

Einblicke in die Handynutzung: Analyse von Dauer und Gewohnheiten

Morris Baude

November 2024

1 Zielsetzung/Abstract

Sehr geehrte Damen und Herren,

bei diesem Projekt handelt es sich um ein Einsteigerprojekt im Bereich der Datenanalyse. Mein Ziel ist es, mir Fähigkeiten anzueignen und Aufgaben zu bewältigen, die den gesamten Prozess der Datenanalyse umfassen – von der Datenaufbereitung über die Datenanalyse bis hin zur Interpretation der Ergebnisse. Das Projekt soll mir dabei helfen, ein grundlegendes Verständnis für die verschiedenen Phasen der Datenanalyse zu entwickeln und praktische Erfahrungen in der Arbeit mit Datensätzen zu sammeln.

Dazu möchte ich zunächst einen Überblick über die Daten geben. Es wird erläutert, um welchen Datensatz es sich handelt und wie ich ihn aufbereitet habe, um damit arbeiten zu können. Im nächsten Schritt folgt die Analyse. Zunächst gewinnen wir einen Überblick, wie die Daten verteilt sind, um anschließend mögliche Zusammenhänge zu erkennen. Dafür betrachten wir die Daten zunächst im Jahresüberblick und analysieren dann, ob es spezifische Zusammenhänge zwischen den Wochentagen und den Daten gibt. Zu guter Letzt untersuchen wir, ob es Korrelationen zwischen den verschiedenen Datenpunkten gibt und überprüfen dies mit Hilfe einer Korrelationsmatrix.

2 Datensatz und Aufbereitung

Zur Vorbereitung dieser Analyse habe ich zwei wesentliche Aspekte bearbeitet. Zum einen die Daten zur zeitlichen Nutzung meiner Handy-Apps, die ich mit der App „StayFree“ erfasst habe. Diese App protokolliert, wie lange ich täglich eine App oder Webseite auf meinem Handy genutzt habe. Da es sich hierbei um persönliche Daten handelt, war eine Anonymisierung erforderlich. Die Daten wurden in den folgenden Schritten vorverarbeitet:

- Löschung nicht genutzter Zeilen und Spalten
- Umwandlung der Spalteneinträge in geeignete Zeitformate

- Löschen von Einträgen, bei denen die Nutzung unter einer Minute lag (-> Halbierung der Datenframegröße)
- Zusammenfassen von Webseiten (-> Reduzierung der Einträge)
- Gruppierung der Apps in Kategorien: ['Spiel', 'Sonstiges', 'System-App', 'Finanzen', 'Navigation/Transport', 'Bildung', 'Gesundheit/Fitness', 'Produktivität', 'Soziale Netzwerke', 'Unterhaltung', 'Gesamte Nutzung']
- Anonymisierung der App-Namen und Speicherung in einer Mapper-Datei zur späteren Zuordnung
- Umwandlung von absoluten Werten in relative Werte

Der anonymisierte Datensatz ist nun unter „Data/ANO_nutzungsdaten_jun_dez.csv“ zu finden (siehe auch Abbildung 1). Der zweite Aspekt umfasst die Erfassung von Daten zu Gewohnheiten. Dazu wurde ein Checkbox-Datensatz erstellt, der für jeden Tag speichert, ob eine bestimmte Gewohnheit durchgeführt wurde. Wurde die Gewohnheit erfüllt, wird der entsprechende Tag mit einer 1 markiert, andernfalls mit einer 0. Die getrackten Gewohnheiten umfassen Training, Lernen und Kreatin.

Der zweite Datensatz enthält den „Gewohnheitsscore“. Dieser Score ist abhängig davon, wie konsequent die Gewohnheit durchgeführt wurde. Wurde die Gewohnheit regelmäßig ausgeführt, erhält der Tag einen hohen Score; andernfalls einen niedrigen. Auf diese Weise wird verhindert, dass zu viele Null-Werte im Datensatz vorhanden sind.

Bei beiden Datensätzen habe ich keine zusätzliche Anonymisierung vorgenommen. Wichtig war lediglich, dass die Gewohnheiten zusammengefasst werden, da sie zunächst einzeln gespeichert wurden.

Um ein besseres Verständnis der Gewohnheitsdatensätze zu ermöglichen, habe ich in Abbildung 2 einen kleinen Ausschnitt der Gewohnheit „Lernen“ im Zeitraum von Januar bis März abgebildet.

Diese Abbildung veranschaulicht das zuvor erläuterte Prinzip: Der Score steigt, wenn die Gewohnheit über einen zusammenhängenden Zeitraum hinweg konsequent ausgeführt wird. Bei einer Unterbrechung der Ausführung sinkt der Score entsprechend.

Die beiden Datensätze sind unter "Data/gewohnheiten_cb_neu.csv" und "Data/gewohnheiten_score.csv" gespeichert und als Snapshot auf Abbildungen 3 und 4 zu sehen.

	2024-01-01	2024-01-02	2024-01-03	...	2023-12-31	Gesamte Nutzung	Gruppe
Unnamed: 0							
A1	0.000000	0.000000	0.282157	...	0.000000	0.004401	Spiel
A2	0.000000	0.000000	0.000000	...	0.000000	0.000047	Spiel
A3	0.000000	0.000000	0.000000	...	0.000000	0.000055	Sonstiges
...
A105	0.000000	0.000000	0.000000	...	0.000000	0.000057	Produktivität
A106	0.000000	0.000000	0.000000	...	0.000000	0.000179	Unterhaltung
A107	0.003779	0.008192	0.099446	...	0.032066	0.053846	Sonstiges

