

Projekt Dokumentation

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

23. Dezember 2023

Inhaltsverzeichnis

1	Einleitung & Motivation	3
1.1	Fragestellungen	3
2	Vorgehen	3
2.1	Eingesetzte Tools	4
3	Daten-Quellen.....	4
3.1	tutti.ch	4
3.2	revendo.com	5
3.3	preisvergleich.ch	5
4	Umsetzung	5
4.1	Tutti	5
4.1.1	Extrahieren	6
4.1.2	Transformieren	6
4.1.3	Laden	7
4.2	Revendo	8
4.2.1	Extrahieren	8
4.2.2	Transformieren	8
4.2.3	Laden	10
4.3	Preisvergleich	10
4.3.1	Extrahieren	10
4.3.2	Transformieren	10
4.3.3	Laden	12
4.4	Daten Zusammenführung.....	12
4.4.1	Transform: Finaler Datensatz aufbereiten	12
4.4.2	Load: Upload auf MariaDB	12
5	Resultate.....	12
5.1	Hauptfragestellung	12
5.2	1. Subfrage: Preis pro iPhone-Modell	15
5.3	2. Subfrage: Einfluss des Zustandes	17
5.4	3. Subfrage: Preisverringerung	18
5.5	4. Subfrage: Preisunterschiede pro Farbe	19
6	Reflektion.....	20
	Referenzliste.....	22
	Tabellen- & Abbildungsverzeichnis	23
	Anhang	24

*Einverständniserklärung: Die vorliegende Arbeit darf anderen Student*innen zur Verfügung gestellt werden.*

1 Einleitung & Motivation

Ein Leben ohne Smartphone ist heutzutage für die meisten Menschen im Alltag und Berufsleben undenkbar. 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. Die Marke Apple, bekannt für ihre Funktionalität und ihren Status als Premiumprodukt, ist besonders im Gebrauchtmittelmarkt beliebt, da gebrauchte iPhones nachhaltige und preiswertere Alternativen bieten. Das Interesse spiegelt sich auch im wachsenden Gebrauchtmittelmarkt wider, beispielsweise bei der Firma «Revendo», für die iPhones das wichtigste Produkt darstellt (Schurter, 2021).

Allerdings herrscht Unsicherheit über den Wert gebrauchter iPhones, da die Preisgestaltung variabel ist und von mehreren Faktoren wie Modell, Zustand, Akkuleistung, Speicherplatz, Originalzubehör und Farbe abhängt. Diese Komplexität erschwert es Käufer*innen, faire Preise zu erkennen, besonders auf privaten Verkaufsseiten mit oft unvollständigen Produktbeschreibungen.

Unser Ziel bestand daher darin, Käufer*innen in ihrer Kaufentscheidung für ein gebrauchtes iPhone zu unterstützen, indem wir die komplexe und zeitaufwendige Aufgabe der Preisermittlung- und des Vergleichs vereinfachen und transparent machen, welche Faktoren den Preis beeinflussen. Dafür wollen wir von verschiedenen Plattformen Angebote scrapen und eine Preisübersicht für die verschiedenen Modellkonstellationen erstellen. Diese soll den Vergleich über verschiedene Plattformen hinweg vereinfachen.

1.1 Fragestellungen

Ausgehend von diesen Überlegungen ergibt sich folgende Fragestellung für den/die potentiellen Käufer*innen, mit der Voraussetzung, dass der/die bereits eine klare Vorstellung vom gewünschten iPhone-Modell und der benötigten Speichergrösse hat:

Welche Plattform bietet das gewünschte iPhone-Modell in welchem Zustand zu welchem Preis an? Wie gross sind Preisunterschiede zwischen Farben?

Mithilfe dieser Übersicht soll der/die potentielle Käufer*in entscheiden können, welches Angebot am besten zu seinen/ihren Bedürfnissen passt. Im Rahmen dieser Fragestellung möchten wir mithilfe der gesammelten Daten ebenfalls folgende, übergreifende Fragen beantworten:

- 1. Subfrage:** Was sind die durchschnittlichen Preise pro iPhone-Modelle ab dem Baujahr 2018 auf den verschiedenen Plattformen (Gebraucht und Erstmarkt)?
- 2. Subfrage:** Wie stark beeinflusst der Zustand (optischer Zustand und Akkukapazität) eines iPhones dessen Preis auf dem Gebrauchtmittelmarkt?
- 3. Subfrage:** In welchem Umfang verringern sich die Preise von gebrauchten iPhones im Vergleich zum aktuellen Verkaufspreis eines neuen iPhones? Wie gross ist der Einfluss der Speichergrösse und Erscheinungsjahr?
- 4. Subfrage:** Welche Preisunterschiede gibt es pro iPhone-Modellen in Abhängigkeit von ihrer Farbe? Gibt es ein allgemeines Muster über alle Modelle hinsichtlich der Preisgestaltung für die verschiedenen Farboptionen?

2 Vorgehen

In unserem Projekt haben wir den ETL-Prozess (Extract, Transform, Load) zur Datengewinnung aus verschiedenen Online-Quellen und deren anschliessender Speicherung in einer Maria-Datenbank strukturiert. Im Extraktionsprozess sammeln wir spezifische Produktdaten von den Plattformen tutti.ch, revendo.com und preisvergleich.ch.

Die gewonnenen Daten durchlaufen dann eine Transformationsphase, in der wir die Daten für die Analyse aufbereiten, indem wir Schreibkonventionen vereinheitlichen, fehlende Werte ergänzen und Duplikate beseitigen. Um die Zusammenführung zu vereinfachen und Schreibkonventionen einzuführen, haben wir für die Modelle sowie Farben ein Codebuch erstellt. Unser Ziel ist es, einen Datensatz zu erstellen, der eine umfassende Übersicht über die durchschnittlichen Preise verschiedener iPhone-Modelle bietet, segmentiert

nach Modell, Farbe, Speichergrösse und Quelle. Die Anreicherung von Daten sind je nach Datensatz unterschiedlich und können im jeweiligen Kapitel im Detail nachgeschlagen werden. Die abschliessende Phase des Prozesses umfasst das Laden der transformierten Daten in die MariaDB-Datenbank, um sie für Analysen verfügbar zu machen. Das nachfolgende Diagramm zeigt den Prozess.

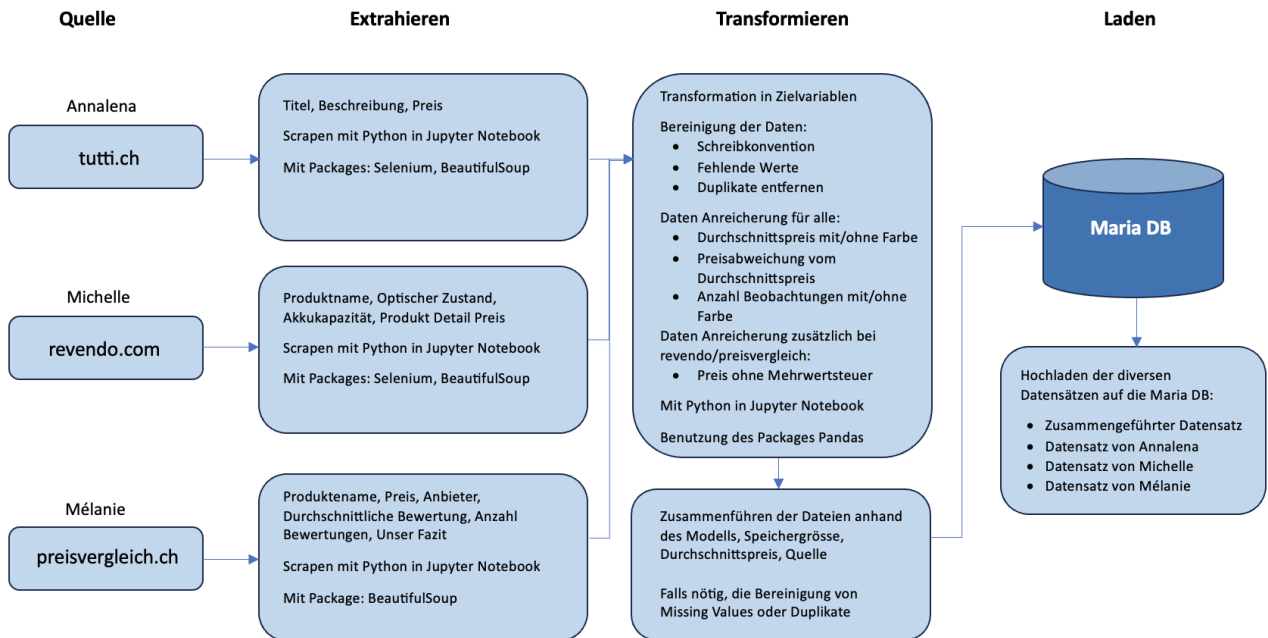


Abbildung 1 - Kontextdiagramm des ETL-Prozesses

2.1 Eingesetzte Tools

Das Kontextdiagramm illustriert die eingesetzten Tools für den Scraping-Prozess, welcher mit Python implementiert wurde. Für das Sammeln der Daten von Tutti und Revendo nutzten wir die Bibliotheken BeautifulSoup und Selenium. Im Gegensatz dazu kam bei der Arbeit mit Preisvergleich ausschliesslich BeautifulSoup zum Einsatz. Als Entwicklungsplattform diente Jupyter Notebook, welches den gesamten ETL-Prozess unterstützt. Dabei waren je nach Datensatz und Anforderung unterschiedliche zusätzliche Python-Pakete notwendig, die jeweils in den Jupyter Notebook Files dokumentiert sind.

Die Bereinigung und Anreicherung der Daten erfolgten hauptsächlich unter Verwendung des Pandas-Pakets, während für numerische Berechnungen das Numpy-Paket zum Einsatz kam. Zum Abschluss wurden die Datensätze mittels Python in die MariaDB hochgeladen.

3 Daten-Quellen

Zur Beantwortung unserer Fragestellungen sammeln wir Daten von drei relevanten Webseiten: tutti.ch, revendo.com und preisvergleich.ch. Diese decken ein breites Spektrum des iPhone-Marktes ab, von privaten Angeboten zu gebrauchten Geräten bis hin zu neuen Geräten auf dem Erstmarkt. Bei unserer Datensammlung achten wir auf Datenschutzkonformität, verwenden keine personenbezogenen Daten und stellen durch die Überprüfung der robots.txt-Dateien der Webseiten sicher, dass unser Datenscraping den Vorgaben entspricht.

3.1 tutti.ch

Tutti.ch ist eine beliebte Online-Plattform in der Schweiz, die als Kleinanzeigenmarkt fungiert. Hier werden viele verschiedene Gebrauchtwaren sowohl von Privatpersonen als auch von professionellen Händlern zum Verkauf oder Tausch angeboten. Die Möglichkeit, Anzeigen kostenlos aufzuschalten, macht die Plattform besonders attraktiv und resultiert in einer grossen Vielfalt an Inseraten. Dies wiederum bietet eine breite Auswahl an möglichen gebrauchten iPhone-Geräten, die wir für unsere Analyse benötigen.

