert werden.

Informationen auf der Webseite (Extract)	Zielvariablen in der Datenbank (Transform)	Beispiel
Produktname	Kategorie	iPhone
	Modell	11
	Speicher in GB	512
	Farbe	Schwarz
Kamera	Kamera	12 Mpx
Optischer Zustand	Optischer Zustand	Schnäppchen

Akkuzustand	Akkukapazität	Gebraucht
Produkt Detail Preis	Preis in CHF	367 CHF

Tabelle 1 - Datenstruktur revendo.com

Wir extrahieren und transformieren möglichst viele Informationen, aber nicht alle sind essentiell für die Auswertung der Frage. Folgende Zielvariablen sollen im Nachgang erarbeitet werden:

- Datenquelle (für diesen Datensatz immer revendo.com)
- Berechnung: Preis ohne Mehrwertsteuer
- Berechnung: Preisabweichung vom durchschnittlichen Preis des Modells
- Berechnung: Durchschnittlicher Preis des Modells
 - Anhand des Modells, Farbe sowie Zustand

3.3 preisvergleich.ch

Um Preisänderungen auf dem Sekundärmarkt (Secondhand) einordnen zu können, werden auch Preise vom Primärmarkt erfasst. Dies ermöglicht einen direkten Vergleich zwischen den aktuellen und den veränderten Preisen. Für die Datenerhebung wird die Website <https://www.preisvergleich.ch/tag/apple-iphone.html> genutzt, auf der bereits ein Filter für Apple-Smartphones aktiviert ist. Wir haben diesen Shop ausgewählt, da er einerseits die günstigsten Marktpreise bietet und andererseits die Möglichkeit bietet, einfach zusätzliche Daten zu Angebotspreisen auf anderen Plattformen zu sammeln.

Die Verwendung der Daten von der Webseite ist grundsätzlich nicht gestattet, wie aus der robots.txt-Datei der Webseite hervorgeht. Dort wird beispielsweise das Crawlen von Produktlisten untersagt (Beispiel aus der robots.txt Datei: *Disallow: /product-list/*). Im Impressum der Webseite wird jedoch angegeben, dass «Fotokopien und Downloads von Web-Seiten dürfen nur für den persönlichen, privaten und nicht kommerziellen Gebrauch hergestellt werden» (preisvergleich.ch, 2023).

Trotz dieser Einschränkungen beabsichtigen wir, die Daten zu nutzen, auch wenn dies eine rechtliche Grauzone darstellt. Da unser Zweck nicht kommerzieller Natur ist und wir keine personenbezogenen Daten erfassen, sind wir bereit, dieses Risiko einzugehen.

Es bestehen zum Zeitpunkt am 16.11.2023 349 Einträge von iPhone Modellen, die gesprapt werden können. Die folgenden Informationen sollen daraus extrahiert werden.

Informationen der Webseite (Extract)	Zielvariable in der Datenbank (Transform)	Beispiel
Produktname	Model	iPhone 13
	Farbe	Blau
	Speicher in GB	512

	Grösse	6.1
	Kamera	12 MP
Preis	Preis in CHF	917.20
Anbieter	Anbieter	Galaxus
Durchschnittliche Bewertung	Durchschnittliche Bewertung von 1-5	4.4
Anzahl Bewertungen	Anzahl Bewertungen	7065
Unser Fazit	Bewertung preisvergleich.ch	5.2

Tabelle 2 - Datenstruktur preisvergleich.ch

Wir extrahieren und transformieren möglichst viele Informationen, aber nicht alle sind essentiell für die Auswertung der Frage. Zudem werden folgenden Zielvariablen im Nachgang hinzugefügt und berechnet, damit diese dann auch mit den anderen gesammelten Datensätze übereinstimmen:

- Optischer Zustand (für diesen Datensatz immer *Ungeöffnet*)
- Akkukapazität (für diesen Datensatz immer *Neu*)
- Datenquelle (für diesen Datensatz immer *preisvergleich.ch*)
- Berechnung: Preis ohne Mehrwertsteuer
- Berechnung: Durchschnittlicher Preis des Modells (basierend auf Speichergrösse & Model)
- Berechnung: Preis-Abweichung vom Durchschnittlichen Preis des Modells

4 Kontextdiagramm

Das Diagramm stellt einen ETL-Prozess (Extract, Transform, Load) dar, bei dem Daten aus verschiedenen Quellen (tutti.ch, revendo.com, preisvergleich.ch) extrahiert, transformiert und schliesslich in eine Maria Datenbank geladen werden. Jeder Schritt wird beschrieben, einschliesslich der verwendeten Tools und Methoden wie Python in Pycharm, Scraping mit Selenium oder/und BeautifulSoup, Datenbereinigung und -anreicherung mit Pandas sowie das Hochladen verschiedener Datensätze in die Datenbank.