107 rows x 368 columns

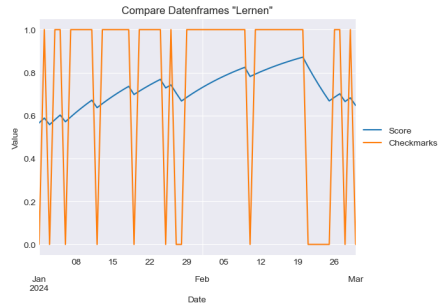


Abbildung 1: Snapshot der anonymisierten Nutzungsdaten

Abbildung 2: Vergleich Score- und Checkboxdatensatz lernen

	2024-08-17	2024-08-16	2024-08-15	...	2023-09-23	2023-09-22	2023-09-21
Unnamed: 0							
Kreatin	0	1	1	...	1	1	1
Lernen	0	0	1	...	1	1	1
Training	0	1	1	...	1	1	1

3 rows x 332 columns

	2024-08-17	2024-08-16	2024-08-15	...	2023-09-23	2023-09-22	2023-09-21
Unnamed: 0							
Kreatin	0.7172	0.7565	0.7431	...	0.1478	0.1011	0.0519
Lernen	0.6949	0.7330	0.7731	...	0.1478	0.1011	0.0519
Training	0.7331	0.7732	0.7608	...	0.1478	0.1011	0.0519

3 rows x 332 columns

Abbildung 3: Snapshot des Checkboxdatensatz für Gewohnheiten

Abbildung 4: Snapshot des Scoredatensatz für Gewohnheiten

3 Übersicht Datennutzung

Nachdem wir nun die Struktur der Daten kennen, liegt der Fokus auf der Verteilung der Nutzung. Dazu betrachten wir zunächst die gesamte Verteilung (siehe Abbildung 5).

Die Abbildung zeigt die Verteilung des Handynutzungsverhaltens für alle Apps, wobei Apps, die weniger als 2 % der Nutzung ausmachen, zusammengefasst und unter der Rubrik „Weitere“ angezeigt werden. Was sofort ins Auge fällt, ist, dass über 50 % der Handynutzung durch die drei Apps A81, A35 und A79 ausgemacht wird. Dies könnte uns später von großer Hilfe sein, da bei diesen drei Apps das Einsparungspotenzial am größten ist. Weiterhin zeigt das Kreisdiagramm, dass nur insgesamt neun Apps eine prozentuale Nutzung von mehr als 2 % aufweisen, was auf eine geringe Diversität in der Handynutzung hinweist. Der verbleibende Anteil von knapp 20 % der Nutzungszeit wird durch 97 Apps ausgemacht.

Dies wirft die Frage auf, warum so viele Apps so wenig genutzt werden. Handelt es sich hierbei möglicherweise um Fehlinstallationen? Um dieser Frage nachzugehen, wurde die Kategorie „Weitere“ näher analysiert (siehe Abbildung 6). In dieser Analyse wurden die Apps nach der Anzahl der Tage eingeteilt, an denen sie genutzt wurden. Diese Tagesintervalle wurden anschließend in die gegebenen Kategorien eingeteilt, um mögliche Muster in Bezug auf die geringe Nutzung zu

identifizieren. Das Diagramm zeigt eine sehr variable Verteilung: Es gibt Apps, die zeitlich nur wenig genutzt werden, aber dennoch an mehr als 100 Tagen aktiv waren, was auf eine gewisse Relevanz hinweist und darauf schließen lässt, dass sie für bestimmte Nutzungszwecke wichtig sein könnten. Gleichzeitig gibt es den entgegengesetzten Trend: Mit 49 Apps, also mehr als der Hälfte, existieren viele Anwendungen, die an weniger als 20 Tagen genutzt wurden. Dies deutet darauf hin, dass es sich bei diesen Apps wahrscheinlich um unwichtige Anwendungen handelt, gerade weil die meisten der Apps unter die Kategorie 'Spiel' und 'Sonstiges' fallen. Sie könnten entweder im Hype installiert worden sein, weil sie von Freunden genutzt wurden oder populär waren, sich aber später als wenig nützlich erwiesen haben. Alternativ könnten es spezialisierte Apps für selten auftretende Aufgaben sein. Ein Beispiel hierfür ist die „McDonald's“-App, die nur an wenigen Tagen sinnvoll, aber auch dann nur kurz genutzt wird.

Wie bereits bemerkt, habe ich im Vorfeld alle Apps in Kategorien eingeteilt, um zusätzliche Erkenntnisse zu erzielen. So habe ich erneut ein Kreisdiagramm für die Kategorien erstellt. Es lässt sich bereits vorab sagen, dass die Gruppen, die A81, A35 oder A79 beinhalten, auch in der neuen Aufteilung die Mehrheit ausmachen werden. In Abbildung 7 ist besonders gut zu erkennen, dass die Kategorien „Unterhaltung“ und „Soziale Netzwerke“ jeweils über 30 % der Nutzung ausmachen. Diese Kategorien umfassen auch die drei am häufigsten genutzten Apps. Interessant ist zudem, dass sich eine weitere Gruppe hervorgehoben hat: die Kategorie „Spiele“, die mit 15 % auffällt. Diese Gruppe umfasst 23 Apps und ist damit nach der Kategorie „Sonstige“ die zweitgrößte Gruppe. Auffällig ist, dass in dieser Kategorie viele kleine Nutzungen zusammen einen bedeutenden Anteil ausmachen. Alle weiteren Kategorien tragen jeweils weniger als 5 % zur Gesamtnutzung bei und werden daher in der weiteren Analyse eher außer Acht gelassen.