Ich entschied mich, direkt eine gefilterte Seite als Basis für die Datenerhebung zu nutzen: <https://www.tutti.ch/de/q/handys/Ak6ZpcGhvbWqY2VsbFBob25lc5SSkqhsYW5ndWFnZaJkZZKn3JnYW5pY6V0dXR0acCRlKVwcmJjZCJkwMA?sorting=newest&page=1&query=iphone>. Bei diesem Link ist auf tutti.ch der Suchbegriff «iphone» eingegeben und die Kategorie Handys ausgewählt, um Zubehör auszuschliessen. Zudem ist die Option «Nur Inserate von tutti.ch» aktiviert und «Deutsch» als Sprache ausgewählt. Zusätzlich wurde ein Mindestpreis von 100 CHF festgelegt. Das direkte Aufrufen des gefilterten URLs ist wesentlich schneller und effizienter, da es unnötige Navigationsschritte und Interaktionen mit der Webseite umgeht. Weniger Interaktionen mit der Webseite bedeuten ausserdem eine geringere Fehleranfälligkeit, was den Scraping-Prozess zuverlässiger macht.

Insbesondere bei tutti.ch ist die Einhaltung der robot.txt-Richtlinien zum Schutz der Nutzer*innen wichtig. Die robots.txt gestattet grundsätzlich den Zugriff auf die Webseite, jedoch mit bestimmten Ausnahmen. Gesperrt sind persönliche Bereiche wie die Favoriten, personalisierte Werbung oder Nutzungsdaten sowie Funktionen wie «Promote» und «Payment». Da wir auf diese Bereiche nicht zugreifen, stellt das Scraping aus Sicht des Datenschutzes keine weiteren Probleme dar.

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.

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 nutzen wir die Daten, auch wenn dies eine rechtliche Grauzone darstellt. Da unser Zweck nicht kommerzieller Natur ist und wir keine personenbezogenen Daten erfassen, gehen wir das Risiko ein.

4 Umsetzung

In diesem Kapitel erklären wir, wie der ETL-Prozess umgesetzt wurde und zeigen Umsetzungs-Probleme sowie deren Lösungsansätze auf.

4.1 Tutti

Die Daten von tutti wurden von Annalena Kreischer scraped, daher ist sie für dieses Kapitel verantwortlich. In diesem Kapitel wird erklärt, wie die Daten von der Quelle tutti.ch extrahiert, transformiert und geladen wurden.

4.1.1 Extrahieren

Als Teil meiner Vorabüberlegung habe ich den Seitenquelltext von tutti.ch untersucht, um zu entscheiden, ob der Einsatz von Selenium notwendig. Daher habe ich explizit nach Hinweisen auf JavaScript gesucht. Der Quelltext enthielt unter anderem einen expliziten Hinweis: "Enable JavaScript and cookies to continue". Dies bestätigte mir, dass es sich um eine dynamisch ladende Webseite handelt. BeautifulSoup ist hierfür ungeeignet, da man damit kein JavaScript auszuführen kann. Zudem werden nur eine bestimmte Anzahl an Ergebnisse pro Seite angezeigt, weshalb Klicken erforderlich ist. Daher war der Einsatz von Selenium notwendig.

Beim Öffnen der Webseite wurde ich von einer Sicherheitsmassnahme von Cloudflare aufgehalten. Eine der Sicherheitsfunktionen von Cloudflare ist die sogenannte "JavaScript Challenge", die ich auch im Quelltext gefunden habe. Cloudflare verwendet diese Technik, um zu überprüfen, ob der/die Besucher*in tatsächlich ein/e Nutzer*in ist und nicht ein automatisiertes Skript. Dies stellte zunächst ein Problem dar, denn ein automatisiertes Skript war das Ziel und Cloudflare hat den automatisierten Zugriff durch Selenium erkannt und blockiert. Die Lösung bestand schliesslich darin, Selenium in Verbindung mit `undetected_chromedriver` zu verwenden, eine Erweiterung des Standard-ChromeDrivers. `undetected_chromedriver` ist dafür konzipiert, automatisierte Browser-Interaktionen weniger erkennbar zu machen. So konnte ich Cloudflare umgehen und Zugriff erhalten.

In einem ersten Ansatz habe ich zunächst den gesamten hinterlegten HTML-Code pro Angebot mit Selenium gescrapt und anschliessend mit BeautifulSoup die einzelnen Elemente in Spalten sortiert. Im finalen Ansatz schliesslich habe ich den Prozess kombiniert: Für jede Seite wurde die URL dynamisch generiert, indem die aktuelle Seitennummer in eine vordefinierte URL-Vorlage eingefügt wurde. Auf jeder Seite wurden alle Angebote identifiziert mittels XPath. Für jedes Angebot wurden die in Elementen gespeicherten Informationen mit BeautifulSoup und XPath extrahiert. Falls eine Seite keine Angebote enthielt, wurde der Prozess beendet. Zum Schluss wurden die gesammelten Daten in einen Pandas DataFrame überführt.

Das Scraping fand am 03.12.2023 stand und es wurden 2656 Angebote gescrapt. Der Code ist in der Datei «tutti_scraping». Ich habe alle pro Angebot verfügbaren Informationen gescrapt, auch wenn letztlich nicht alle nötig waren:

- Anzahl an hochgeladenen Fotos
- Details zu Ort und Zeit des Hochladens des Angebots
- Titel des Angebots
- Beschreibung des Angebots
- Preis des Angebots

4.1.2 Transformieren

Im nächsten Schritt wurden dann die gescrapteten Daten transformiert. Der Code ist in der Datei «tutti_transforming». Zentral war hierbei, die relevanten Informationen aus dem Titel sowie aus der Beschreibung zu extrahieren. Als eher einfache Schritte habe ich zunächst das Datum und den Ort in einzelne Spalten extrahiert, da diese zur Kontrolle und Behandlung von Duplikaten relevant waren. Das älteste Angebot stammt vom 04.10.2023, was darauf hindeutet, dass die Daten sehr aktuell sind. Beim Preis habe ich die Zahlen extrahiert und jeweils zum Datentyp «float» umgewandelt. Die Anzahl an hochgeladenen Fotos war nicht weiter relevant und wurde darum entfernt.

Anschliessend habe ich den Datensatz für die Extraktion der relevanten Spalten vorbereitet, das sind das Modell, der Speicher in GB, die Farbe des iPhones, die Akkukapazität sowie der optische Zustand. Dafür habe ich den Titel und die Beschreibung normalisiert durch Umwandlung in Kleinbuchstaben. Anschliessend habe ich aus den Zeilen, in denen das Stichwort «iphone» nicht austaucht, Wörterbücher generiert. Mithilfe des ersten werden Angebote ausgeschlossen, die Smartphones anderer Marken anbieten. Durch das zweite wurden die Schreibvariationen von «iphone» vereinheitlicht.

Danach habe ich mit einer Stichwortsuche die Angebote auf Hinweise von Defekten untersucht, die die Kaufentscheidung sehr wahrscheinlich beeinflussen (beispielsweise eine kaputte Gesichtserkennung). Basierend darauf habe ich ein weiteres Wörterbuch erstellt und Angebote mit solch schwerwiegenden Defekten ausgeschlossen. Anschliessend habe ich den Datensatz auf Shops geprüft, da wir nur private Angebote scrapen wollten, sowie auf mehrfache Einträge, da diese das Ergebnis wie Durchschnittspreise verzerren würden. Dafür habe ich nach gleichen Anfängen bzw. Enden beim Titel und der Beschreibung von Angeboten gesucht und diese dann manuell geprüft. Ebenso habe ich Sets, Duplikate und nicht-deutschsprachige Angebote ausgeschlossen. Als finaler Schritt der Vorbereitung habe ich mithilfe eines

Wörterbuchs Angebote zu für uns irrelevanten iPhone-Modellen ausgeschlossen und mit einem weiteren Wörterbuch die relevanten Modelle hinsichtlich der Schreibweise in Titel und Beschreibung standardisiert. Herausfordernd war, die Abhängigkeiten zu bedenken, sodass durch den Ausschluss des iPhones X (zu alt) nicht auch die relevanten Modell iPhone XS aus Versehen ausgeschlossen wurden. War kein Modell zuordenbar, wurde das Angebot entfernt.

Die tatsächliche Extraktion habe ich schrittweise vorgenommen (Modell, Speicher, Farbe, optischer Zustand, Akkukapazität) und mir vorab Kontrollschleifen überlegt, die ich beispielhaft an der Modellextraktion erklären möchte. Als erstes extrahiere ich die Modellnamen aus dem Titel und aus der Beschreibung in jeweils eigene Spalten, die nur zeitweise benötigt werden. Dann vergleiche ich die Werte dieser beiden Spalten miteinander. Sollte nur der Titel oder die Beschreibung einen Wert haben oder der Wert in beiden Spalten übereinstimmen, so wird dieser in die finale Spalte übernommen. Widersprechen sich die Werte, wird der Wert «Exemption» übernommen und muss manuell geprüft werden. Ebenso prüfe ich Angebote, bei denen weder im Titel noch in der Beschreibung ein Modellname gefunden wurde (beim Modell ist dies irrelevant, weil diese bereits ausgeschlossen wurden). Als letztes prüfe ich Angebote, bei denen in der finalen Spalte zwei Werte stehen (beispielsweise, weil zwei Modellnamen gefunden wurden).

Beim Modell und Speicher bestand die Besonderheit darin, dass diese Angaben vorhanden sein mussten, sonst wurden die Angebote ausgeschlossen. Bei der Farbextraktion habe ich mich dafür entschieden, ein modellspezifisches Wörterbuch für die möglichen Farbvariationen zu erstellen. Konnte keine Farbe gefunden werden wurde «Unbekannt» eingesetzt. Beim optischen Zustand bestand die Herausforderung darin, ein spezifisches Wörterbuch zu erstellen, dass nicht aus Versehen Angeboten den Zustand «Ungeöffnet» zuordnet, bei denen nur ein ungeöffnetes Ladekabel dabei ist. Zudem mussten die verschiedenen Zustandsbeschreibungen in die drei vorab festgelegten Kategorien eingeordnet werden. Bei der Akkukapazität konnte ich mit regulären Ausdrücken arbeiten und anschliessend die im Codebuch festgelegte Dichotomisierung in «Neu» (> 97%) und «Gebraucht» vornehmen. Bei iPhones mit optischem Zustand «Ungeöffnet» wurde «Neu» als Akkukapazität zugeordnet. War beim optischen Zustand oder bei der Akkukapazität kein Wert zu finden, wurde der schlechtmöglichste Zustand angegeben.

Anschliessend habe ich die einzelnen Modell-Speicherkombinationen (z.B. iPhone 13 Pro Max mit 128 GB) auf preisliche Ausreisser überprüft. Ich habe mich entgegen dem ursprünglichen Plan gegen die Verwendung von Rosner's Test entschieden, weil die statistische Signifikanz inhaltlich keine Relevanz für die Beurteilung der Ausreisser hatte. Die grafische Analyse mittels Boxplots diente daher zur Identifikation für eine manuelle Prüfung. Sollten jedoch keine Auffälligkeiten im Angebot sein, wurden die Ausreisser nicht ausgeschlossen. Diese Entscheidung war relevant für die Datenanreicherung. Wie meine beiden Teamkolleginnen habe auch ich den Datensatz um den Durchschnittspreis ohne Berücksichtigung der Farbe sowie unter Berücksichtigung der Farbe erweitert und anschliessend die Abweichung dieser beiden Werte berechnet. Aufgrund der Ausreisser habe ich jedoch den Median anstelle des Mittelwerts verwendet. Der finale Datensatz umfasste 1539 Angebote. Ich habe mich gegen die geplante Erweiterung durch Umrechnung von CHF in Euro entschieden, da es doch keine weitere Relevanz hatte und selbst wenn, beim zusammengeführten Datensatz sinnvoller gewesen wäre. Dann hab ich die Spalte Datenquelle hinzugefügt, die für alle Zeilen den Wert «tutti» enthält. Abschliessend habe ich den Datensatz aggregiert, sodass alle Kombinationen (Modell x Speicher in GB x Farbe x Akkukapazität x Optischer Zustand) einmal vorhanden sind (mit Angabe der zugrundeliegenden Anzahl an Beobachtungen). Der finale Datensatz für die Zusammenführung bestand aus 696 Zeilen.

