# Marketing-KPI-Analyse und Modellinterpretation mit SHAP

Ein datenanalytisches Lernprojekt zur Bewertung von Kampagnenerfolgen

**Autor**: Sefer Adiyaman  
**Projektzeitraum:** 08. Juli – 22. Juli 2025  
**Verwendete Tools:** Python (pandas, matplotlib, shap), Excel, Jupyter Notebook

# 🔹 1. Zielsetzung

Ziel dieser Analyse ist die Entwicklung einer künstlich konstruierten, datengetrieben gewichteten Gesamtkennzahl (Performance Score). Dieser Score vereint mehrere zentrale Marketing-KPIs in einem einzigen Wert und ermöglicht damit eine objektive, vergleichbare und handlungsrelevante Bewertung von Kampagnen – unabhängig von isolierten Einzelmetriken oder subjektiver Einschätzung.  
  
Gleichzeitig bildet der Score die Grundlage für ein Executive Dashboard, das speziell für Personen konzipiert ist, die beim durchführenden Unternehmen Kampagnen betreuen und Kundenverantwortung tragen („Executives“).  
  
Dieses Dashboard erlaubt es:  
• einen schnellen Überblick über Kampagnen zu gewinnen,  
• Detailanalysen der KPIs und Scores durchzuführen,  
• und direkt mit dem Kunden in Kontakt zu treten.  
  
Das Projekt verfolgt somit zwei zentrale Ziele:  
1. Konstruktion eines datenbasierten, künstlichen Performance Scores  
2. Entwicklung eines benutzerfreundlichen Dashboards zur Entscheidungsunterstützung im Kampagnenmanagement

# 🔹 2. Datenbasis und Kontext

Der Datensatz enthält Informationen zu 1.000 simulierten Marketingkampagnen. Er kombiniert:

* Metadaten (z. B. Kampagnenname, Laufzeit, Kanal)
* Kundendaten (z. B. Ansprechpartner, Kontaktdaten)
* Leistungskennzahlen (z. B. Views, Clicks, Conversions, Expense)

Eine Auswahl zentraler Spalten:

|  |  |
| --- | --- |
| Spalte | Beschreibung |
| Campaign Number | Eindeutige ID jeder Kampagne |
| Name | Kampagnenbezeichnung |
| Start/End Date | Kampagnenzeitraum |
| Channel | Verwendeter Marketingkanal |
| Client | Name des Kunden |
| Address | Adresse des Kunden |
| Email/Phone | Kontaktdaten des Kunden |
| Contact | Kundenansprechpartner |
| Audience | Zielgruppe der Kampagne |
| Location | Zielland der Kampagne |
| Views, Likes | Sichtkontakte, Engagements |
| Clicks, Conversions | Reaktionen und Zielerreichung |
| Expense | Kampagnenkosten in USD |
| Executive | Zuständige interne Kontaktperson |
| CR | Conversion Rate (gegeben: Conversions / Clicks) |

# 🔹 3. Datenbereinigung

Die Daten wurden vor der Analyse in Google Spreadsheets systematisch bereinigt, um eine konsistente Weiterverarbeitung zu ermöglichen. Die Maßnahmen umfassten:

|  |  |
| --- | --- |
| Bereinigungsschritt | Zweck / Wirkung |
| Entfernen des Zeitstempels | Datumfelder enthielten stets 00:00:00 |
| Vereinheitlichung der Clientnamen | Für konsistente Gruppierung und Zuordnung |
| Atomisierung der Adressdaten | Zerlegung in Street, City, ZIP, Country |
| Vereinheitlichung von Telefonnummern | Verbesserung der Lesbarkeit und Verarbeitung |
| Trennung der Zielgruppe | Audience wurde in Age und Gender unterteilt |

Diese Bereinigung diente als Basis für die spätere KPI-Konstruktion sowie für Filter- und Kontaktfunktionen im Executive Dashboard.

