

Projektbericht zum Modul Information Retrieval und
Visualisierung Sommersemester 2022

Smartphone Analyse Tool

Tom Schindler

20. August 2024

Inhaltsverzeichnis

1	Einleitung	3
1.1	Anwendungshintergrund	3
1.2	Zielgruppen	4
1.3	Überblick und Beiträge	4
2	Daten	5
3	Visualisierungen	7
3.1	Analyse der Anwendungsaufgaben	7
3.2	Anforderungen an die Visualisierungen	8
3.3	Präsentation der Visualisierungen	8
3.3.1	Visualisierung Eins	8
3.3.2	Visualisierung Zwei	8
3.3.3	Visualisierung Drei	8
3.4	Interaktion	8
4	Implementierung	9
5	Anwendungsfälle	9
5.1	Anwendung Visualisierung Eins	9
5.2	Anwendung Visualisierung Zwei	9
5.3	Anwendung Visualisierung Drei	9
6	Verwandte Arbeiten	9
7	Zusammenfassung und Ausblick	9

1 Einleitung

In der wettbewerbsintensiven Welt der Smartphone-Technologie ist es für Unternehmen entscheidend, die Präferenzen ihrer Kunden genau zu verstehen. Die Analyse von Smartphone-Hardwaredaten bietet wertvolle Einblicke in das Nutzerverhalten und die Vorlieben der Kunden hinsichtlich verschiedener Gerätefunktionen. Diese Erkenntnisse können genutzt werden, um Marketingstrategien gezielt auszurichten und maßgeschneiderte Produkte und Dienstleistungen zu entwickeln, die den spezifischen Bedürfnissen der Nutzer entsprechen.

Ziel dieses Berichts ist es, ein speziell entwickeltes Analyse-Tool vorzustellen, das dazu dient, die Hardwaredaten von Smartphones zu visualisieren und zu analysieren. Im Fokus stehen dabei Fragestellungen wie: Welche Hardwaremerkmale sind für Kunden besonders wichtig? Wie beeinflussen diese Präferenzen die Kaufentscheidungen (Bewertung)? Und wie können diese Informationen genutzt werden, um gezielte Marketingstrategien zu entwickeln und Produkte zu optimieren?

Mithilfe von Techniken der Informationsvisualisierung wird im weiteren Verlauf des Berichts gezeigt, wie diese Fragestellungen beantwortet werden können. Dies unterstützt Unternehmen dabei, tiefere Einblicke in die Präferenzen ihrer Kunden zu gewinnen, um so den Erfolg auf dem Markt zu steigern und innovativere, auf die Bedürfnisse der Kunden abgestimmte Produkte zu entwickeln.

1.1 Anwendungshintergrund

Die Analyse von Smartphone-Hardwaredaten spielt eine zentrale Rolle, wenn es darum geht, Kundenpräferenzen zu verstehen und darauf basierend gezielte Marketingstrategien zu entwickeln. Moderne Smartphones bestehen aus einer Vielzahl von Hardwarekomponenten, wie Prozessoren, Speichereinheiten, Batterien und Bildschirmgröße, die alle unterschiedlich wahrgenommen und genutzt werden. Diese Daten können durch geeignete Analysetools extrahiert und visuell aufbereitet werden, um Muster und Trends zu erkennen, die für Unternehmen von großem Interesse sind.

Ein erfolgreiches Analyse-Tool muss daher in der Lage sein, diese komplexen und umfangreichen Daten in einer Form darzustellen, die für die Entscheidungsfindung nützlich ist. Dies erfordert nicht nur eine leistungsfähige Datenverarbeitung, sondern auch effektive Visualisierungstechniken, die es ermöglichen, wichtige Erkenntnisse schnell und klar zu erfassen.