Es ist wichtig zu erwähnen, dass die Datenqualität trotz sorgfältiger und teilweise manueller Aufbereitung kritisch zu betrachten ist. Erstens ist die Zustandszuordnung problematisch, da unterschiedlich empfunden wird, was beispielsweise "wie neu" bedeutet. Diese Subjektivität erschwerte eine einheitliche Kategorisierung und kann die Vergleichbarkeit der Daten beeinträchtigen. Zweitens könnte der Preis in einigen Fällen auch Zubehör oder zusätzliche Leistungen umfassen. Solche Variationen können zu Preisschwankungen führen, die nicht allein auf das iPhone-Modell oder seine Eigenschaften zurückzuführen sind. Zudem kann es auch nach wie vor irreführende oder betrügerische Angebote geben, die die Analyse verfälschen. Es war nicht möglich, jedes einzelnen Angebots dahingehend zu prüfen. Aufgrund der Komplexität und Variabilität der Angebotsdaten besteht außerdem die Möglichkeit, dass während des Transformationsprozesses Fehler aufgetreten sind wie falsch zugeordnete Daten, ausgelassene relevante Angebote oder inkorrekt extrahierte Informationen. All dies kann die Aussagekraft und Zuverlässigkeit unserer Analyse beeinflussen.

4.1.3 Laden

Abschliessend habe ich den Datensatz als Nutzerin «cip_annalena» auf die MariaDB hochgeladen, spezifisch in eine zuvor angelegte Datenbankstruktur «db_tutti» in die Tabelle «Tabelle_tutti». Anschliessend habe

ich mit Abfragen überprüft, ob die Anzahl an Zeilen übereinstimmt und mir die ersten Zeilen angeschaut, um den Erfolg des Ladevorgangs zu prüfen. Der Code für das Hochladen und Prüfen ist in dem Notebook «tutti_load_db».

4.2 Revendo

In diesem Abschnitt wird der ETL-Prozess von Revendo detailliert beschrieben. Dabei werden die einzelnen Teilschritte des ETL-Prozesses eingehend erläutert. Für jeden Teilschritt werden entsprechende Python-Skripte im Anhang bereitgestellt, um eine umfassende und detaillierte Erklärung zu gewährleisten. Michelle Peter ist für die Ausarbeitung und Implementierung dieses Abschnitts verantwortlich.

4.2.1 Extrahieren

Nach mehreren Anläufen wurde schliesslich am 6. Dezember 2023 die Daten von revendo.com mit folgenden Informationen gescrapt:

- Produktname: Apple iPhone XR 64 GB in Schwarz
- Optischer Zustand: Schnäppchen, Gebrauchte, Sehr gut, Wie neu, Neu
- Akkuzustand: Gebrauchte, Neu
- Produktdetailpreis: CHF 304

Ursprünglich war geplant, auch die Kamerainformationen zu scrapen. Aufgrund grösserer Herausforderungen beim Scrapen der Webseite wurde jedoch auf diese Information verzichtet. Dies ist für unsere Fragestellung ebenfalls nicht von Bedeutung.

Die Webseite von Revendo ist dynamisch gestaltet, mit BeautifulSoup konnten nur jeweils die Anfangsseite gescrapt werden aber nicht die weiteren. Weshalb zum Scrapen die beiden Pakete BeautifulSoup und Selenium verwendet wurden. Eine grosse Herausforderung war das Scrapen der verschiedenen Zustände (Optisch und Akku). Während des Scraping-Prozesses wurde klar, dass es nicht einfach ist, die diversen Zustandskombinationen und die zugehörigen Preise zu erfassen. Ursprünglich sollte auf jedes einzelne Modell über die Produkt-URL zugegriffen und mit verschiedenen Befehlen die einzelnen Zustände pro Angebot abgefragt werden. Es stellte sich jedoch heraus, dass die Produktseiten teilweise fehlerhafte Verfügbarkeitsangaben aufwiesen. Obwohl ein Button für den optischen Zustand (z.B. Schnäppchen) anklickbar war, gab es keine Produktverfügbarkeit. Selenium hatte Schwierigkeiten, die Preise korrekt zu scrapen. Zudem benötigte das Scrapen in dieser Variante für 270 Produkte über drei Stunden. Daher wurde die Vorgehensweise geändert und stattdessen auf der Produktübersichtsseite gescrapt. Mittels der Filter 'Optischer Zustand' und 'Akkuzustand' konnten so die jeweiligen Kombinationen gefiltert werden. Daraus ergaben sich 9 verschiedene URL's zum scrapen. Je nach Kombination waren die Anzahl Seiten unterschiedlich, Selenium hat jeweils so lange gescrapt bis kein Angebot mehr auffindbar war. Allerdings wurde zu spät bemerkt, dass es sich bei der Produktübersicht Seite jeweils um ab-Preise handeln, sprich um die niedrigsten Preise handelt. Selbst wenn der Filter «Wie neu» beim optischen Zustand und «Neu» beim Akkuzustand eingestellt wurde, zeigte es, falls vorhanden, den günstigsten Preis der niedrigsten Kombination an.

Da wir eine dynamische Anzahl von Seiten scrapen mussten, wurde der Befehl so eingestellt, dass sich die Seitenzahl je nach Filter automatisch anpasst. Mit dem DevTool in Chrome wurden die passenden Selektoren ausgewählt. Obwohl die Preise nicht präzise gescrapt werden konnten, war die Abfrage insgesamt deutlich schneller als der erste Ansatz und ist im Detail in der Datei «01_extract_revendo.ipynb» einsehbar.

4.2.2 Transformieren

Nachdem die relevanten Informationen erfasst wurden, folgte der Prozess, die Daten in das gewünschte Format zu transformieren. Dabei wurde mit den nachstehenden Variablen im Datensatz gearbeitet. Zwischen diesen Variablen und jenen, die im Prozessplan vordefiniert wurden, besteht eine Diskrepanz. Da die Kamera, Bildschirmgrösse sowie das Betriebssystem für die Fragestellung nicht relevant waren, wurden diese beim Scraping nicht mehr berücksichtigt.

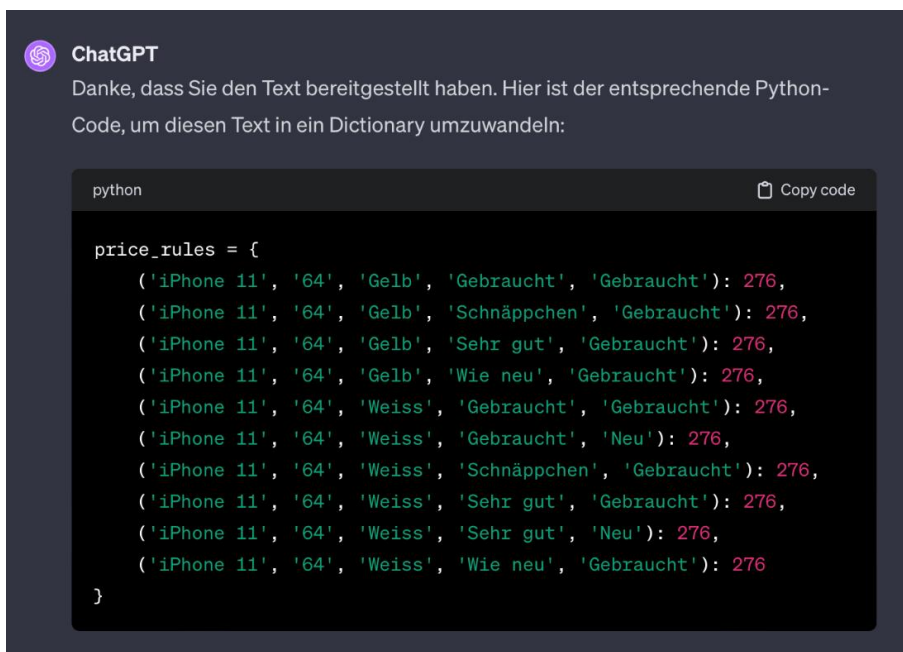
Informationen auf der Webseite (Extract)	Zielvariablen in der Datenbank (Transform)	Beispiel
Produktname	Modell	iPhone 11
	Speicher in GB	512

	Farbe	Schwarz
Optischer Zustand	Optischer Zustand	Schnäppchen
Akkuzustand	Akkukapazität	Gebraucht
Produkt Detail Preis	Preis in CHF	367

Tabelle 1 - Datenstruktur revendo.ch

Dabei wurde als eines der ersten Schritte, die Spaltennamen korrekt umbenannt und der Datensatz nach Modell sowie Preis sortiert – dies erleichterte die Übersicht und half bei der Anpassung der Preise. Die grösste Herausforderung bei der Datenbereinigung bestand darin, die korrekten Preise zu ermitteln und im Datensatz zu aktualisieren, da sie, wie bereits erwähnt, nicht immer zuverlässig gescrapet wurden. Zur Lösung dieses Problems wurde ein umfangreiches Dictionary angelegt, dass jedes Modell detailliert auflistet. Jeder Schlüssel im Dictionary repräsentierte eine einzigartige Kombination aus Modell, Speichergrösse in GB, Farbe, optischem Zustand und Akkukapazität. Der zugehörige Wert wurde durch eine sorgfältige manuelle Überprüfung auf der Revendo-Webseite festgelegt. Ein Beispiel hierfür ist: ('iPhone 15 Plus', '512', 'Schwarz', 'Neu', 'Neu'): 1254. Diese zeitaufwendige Aufgabe erforderte die Überprüfung und Anpassung von insgesamt 875 Datensatzzeilen.

Um diesen Prozess effizienter zu gestalten, wurde mit ChatGPT zusammengearbeitet. Dies ermöglichte es, jedes Modell schnell und korrekt ins vorgesehene Format zu bringen. Durch diese Kooperation konnte die Datenqualität erheblich gesteigert und der Zeitaufwand minimiert werden.



The screenshot shows a ChatGPT interface with a message from the user asking for a Python code to convert text into a dictionary. The ChatGPT response provides a Python dictionary named 'price_rules' with 12 entries. Each entry is a tuple of five strings representing product details, followed by a price value. The entries are as follows:

```
python
price_rules = {
    ('iPhone 11', '64', 'Gelb', 'Gebraucht', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Gelb', 'Schnäppchen', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Gelb', 'Sehr gut', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Gelb', 'Wie neu', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Weiss', 'Gebraucht', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Weiss', 'Gebraucht', 'Neu'): 276,
    ('iPhone 11', '64', 'Weiss', 'Schnäppchen', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Weiss', 'Sehr gut', 'Gebraucht'): 276,
    ('iPhone 11', '64', 'Weiss', 'Sehr gut', 'Neu'): 276,
    ('iPhone 11', '64', 'Weiss', 'Wie neu', 'Gebraucht'): 276
}
```

Abbildung 2 - Output von ChatGPT mit dem Dictionary Format

Für das Modell und die Farbe wurde derselbe Ansatz verfolgt, bei dem zunächst die korrekte Schreibweise gemäss des Codebuchs gesucht wurde. Anschliessend wurde der korrekte Modellname oder die entsprechende Farbe eingefügt. Beim Speicher wurde mittels Regex nach einer Zahl gefolgt von «TB» oder «GB» gesucht und anschliessend diese entsprechend ersetzt.