Nachdem wir nun einen Überblick über das Nutzungsverhalten gewonnen haben, richten wir den Fokus auf die Gewohnheiten. Das Datenframe enthält, wie bereits beschrieben, die Gewohnheiten „Lernen“, „Sport“ und „Kreatin“. Dabei steht „Kreatin“ für ein Nahrungsergänzungsmittel, das häufig nach dem Sport eingenommen wird, um positive Effekte zu erzielen. Da eine konsequente Einnahme entscheidend ist, wurde diese Gewohnheit separat erfasst und getrackt. Die Daten umfassen 332 Tage, was die maximale Anzahl an möglichen Tagen für die Ausführung der Gewohnheiten darstellt. Betrachtet man Abbildung 8, zeigt sich, dass die Gewohnheit „Lernen“ mit 224 Tagen die häufigste ist. Das bedeutet, dass sie an zwei Dritteln aller Tage ausgeführt wurde. Dieser Wert lässt sich auch im Kontext einer Woche interpretieren: Mit einem Vollzeitstudium, bei dem an 5 von 7 Tagen pro Woche gelernt werden sollte, ist eine Quote von zwei Dritteln aller Tage durchaus akzeptabel. Allerdings bietet der Datensatz keine Informationen darüber, wie lange tatsächlich gelernt wurde. Daher lässt sich der Erfolg der Gewohnheit lediglich anhand der Häufigkeit der erledigten Tage beurteilen. Ein interessanter Fakt ist, dass Kreatin an 192 Tagen eingenommen wurde, obwohl nur an 187 Tagen Training absolviert wurde. Eigentlich könnte man erwarten, dass Kreatin seltener verwendet wird, da es vor allem an Trainingstagen seine volle Wirkung entfaltet. Die Diskrepanz lässt sich jedoch

vermutlich dadurch erklären, dass eine konsequente Einnahme wichtiger ist als die reine Verwendung an Trainingstagen. Daher wurde Kreatin wahrscheinlich auch an Regenerationstagen konsumiert.

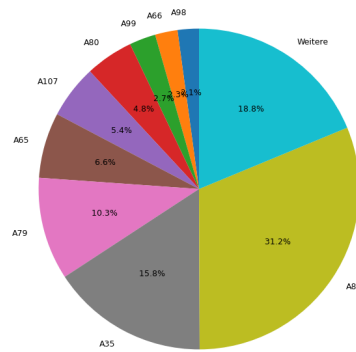


Abbildung 5: Nutzungsverteilung der Apps

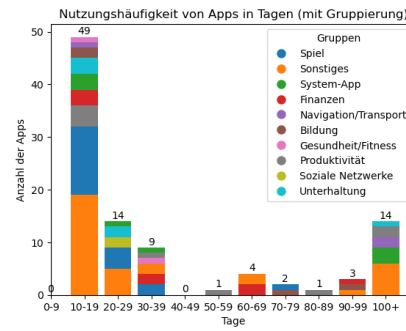


Abbildung 6: Nutzungsanzahl der Apps aus Kategorie 'Weitere'

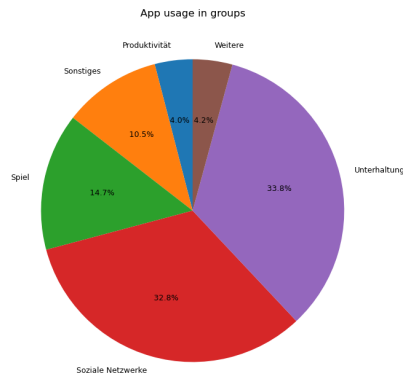


Abbildung 7: Nutzungsverteilung der Kategorien

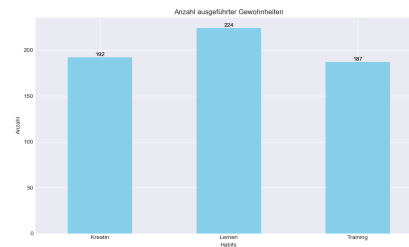


Abbildung 8: Jahresanzahl Gewohnheiten

4 zeitliche Entwicklungstrends

Legen wir nun den Fokus auf die Verteilung der Daten im Verlauf eines Jahres. In dem Boxplot (siehe Abbildung 9) sind die täglichen Gesamtzeiten der

Handynutzung über ein Jahr dargestellt. Auch wenn es ohne konkrete Zahlen schwierig ist, detaillierte Aussagen zu treffen, ermöglicht der Boxplot eine schrittweise Analyse der Verteilung. Es ist gut zu erkennen, dass die Daten eine hohe Streuung aufweisen. Nach unten hin sind sie durch 0 Minuten begrenzt – dies entspricht Tagen, an denen das Handy überhaupt nicht genutzt wurde. Auffällig ist, dass 50 % aller Nutzungszeiten relativ dicht beieinander liegen, was durch die kompakte Länge der Box im Diagramm verdeutlicht wird. Nach oben hin zieht sich der Wertebereich deutlich weiter auseinander als nach unten. Dies deutet darauf hin, dass es zwar weniger Tage gibt, an denen das Handy besonders lange genutzt wurde, die Variation der Nutzungsdauer an diesen Tagen jedoch groß ist. Dieses Phänomen lässt sich durch den größeren Spielraum nach oben erklären: Während die Nutzung nach unten durch 0 Minuten begrenzt ist, gibt es nach oben nur die natürliche Grenze von 24 Stunden pro Tag.

Nachdem die einzelnen Programme im Vorfeld kategorisiert wurden, ist es nun einfacher, aussagekräftige Analysen über größere App-Mengen durchzuführen. Aus diesem Grund konzentriert sich die weitere Analyse der Nutzungsdaten auf die Kategorien.