Besonders relevant ist dabei, wie die Visualisierung von Hardwaredaten auf die Präferenzen der Kunden hinweist und wie sich diese Informationen in die Entwicklung maßgeschneiderter Produkte und Dienstleistungen umsetzen lassen. Dabei ist es entscheidend, dass die Visualisie-

rungen so gestaltet sind, dass sie sowohl detaillierte Einblicke ermöglichen als auch übersichtlich genug bleiben, um auf den ersten Blick verständlich zu sein.

1.2 Zielgruppen

Die Zielgruppe für das Smartphone Analyse Tool umfasst hauptsächlich Produktmanager, Marketingexperten und Unternehmensanalysten in der Smartphone Technologiebranche. Diese Personen sind daran interessiert, tiefere Einblicke in die Kundenpräferenzen zu gewinnen, um fundierte Entscheidungen bezüglich Produktentwicklungen und Marketingstrategien zu treffen. Darüber hinaus könnten auch Führungskräfte und Entscheidungsträger aus dem Bereich strategische Planung von den Ergebnissen profitieren, da sie auf Basis dieser Erkenntnisse marktgerechte Produkte entwickeln möchten.

Bei dieser Zielgruppe kann von einem soliden Vorwissen in den Bereichen Marktanalyse, Produktentwicklung und Kundenforschung ausgegangen werden. Viele der Anwender sind bereits mit grundlegenden Konzepten der Datenanalyse vertraut und haben ein Verständnis für die Bedeutung von Hardwaremerkmalen bei der Kaufentscheidung von Konsumenten. Sie benötigen jedoch spezialisierte Werkzeuge, die es ihnen ermöglichen, komplexe Daten auf intuitive Weise zu visualisieren und so gezielt Erkenntnisse über Kundenbedürfnisse und -verhalten zu gewinnen.

Die durch die Visualisierungen adressierten Informationsbedürfnisse betreffen insbesondere:

- **Identifizierung von Kundenpräferenzen:** Welche Hardwarekomponenten sind für die Kunden am wichtigsten, und wie wirken sich diese Präferenzen auf Kaufentscheidungen aus?
- **Produktentwicklung:** Welche Features sollten in zukünftigen Produktgenerationen stärker berücksichtigt oder optimiert werden, um die Wettbewerbsfähigkeit zu erhöhen?

Die Visualisierungen bieten den Anwendern somit eine klare und übersichtliche Darstellung der relevanten Daten, die es ihnen ermöglicht, ihre strategischen Ziele effektiver zu erreichen.

1.3 Überblick und Beiträge

Mit Hilfe des Analyse-Tools ist es möglich, aus einer großen Datenmenge[1], die zahlreiche moderne Smartphone-Modelle umfasst, gezielte Vergleiche und Analysen durchzuführen. Dabei werden unter anderem Daten zu den Hardware-Komponenten, wie verbauter Prozessor, vorhandener Speicher und Batteriekapazität, vergleichbar gemacht. Darüber hinaus werden Preis-Leistungs-Aspekte anhand von Kundenbewertungen veranschaulicht, was einen wichtigen Beitrag zur Identifizierung von Kundenpräferenzen leistet.

Die Daten werden in vier verschiedenen Visualisierungen dargestellt, um die Informationsbedürfnisse der Zielgruppe umfassend zu erfüllen. Diese Visualisierungen sind:

- **Scatterplot:** Stellt jeweils zwei numerische Daten auf zwei Achsen dar und ermöglicht so die Analyse von Korrelationen zwischen verschiedenen Variablen.
- **Parallel-Korrdinaten-Plot:** Zeigt alle numerischen Daten der Datenmenge in einer mehrdimensionalen Visualisierung, um komplexe Zusammenhänge übersichtlich darzustellen.
- **Baum-Datstellung:** Visualisiert alle Smartphone-Hersteller und deren Modelle aus der Datenmenge in einer hierarchischen Struktur.
- **Starplot:** Stellt alle numerischen Daten zu einem Smartphone in einer sternförmigen Visualisierung dar und bietet so einen kompakten Überblick über die Leistungsmerkmale eines Geräts.