Anschliessend wurden die optischen Zustände gemäss des Codebuchs vereinheitlicht. Auch hier kam ein Dictionary zum Einsatz, dass die Werte zusammenfasst und unterschiedliche Schreibweisen der Zustände berücksichtigt. Nun zeigt der Datensatz für jedes Modell mehrere identische Zustandskombinationen auf, allerdings mit jeweils unterschiedlichen Preisen.

Bevor es zur Bereicherung mit neuen Daten übergang, wurde eine Überprüfung auf Duplikate durchgeführt. Es zeigte sich, dass bei mehreren Modellen unter der Farbangabe «NaN» stand. Da bei der Kontrolle nicht nachvollzogen werden konnte, um welche Farbe es sich handelt, wurden die Duplikate entfernt. Der vollständige Bereinigungsprozess kann in im File «02_transform_part1_revendo.ipynb» gefunden werden.

Zum Abschluss des Datenverarbeitungsprozesses erstellen wir eine aggregierte Darstellung der Daten, die in der Datei «03_transform_part2_revendo.ipynb» gesichert wird. In diesem Schritt ergänzen wir den Datensatz gezielt um spezifische Zielvariablen, wie Datenquellen und führen wichtige Berechnungen, wie Durchschnittlicher Preis des Modells, durch. Die zusätzlichen Berechnungen zielen darauf ab, uns bei der Beantwortung unserer Fragestellung zu unterstützen, mit Ausnahme derjenigen, die die Mehrwertsteuer betreffen:

- Datenquelle (für diesen Datensatz immer *Revendo*)
- Berechnung: Preis ohne Mehrwertsteuer
- Berechnung: Durchschnittlicher Preis des Modells (basierend auf Speicher & Modell)
- Berechnung: Durchschnittlicher Preis des Modells (basierend auf Speicher, Modell & Farbe)
- Berechnung: Preis-Abweichung vom Durchschnittlichen Preis ohne Farbe

4.2.3 Laden

Nach der Bereinigung des Datensatzes, wurde der Datensatz auf MariaDB übertragen. Zuerst wird eine neue Benutzerin namens «cip_mp» angelegt. Gleichzeitig wird eine neue Datenbank mit dem Namen «iPhoneRevendo» erstellt, für die entsprechende Zugriffsrechte vergeben werden. Die Details zur Erstellung dieser Datenbank sind in der Text-Datei «load_mariadb» festgehalten. Anschliessend erfolgt der Hochladevorgang der Daten mittels eines Python-Skripts, dokumentiert in «04_loading_mariadb.ipynb». Dieses Skript leitet den gesamten Prozess der Datenübertragung und -integration in MariaDB, wobei sämtliche Schritte und Methoden ausführlich beschrieben sind.

4.3 Preisvergleich

In diesem Kapitel wird erklärt, wie die Daten von der Quelle preisvergleich.ch extrahiert, transformiert und geladen wurden. Der Code dazu ist in der Beilage zum Projekt, in den jeweiligen Subkapiteln wird jeweils das passende File angegeben. Dieses Kapitel wurde durch Mélanie Bigler bearbeitet und ausgeführt.

4.3.1 Extrahieren

Die Daten von Preisvergleich.ch wurden am 26. November 2023 gescraped. Es wurden 353 Einträge extrahiert mit den folgenden Informationen:

- Produkttitel – Beispiel: *Apple - iPhone 11 (128 GB, 6.1", 12 MP, Schwarz)*
- Unser Fazit – Beispiel: *5.5*
- Durchschnittliche Bewertung – Beispiel: *4.6*
- Anzahl Bewertungen – Beispiel: *3716*
- Anbieter - Beispiel: *microspot.ch*
- Preis – Beispiel: *CHF 499.00*

Bemerkung: Nicht alle iPhone Modelle haben eine Durchschnittliche Bewertung und dazu die Anzahl Bewertungen. Da dies aber nicht von direkter Relevanz ist für unsere Fragestellungen, werden fehlende Daten mit «None» hinterlegt.

Die Webseite ist nicht dynamisch, respektive alle die oben genannten Informationen können über den Seitenquelltext gefunden werden. Dazu wird das Python Package *Beautiful Soup* genutzt. Für jedes Informationsteil wurde der passende Selector über das DevTool gesucht. Anschliessend musste ebenfalls eine Regel eingeführt werden, welche die nächste Seite aufruft, da die 353 iPhone Einträge über 5 Seiten verteilt waren.

Schlussendlich können alle Informationen ohne grosse Probleme eingelesen werden. Der detaillierte Code ist im Skript mit dem Namen «01_extract_scraping-preisvergleich.py» einsehbar.

4.3.2 Transformieren

Nachdem alle relevanten Informationen eingelesen wurden, geht es darum die Daten in die richtige Form zu bringen. Dies geschieht mit der folgenden Logik:

Informationen der Webseite (Extract)	Zielvariable in der Datenbank (Transform)	Beispiel
Produkttitel	Modell	iPhone 13

	Farbe	Blau
	Speicher in GB	512
Preis	Preis in CHF	1200.00

Tabelle 2 - Datenstruktur preisvergleich.ch

Wir haben uns bewusst bei diesem Schritt nur auf die relevanten Infos für diese Arbeit fokussiert (Modell, Farbe, Speicher in GB und Preis in CHF). Der schwierigste Part waren aus dem Produkttitel, die richtigen Infos zu erhalten. Hier die wichtigsten Schritte:

1. *Preis*: Anpassung, damit nur die Zahl am Ende in der Tabelle steht
2. *Modell*: Richtige Schreibweise wie im Codebuch definiert, dazu wurde der Produkttitel nach den iPhone Modell Namen durchsucht, bei Treffern wurde der korrekte Modell Name eingefügt.
3. *Farbe*: Das selbe wurde bei der Farbe gemacht. Der Produkttitel wurde mit der Farben Konventionen im Codebuch verglichen und bei einem Treffer zum richtigen Namen ersetzt.
4. *Speicher*: Hier wurde mithilfe von Regex die Nummern gefolgt von TB oder GB gesucht und mit diesen Zahlen ersetzt.

Nach diesem Vorgang wurden die Spalten *Preis* und *Speicher in GB* in Zahlen umgewandelt.

Anschliessend wurde überprüft, welche Reihen NAs ausweisen. Es waren insgesamt 41 Zeilen, besonders im Bereich der Modelle gab es Fehler. Dies aus dem Grund das einige Modelle zu alt waren (vor 2018 auf den Markt gekommen), sowie dass die iPhone SE Modellnamen nicht erkannt wurden. Daher wurde eine Liste von diesen Zeilen extrahiert um sie qualitativ zu überprüfen und anschliessend manuell anzupassen. Alle zu alten iPhone Modelle sowie ein Eintrag mit fehlenden Informationen wurden gelöscht.

Ein weiterer Schritt in der Datenqualitätsprüfung ist das Finden von möglichen Ausreissern (Outliers). Dazu wird ein Boxplot gemacht, um zu sehen, ob es grosse Unterschiede gibt. Dies wurde mithilfe von Boxplots für die numerischen Spalten Preis und Speicher in GB überprüft.

Im Boxplot sind keine Ausreisser zu erkennen, daher setzen wir die Arbeit mit diesen Daten fort.

Da die Daten auf dieser Plattform sehr sauber sind, werden diese extra verunreinigt. Dazu wird ChatGPT verwendet. Es sollen 10% der Beobachtungen gezielt verunreinigt werden, damit diese anschliessend unter Verwendung von Pandas richtiggestellt werden (Prompt für ChatGPT kann in der nächsten Abbildung nachvollzogen werden). Daher wird zu diesem Zeitpunkt ein CSV generiert. Bis hier kann der Code [*02_transforming_preisvergleich_part1.ipynb*](#) nachvollzogen werden.

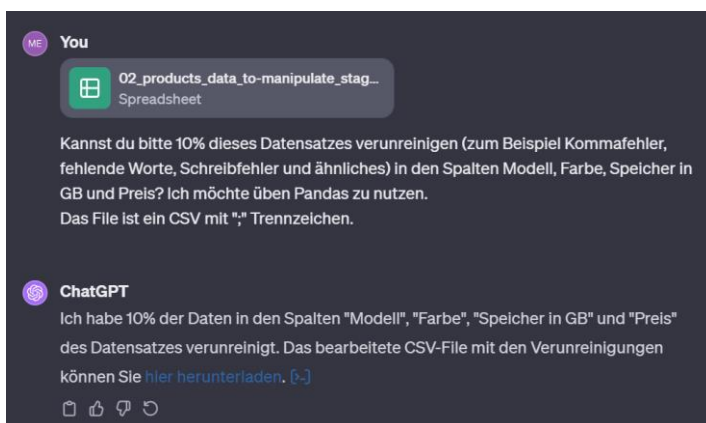


Abbildung 3 - Prompt an ChatGPT zur Datenverunreinigung

Der verunreinigte Datensatz von ChatGPT wurde im File «*03_transforming_preisvergleich_part2.ipynb*» weiterverarbeitet. Dazu wurden Schreibfehler manuell bereinigt. Beim Bereinigen konnte leider die Spalte mit dem Preis schwer bereinigt werden, da die Daten nicht auffällig verändert wurde, es gab auch keine fehlenden Werte. Daher wurden die alten Werte wieder hinzugefügt, um möglichst genaue Daten zu haben.

Nachdem die Daten wieder bereinigt waren, wurden die folgenden Spalten hinzugefügt und/oder berechnet:

- 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 Speicher & Modell)
- Berechnung: Durchschnittlicher Preis des Modells (basierend auf Speicher, Modell & Farbe)
- Berechnung: Preis-Abweichung vom Durchschnittlichen Preis ohne Farbe

Schlussendlich wird ein aggregiertes Dataframe erstellt, das dann mit den anderen Datenquellen zusammengeführt werden soll:

Modell	Speicher (in GB)	Farbe	Datenquelle	Optischer Zustand	Akkukapazität	Durchschnittspreis mit Farbe (CHF)	Durchschnittspreis ohne Farbe (CHF)	Abweichung Durchschnittspreis ohne Farbe

4.3.3 Laden

Der finale Datensatz «04_preisvergleich_daten_final_stage3.csv» wurde nun auf die MariaDB hochgeladen. Dazu wird zuerst eine neue Benutzer*in namens «cip_mb» eröffnet sowie eine neue Datenbank namens *iPhonePreisvergleich* erstellt und die Berechtigung vergeben. Details zum Aufsetzen können im File «04_Load_create-db.txt» nachgeschaut werden. Anschliessend werden die Daten hochgeladen via Python Code. Dieser ist in der Datei «05_load_mariaDB.ipynb» einsehbar. Die Daten können nun von dieser Datenbank bezogen werden.

4.4 Daten Zusammenführung

Alle Datensätze sind nun auf den einzelnen MariaDBs der Autorinnen. Um die Fragestellungen beantworten zu können, sollen nun alle Daten zusammengeführt werden und ebenfalls auf eine MariaDB geladen werden. Die Zusammenführung wurde durch Studentin C, Mélanie Bigler, durchgeführt, da sie den einfachsten Datensatz verarbeitet hatte. Der Code «06_data-merge.ipynb» liegt daher im Ordner der Studentin C und im Unterordner «data/data merge» können alle Datensätze für die Zusammenführung gefunden werden.