Zunächst wurde das Nutzungsverhalten der drei Hauptkategorien „Unterhaltung“, „Soziale Netzwerke“ und „Spiele“ als Liniendiagramm dargestellt (siehe Abbildung 10). Dabei fällt auf, dass es im November einen deutlichen Einschnitt gibt, mit einer kurzen Periode von Tagen, an denen das Handy gar nicht genutzt wurde. Wie zu erwarten war, wechseln sich die Kategorien „Unterhaltung“ und „Soziale Netzwerke“ als die meistgenutzten Kategorien am Tag ab, da sie zusammen 60 % der Gesamtnutzung ausmachen. Eine weitere Beobachtung zeigt, dass zwischen August und Dezember die Kategorie „Unterhaltung“ vorrangig am meisten genutzt wurde, während sich dies ab Januar zugunsten der „Sozialen Netzwerke“ veränderte. Interessant ist, dass dieser Trend nicht stabil bleibt: Ab Juli 2024 liegen die Unterhaltungsprogramme wieder vorne. Eine mögliche Erklärung könnte in der saisonalen Nutzung liegen: Während im Frühling und Sommer Soziale Netzwerke vermehrt genutzt werden – möglicherweise für das Teilen von Fotos und Erlebnissen –, scheint im Herbst und Winter die Nutzung von Unterhaltungsprogrammen zuzunehmen, da die Tage trüber werden und Nutzer sich eher gemütlich unterhalten lassen möchten. Für eine präzise Interpretation reicht jedoch das Tracking von lediglich einem Jahr nicht aus.

Im nächsten Schritt wird ein genauerer Blick auf die drei Hauptkategorien geworfen, um zu untersuchen, ob sich innerhalb der Gruppen ebenfalls Trends erkennen lassen und wie sich das Nutzungsverhalten möglicherweise dadurch erklären lässt.

Beziehen wir uns zunächst auf die Kategorie „Unterhaltung“, die mit 33,8 % die meistgenutzte war. In Abbildung 11 ist deutlich zu erkennen, dass eine App (blau markiert) die Kategorie dominiert. Ihre Nutzung zeigte im Verlauf des Jahres starke Schwankungen, und ab Anfang Mai wurde eine zeitliche Beschränkung integriert. Wie man im Diagramm sieht, erhält der Graph ab diesem Zeitpunkt eine Art obere Schranke, die nicht mehr überschritten wird, wodurch die Nutzung der App nach Mai deutlich abflacht. Ab Mitte Juli kam dann eine neue App (orange markiert) hinzu, was auch den starken Anstieg in der Kate-

gorie „Unterhaltung“ ab diesem Zeitpunkt erklärt. Generell scheint eine zeitliche Beschränkung einen signifikanten Einfluss auf die Nutzung zu haben, da sie relativ konstant eingehalten wird. Es ist jedoch wichtig, darauf zu achten, dass der zeitliche Gewinn nicht einfach durch den Einsatz einer anderen App ersetzt wird.

Kommen wir nun zum Jahresüberblick der Kategorie „Soziale Netzwerke“ (siehe Abbildung 13). Es fällt auf, dass sich die Nutzung der Apps bis Anfang Dezember stark in Grenzen hielt, da zu diesem Zeitpunkt nur eine App verwendet wurde, die einen großen Anteil an der Nutzung der Kategorie ausmachte (orange markiert). Ab Dezember kam eine weitere App (blau) hinzu, die eine signifikante Nutzungszeit aufwies. Diese App erreichte Ende Januar den Höhepunkt der gesamten Kategorie. Die Nutzung der blauen App war generell höher als die der orangenen App, flachte jedoch ab Mitte April wieder ab. Auch hier scheint die zeitliche Beschränkung für die blaue App gut funktioniert zu haben, sodass es den Anschein hat, dass das hohe Nutzungsverhalten von Dezember bis April danach wieder unter Kontrolle gebracht wurde. Die letzte App, die grün markiert ist, begann im Februar ihre Nutzung und erreichte dort ebenfalls ihren Höhepunkt. Danach pendelte sich die Nutzung auf einem mittleren Niveau ein. Trotz der anfangs hohen Nutzungswerte, die gegen Ende der Aufzeichnungen sanken, sieht die Entwicklung insgesamt positiv aus. Es ist jedoch wichtig, sich nicht täuschen zu lassen, da am Ende drei Apps mit mittlerer Aktivität zu sehen sind, was im Vergleich zu Beginn, als die Nutzung nur weniger Apps ausgeprägt war, eine Veränderung darstellt. Mit Rückblick auf Abbildung 10 zeigt sich trotzdem, dass insgesamt ein erfolgreicher zeitlicher Rückgang der Nutzung in der Kategorie „Soziale Netzwerke“ stattgefunden hat.

Die letzte Kategorie, die wir näher beleuchten, ist die „Spiel“-Kategorie (siehe Abbildung 12). In dieser Kategorie war der größte Nutzen der Vorverarbeitung sichtbar, als die Anzahl der angezeigten Spiele auf die sechs am meisten gespielten begrenzt wurde. Es gab jedoch auch zahlreiche weitere Spiele, die nach einem starken „Nutzungshype“ fast nicht mehr genutzt wurden. Diese Tendenz lässt sich in fast jedem Spiel im Diagramm erkennen: Meistens starten die Apps mit einem sehr starken Anstieg der Nutzungsdauer (siehe Rot, Lila, Braun, Blau), flachen dann aber im Laufe der Zeit ab, bis sie nahezu gar nicht mehr genutzt werden. Es fällt auch auf, dass es keine Apps gibt, die sich – wie bei den „Sozialen Netzwerken“ – über einen langen Zeitraum hinweg halten und kontinuierlich genutzt werden. Stattdessen werden die Spiele hauptsächlich phasenweise gespielt. Dies zeigt, dass in der Kategorie „Spiele“ großes Einsparungspotenzial besteht. Man sollte sich im Vorfeld ernsthaft fragen, ob es sich lohnt, eine App herunterzuladen, da diese zu Beginn oft ein starker Zeitfresser ist, während das Interesse anscheinend schnell wieder verfliegt. Kaum ein Spiel hält sich länger als sechs Monate, mit Ausnahme vielleicht des violetten Spiels, welches jedoch nur eine geringe Nutzung zwischen Mai und Juni aufweist.