# 🔹 4. Feature Engineering

Zur Vorbereitung der Analyse wurden auf Basis der bereinigten Daten fünf zentrale Marketing-KPIs berechnet. Diese wurden mithilfe der Programmiersprache Python (pandas) erzeugt und dienen als Grundlage für den späteren datengetriebenen Performance Score.

|  |  |  |
| --- | --- | --- |
| KPI | Formel | Beschreibung |
| ER (Engagement Rate) | (Likes + Conversions) / Views | Anteil der Views, die zu Interaktionen führten |
| CTR (Click-Through Rate) | Clicks / Views | Anteil der Views, die zu Klicks führten |
| CPC (Cost per Click) | Expense / Clicks | Durchschnittliche Kosten pro Klick |
| CPA (Cost per Acquisition) | Expense / Conversions | Durchschnittliche Kosten pro Conversion |
| CR (Conversion Rate) | Conversions / Clicks | Anteil der Klicks, die zu Conversions führten (bereits im Originaldatensatz vorhanden) |

Die KPI-Konstruktion erfolgte vollständig in Python mittels pandas-Operationen. Die resultierenden Werte wurden als neue Spalten im Datensatz ergänzt und dienen in den folgenden Schritten als Inputvariablen zur Gewichtungsanalyse und Score-Erstellung.

# 🔹 5.1 Künstliche Gewichtung mit Random Forest

Da im Rahmen dieser Analyse kein echter, beobachteter Zielwert wie etwa Umsatz oder ROAS zur Verfügung steht, wurde eine künstliche Zielgröße verwendet: die Anzahl an Conversions. Diese dient rein funktional dazu, durch modellgestützte Verfahren eine datengetriebene Gewichtung der KPIs zu ermöglichen.  
  
Dabei ist wichtig zu betonen, dass zwei der unabhängigen Variablen – CPA (Cost per Acquisition) und CR (Conversion Rate) – rechnerisch direkt mit der Zielgröße Conversions verknüpft sind:  
• CPA = Expense / Conversions  
• CR = Conversions / Clicks  
  
Dieser Zusammenhang ist methodisch bekannt und entspricht der sogenannten "Target-Leakage"-Problematik – also der Verwendung von Merkmalen, die die Zielgröße direkt enthalten oder beeinflussen. In dieser Analyse wird dieses Vorgehen bewusst in Kauf genommen, da kein echter Zielwert vorliegt und der Fokus auf der gewichteten Bewertung der KPIs liegt.  
  
Ziel ist es nicht, ein unabhängiges Vorhersagemodell zu entwickeln, sondern zu untersuchen, welche Methode sich für die Konstruktion einer objektiven KPI-Gesamtbewertung (Performance Score) am besten eignet.  
  
Im weiteren Verlauf wird daher das Ensemble-Verfahren Random Forest eingesetzt, um aus den fünf zentralen KPIs einen gewichteten, datengetriebenen Score abzuleiten.

# 🔹 **5.2 Random Forest Regressor – Feature-Bedeutung**

Um die Bedeutung einzelner KPIs für die Zielgröße Conversions datengetrieben zu bestimmen, wurde ein Random Forest Regressor eingesetzt. Diese Methode zählt zu den nichtlinearen Ensemble-Verfahren und eignet sich besonders, auch komplexe, nichtlineare Zusammenhänge zwischen Variablen abzubilden.  
  
Die Umsetzung erfolgte in Python mit der Bibliothek scikit-learn. Das Modell wurde mit 80 % der Daten trainiert und auf 20 % getestet. Die Modellgüte ist mit:  
  
R² = 0.86  
  
bemerkenswert hoch. Dies zeigt, dass ein Random Forest die Zielgröße Conversions deutlich besser abbildet.  
  
Die geschätzten Feature Importances wurden ebenfalls in Python berechnet und als Balkendiagramm visualisiert. Die folgende Tabelle zeigt die Ergebnisse:

A screen shot of a computer

AI-generated content may be incorrect.