4.4.1 Transform: Finaler Datensatz aufbereiten

Nachdem alle Daten der drei Quellen geladen sind, wurden diese mithilfe des *pandas* Befehl *concat* in eine Tabelle zusammengeführt. Da die Tabellen genau gleich aufgebaut wurden, stellte dies kein Problem dar. Nach der Zusammenführung sind 1512 Reihen Daten verfügbar. Es wurde ein kurzer Check zu möglichen Nas gemacht, welcher keine Auffälligkeiten zeigt. Die Tabelle wird als «05_full_data_iPhonePreise_merged.csv» extrahiert. Zur Beantwortung der Unterfragen sollen zudem Visualisierungen und Modelle eingesetzt werden. Für deren Erstellung ist eine Konvertierung der kategorialen Werte in numerische Werte erforderlich. Hierfür kamen Methoden des *sklearn*-Pakets zum Einsatz, insbesondere der «LabelEncoder» und der «OrdinalEncoder». Das Ergebnis ist der numerisch encodierte Datensatz «05_full_data_iPhonePreise_encoded_merged.csv» sowie JSON-Dateien, die aufzeigen, welche Zahl welcher Kategorie oder welchem Namen entspricht (also Legenden).

4.4.2 Load: Upload auf MariaDB

Schlussendlich werden beide Datensätze als separate Tabelle auf die MariaDB geladen. Hier wurde erneut der Nutzernamen *cip_mb* und die Datenbank *iPhonePreisvergleich* von Mélanie Bigler genutzt. Der kodierte Datensatz ist in der Tabelle namens *TabelleiPhonePreiseEncoded* geladen und der volle Datensatz ohne Kodierung ist mit dem Namen *TabelleiPhonePreiseFull* geladen.

5 Resultate

Nachdem die Daten nun bereit sind, um verarbeitet zu werden, möchten wir die im Kapitel 1.1 aufgeführten Fragestellungen beantworten. Die nachfolgenden Auswertungen, in Form von Plots und Screenshots, können im Jupyter Notebook «07_data-analysis.ipynb» eingesehen und reproduziert werden.

5.1 Hauptfragestellung

Welche Plattform bietet das gewünschte iPhone-Modell in welchem Zustand zu welchem Preis an? Wie gross sind Preisunterschiede zwischen Farben?

Um zu bestimmen, welche Plattform ein gewünschtes iPhone-Modell in welchem Zustand und zu welchem Preis anbietet und wie sich die Preise zwischen verschiedenen Farben unterscheiden, nutzen wir den Datensatz «full_data_iPhonePreise.csv». Als Antwort auf die Frage dient ein interaktives Dashboard, das Nutzer*innen ermöglicht, das Modell und eine Speichergrösse auszuwählen (siehe Screenshot, Beispiel für iPhone 12 mit 256 GB). Sie erhalten dann vier Ergebnisse

- Tabelle zu allen Farb- und Zustandskombinationen auf allen Plattformen, absteigend sortiert nach Durchschnittspreis ohne Berücksichtigung der Farbe
- Tabelle zu allen Zustandskombinationen auf alle Plattformen ohne Farbvariationen, absteigend sortiert nach Durchschnittspreise ohne Berücksichtigung der Farbe
- Boxplot zu den Durchschnittspreisen auf den drei Plattformen über alle Zustände und Farben hinweg
- Boxplot zu den Preisvariationen hinsichtlich der Farben über alle Zustände und Plattformen hinweg

The screenshot shows an interactive dashboard. On the left, there is a dropdown menu for 'Modell:' with a list of iPhone models including iPhone 11, iPhone 11 Pro, iPhone 11 Pro Max, iPhone 12, iPhone 12 Mini, iPhone 12 Pro, iPhone 12 Pro Max, iPhone 13, iPhone 13 Mini, iPhone 13 Pro, iPhone 13 Pro Max, iPhone 14, iPhone 14 Plus, iPhone 14 Pro, iPhone 14 Pro Max, iPhone 15, iPhone 15 Plus, iPhone 15 Pro, iPhone 15 Pro Max, and iPhone SE (2020). Below the dropdown, there is a table with columns: 'Modell', 'Speicher', 'Update On', 'Datenquelle', 'Optischer Zustand', 'Akkukapazität', 'Durchschnittspreis mit Farbe', 'Anzahl der Beobachtungen mit Farbe', 'Durchschnittspreis ohne Farbe', and 'Anzahl der Beobachtungen ohne Farbe'. The table contains data for various iPhone models and their prices.

Abbildung 4 – Screenshot interaktives Dashboard

Wir möchten die Ergebnisse am Beispiel des iPhone 12 mit 256 darstellen. Die tabellarischen Übersichten sind besonders interessant für Personen, die zwischen einem neuen oder gebrauchten Kauf schwanken. Abbildung 5 zeigt die erste Tabelle, die ausgegeben wird (sie berücksichtigt Farben).

	Modell	Speicher in GB	Farbe	Datenquelle	Optischer Zustand	Akkukapazität	Durchschnittspreis mit Farbe	Anzahl der Beobachtungen mit Farbe	Durchschnittspreis ohne Farbe	Anzahl der Beobachtungen ohne Farbe
33	iPhone 12	256	Grün	Preisvergleich	Ungeöffnet	Neu	998.00	1	768.491667	
35	iPhone 12	256	Weiss	Preisvergleich	Ungeöffnet	Neu	878.95	1	768.491667	
34	iPhone 12	256	Schwarz	Preisvergleich	Ungeöffnet	Neu	709.00	1	768.491667	
32	iPhone 12	256	Blau	Preisvergleich	Ungeöffnet	Neu	694.00	2	768.491667	
31	iPhone 12	256 (PRODUCT)RED	(PRODUCT)RED	Preisvergleich	Ungeöffnet	Neu	637.00	1	768.491667	
393	iPhone 12	256 (PRODUCT)RED	(PRODUCT)RED	Revendo	Sehr gut	Neu	527.00	2	527.000000	
388	iPhone 12	256 (PRODUCT)RED	(PRODUCT)RED	Revendo	Sehr gut	Gebraucht	506.00	2	490.000000	
392	iPhone 12	256	Weiss	Revendo	Sehr gut	Gebraucht	506.00	2	490.000000	
946	iPhone 12	256	Unbekannt	tutti	Ungeöffnet	Neu	500.00	1	500.000000	

Abbildung 5 - Filterung nach Modell und Speichergrösse (Ausschnitt der Tabelle)

Aus dieser Tabelle lassen sich folgende Erkenntnisse gewinnen:

- Preisvarianz nach Farbe und Zustand: Es gibt deutliche Preisunterschiede zwischen den Farben und Zuständen. Grüne iPhone 12 sind am teuersten, während (PRODUCT)RED günstiger ist.
- Einfluss der Datenquelle: Die Preise variieren je nach Plattform, wobei Angebote auf Preisvergleich.ch tendenziell teurer sind als auf Revendo oder tutti.ch.
- Bedeutung von optischem Zustand und Akkukapazität: Diese Faktoren beeinflussen den Preis wesentlich. Ungeöffnete Geräte und solche mit neuem Akku sind teurer.

Die zweite Tabelle (siehe Abbildung 6) vereinfacht die Ansicht durch Ignorieren der Farben. Hierbei fällt auf, dass ungeöffnete iPhones mit neuem Akku auf tutti.ch deutlich günstiger angeboten werden als auf Preisvergleich.ch, obwohl dort die Anzahl Beobachtungen grösser ist. Somit kann es sich lohnen, auf dem Gebrauchtwarenmarkt nach ungeöffneten iPhones zu suchen und so zu sparen. Limitierend ist, dass den Kombinationen vor allem auf tutti.ch zum Teil nur eine Beobachtung zugrunde liegt.

index	Modell	Speicher in GB	Akkukapazität	Optischer Zustand	Datenquelle	Durchschnittspreis ohne Farbe	Anzahl der Beobachtungen ohne Farbe
7	7 iPhone 12	256	Neu	Ungeöffnet	Preisvergleich	768.49	30
5	5 iPhone 12	256	Neu	Sehr gut	Revido	527.00	2
8	8 iPhone 12	256	Neu	Ungeöffnet	tutti	500.00	1
2	2 iPhone 12	256	Gebraucht	Sehr gut	Revido	490.00	45
4	4 iPhone 12	256	Neu	Gebraucht	Revido	455.00	9
6	6 iPhone 12	256	Neu	Sehr gut	tutti	455.00	1
1	1 iPhone 12	256	Gebraucht	Gebraucht	tutti	425.00	8
0	0 iPhone 12	256	Gebraucht	Gebraucht	Revido	422.67	24
3	3 iPhone 12	256	Gebraucht	Sehr gut	tutti	350.00	1

Abbildung 6 - Kumulierte Übersicht zu den iPhone 12 256GB Preisen pro Plattform

Die Muster zeigen sich auch bei anderen Modellen (stichprobenartige Analyse). Für die Beantwortung unserer Frage bedeutet es, dass die Plattformwahl einen erheblichen Einfluss auf den Preis hat, insbesondere wenn es um den optischen Zustand des Geräts geht. Während Plattformen wie Preisvergleich.ch tendenziell höhere Preise für ungeöffnete Geräte haben, können Plattformen wie tutti.ch günstigere Angebote bieten, allerdings mit geringerer Verfügbarkeit und möglicherweise weniger Preisstabilität. Käufer*innen sollten daher verschiedene Plattformen vergleichen, um das beste Angebot zu finden. Es gibt iPhones keine «goldene Regel», wo man am besten das iPhone kaufen kann, da es vor allem auf die persönlichen Kaufkriterien ankommt. Daher kann diese Übersicht einen Anhaltspunkt bieten.

Das dritte Ergebnis, das Nutzer*innen erhalten, ist ein Boxplot, der die Preisspannen auf den Plattformen zeigen über alle Zustände hinweg. Dieser soll einen Anhaltspunkt geben, wie unterschiedlich die Preise sind, um einen Anhaltspunkt zu bieten, ob sich der Suchaufwand «lohnt». Diese Entscheidung muss individuell getroffen werden. Abbildung 7 zeigt den Boxplot für das iPhone 12 mit 256 GB.

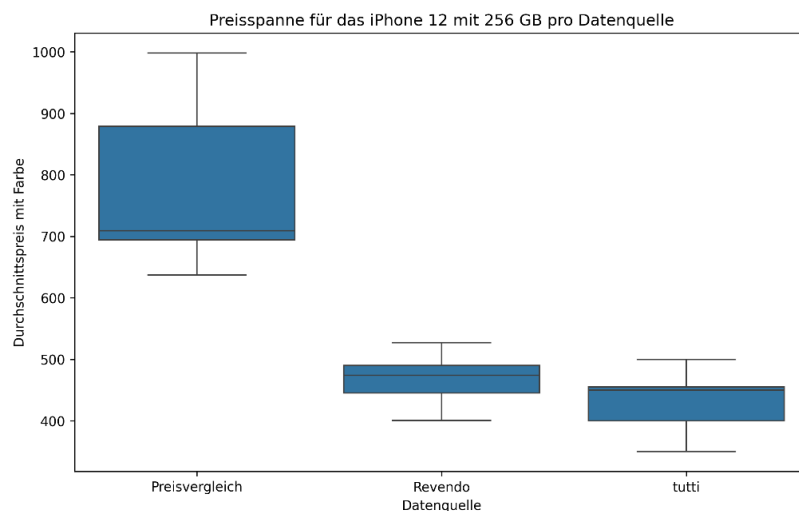


Abbildung 7 - Boxplot iPhone 12 Preisspanne pro Plattform

- Preisvergleich.ch: Der Boxplot zeigt eine sehr grosse Varianz der Preise. Der Median liegt deutlich höher als bei den anderen Plattformen, was darauf hindeutet, dass Preisvergleich.ch tendenziell die höchsten Preise hat - was auch Sinn macht, da hier nur ungeöffnete Geräte angeboten werden.
- Revido: Hier ist der Interquartilsabstand kleiner als bei Preisvergleich.ch, was auf eine kleinere Preisvarianz hinweist. Dies liegt wohl daran, dass die Preise auf Revido wohl klaren Regeln folgen, wie sich der Preis zusammenstellt. Die Preise zeigen das kleinste Preis-Spektrum der drei Datenquellen, wobei der Median niedriger ist als bei Preisvergleich.ch, aber minim höher als bei tutti.