Schauen wir uns nun den Verlauf der Gewohnheitsscores im Jahresüberblick an (siehe Abbildung 14). Ein hoher Score deutet hierbei auf eine konsequente Ausführung der Gewohnheit hin, während ein niedriger Score auf weniger regelmäßige Einnahme hindeutet. Am auffälligsten ist der Kreatin-Graph (blau),

da er die größten Schwankungen aufweist. Er war am höchsten Mitte Februar, bevor er wieder stark einbrach. Anfang Juli erreichte er dann seinen absoluten Tiefpunkt, bei dem fast keine Kreatin-Einnahme stattfand. Interessanterweise hätte ich erwartet, dass dieser Einbruch mit einer geringeren Trainingsfrequenz einhergeht, jedoch war dies nicht der Fall. Abgesehen von einem Einbruch im Dezember zeigt der Trainings-Graph eine starke und phasenweise gut ausgeprägte Ausführung der Gewohnheit, wobei auch hier viele Schwankungen zu beobachten sind. Der 'Lernen'-Graph (orange) weist weniger starke Schwankungen auf. Mit einem Score von meist über 50 % bleibt der Graph kontinuierlich in der oberen Hälfte des Diagramms. Eine mögliche Erklärung hierfür ist, dass Lernen durch Uni und Arbeit eine verpflichtende Aufgabe ist, während Training und Kreatin eher freizeitliche Verpflichtungen darstellen. Positiv zu vermerken ist, dass sich alle Gewohnheiten gegen Ende des Zeitraums stark verbessert haben, was für einen gesunden Lebensstil und eine positive Entwicklung spricht. Sollte dieser Trend anhalten, bin ich mit der Ausführungsquote aller Gewohnheiten sehr zufrieden.

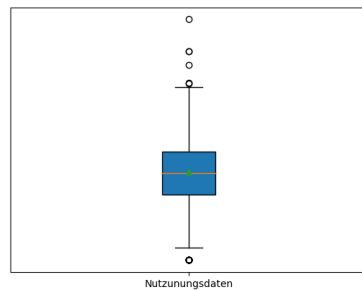


Abbildung 9: Boxplot Nutzungsdaten

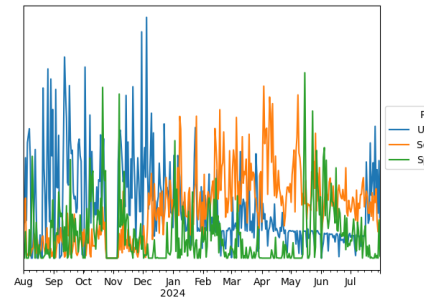


Abbildung 10: Jahresübersicht Topkategorien

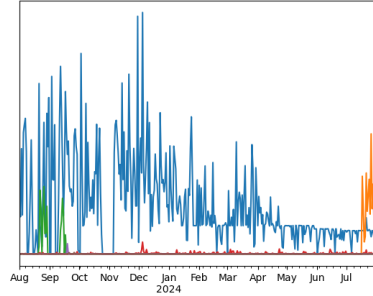


Abbildung 11: Jahresübersicht Kategorie 'Unterhaltung'

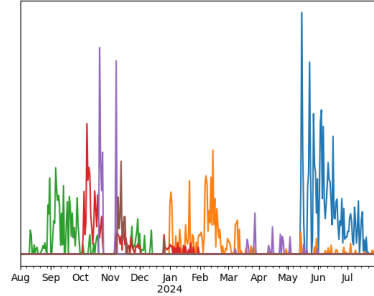


Abbildung 12: Jahresübersicht Kategorie 'Spiel'

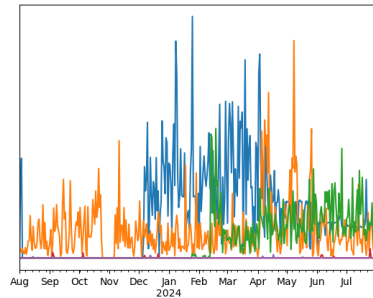


Abbildung 13: Jahresübersicht Kategorie 'Soziale Netzwerke'

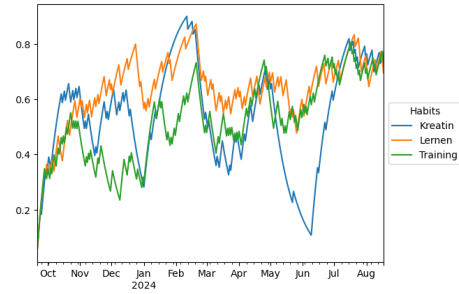


Abbildung 14: Jahresübersicht Gewohnheiten

5 Zusammenhang mit Wochentagen

Um ein tieferes Verständnis für die vielen, kleinen Schwankungen im Jahresverlauf zu erhalten, fokussiert sich der nachfolgende Teil auf den Zusammenhang zwischen Wochentagen und der zeitlichen Nutzung sowie zwischen Wochentagen und den Gewohnheiten.

Dazu betrachten wir in Abbildung 15 die zeitliche Nutzung der Kategorien über eine gesamte Woche. Hierbei können nur Aussagen über die ersten vier Graphen getroffen werden, da die anderen so geringe Schwankungen aufweisen, dass es schwer ist, darüber passende Aussagen zu treffen. Die Kategorie 'Unterhaltung' wird an den Tagen Montag, Dienstag und Sonntag am meisten genutzt und erreicht ihren Höhepunkt am Dienstag. Wahrscheinlich sind dies auch die Tage, an denen weniger soziale Aktivitäten stattfinden und man sich lieber ausruht,