In den folgenden Abschnitten werden die verwendeten Daten, die erstellten Visualisierungstechniken und deren Umsetzung und deren Anwendungsfälle detailliert beschrieben.

2 Daten

Die verwendeten Daten stammen von Kaggle[1] und liegen im CSV-Format vor. Diese Datensammlung umfasst Informationen zu verschiedenen Hardware-Komponenten, Preisen, Bewertungen und weiteren Merkmalen einer großen Anzahl älterer und aktueller Smartphone-Modelle (insgesamt 984 Einträge in der CSV-Datei). Die Datenspalten der CSV-Datei[1] umfassen:

- Brand - Marke des Smartphones
- Model - Model des Smartphones
- Price - Price des Smartphones
- Rating - durchschnittliche Kundenbewertung
- 5G - hat 5G oder nicht
- NFC - hat NFC oder nicht
- IR Blaster - hat IR Blaster oder nicht
- Processor Name - Name des Prozessors
- Processor Brand - Marke des Prozessors
- NumCores - Kernanzahl
- Processor Speed - Anzahl der Zyklen pro Sekunde, die eine CPU ausführen kann
- Ram and internal storage (ram) - RAM and interner Speicher

- Battery - Kapazität der Battery
- Fast Charging - hat Fast Charging oder nicht, und falls ja Kapazität
- Internal Mem - interner Speicher des Smartphones
- RAM - RAM des Smartphones
- Screen Size - Bildschirmgröße des Smartphones
- Resolution - Auflösung des Smartphones
- Refresh Rate - Bildwiederholfrequenz des Smartphones
- Camera - Kameraqualität der Rück- und Frontkamera
- Card - Speicherkarte unterstützt oder nicht und falls ja Speicherkapazität
- OS - Betriebssystem des Smartphones

Die meisten Datenspalten aus der CSV-Datei wurden in die Datenvorverarbeitung der Anwendung integriert, die in Elm programmiert wurde. Ausgenommen davon ist die Spalte „Ram and internal storage (ram)“, da hierfür bereits separate Spalten vorhanden sind. Einige weitere Spalten wie „Camera“, „Card“, „OS“ und „Resolution“ wurden bei der Vorverarbeitung lediglich als Zeichenketten gespeichert und werden in den Visualisierungen der Anwendung nicht weiter berücksichtigt. Diese Spalten weisen Inkonsistenzen in ihren Datenwerten auf (meist in Textform), was es schwierig machte, daraus numerisch vergleichbare Werte zu erzeugen. Diese Daten könnten in zukünftigen Arbeiten weiter aufbereitet und in die Visualisierungen integriert werden.

Alle anderen Daten fließen entweder direkt oder indirekt (z.B. als Filteroption oder Beschriftung) in die Visualisierungen ein. Bei der Vorverarbeitung wurden bei den meisten numerischen Werten bestimmte Zeichen oder Wörter entfernt, um den reinen numerischen Wert zu extrahieren (z.B. Screen Size = „6.6 inches“ zu „6.6“ oder Processor Speed = „2.2 GHz Processor“ zu „2.2“). Daten, die das Vorhandensein eines Merkmals beschreiben, wurden in Elm in Wahrheitswerte übersetzt (z.B. „5G“), welche meist als Filteroption dienen. Alle weiteren beschreibenden (textuellen) Daten wurden in ihrer ursprünglichen Form als Zeichenketten gespeichert und dienen hauptsächlich der Beschriftung oder als Filter. Fehlende Datenwerte wurden in Elm als Nothing-Werte behandelt.