- tutti.ch: Dieser Boxplot hat den niedrigsten Median, was auf im Allgemeinen niedrigere Preise hinweist. Die Preise sind generell etwas geringer als auf Revendo, haben aber eine ähnliche Varianz.

Der Median von tutti.ch und Preisvergleich haben eine Differenz von über 200 CHF. Somit kann ein*e Käufer*in nun entscheiden, ob sich diese Differenz für ihn*sie lohnt, auf tutti zu suchen.

Abbildung 8 zeigt den vierten Boxplot. Hier werden über alle Zustände und Plattformen hinweg die Preisspannen nach Farben aufgeschlüsselt gezeigt. Dies ist relevant für Nutzer*innen, die entweder Wert auf die Farbe legen und das Verhältnis sehen möchten oder für preisorientierte Nutzer*innen, die gerne die günstigste Farbvariante nehmen wollen. Dies kann beispielsweise die Suche auf tutti.ch vereinfachen.

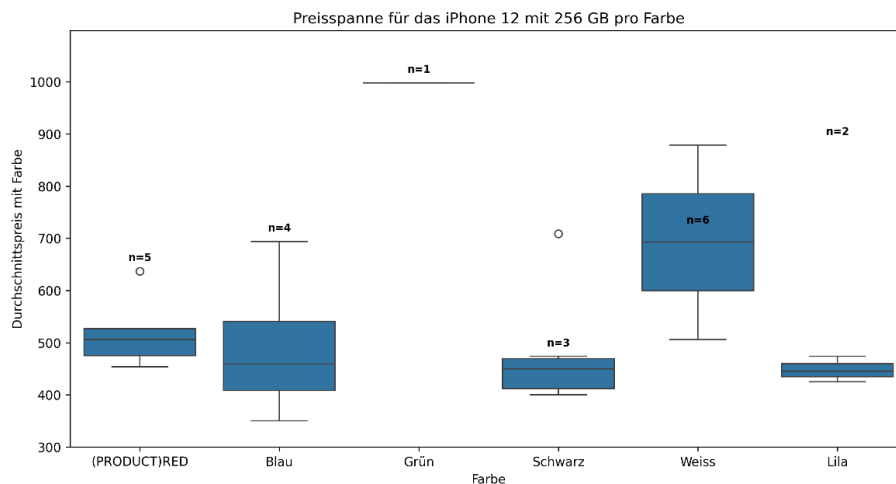


Abbildung 8 - Boxplot für iPhone 12 mit Preis pro Farbe

Der Boxplot zeigt, dass es Preis- und Variabilitätsunterschiede zwischen den Farben gibt. So haben Lila und Blau beispielsweise höhere Variabilitäten beim iPhone 12 mit 256 GB. Es lassen sich hier keine Allgemeinaussagen treffen, da die Datengrundlage pro Modell-Speicherkombination zu gering ist (z.B. Grün liegt nur ein Angebot zugrunde).

5.2 1. Subfrage: Preis pro iPhone-Modell

Was sind die durchschnittlichen Preise pro iPhone-Modelle ab dem Baujahr 2018 auf den verschiedenen Plattformen (Gebraucht und Erstmarkt)?

Nachdem wir mit dem Dashboard eine individuelle Lösung für Nutzer*innen entwickelt haben, wollen wir nun den gesamten Datensatz analysieren und unsere Unterfragen beantworten. Dies soll den Nutzer*innen eine Orientierung bieten, was eine gute Preis-Leistungs-Entscheidung beim Kauf eines iPhones sein könnte. Die Frage wollen wir gerne mit zwei Graphen beantworten.

Der erste Graph ist ein Boxplot, der die Verteilung der durchschnittlichen Preise über alle Modelle hinweg auf den verschiedenen Plattformen zeigt (d.h. wie der erste Boxplot vom Dashboard, nur über alle Modelle hinweg). Es zeigt sich, dass die Preise auf dem Erstmarkt tendenziell höher sind als auf dem Gebrauchtmkt. Auf dem Gebrauchtmkt von Tutti gibt es jedoch eine grössere Preisvarianz. Im Durchschnitt liegen die Preise bei beiden Quellen (Revendo und Tutti) ähnlich.

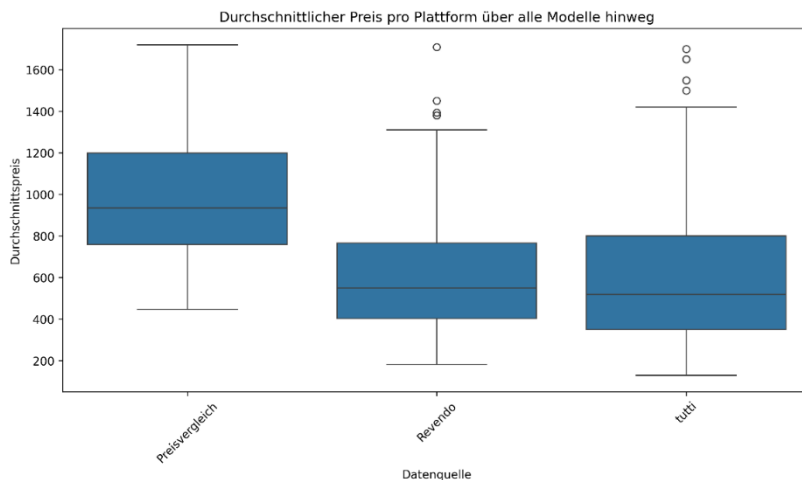


Abbildung 9 - Boxplot durchschnittliche Preise über alle Modelle pro Plattform

Der nächste Plot gibt einen Überblick über die durchschnittlichen Preise pro iPhone-Modell. Auf diesen Aspekt gehen wir später noch ein. Die Übersicht zeigt, bei welchen iPhone-Modellen auf dem Gebrauchtmkt erheblich gespart werden kann, insbesondere bei den älteren Modellen iPhone 11 Pro, iPhone 11 Pro Max, iPhone XS, iPhone XS Max und iPhone 12 Pro Max. Der Neupreis ist hier teilweise doppelt so hoch wie auf dem Gebrauchtmkt. Im Gegensatz dazu zeigen neuere Modelle tendenziell geringere Preisunterschiede zwischen Neu- und Gebrauchtpreisen. Das liegt unter anderem daran, dass es meist weniger Angebote für neuere Modelle auf dem Gebrauchtmkt gibt, da weniger Personen bereit sind, die neuesten Geräte zu verkaufen, was die Preise hoch hält.

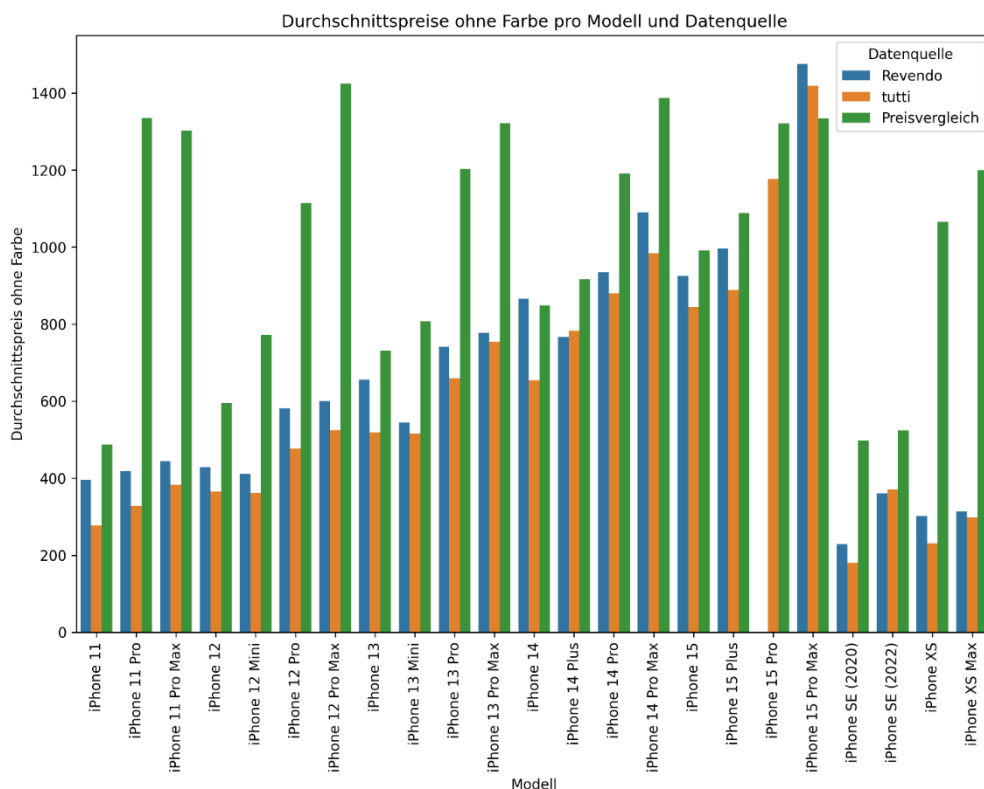


Abbildung 10 - Barplot durchschnittliche Preise ohne Farbe pro Modell und Datenquelle

Interessanterweise sind die Preise für das iPhone 15 Pro Max bei Revendo und Tutti höher als auf dem Neuwarenmarkt. Dies könnte an der geringen Anzahl an Angeboten, an «Schnäppchen» bei Preisvergleich-Anbietern oder an Zubehör liegen. Eine weitere interessante Beobachtung ist, dass die Standard-iPhone-Modelle (ohne Zusatzbezeichnungen wie Max oder Pro) auf dem Neuwarenmarkt deutlich an Wert verlieren. Im Gegensatz dazu halten die spezielleren Modelle ihren Wert besser im Vergleich zu ihren Nachfolgern (z.B. iPhone 11 Pro). Zusammenfassend lässt sich sagen, dass ältere Modelle auf dem Gebrauchtmärkte günstiger werden, während die Preise auf dem Neuwarenmarkt weniger stark sinken. Dies könnte auch daran liegen, dass der Zustand älterer iPhone-Modelle aufgrund längerer Nutzungsdauer nicht mehr so gut ist und darum der Durchschnittspreis sinkt.

5.3 2. Subfrage: Einfluss des Zustandes

Wie stark beeinflusst der Zustand (optischer Zustand und Akkukapazität) eines iPhones dessen Preis auf dem Gebrauchtmärkte?