während man sich unterhalten lässt. Von Mittwoch bis Samstag übernimmt dann die Kategorie 'Soziale Netzwerke' die Rolle als meistgenutzte Kategorie. Diese Tage liegen näher am Wochenende, was damit zusammenhängen könnte, dass es um die Planung des Wochenendes oder andere soziale Aktivitäten geht. Zudem schwankt die Nutzung dieser Kategorie über die Woche weniger stark, was auf eine gleichmäßige und konstante Nutzung der Apps im Verlauf der Woche hindeutet. Bei dem grauen Graphen, der die Kategorie 'Spiel' darstellt, sieht man einen ähnlichen Trend wie bei der 'Unterhaltung'. Während die Nutzung von Montag bis Donnerstag ansteigt, fällt sie am Wochenende wieder auf ein Minimum, wobei der Tiefpunkt am Samstag liegt. Dies könnte ähnliche Gründe wie bei 'Unterhaltung' haben, da man an den Wochenendtagen eher anderen Aktivitäten nachgeht, anstatt am Handy zu spielen. Die letzte Kategorie, 'Sonstiges', ist besonders interessant, da sie viele Schwankungen im Verlauf der Woche verzeichnet. Es wäre hilfreich zu erfahren, welche App für diese starken Schwankungen verantwortlich ist, da ohne diese Information keine verlässlichen Aussagen über die Ursachen der Schwankungen getroffen werden können. Schauen wir uns nun die Gewohnheiten und ihre Veränderung der Ausführung über die Woche in Abbildung 16 an. Was hier besonders auffällt, ist der orangefarbene Graph, der das 'Lernen' repräsentiert. Diese Gewohnheit wurde von Montag bis Donnerstag sehr häufig ausgeführt, sinkt jedoch am Samstag auf ihr Minimum, mit weniger als 15 ausgeführten Tagen. Dies zeigt einen deutlichen Unterschied zwischen der Woche und dem Wochenende. Dieser Rückgang lässt sich leicht damit erklären, dass während der Woche Uni und Arbeit anstanden, während man sich am Wochenende eher ausgeruht hat. Die Schwankungen im Graphen über das Jahr hinweg können ebenfalls durch die Wochenendpausen erklärt werden, da die Einbrüche vermutlich durch die Wochenenden stattfanden. Der 'Training'-Graph (grün) weist seine Tiefpunkte am Mittwoch und Freitag auf, während an den anderen Tagen häufiger trainiert wurde. Dies war im Vorfeld eigentlich schon zu erwarten, da mein Fußballtraining am Dienstag und Donnerstag stattfand und die Spieltage auf Samstag oder Sonntag fallen. In den ersten drei Wochentagen orientiert sich der Kreatin-Graph stark an den Trainingsausführungen, was darauf zurückzuführen ist, dass man nach dem Training Kreatin eingenommen hat. Danach spalten sich jedoch die beiden Graphen, was schwer zu interpretieren ist. Besonders auffällig ist, dass am Donnerstag der Tag war, an dem am meisten Kreatin eingenommen wurde, während die Nutzung bis Sonntag stark zurückging. Es bleibt unklar, warum gerade am Donnerstag ein so hoher Anstieg der Kreatineinnahme zu verzeichnen war und warum der Wert bis zum Sonntag drastisch fiel. Interessanterweise war die Häufigkeit der Kreatineinnahme am Samstag genauso hoch wie am Mittwoch, jedoch wird dieser Punkt durch den starken Rückgang nach Donnerstag weniger sichtbar.

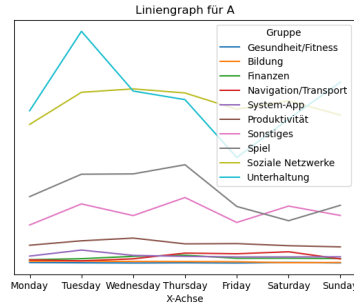


Abbildung 15: Wochenübersicht Nutzungskategorien

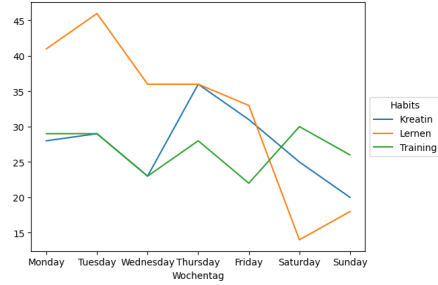


Abbildung 16: Wochenübersicht Gewohnheiten

6 Korrelationsprüfung

Schauen wir uns nun an, wie der Zusammenhang zwischen den Kategorien und den Gewohnheiten aussieht. Dazu berechnen wir die Korrelation zwischen jeder Gewohnheit und jeder Kategorie. Der Korrelationswert beschreibt, wie stark der Zusammenhang zwischen zwei Variablen ist. Man unterscheidet dabei zwischen positiven, negativen und keinen Relationen:

- **Positive Korrelation:** Wenn eine Variable steigt, steigt auch die andere. Eine perfekte positive Korrelation wird mit 1 abgebildet und bedeutet, dass beide Variablen gleichzeitig steigen oder fallen.
- **Negative Korrelation:** Wenn eine Variable steigt, sinkt die andere. Eine perfekte negative Korrelation wird mit -1 abgebildet.
- **Keine Korrelation:** Wenn die Variablen keinerlei Zusammenhang haben, ergibt sich ein Wert von 0.

Die gängige Methode zur Berechnung der Korrelation ist die Pearson-Korrelation, die in Python standardmäßig verwendet wird. Sie wird durch die folgende Formel berechnet:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Dabei sind:

- x_i und y_i die einzelnen Werte der beiden Variablen.
- \bar{x} und \bar{y} die Mittelwerte der Variablen x und y .

Die Pearson-Korrelation misst also, wie gut die lineare Beziehung zwischen zwei Variablen ist. Ein Wert nahe 1 oder -1 zeigt einen starken Zusammenhang an, während ein Wert nahe 0 auf keinen linearen Zusammenhang hinweist. Nun bleibt nur noch die Frage, ob es Kategorien gibt, die die Ausführung von bestimmten Gewohnheiten begünstigen oder sogar entgegenwirken.