Die Conversion Rate (CR) hat mit Abstand den größten Einfluss auf die Modellvorhersage – was plausibel ist, da sie rechnerisch eng mit Conversions verknüpft ist. Dennoch zeigt das Modell auch, dass weitere KPIs wie CPA, Engagement Rate, CTR und CPC messbare Beiträge leisten.  
  
Visualisierung:  
Die folgende Abbildung zeigt die Feature-Wichtigkeit im Balkendiagramm, erzeugt mit matplotlib und seaborn in Python:

A graph with blue squares

AI-generated content may be incorrect.

# 🔹 6.1 Einleitung zu SHAP

Um die im Random Forest ermittelten Feature-Gewichte noch differenzierter und individueller nachvollziehen zu können, wurde ergänzend das Verfahren SHAP (SHapley Additive exPlanations) eingesetzt. Dieses erlaubt eine modellagnostische Analyse auf Basis spieltheoretischer Prinzipien und quantifiziert, wie stark jedes Merkmal im Einzelfall zur Modellentscheidung beiträgt.  
  
Obwohl das Ziel dieser Analyse nicht in der Entwicklung eines produktiven Vorhersagemodells besteht, lässt sich ein trainierter Random Forest dennoch sinnvoll nutzen, um die relative Bedeutung der KPIs datengetrieben zu untersuchen. Zu diesem Zweck wurde das Verfahren SHAP (SHapley Additive ExPlanations) eingesetzt, das auf spieltheoretischen Prinzipien basiert und den Beitrag jedes einzelnen Merkmals zur Modellvorhersage quantifiziert – unabhängig davon, ob das Modell produktiv eingesetzt wird.  
  
Im Fokus steht die Frage: Welche Features tragen wie stark zur Modellentscheidung bei? Die SHAP-Analyse bestätigt, die bereits im Random Forest beobachtete, Dominanz der Conversion Rate (CR) und der Cost per Acquisition (CPA). Gemeinsam erklären diese beiden Merkmale einen Großteil des Gesamtbeitrags zum Modelloutput.

Merkmale über 73 % des Gesamtbeitrags zum Modelloutput.  
  
Zur Bewertung der Modellgüte wurde der Random Forest mithilfe klassischer Metriken überprüft:  
• R² (Modell vs. echte Conversions): 0.818  
• MAE (Modell vs. echte Conversions): 91.29  
• R² (Modell vs. SHAP-Rekonstruktion): 1.000  
• MAE (Modell vs. SHAP-Rekonstruktion): 0.00  
  
✅ Diese perfekten Werte zeigen, dass die SHAP-Zerlegung präzise die Modellentscheidungen abbildet. Das macht sie zu einem idealen Werkzeug für die anschließende visuelle Analyse.

# 🔹 6.2 Modellinterpretation mit SHAP Zur vertieften Interpretation der im Random Forest ermittelten KPI-Gewichte wurde das Verfahren SHAP (SHapley Additive Explanations) eingesetzt. SHAP basiert auf spieltheoretischen Prinzipien und analysiert, wie stark jedes einzelne Merkmal zur Vorhersage des Modells beiträgt – und zwar nicht nur global, sondern auch individuell für jede einzelne Beobachtung. Damit bietet SHAP eine transparente Möglichkeit zur Modellinterpretation, selbst wenn das Modell nicht zur produktiven Vorhersage, sondern rein zur erklärenden Gewichtung eingesetzt wird.

🔎 SHAP-basierte Feature-Wichtigkeit & Modellgüte:

A screenshot of a computer

AI-generated content may be incorrect.

Die Tabelle zeigt die mittleren SHAP-Werte (Mean Absolute SHAP) pro Feature und deren relative Bedeutung in Prozent.  
Dabei dominieren – wie bereits im Random Forest – die Conversion Rate (CR) und die Cost per Acquisition (CPA) das Modell.  
SHAP bestätigt somit die Gewichtung aus dem Random Forest mit hoher Genauigkeit.