Verarbeitungsschritte:

1. Auswahl der Daten (nach Sinnhaftigkeit für Projekt)
2. Zuordnung der Daten zu einer Art (numerische vergleichbare Werte, Wahrheitswerte, textuelle Beschreibungen)
3. Falls nötig manuelle Anpassung einzelner (fehlerhafter) Werte

4. Parsen der Datenwerte von der CSV-Datei:

- Falls numerischer Wert in Zeichenkette - extrahieren durch entfernen der Textinhalte und parsen zu Datentyp
- Falls numerischer Wert - extrahieren und parsen zu Datentyp
- Falls Wahrheitswert - in Elm Wahrheitwert übersetzen
- Falls Wert nicht vorhanden (egal welcher Art) - in Elm Nothing übersetzen
- sonst - als Zeichenkette speichern

5. Speicherung der Daten im Model

Anmerkung: Da bei einigen wenigen Smartphone-Modellen in der CSV-Datei falsche Datenwerte für den RAM und den internen Speicher vorlagen, wurden diese Daten korrigiert oder ergänzt. Beispielsweise waren bei einigen Modellen RAM und interner Speicher vertauscht oder der interne Speicher wurde fälschlicherweise unter RAM eingetragen (z.B. RAM = „512GB“).

Die ausgewählten Daten eignen sich gut (bzw. ausreichend) für die Analyse der Smartphone-Daten in Bezug auf die Kundenzufriedenheit und erfüllen die Informationsbedürfnisse der definierten Zielgruppe. Aus den Daten lassen sich Zusammenhänge zwischen verbauter Hardware und Preis sowie den Kundenbewertungen ableiten und auswerten. Diese Erkenntnisse ermöglichen es, Schwachstellen und Erfolge der verschiedenen Modelle zu identifizieren. Die gewonnenen Informationen können von Produktmanagern, Marketingexperten und Unternehmensanalysten in die zukünftige Produktentwicklung und -planung neuer Modelle einfließen, um wirtschaftlich erfolgreiche und kundenorientierte Smartphones zu entwickeln.

Leider fehlten in der ausgewählten Datenmenge zusätzliche Informationen wie das Erscheinungsdatum und die Anzahl der verkauften Einheiten. Diese Daten wären jedoch äußerst wertvoll, um die Informationsbedürfnisse der definierten Zielgruppe noch besser zu unterstützen. Das Erscheinungsdatum eines Smartphones könnte Rückschlüsse auf die in einem bestimmten Jahr oder einer bestimmten Zeit verwendete Hardware ermöglichen, was helfen würde, ältere Modelle bei der Produktentwicklung und -planung auszuschließen. Die Daten zu den verkauften Einheiten könnten zudem den tatsächlichen Erfolg eines Modells verdeutlichen, anstatt sich ausschließlich auf Kundenbewertungen zu stützen.

3 Visualisierungen

3.1 Analyse der Anwendungsaufgaben

Analysieren sie die konkreten Anwendungsaufgaben, die die Lösung des Zielproblems durch die Anwender:innen bearbeitet werden müssen. Welche sinnvollen mentale Modelle helfen den Perso-

nen bei der Bearbeitung. Welche Visualisierungen helfen den Personen, die die Software verwenden, sinnvolle mentale Modelle aufzubauen. Sind diese mentalen Modelle für sie notwendig, um die Aufgaben lösen zu können? Gehen sie bei ihrer Argumentation von den Anwendungsaufgaben aus und kommen sie dann zu den mentalen Modellen, deren Aufbau durch Visualisierungen unterstützt wird.

3.2 Anforderungen an die Visualisierungen

Leiten sie Anforderungen an das Design der Visualisierungen ab, die sich durch ihre Analyse des Zielproblems ergeben.