Kleiner Reminder: Nur weil es eine Korrelation gibt, heißt es nicht, dass es auch die Kausalität ist. Daher stellen die nachfolgenden Aussagen eher Behauptungen als Erklärungen dar.

Bei der Erstellung der Korrelationsmatrix habe ich den Score der Gewohnheiten verwendet und alle möglichen Kombinationen der Gewohnheiten berücksichtigt. Dies wurde mit der Intention gemacht, dass beispielsweise die Gewohnheit „Lernen“ alleine keine Korrelation mit einer Kategorie aufweist. Wenn man jedoch diese Gewohnheit mit einer anderen kombiniert, kann eine Korrelation entstehen. Zu diesem Zweck habe ich den Durchschnitt der jeweiligen Kombinationen berechnet, wobei nur die Gewohnheiten berücksichtigt wurden, die darin enthalten sind. Die Matrix war aufgrund der 7 Gewohnheiten und 10 Kategorien sehr unübersichtlich, was zu insgesamt 70 Einträgen führte, die ausgewertet werden mussten. Auffällig war, dass bei den meisten Kombinationen keine signifikante Korrelation auftrat. Deshalb wurden nur die Zeilen und Spalten berücksichtigt, die einen Korrelationswert von mindestens 0,4 aufwiesen. Wichtig ist, dass es sich bei den Kategorien um relative Daten handelt. Das bedeutet, dass eine Korrelation zwischen einer Gewohnheit und einer Kategorie nicht darauf hinweist, dass die Nutzungszeit der Kategorie steigt, wenn die Gewohnheit häufiger ausgeführt wird. Vielmehr bezieht sich dies auf eine prozentuale Zunahme der Nutzung der Kategorie im Verhältnis zur Gesamt-Nutzung.

So führt die Abbildung 17 zu überraschenden Erkenntnissen. Die stärkste negative Korrelation besteht zwischen der Gewohnheit „Lernen“ und den produktiven Apps. Auch in Kombination mit „Training“ entsteht eine negative Korrelation von -0,46, was darauf hindeutet, dass an den Tagen, an denen man trainiert oder lernt, weniger mit produktiven Apps arbeitet. Im Gegensatz dazu zeigt sich eine positive Korrelation zwischen der Nutzung von sozialen Netzwerken und den Gewohnheiten „Training“ sowie „Lernen und Training“. Dies ist ebenfalls sehr interessant, da sich dies ebenfalls schwer erklären lässt.

In der Folge habe ich noch einmal die absoluten Werte auf ihre Korrelation mit den Gewohnheiten geprüft, jedoch konnten hier keine neuen Erkenntnisse gewonnen werden.

Mit der Prüfung auf Korrelationen zwischen den Kategorien konnten ebenfalls keine aussagekräftigen Ergebnisse erzielt werden. Bei den relativen Werten ist klar, dass eine Erhöhung der relativen Nutzung einer Kategorie zwangsläufig dazu führen muss, dass sich die Nutzung aller anderen Kategorien verringert, da es nur einen Gesamt-nutzungsanteil von 100 Prozent gibt. Da die Kategorien „Unterhaltung“ und „Soziale Netzwerke“ den größten prozentualen Anteil an der Nutzung ausmachen, weisen sie auch die stärkste negative Korrelation auf. Dies liegt daran, dass die Zeit für die Handynutzung an einem Tag begrenzt ist. Wenn also die Nutzung von Unterhaltungs-Apps steigt, sinkt häufig der Anteil der Nutzung von sozialen Netzwerken. Auch bei über Korrelationen zwischen

den absoluten Nutzungswerten konnte keine neuen Erkenntnisse gewonnen werden.

Zum Schluss habe ich die Korrelationen zwischen den Gewohnheiten näher untersucht. Ich hatte erwartet, dass die größte positive Korrelation zwischen der Einnahme von Kreatin und Sport besteht, während beim Lernen viele verschiedene Ergebnisse möglich wären. Die Korrelationsmatrix (Abbildung 18) hat mich jedoch eines Besseren belehrt: Entgegen meinen Erwartungen war die Korrelation zwischen Kreatin und Training die schwächste der drei, während die stärkste Korrelation zwischen Kreatin und Lernen besteht. Das bedeutet, je regelmäßiger ich Kreatin einnehme, desto konstanter bin ich auch beim Lernen. Positiv hervorzuheben ist, dass alle Korrelationen positiv sind, was darauf hindeutet, dass keine negativen Einflüsse zwischen den Gewohnheiten bestehen. Man kann also davon ausgehen, dass, wenn man eine Gewohnheit an einem Tag ausführt, die Wahrscheinlichkeit hoch ist, dass man auch eine weitere Gewohnheit abschließt. Insgesamt bin ich mit den Erkenntnissen aus den Korrelationen nicht ganz zufrieden, da sich keine tiefgehenden Aussagen oder Erklärungen treffen lassen. Dies könnte auch an dem begrenzten Datensatz liegen.

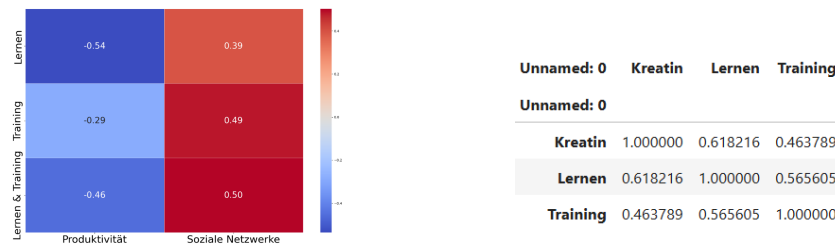


Abbildung 17: Ausschnitt Korrelationsmatrix Nutzungskategorien X Gewohnheiten

Abbildung 18: Korrelationsmatrix Gewohnheiten

7 Fazit

Kommen wir nun zum Fazit: Was kann man aus der Analyse mitnehmen, welche nützlichen Erkenntnisse konnten gewonnen werden und was lässt sich für die Zukunft ableiten?