Zur Beantwortung dieser Frage haben wir ein lineares Modell entwickelt, das den Einfluss des optischen Zustands und der Akkukapazität auf den Preis eines iPhones auf dem Gebrauchtmärkte untersucht. Dieses Modell basiert ausschliesslich auf Daten des Gebrauchtmärkte (Datenquellen: Tutti und Revendo), um die Auswirkungen dieser Faktoren auf die Preisgestaltung zu analysieren. Zunächst betrachten wir die Verteilung dieser Zustände in den Daten (Abbildung 11 und 12).

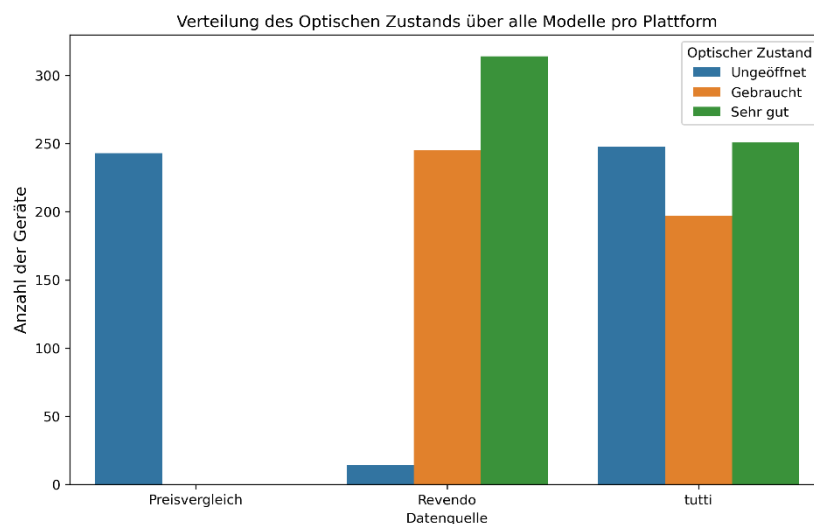


Abbildung 11 - Verteilung des Optischen Zustands über alle Modelle pro Plattform

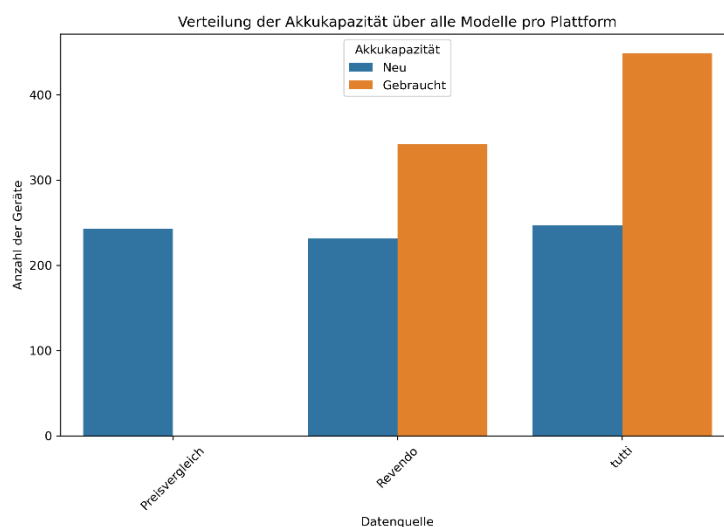


Abbildung 12 - Verteilung der Akkukapazität über alle Modelle pro Plattform

Aus den Daten geht hervor, dass bei Revendo kaum ungeöffnete iPhones angeboten werden, wohingegen wir eine ähnliche Anzahl an Beobachtungen in den anderen optischen Zuständen und bei der Akkukapazität vorfinden. Bei Tutti sind die Beobachtungen zum optischen Zustand vergleichbar, jedoch weist die Akkukapazität nahezu doppelt so viele gebrauchte Zustände auf. Grundsätzlich scheinen wir aber eine genügend gute Verteilung zu haben der beiden Variablen, abgesehen von ungeöffneten iPhones bei Revendo.

Anschliessend haben wir das Modell berechnet (Abbildung 13). Das Modell zeigt, dass eine vollständige Akkukapazität den Preis durchschnittlich um 112,88 CHF steigert. Bei den optischen Zuständen - gebraucht, Sehr gut und Ungeöffnet - beobachten wir Folgendes: Ein iPhone in sehr gutem Zustand ist im Durchschnitt 79,78 CHF teurer als ein gebrauchtes. Der Preisunterschied zwischen sehr gutem Zustand und ungeöffnetem Zustand ist jedoch marginal: Ungeöffnete Geräte sind nur 84,17 CHF teurer als gebrauchte.

OLS Regression Results							
Dep. Variable:	Durchschnittspreis		R-squared:	0.064			
Model:	OLS		Adj. R-squared:	0.061			
Method:	Least Squares		F-statistic:	28.68			
Date:	Sat, 16 Dec 2023		Prob (F-statistic):	6.08e-18			
Time:	09:50:13		Log-Likelihood:	-8952.7			
No. Observations:	1269		AIC:	1.791e+04			
Df Residuals:	1265		BIC:	1.793e+04			
Df Model:	3						
Covariance Type:	nonrobust						
		coef	std err	t	P> t	[0.025	0.975]
	const	505.5505	14.053	35.975	0.000	477.981	533.120
	Akkukapazität_1.0	112.8800	17.410	6.484	0.000	78.724	147.036
	Optischer Zustand_1.0	79.7782	17.871	4.464	0.000	44.719	114.838
	Optischer Zustand_2.0	84.1740	23.304	3.612	0.000	38.455	129.893
	Omnibus:	102.809	Durbin-Watson:	0.222			
	Prob(Omnibus):	0.000	Jarque-Bera (JB):	126.859			
	Skew:	0.739	Prob(JB):	2.84e-28			
	Kurtosis:	3.466	Cond. No.	4.18			

Abbildung 13 - Screenshot Lineares Modell für die iPhone Zustände

Zusammenfassend hat die Akkukapazität einen grösseren Einfluss auf den Preis als der optische Zustand. Dennoch macht der Zustand mit rund 80 CHF einen beachtlichen Unterschied, wobei es kaum relevant ist, ob ein iPhone im sehr guten oder ungeöffneten Zustand gekauft wird. Ein signifikanter Preisunterschied ist hauptsächlich bei gebrauchten iPhones festzustellen.

5.4 3. Subfrage: Preisverringern

In welchem Umfang verringern sich die Preise von gebrauchten iPhones im Vergleich zum aktuellen Verkaufspreis eines neuen iPhones? Wie gross der ist Einfluss der Speichergrösse und Erscheinungsjahr?

In der Rückschau erkennen wir, dass diese Fragestellung nicht gut zu unseren Daten passt. Um festzustellen, wie viel der Preis eines Modells seit seiner Einführung gesunken ist, benötigen wir die ursprünglichen Verkaufspreise der iPhones zum Zeitpunkt ihrer Markteinführung. Eine Analyse des Preisverfalls würde zudem einen Datensatz brauchen, der Preisinformationen über einen längeren Zeitraum hinweg umfasst. Dies würde es ermöglichen, Trends zu erkennen, wie schnell und in welchem Umfang die Preise sinken. Wir haben uns auf die aktuellen Preise von gebrauchten und neuen iPhones beschränkt, ohne die historischen Verkaufspreise oder andere relevante Daten zu berücksichtigen. Um dennoch Erkenntnisse aus den vorhandenen Daten zu gewinnen, haben wir den zweiten Teil der Frage abgekoppelt und separat beantwortet.

Wir möchten daher nun den Einfluss des Erscheinungsjahres sowie der Speichergrösse auf den Preis der verschiedenen iPhone-Modelle untersuchen. Dafür haben wir unserem Datensatz die entsprechenden Erscheinungsjahre hinzugefügt. Diese waren bereits im Codebuch festgehalten, da wir damit bestimmt haben, welche Modelle wir untersuchen wollen. Anschliessend trainieren wir ein lineares Modell, um den Einfluss dieser beiden Faktoren auf den Preis zu analysieren (Abbildung 14). Hierfür haben wir die Daten vom Gebrauchtmrkt verwendet. Neupreise sind in der Regel festgesetzt und spiegeln den vom Hersteller

empfohlenen Verkaufspreis wider. Das bedeutet, dass es weniger Preisvariationen gibt, die analysiert werden können, und daher weniger aussagekräftige Daten für ein Modell, das den Einfluss von Variablen wie Speichergrösse und Erscheinungsjahr auf den Preis untersucht. Der Gebrauchtmärkte ist zudem dynamischer und bietet ein realistischeres Bild davon.

OLS Regression Results							
Dep. Variable:	Durchschnittspreis		R-squared:	0.759			
Model:	OLS		Adj. R-squared:	0.758			
Method:	Least Squares		F-statistic:	793.4			
Date:	Sat, 16 Dec 2023		Prob (F-statistic):	0.00			
Time:	09:50:13		Log-Likelihood:	-8092.9			
No. Observations:	1269		AIC:	1.620e+04			
Df Residuals:	1263		BIC:	1.623e+04			
Df Model:	5						
Covariance Type:	nonrobust						
		coef	std err	t	P> t 	[0.025	0.975]
const	-3.042e+05	6780.119	-44.868	0.000	-3.18e+05	-2.91e+05	
Erscheinungsjahr	150.7936	3.357	44.914	0.000	144.207	157.380	
Speicher in GB_128	47.4975	12.942	3.670	0.000	22.107	72.888	
Speicher in GB_256	154.1050	12.531	12.298	0.000	129.522	178.688	
Speicher in GB_512	279.7190	15.979	17.505	0.000	248.371	311.067	
Speicher in GB_1000	445.6231	22.691	19.639	0.000	401.107	490.139	
Omnibus:	18.342	Durbin-Watson:	0.642				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28.239				
Skew:	-0.115	Prob(JB):	7.38e-07				
Kurtosis:	3.693	Cond. No.	3.42e+06				

Abbildung 14 - Screenshot Lineares Modell für die iPhone Zustände

Die Analyse ergibt, dass jedes Jahr den Preis eines iPhones durchschnittlich um 150,79 CHF erhöht. Das bedeutet, dass ein iPhone-Modell durchschnittlich um 150 Franken teurer ist als ein Modell aus dem Vorjahr. Bei der Speichergrösse dienen 64 GB Speicherplatz als Basis. Steigt die Speichergrösse auf 128 GB, erhöht sich der Preis durchschnittlich um 47,50 CHF im Vergleich zu einem 64 GB Modell. Diese Tendenz setzt sich bei den grösseren Speichervarianten fort: Der Preis steigt jeweils um rund 100 CHF beim Wechsel von 128 GB zu 256 GB (154,11 – 47,5) und von 256 GB zu 512 GB (279,71 – 154,11). Bei einer Speichergrösse von 1 TB verdoppelt sich der Preiszuwachs (445,62 – 279,71). Die Preissteigerung ist somit einigermaßen proportional zur GB-Anzahl.

5.5 4. Subfrage: Preisunterschiede pro Farbe

Welche Preisunterschiede gibt es pro iPhone-Modellen in Abhängigkeit von ihrer Farbe? Gibt es ein allgemeines Muster über alle Modelle hinsichtlich der Preisgestaltung für die verschiedenen Farboptionen?