1. SHAP Summary Plot (Streuung):

A graph of blue and red dots

AI-generated content may be incorrect.

Dieser Streu-Plot zeigt den Einfluss jedes Features auf die Modellvorhersage.   
Jeder Punkt stellt eine Kampagne dar. Blau steht für niedrige, Pink für hohe Feature-Werte.  
Die horizontale Lage der Punkte gibt an, ob ein Wert die Vorhersage positiv oder negativ beeinflusst hat.  
Beispiel: Hohe CR-Werte (pink rechts) erhöhen stark die Conversions, hohe CPA-Werte (pink links) senken diese.

2. SHAP Waterfall Plot (Kampagne 100):

A graph with numbers and a number

AI-generated content may be incorrect.  
Dieser Plot zeigt für eine konkrete Kampagne, wie sich die einzelnen Merkmalsausprägungen additiv zur finalen Modellvorhersage zusammensetzen.  
Der Baseline-Wert liegt hier bei etwa 504, die finale Vorhersage beträgt 920.  
**Die Conversion Rate (+163), CPA (+163), ER (+92) und CTR (+54)** wirken steigernd, **nur CPC (-0.08)** wirkt leicht senkend auf die Vorhersage.

## 🔹 7. Entscheidung: Verwendung der SHAP-Gewichte

In dieser Analyse haben wir zwei Methoden verwendet, um herauszufinden, welche KPIs (Kennzahlen) den größten Einfluss auf die Conversions haben: 1) die Feature-Wichtigkeit aus dem Random Forest Modell und 2) die SHAP-Werte (SHapley Additive Explanations).

Beide Methoden liefern eine ähnliche Reihenfolge der wichtigsten Merkmale. Aber SHAP hat entscheidende Vorteile, weil es genauer zeigt, wie stark jedes einzelne Merkmal – also jeder KPI – wirklich zur Vorhersage beiträgt. Und zwar für jede Kampagne einzeln und für das Modell insgesamt.

Vergleich der Methoden:

|  |  |  |
| --- | --- | --- |
| Kriterium | Random Forest | SHAP |
| Erklärung einzelner Kampagnen möglich | Nein | Ja |
| Globale UND lokale Analyse | Nur global | Global + lokal |
| Nachvollziehbar und additiv | Begrenzt | Ja |
| Modell genau rekonstruierbar | Nein | Ja (R² = 1.00) |

Deshalb haben wir uns entschieden, die SHAP-Werte für die Gewichtung der KPIs zu verwenden. Sie liefern eine objektivere und besser nachvollziehbare Grundlage für unseren Performance Score.

Die endgültigen KPI-Gewichte sehen so aus:

|  |  |
| --- | --- |
| KPI | SHAP-Gewicht |
| CR | 43.06 % |
| CPA | 32.06 % |
| ER | 11.7 % |
| CTR | 8.11 % |
| CPC | 5.07 % |

Diese Gewichte nutzen wir nun im nächsten Schritt zu Ermittlung unseres Performance Scores.

# Kapitel 8: Berechnung des Performance Scores

Nachdem mithilfe von SHAP (SHapley Additive Explanations) die relativen Bedeutungen der fünf zentralen KPIs datengetrieben bestimmt wurden, erfolgte im nächsten Schritt die Berechnung eines einheitlichen Performance Scores für jede Kampagne.  
  
Ziel war es, die verschiedenen KPIs – die jeweils unterschiedliche Einheiten und Skalen besitzen – in einem einzigen, vergleichbaren Score zwischen 0 und 100 zu bündeln.  
  