Im ersten Abschnitt wurde deutlich, dass mehr als 50 Prozent der Nutzung auf drei Apps zurückzuführen sind. Bei diesen Apps sollte nun überlegt werden, wie deren Nutzungsdauer eingeschränkt werden kann. Wie später gezeigt, wäre eine Nutzungsbeschränkung sinnvoll, da sie einen positiven Effekt hat. Des Weiteren gab es fast 50 Apps, die über das Jahr hinweg kaum genutzt wurden. Es ist ratsam, sich einen Überblick über diese Apps zu verschaffen und zu überlegen, ob sie gelöscht werden können, falls dies noch nicht geschehen ist. Für die Zukunft sollte man sich vor der Installation einer neuen App generell die Frage stellen, ob sie wirklich notwendig ist. Außerdem könnte man überlegen, ob es

zwischen Sonntag und Mittwoch eine Möglichkeit gibt, sich zu erholen, ohne auf Unterhaltungs-Apps zurückzugreifen. Die Einschränkung von sozialen Netzwerken wäre für mich jedoch nicht der richtige Weg, da der Kontakt und Austausch mit anderen Menschen wichtig sind. Es wäre höchstens sinnvoll zu prüfen, ob die sozialen Netzwerke hauptsächlich für den Kontakt genutzt werden oder ob hier Einsparmöglichkeiten bestehen.

Bei den Gewohnheiten kann man durchaus stolz auf den Fortschritt sein. Besonders der Aufwärtstrend aus Abschnitt zwei zeigt deutlich, dass man hier zufrieden sein kann. Wichtig ist nun, diese Konstanz beizubehalten und vielleicht auch an den Wochenenden zu überlegen, ob man dort der ein oder anderen Gewohnheit nachgehen sollte, um die Konstanz weiter zu steigern. Aus den Korrelationen zwischen den Gewohnheiten lässt sich außerdem ableiten, dass es vorteilhaft ist, mindestens einer Gewohnheit pro Tag nachzugehen, da dies wahrscheinlich auch zu einer verstärkten Ausführung anderer Gewohnheiten führt. Wie bereits erwähnt, lassen sich jedoch anhand der Korrelationen keine weiteren tiefgehenden Aussagen treffen, sodass wir am Ende der gewonnenen Erkenntnisse angekommen sind.

Ein Ausblick auf mögliche zukünftige Erweiterungen für dieses Projekt bietet verschiedene interessante Ansätze. Eine naheliegende Möglichkeit ist die Erweiterung des Datensatzes, insbesondere durch das Tracking der tageszeitlichen Nutzung. Dies würde es ermöglichen, herauszufinden, zu welchen Uhrzeiten welche Apps am meisten genutzt werden, was wiederum weitere Optimierungsmöglichkeiten aufzeigt. In Kombination mit den Wochentagen könnten hier möglicherweise sehr interessante Muster und Zusammenhänge entdeckt werden. Ein weiterer Ansatz ist die Einbeziehung externer Faktoren. Dies könnte auf einfache Weise durch die Erweiterung des Gewohnheitsdatensatzes um weitere Gewohnheiten geschehen, um so bessere Erklärungen für die Nutzung bestimmter Apps zu erhalten. Eine komplexere Variante wäre es, Faktoren wie das Wetter oder die Stimmung einer Person zu berücksichtigen, um zu prüfen, wie diese das Nutzungsverhalten beeinflussen. Darüber hinaus wäre es interessant, einen Vergleichsdaten-Satz zu verwenden, um zu prüfen, ob das eigene Nutzungsverhalten in bestimmten Bereichen Ähnlichkeiten mit dem anderer aufweist oder ob es signifikante Unterschiede gibt. Hierfür würde bereits der Datensatz einer weiteren Person ausreichen, aber auch ein Vergleich mit einem globalen Durchschnitt, etwa in Bezug auf App-Nutzung oder Handynutzung, könnte aufschlussreiche Ergebnisse liefern. Solche Daten werden zum Beispiel von Apps wie StayFree zur Verfügung gestellt. Schließlich könnte auch eine genauere Analyse des Nutzungsmusters weitere interessante Einsichten bieten. Vorausgesetzt, man kann die tageszeitliche Nutzung tracken, lässt sich untersuchen, ob bestimmte Apps als „Vorreiter“ oder „Ursache“ für die Nutzung anderer Apps fungieren. Zum Beispiel könnte man herausfinden, ob die Nutzung einer produktiven App häufig von der Nutzung einer Unterhaltungs-App gefolgt wird, was auf kleine, wiederkehrende Gewohnheiten hinweisen könnte.

Persönliches Fazit:

Abgesehen von den inhaltlichen Erkenntnissen hat mich dieses Projekt auch

in anderen Bereichen weitergebracht. Ich habe gelernt, wie man Daten effektiv vorverarbeitet und bereinigt, um sensible Informationen aus dem Datensatz auszuschließen. Zudem konnte ich meine Fähigkeiten in der Datenvisualisierung erweitern, indem ich mich mit verschiedenen Diagrammtypen auseinandergesetzt habe. Dadurch konnte ich besser beurteilen, bei welchen Arten von Daten unterschiedliche Visualisierungen hilfreich und aussagekräftig sind und wann sie möglicherweise keine klaren Aussagen liefern. Auch meine Kenntnisse in der Datenanalyse und -interpretation haben sich dadurch verbessert. Neben den technischen Aspekten habe ich auch gelernt, wie man ein Projekt strukturiert und plant, worauf man achten muss und wie man einen „roten Faden“ entwickelt, an dem man sich im Verlauf des Projekts orientieren kann. Alles in allem war es eine sehr interessante und herausfordernde Aufgabe, die mir wertvolle Erfahrungen im Bereich der Datenanalyse vermittelt hat.