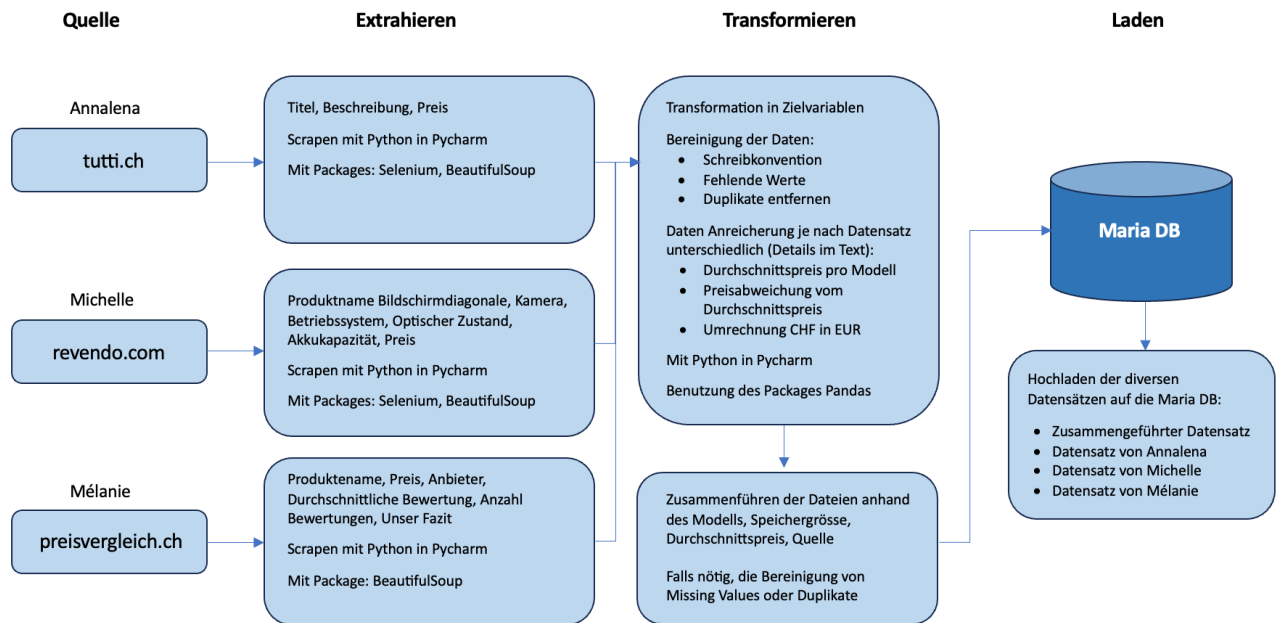


Abbildung 1 - Kontextdiagramm des ETL-Prozesses

Da wir die Daten von verschiedenen Plattformen sammeln, könnte die Vereinheitlichung der Datenformate- und -strukturen eine Herausforderung darstellen.

Schlussendlich möchten wir einen Datensatz der basierend auf das iPhone Modell, Farbe und Speichergrösse die durchschnittlichen Preise angibt pro Quelle. Dies soll mithilfe eines Full Joins gemacht werden.

Nachfolgend die geplante Kopfzeile für den Datensatz:

Modell	Farbe	Speicher in GB	Durchschnittspreis	Quelle
--------	-------	-------------------	--------------------	--------

Es wird in Erwägung gezogen, weitere Variablen aus den drei individuellen Datensätzen bei Bedarf zu ergänzen, um die Vollständigkeit und Aussagekraft unserer Datenanalyse zu optimieren. Beispielsweise sofern es die Datenqualität der gescrapten Inhalte von tutti zulassen, soll der Datensatz zusätzlich mit dem optischen Zustand und/oder der Akkukapazität erweitert werden. Die Kopfzeile für den zusammengezogenen Datensatz würde dann wie folgt aussehen:

Modell	Farbe	Speicher in GB	Durchschnittspreis	Quelle	Optischer Zustand	Akkukapazität
--------	-------	-------------------	--------------------	--------	-------------------	---------------

5 Risiken und Herausforderungen

Bisher haben wir den geplanten ETL-Prozess (Extract, Transform, Load) erarbeitet und die Webseiten oberflächlich geprüft. Eine Herausforderung könnte sein, dass das Scrapen der Daten nicht wie geplant verläuft. Unser Hauptziel ist die Sammlung der geplanten Daten, die konkrete Quelle dieser Daten ist dabei

von untergeordneter Bedeutung. Sollten wir auf unüberwindbare Schwierigkeiten beim Scrapen der Daten von den ausgewählten Webseiten stossen, behalten wir uns daher die Möglichkeit vor, auf alternative Webseiten mit ähnlichen Daten auszuweichen, von denen es bei unserem Thema viele gibt.