Die folgenden Schritte wurden vollständig in Python mit Hilfe von pandas und matplotlib umgesetzt:

## 1. Normalisierung der KPIs (Min-Max-Skalierung)

Zunächst wurden alle fünf KPIs auf einen Wertebereich zwischen 0 und 1 transformiert, damit sie vergleichbar werden. Dafür wurde die Min-Max-Skalierung verwendet:  
  
Normierter Wert = (Wert - Minimum) / (Maximum - Minimum)  
  
Dies erfolgte für:  
- CTR (Click-Through Rate)  
- CR (Conversion Rate)  
- ER (Engagement Rate)  
- CPC (Cost per Click)  
- CPA (Cost per Acquisition)

## 2. Gewichtung der normalisierten KPIs

Anschließend wurden die normalisierten KPI-Werte mit den SHAP-basierten Gewichten multipliziert und zu einem Gesamtwert (Index) aufsummiert. Die Gewichtung beruht auf den in SHAP identifizierten relativen Bedeutungen der einzelnen KPIs.  
  
**SHAP-Gewichte:**

|  |  |
| --- | --- |
| KPI | SHAP-Gewicht (%) |
| CR | 43.06 % |
| CPA | 32.06 % |
| ER | 11.07 % |
| CTR | 8.11 % |
| CPC | 5.07 % |

Formel (vereinfacht):  
Performance Index = (w1 \* CTR\_Norm + w2 \* CR\_Norm + ...) \* 100

## 3. Median-Verschiebung & Clipping

Der resultierende Score wurde so verschoben, dass der Median der Kampagnen genau 50 beträgt. Zudem wurde der Score auf den Wertebereich 0 bis 100 beschränkt („Clipping“), um Ausreißer zu begrenzen und eine einheitliche Skala zu garantieren.  
  
Performance Score = clip(Index - Median + 50, 0, 100)

## 4. Runden der Werte

Die berechneten Scores wurden auf fünf Nachkommastellen gerundet, um eine konsistente Darstellung zu gewährleisten.

## 5. Visualisierung der Verteilung

Zum Abschluss wurde die Verteilung der Scores analysiert:  
- Histogramm zur Übersicht über die Häufigkeit der Scores

A graph of performance score

AI-generated content may be incorrect.

- Boxplot zur Erkennung von Ausreißern und Spannweite

A graph of a performance score

AI-generated content may be incorrect.  
  
- Output:

A black background with white text

AI-generated content may be incorrect.  
  
Diese Verteilung zeigt, dass der Score gut ausbalanciert ist.

Diese berechneten Performance Scores werden im letzten Schritt in das finale Dashboard integriert und ermöglichen eine schnelle, objektive Bewertung von Marketingkampagnen.

## 6. Skalierung des Performance Scores

**Begründung:** Zur besseren Interpretierbarkeit wurde der berechnete Performance Score auf eine standardisierte Skala von 0 bis 100 übertragen. Dabei wurde der Median der ursprünglichen Index-Werte auf den Wert 50 verschoben. Dies sorgt dafür, dass sich die Hälfte der Kampagnen unterhalb und die andere Hälfte oberhalb des Werts 50 befinden.  
  
Ein solcher symmetrischer Referenzpunkt erleichtert die Einordnung und Visualisierung der Ergebnisse – insbesondere bei der Kommunikation gegenüber Stakeholdern. Zudem verhindert das Clipping auf den Bereich [0, 100] extreme Ausreißer und sorgt für ein stabiles, vergleichbares Score-Niveau.

# 🔹 Kapitel 9 Kampagnenanalyse im interaktiven Dashboard

Das interaktive Tableau-Dashboard erlaubt eine tiefgreifende Analyse der Kampagnenperformance für einzelne Kunden. In diesem Beispiel wurde der Filter oben rechts auf den Kunden **Edwards LLC** gesetzt, wodurch automatisch nur dessen Kampagnendaten angezeigt werden. Durch diesen Mechanismus können „Executives“ gezielt Kunden analysieren, ohne andere Daten zu sehen.