3.3 Präsentation der Visualisierungen

Präsentieren sie die visuelle Abbildungen und Kodierungen der Daten und Interaktionsmöglichkeiten. Sie müssen begründen, warum und wie gut ihre Designentscheidungen die erstellten Anforderungen erfüllen. Weiterhin müssen sie begründen, warum die gewählte visuelle Kodierung der Daten für das zulösenden Problem passend ist. Typische Argumente würden hier auf Wahrnehmungsprinzipien und Theorie über Informationsvisualisierung verweisen. Die besten Begründungen diskutieren explizit die konkrete Auswahl der Visualisierungen im Kontext von mehreren verschiedenen Alternativen. Machen sie hier nicht den Fehler, einfach nur Visualisierung aus den vorgegebenen Bereichen zu diskutieren, weil das in der Regel nicht sinnvoll ist. Wenn sie sich für einen Scatterplot entschieden haben, ist ein Zeitreihendiagramm in der Regel keine Alternative. Diskutieren sie also nicht einfach Zeitreihendiagramme, weil sie in den Anforderungen an das Projekt neben Scatterplots stehen, sondern suchen sie nach echten alternativen Visualisierungen, die zum Aufbau eines vergleichbaren mentalen Modells führen. Diskutieren sie die Expressivität und die Effektivität der einzelnen Visualisierungen.

Die eben beschriebenen Präsentationen und Begründungen sollen für jede der drei folgenden Visualisierungen durchgeführt werden.

3.3.1 Visualisierung Eins

3.3.2 Visualisierung Zwei

3.3.3 Visualisierung Drei

3.4 Interaktion

Die präsentierten Visualisierungstechniken müssen interaktiv zu einer Anwendung verknüpft werden. Die Interaktion mit einer Visualisierung soll in den anderen Visualisierungen zu einer Änderung führen. Erklären sie die möglichen Interaktionen mit den einzelnen Visualisierungen und die möglichen Verknüpfungen zwischen ihnen. Begründen Sie warum die konkreten Interaktionen umgesetzt wurden und welche Zwecke für die Anwenderinnen mit ihnen unterstützt

werden. Begründen sie ebenfalls warum sie andere Interaktionsmöglichkeiten nicht umgesetzt haben. Wenn sie keine der geforderten Interaktionen umsetzen, erhalten Sie im gesamten Projekt deutlichen Punktabzug.

4 Implementierung

Beschreiben Sie die Implementierung ihrer Visualisierungsanwendung in Elm. Stellen die Gliederung ihres Quellcodes vor. Haben Sie verschiedene Elm-Module erstellt. Was war aufwändig umzusetzen, was ließ sich mit dem vorhandenen Code aus den Übungen relativ einfach umsetzen?

Wie sieht die Elm-Datenstruktur für das Model aus, in dem die verschiedenen Zustände der Interaktion gespeichert werden können.

5 Anwendungsfälle

Präsentieren sie für jede der drei Visualisierungen einen sinnvollen Anwendungsfall in dem ein bestimmter Fakt, ein Muster oder die Abwesenheit eines Musters visuell festgestellt wird. Begründen sie warum dieser Anwendungsfall wichtig für die Zielgruppe der Anwenderinnen ist. Diskutieren sie weiterhin, ob die oben beschriebene Information auch mit anderen Visualisierungstechniken hätte gefunden werden können. Falls dies möglich wäre, vergleichen sie die den Aufwand und die Schwierigkeiten ihres Ansatzes und der Alternativen.

5.1 Anwendung Visualisierung Eins

5.2 Anwendung Visualisierung Zwei

5.3 Anwendung Visualisierung Drei

6 Verwandte Arbeiten

Führen sie eine kurze Literatursuche in der wissenschaftlichen Literatur zu Informationsvisualisierung und Visual Analytics nach ähnlichen Anwendungen durch. Diskutieren sie mindestens zwei Artikel. Stellen sie Gemeinsamkeiten und Unterschiede dar.

7 Zusammenfassung und Ausblick

Fassen sie die Beiträge ihre Visualisierungsanwendung zusammen. Wo bietet sie für die Personen der Zielgruppe einen echten Mehrwert.

Was wären mögliche sinnvolle Erweiterungen, entweder auf der Ebene der Visualisierungen und/oder auf der Datenebene?

Anhang: Git-Historie

Literatur

- [1] Nitya verma. „Smartphone Dataset (Cleaned)“. In: (2024). URL: <https://www.kaggle.com/datasets/nityaverma19/smartphone-dataset-cleaned>.