Eine weitere Herausforderung ist, dass sich die Produkte beziehungsweise Preise auf allen drei Webseiten täglich ändern können. Daher ist es wichtig, dass wir das Scraping an einem Zeitpunkt durchzuführen statt an mehreren, um eine Momentaufnahme der verfügbaren Daten zu erfassen. Dieser Ansatz spiegelt die Erfahrung eines potenziellen Käufers wider, der die Plattform an einem bestimmten Tag besucht und auf die zu diesem Zeitpunkt verfügbaren Informationen angewiesen ist. Es bietet daher ein realistisches Bild der Marktsituation. Der gewählte Zeitpunkt selbst ist weniger entscheidend, da der Schwerpunkt auf der Erfassung eines authentischen und zeitlich relevanten Datenstands liegt.

Darüber hinaus werden bei tutti.ch vor allem unterschiedliche Schreibkonventionen, fehlende Informationen und Rechtschreibfehler problematisch werden (Beispiel: neu, ungebraucht, ungeöffnet, GB, Gigabyte, ...). Diese Faktoren können die Qualität und Konsistenz der gesammelten Daten beeinträchtigen und erfordern daher besondere Aufmerksamkeit während des Scraping- und Datenverarbeitungsprozesses. Insbesondere sind Qualitätskontrollen erforderlich, um Daten nicht versehentlich zu löschen oder falsch zuzuordnen.

Eine weitere Herausforderung wird beim Zusammenfügen unserer Datensätze auftreten. Präventiv ist es daher zentral, Standardisierungen für Variablennamen sowie modellspezifische Inhalte (z. B. Kodierungen) festzulegen. Ebenso sollten Datenformate vereinheitlicht sein, um Inkonsistenzen zu vermeiden. Hier wird ein Codebuch erarbeitet, das alle iPhone-Modelle beinhaltet (siehe Anhang), eine detaillierte Beschreibung der Variablen und ihrer möglichen Werte (Akkukapazität und optischer Zustand) sowie alle Farben nach Modell aufgeschlüsselt. Dies dient dem Verständnis und unterstützt das Zusammenführen unserer Datensätze. Zudem können wir so überprüfen, ob sämtliche Modelle in allen Farbvarianten im Datensatz repräsentiert sind.

6 Referenzlist

Maciej, M. (2020, May 12). iOS-Updates: Wie lange noch für iPhone 6s, 7 und 8? Deinhandy.De.
<https://blog.deinhandy.de/iphone-updates-wie-lange-noch-fuer-iphone-6s-7-und-8>

preisvergleich.ch. (11. November 2023). Impressum. <https://www.preisvergleich.ch/impressum.html>

Schurter, D. (2021, November 18). Der Handel mit gebrauchten, älteren iPhones boomt in der Schweiz.
Watson. <https://www.watson.ch/digital/schweiz/375400456-der-handel-mit-gebrauchten-aelteren-iphones-boomt-in-der-schweiz>

Statista. (2023a, November 16). Schweiz - Besitz von Mobiltelefonen 2022 | Statista.
<https://de.statista.com/statistik/daten/studie/537944/umfrage/besitz-von-smartphone-bzw-tablet-in-der-schweiz/>

Statista. (2023b, November 16). Schweiz - Nutzeranteile der Smartphone-Hersteller 2022 | Statista.
<https://de.statista.com/statistik/daten/studie/326565/umfrage/marktanteile-von-mobiltelefon-herstellern-in-der-schweiz/>

7 Tabellen- & Abbildungsverzeichnis

Tabelle 1 - Datenstruktur revendo.com	7
Tabelle 2 - Datenstruktur preisvergleich.ch	8
Abbildung 1 - Kontextdiagramm des ETL-Prozesses	9

Anhang – Codebuch

Übersicht iPhone-Modelle seit 2018

Modell	Erscheinungsjahr
iPhone 15	September 2023
iPhone 15 Plus	September 2023
iPhone 15 Pro	September 2023
iPhone 15 Pro Max	September 2023
iPhone 14	September 2022
iPhone 14 Plus	September 2022
iPhone 14 Pro	September 2022
iPhone 14 Pro Max	September 2022
iPhone SE (2022)	März 2022
iPhone 13	September 2021
iPhone 13 Pro	September 2021
iPhone 13 Pro Max	September 2021
iPhone 13 Mini	September 2021
iPhone 12	Oktober 2020
iPhone 12 Pro	Oktober 2020
iPhone 12 Pro Max	Oktober 2020
iPhone 12 Mini	Oktober 2020
iPhone SE (2020)	April 2020
iPhone 11	September 2019
iPhone 11 Pro	September 2019
iPhone 11 Pro Max	September 2019
iPhone XS	September 2018
iPhone XS Max	September 2018

Quelle: <https://www.sparhandy.de/apple/info/iphone-modelle/>