Dashboard 1 – Übersicht zur Funnel-Performance (Edwards LLC):

A screenshot of a computer

AI-generated content may be incorrect.

**Score-Verständnis und Gewichtung**

Der Performance Score ist ein zusammengefasster Kennwert zwischen 0 und 100, der auf den wichtigsten KPIs basiert:

* **Conversion Rate (CR):** 45 % Gewichtung
* **Cost per Acquisition (CPA):** 28 % Gewichtung
* **Engagement Rate (ER):** 12 % Gewichtung
* **Click-Through-Rate (CTR):** 8 % Gewichtung
* **Cost per Click (CPC):** 6 % Gewichtung

**Je höher der Score**, desto besser ist die kombinierte Performance der Kampagne. Niedrige CPC- und CPA-Werte wirken sich positiv aus, da sie auf geringe Kosten hindeuten. Gleichzeitig tragen hohe Werte bei CR, CTR und ER ebenfalls zu einem besseren Score bei. Die Gewichtungen spiegeln wider, welche KPIs für effiziente Kampagnen am wichtigsten sind.

Im Fall von *Edwards LLC* ergibt sich ein Score von **49,6 Punkten**, was unter dem Durchschnitt liegt. Der hohe CPA von 58 € und die geringe Conversion Rate von 0,7 % dämpfen den Score, obwohl der CPC mit 9 € relativ niedrig ist.

**Navigation zur Detailansicht**

Links neben dem Titel "Campaign Conversion Funnel" befindet sich ein schwarzer **Navigation-Button** mit einem Pfeil. Klickt man auf diesen, gelangt man zur **Detailansicht der Kampagnenliste** für den ausgewählten Kunden. Diese Ansicht listet alle Kampagnen von *Edwards LLC* inklusive der zugehörigen KPIs (CPA, CTR, CPC, CR, ER).

Dashboard 2 – Kampagnendetails und Kundenkontakt (inkl. Tooltip-Beispiel):

A screenshot of a computer

AI-generated content may be incorrect.

Zusätzlich ist jede Kampagne farblich nach **Score-Kategorie** eingeteilt:

* 🟢 **Grün:** Top 25 % der Kampagnen (exzellente Performance)
* 🟡 **Gelb:** Mid-High 25 % (oberer Durchschnitt)
* ⚪️ **Weiß:** Median
* 🟠 **Orange:** Mid-Low 25 % (unterer Durchschnitt)
* 🔴 **Rot:** Bottom 25 % (optimierungsbedürftig)

Diese farbliche Kodierung hilft "Executives" sofort zu erkennen, welche Kampagnen sehr gut oder schlecht performen. So kann schnell priorisiert werden, auf welche Kunden oder Kampagnen man sich fokussieren sollte.

**Tooltip-Erklärung (rechte Box bei Hover)**

Fährt man mit der Maus über eine Kampagne, erscheint rechts ein **Tooltip-Fenster**, das zentrale Kennzahlen zusammenfasst:

* 👀 views
* ❤️ Likes
* 🖱️ Clicks
* 📈 Conversions
* 💰 Expense (Gesamtkosten)
* 🖱️ Score

Diese Sofortübersicht ermöglicht eine schnelle Einschätzung der Kampagnenleistung ohne tiefere Analyse.

**Kontaktaufnahme per Button**

Im unteren Bereich der Seite findet sich der Bereich **Client Details**. Neben den Kontaktdaten der Kundenperson (Name, Telefonnummer, E-Mail, Adresse) gibt es einen **E-Mail-Button**. Mit einem Klick darauf kann direkt eine Mail an die Kundenperson gesendet werden, was eine schnelle Kommunikation ermöglicht.

