

Beispiel Hausarbeit Data Science 1

Offene Fallstudie: Analyse von Kundendaten

Ausgangssituation

Sie arbeiten als **Junior Data Scientist** bei einem wachsenden **Streaming-Anbieter für Filme und Serien**. Das Unternehmen bietet verschiedene Abonnementmodelle an, die monatlich abgerechnet werden und sich in Preis und Leistungsumfang unterscheiden. In den vergangenen Jahren hat der Streaming-Dienst umfangreiche Daten über seine Nutzerinnen und Nutzer gesammelt. Diese Daten umfassen unter anderem Informationen zum Abonnement und zur Vertragsdauer, zum Nutzungsverhalten (z. B. Sehdauer oder Anzahl genutzter Geräte), zu Zahlungs- und Tariffinformationen, zu Supportanfragen sowie zum bisherigen Verlauf des Abonnements.

Das Management steht aktuell vor mehreren strategischen Herausforderungen: Die Anzahl der Kündigungen ist in den letzten Monaten gestiegen. Marketing- und Rabattaktionen sollen gezielter eingesetzt werden. Zudem ist unklar, welche Kundengruppen langfristig besonders wertvoll für das Unternehmen sind.

Bisher wurden Entscheidungen überwiegend auf Basis aggregierter Kennzahlen und Erfahrungswerte getroffen. Die Geschäftsführung möchte künftig stärker auf datenbasierte Analysen setzen und hat das Data-Science-Team beauftragt, konkrete Handlungsempfehlungen zu entwickeln.

Als Teil dieses Teams sollen Sie eine erste Analyse der vorhandenen Kundendaten durchführen und Ihre Ergebnisse für das Management aufbereiten.

Aufgabenstellung

Ihre Aufgabe ist es, den bereitgestellten Datensatz des Streaming-Dienstes zu analysieren und auf dieser Basis eine wirtschaftlich relevante Fragestellung zu beantworten, die das Management bei zukünftigen Entscheidungen unterstützt.

Dabei sollen Sie:

- aus den vorhandenen Daten eine klare Analysefrage ableiten,
- definieren, welche Zielgröße für diese Fragestellung geeignet ist,
- entscheiden, welcher Typ der Datenanalyse zur Beantwortung der Fragestellung passt,
- ein entsprechendes Modell entwickeln und bewerten,
- die Ergebnisse verständlich und wirtschaftlich sinnvoll aufbereiten.

Sie sind in Ihrer Vorgehensweise frei. Entscheidend ist nicht die Komplexität des Modells, sondern die Nachvollziehbarkeit Ihrer methodischen Entscheidungen sowie die Interpretation der Ergebnisse.

Datensatz/Variablenbeschreibung

Der Datensatz steht als CSV-Datei unter folgendem Link zur Verfügung:

<https://data.hsbo.de/use-case-streaming.csv>

Die folgende Tabelle beschreibt die im Datensatz enthaltenen Variablen. Eine Beobachtung entspricht jeweils einem Kunden des Streaming-Dienstes.

Variable	Typ	Beschreibung
user_id	ID	Eindeutige Kennung eines Nutzers
subscription_months	numerisch	Dauer des Abonnements in Monaten
monthly_fee	numerisch	Monatliche Abonnementgebühr in Euro
total_revenue	numerisch	Bisher vom Kunden generierter Gesamtumsatz in Euro
plan_type	kategorial	Abo-Typ (Basic, Standard, Premium)
payment_method	kategorial	Zahlungsmethode (z. B. CreditCard, PayPal, Invoice)
avg_watch_hours	numerisch	Durchschnittliche Sehdauer pro Woche in Stunden
num_profiles	numerisch	Anzahl angelegter Nutzerprofile im Konto
device_count	numerisch	Anzahl unterschiedlicher Geräte, die für das Streaming genutzt werden
support_requests	numerisch	Anzahl gesendeter Supportanfragen
churn	binär	Kündigungsstatus des Kunden (Yes = gekündigt, No = aktiv)
favorite_genre	kategorial	Bevorzugtes Film- oder Seriengenre des Nutzers
last_login_hour	numerisch	Uhrzeit (0–23) des letzten Logins
email_domain	kategorial	Domain der hinterlegten E-Mail-Adresse
promo_code_used	binär	Gibt an, ob bei der Anmeldung ein Promo-Code verwendet wurde (Yes/No)

Abgabe- und Prüfungsform

Schreiben Sie für ihr Projekt eine Hausarbeit im Umfang von maximal 6 Seiten (z.B. mit Word, LaTeX). Die Hausarbeit soll die wesentlichen Schritte ihrer Analyse und ihrer Entscheidungen enthalten, sowie eine Zusammenfassung der Ergebnisse.

Als Dateiformat für die Hausarbeit wird nur die Abgabe im PDF Format akzeptiert. Die Analyse soll mit einem Jupyter Notebook durchgeführt werden, das zusätzlich mit abgegeben werden muss.

Die Bewertung ihrer Abgabe erfolgt anhand der abgegebenen Hausarbeit sowie einer mündlichen Prüfung/Fachgespräch dazu. Sie können die Hausarbeit in einer Gruppe mit maximal 3 Studierenden erstellen, die mündliche Prüfung erfolgt in Einzelgesprächen.

Wichtig: Die Hausarbeit, sowie das abgegebene Jupyter Notebook muss die Namen und Matrikelnummern aller an der Erstellung beteiligten Personen enthalten.