Damit wir die letzte Fragestellung beantworten können, haben wir eine Heatmap erstellt, die die Preisabweichung in CHF eines iPhone-Modells mit einer bestimmten Farbe vom Durchschnittspreis des Modells über alle Farben hinweg zeigt. Rote Werte bedeuten, dass der Preis über dem allgemeinen Durchschnitt für dieses Modell liegt, blau bedeutet, dass er darunter liegt. Es ist wichtig zu beachten, dass die Daten von tutti mit medianbasierten Preisen berechnet wurde, während die anderen Quellen Mittelwerte verwenden. Aufgrund der unterschiedlichen Berechnungsgrundlagen sind die dargestellten Abweichungen nicht als absolute, sondern als relative Indikatoren zu interpretieren, die aufzeigen, in welchem Verhältnis die Preise der einzelnen Farbvarianten zum allgemeinen Preisniveau stehen.

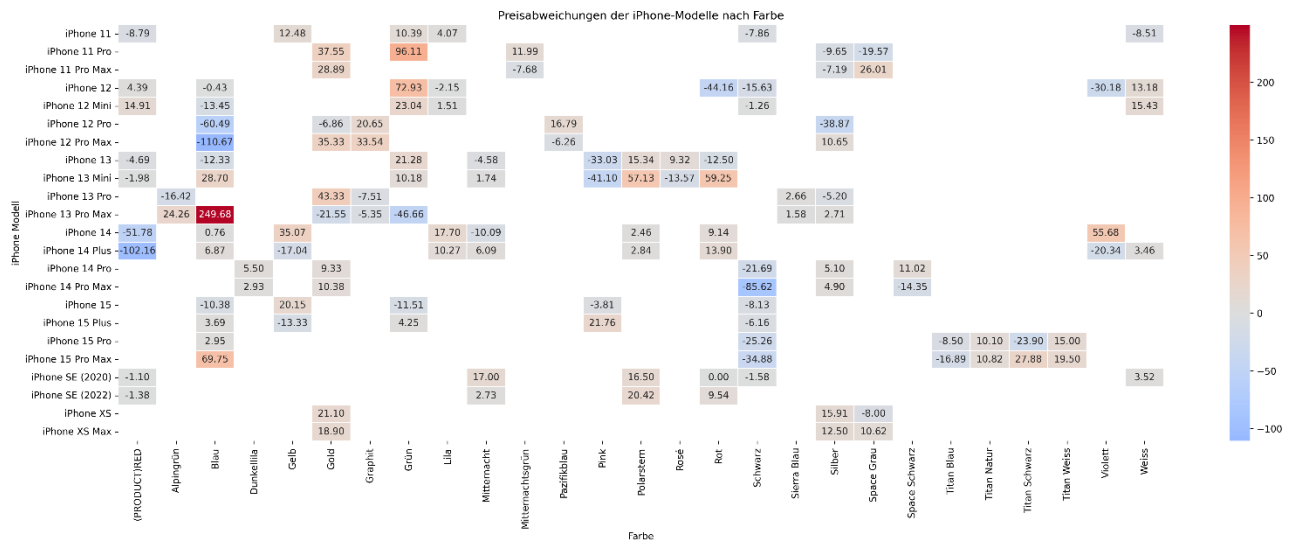


Abbildung 15 - Heatmap Preise pro Modell und Farbe

Es können aus der Visualisierung folgende Trends erkannt werden:

- Einige Farben scheinen generell höhere oder niedrigere Preise über die verschiedenen Modelle hinweg zu haben. Zum Beispiel die Farbe Schwarz hat fast bei jedem Modell eine negative Abweichung, während die Farben Gold oder Polarstern fast durchwegs positive Abweichungen aufzeigen. Dies könnte darauf hindeuten, dass bestimmte Farben seltener oder bei den Konsument*innen beliebter sind und daher zu höheren Preisen verkauft werden.
- Die Preisabweichungen sind bei den neueren Modellen (wie iPhone 13 Pro Max und iPhone 14 Pro Max) ausgeprägter als bei älteren Modellen. Dies könnte darauf hinweisen, dass bei neueren Modellen die Farbwahl eine grössere Rolle für den Wert des Geräts spielt oder sich die Preise bei den älteren Modellen angeglichen haben.
- Innerhalb der einzelnen Modelle können die Preisabweichungen je nach Farbe variieren. Beispielsweise ist bei der Farbe Blau zu beobachten, dass der Durchschnittspreis je nach Modell über oder unter dem allgemeinen Durchschnitt liegt.
- Bei einigen Farboptionen sind auffällig hohe Preisabweichungen festzustellen. Eine genauere Untersuchung wäre notwendig, um zu bestimmen, ob diese auf Datenqualität zurückzuführen sind oder ob sie auf die Beliebtheit oder geringe Nachfrage bestimmter Modelle hindeuten.

Zusammenfassend lässt sich festhalten, dass kein einheitliches Muster für alle Farboptionen über alle Modelle hinweg zu erkennen ist, da die Preisabweichungen je nach Modell und Farbe variieren. Dennoch scheinen bestimmte Farben tendenziell entweder zu höheren oder niedrigeren Preisen zu führen, was auf eine generelle Marktpräferenz oder -verfügbarkeit dieser Farben hindeuten könnte. Es zeigt sich somit insgesamt, dass die Preisunterschiede von Modell und Farbe abhängen.

6 Reflektion

In diesem abschliessenden Kapitel reflektieren wir über die Herausforderungen, denen wir im Verlauf des Projekts begegnet sind und was unsere Erkenntnisse daraus sind. Inhaltlich sind wir mit den Ergebnissen unserer Analyse zufrieden. Wir konnten interessante Einsichten rund um iPhones auf dem Gebrauchtmärkte gewinnen und unsere spezifischen Fragestellungen beinahe vollständig beantworten.

Das wesentliche Learning aus diesem Projekt ist die Erkenntnis, dass eine gründliche Planung essentiell ist, bevor man Daten sammelt und analysiert. Es ist wichtig, bereits im Vorfeld genau zu überlegen, welche Fragen beantwortet werden sollen und welche Methoden – ob Visualisierungen oder statistische Modelle – dafür benötigt werden (beispielsweise Vor- und Nachteile von Median und Mittelwert). Dies beinhaltet auch eine klare Definition der Anforderungen an die Daten. So haben uns für eine Fragestellung bestimmte Arten von Daten gefehlt, während wir gleichzeitig zunächst Informationen gesammelt bzw. hinzugefügt haben, die irrelevant waren. Ebenso haben wir geplante Anreicherungen nicht mehr durchgeführt, weil wir während des Prozesses gemerkt haben, dass wir sie gar nicht brauchen. Auch der Umfang der verfügbaren und der erforderlichen Datenmenge für die gewünschte Analyse sollte vorher klar sein, um sicherzustellen, dass die Datenbasis robust genug für aussagekräftige Auswertungen ist. Die Vielzahl an Kombinationsmöglichkeiten von Modellvarianten, Akkuzuständen, optischen Zuständen, Speichergrößen und Farben führte bei uns

dazu, dass es zum Teil pro Kombination nur wenige Datenpunkte gab. Für künftige Projekte könnte eine Reduzierung der Variablen – zum Beispiel durch das Auslassen der Farben – die Anzahl der Datenpunkte pro Kategorie erhöhen und so zu aussagekräftigeren Ergebnissen führen.

Eine wichtige Erkenntnis aus unserem Projekt war die Bedeutung des Transformations- und Datenbereinigungsprozesses, der einen erheblichen Teil der gesamten Arbeitszeit in Anspruch nahm. Die Aufbereitung der Daten von Plattformen wie tutti.ch oder revendo.com erforderte mehr Zeit als erwartet, hauptsächlich wegen umfangreicher Nachbereitungen und Datenbereinigungen. Obwohl wir zudem ursprünglich das Ziel verfolgten, unseren Code so allgemein wie möglich zu halten, um Flexibilität und Wiederverwendbarkeit zu maximieren, stellten wir fest, dass eine beträchtliche Menge an datensatzspezifischem Code erforderlich war. Diese Erfahrung unterstreicht die Wichtigkeit eines realistischen Zeitplans und einer sorgfältigen Prüfung der Datenqualität bei der Auswahl von Datenquellen. Künftig könnte beispielsweise der Einsatz von Datenquellen, die strukturierte Informationseingaben bieten als beispielsweise tutti, die Qualität und Verwertbarkeit der Daten erhöhen und somit die Effizienz steigern sowie den Arbeitsaufwand besser managen.

Sehr hilfreich waren die Erstellung des Codebuchs und die Vorabklärung im Umgang mit fehlenden Daten. Dies erwies sich besonders beim Zusammenführen (Merging) der Daten als vorteilhaft, da Inkonsistenzen reduziert und der Prozess beschleunigt wurden. Dies unterstreicht die Wichtigkeit einer strukturierten Datenvorbereitung, insbesondere wenn Daten aus verschiedenen Quellen oder von unterschiedlichen Teammitgliedern zusammengeführt werden sollen. Dies nehmen wir so für zukünftige Projekte mit.

Ein besonders wertvolles technisches Learning dieses Projekts war die Entwicklung unserer Fähigkeiten im Bereich des Daten-Scrapings, beispielsweise auch im Umgang mit Herausforderungen wie Cloudflare. Durch die Arbeit mit Python und verschiedenen Datenverarbeitungspaketen, insbesondere Pandas, konnten wir zudem unsere analytischen Fähigkeiten erheblich verbessern.

Abschliessend lässt sich festhalten, dass uns dieses Projekt eine Vielzahl von Lernmöglichkeiten geboten hat und unser Verständnis für den ETL-Prozess vertieft wurde. Insgesamt war das Projekt eine wertvolle Erfahrung, die unser Wissen und unsere Fähigkeiten im Bereich der Datenverarbeitung und -analyse bereichert hat.

Referenzliste

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/>

Tabellen- & Abbildungsverzeichnis

Tabelle 1 - Datenstruktur revendo.ch	9
Tabelle 2 - Datenstruktur preisvergleich.ch	11
Abbildung 1 - Kontextdiagramm des ETL-Prozesses	4
Abbildung 2 - Output von ChatGPT mit dem Dictionary Format.....	9
Abbildung 3 - Prompt an ChatGPT zur Datenverunreinigung.....	11
Abbildung 4 - Screenshot interaktives Dashboard.....	13
Abbildung 5 - Filterung nach Modell und Speichergrösse (Ausschnitt der Tabelle)	13
Abbildung 6 - Kumulierte Übersicht zu den iPhone 12 256GB Preisen pro Plattform.....	14
Abbildung 7 - Boxplot iPhone 12 Preisspanne pro Plattform.....	14
Abbildung 8 - Boxplot für iPhone 12 mit Preis pro Farbe	15
Abbildung 9 - Boxplot durchschnittliche Preise über alle Modelle pro Plattform	16
Abbildung 10 - Barplot durchschnittliche Preise ohne Farbe pro Modell und Datenquelle	16
Abbildung 11 - Verteilung des Optischen Zustand über alle Modelle pro Plattform	17
Abbildung 12 - Verteilung der Akkukapazität über alle Modelle pro Plattform.....	17
Abbildung 13 - Screenshot Lineares Modell für die iPhone Zustände.....	18
Abbildung 14 - Screenshot Lineares Modell für die iPhone Zustände.....	19
Abbildung 15 - Heatmap Preise pro Modell und Farbe.....	20

Anhang

Die Unterlagen zu allen Projektteilen sind im Ordner abgelegt. Der Ordner hat folgende Struktur:

- Ordner mit allen individuellen Codes, Videos und Daten
 - o Bigler_melanie_studentC
 - o Kreischer_annalena_studentA
 - o Peter_michelle_studentB
- Codebuch
- Gruppenvideo