Der Score oben links gibt an, wie erfolgreich eine Kampagne im Vergleich zu anderen performt. Er setzt sich aus folgenden KPIs mit entsprechender Gewichtung zusammen:

* • Conversion Rate (CR): 45 %
* • Cost per Acquisition (CPA): 28 %
* • Engagement Rate (ER): 12 %
* • Click-Through Rate (CTR): 8 %
* • Cost per Click (CPC): 6 %

Je höher die CR und ER, desto besser. Geringe Kostenkennzahlen (CPA, CPC) wirken sich ebenfalls positiv auf den Score aus. Der Score liegt zwischen 0 und 100 Punkten.

Über den Zurück-Button (oben rechts im Dashboard) gelangt man zur Detailansicht, die alle Kampagnen des gewählten Kunden auflistet. Die Liste enthält zusätzlich demografische Daten (Alter, Geschlecht), den Zeitraum sowie alle KPIs. Die Spalte „Score Cat.“ visualisiert die Leistung: 🔴 = schwach, 🟠 = unterdurchschnittlich, 🟡 = mittel, 🟢 = sehr gut.

Ein Klick auf eine Kampagne öffnet einen Tooltip mit wesentlichen Kennzahlen: Views, Likes, Clicks, Conversions, Expenses und dem berechneten Score. Dies erleichtert die schnelle Bewertung einzelner Kampagnen.

Im unteren Abschnitt werden die Kontaktdaten des Kunden angezeigt. Über das 📧-Symbol kann direkt eine E-Mail an die verantwortliche Kontaktperson gesendet werden.

# 🔹 10. Reflexion

Im Verlauf dieses Projekts wurde ein simuliertes Marketing-Datenset verwendet, um die Rolle zentraler KPIs wie **CTR**, **CPC**, **CPA**, **CR** und **ER** zu untersuchen. Im Fokus stand dabei nicht die Vorhersage echter Kampagnenergebnisse, sondern das Verständnis für die **Zusammenhänge zwischen Metriken** und deren **Erklärbarkeit** mithilfe datenanalytischer Methoden – insbesondere durch **SHAP (SHapley Additive exPlanations)**.

In der realen Marketingwelt fließen jedoch weitere Faktoren in die Bewertung von Kampagnenerfolg ein:  
**Umsatzgrößen, ROAS (Return on Ad Spend), Customer Lifetime Value**, sowie **weiche Interaktionen** wie **Likes, Kommentare und Shares** – insbesondere bei Branding-Kampagnen. Auch der tatsächliche Einfluss von Klicks und Engagement auf Conversions ist komplexer, als es ein rein numerischer Datensatz abbilden kann.

Trotz dieser Einschränkungen ermöglichte das Projekt einen praxisnahen Einstieg in die Anwendung statistischer und datenanalytischer Methoden:

* **Statistische Verfahren:**  
  • Min-Max-Normalisierung zur Vergleichbarkeit von KPIs  
  • Median-Verschiebung zur robusten Mittelwertanpassung  
  • Skalierung zur Verhinderung von Verzerrungen bei Modellinterpretationen
* **Technische Tools:**  
  • **Python** (insbesondere pandas, matplotlib, shap) zur Datenverarbeitung, Visualisierung und Modellinterpretation  
  • **Excel** zur transparenten KPI-Berechnung  
  • **Jupyter Notebook** zur strukturierten Analyse und Dokumentation – ein zentrales Werkzeug, das Code, Kommentare und Ausgaben in einem interaktiven Format verbindet
* **Erlernte Kompetenzen:**  
  • Datenbereinigung und Vorbereitung  
  • Entwicklung und Anwendung von KPIs im Marketing  
  • Interpretation und Visualisierung von Kennzahlen  
  • Kritische Bewertung künstlicher Datenstrukturen  
  • Anwendung von SHAP zur erklärbaren Modellanalyse

Insgesamt hat das Projekt nicht nur meine technischen Fähigkeiten geschärft, sondern auch mein Verständnis dafür vertieft, wie **datengetriebene Entscheidungen im Marketing vorbereitet werden** – und wie wichtig es ist, Ergebnisse immer im **Kontext ihrer Entstehung** zu hinterfragen.