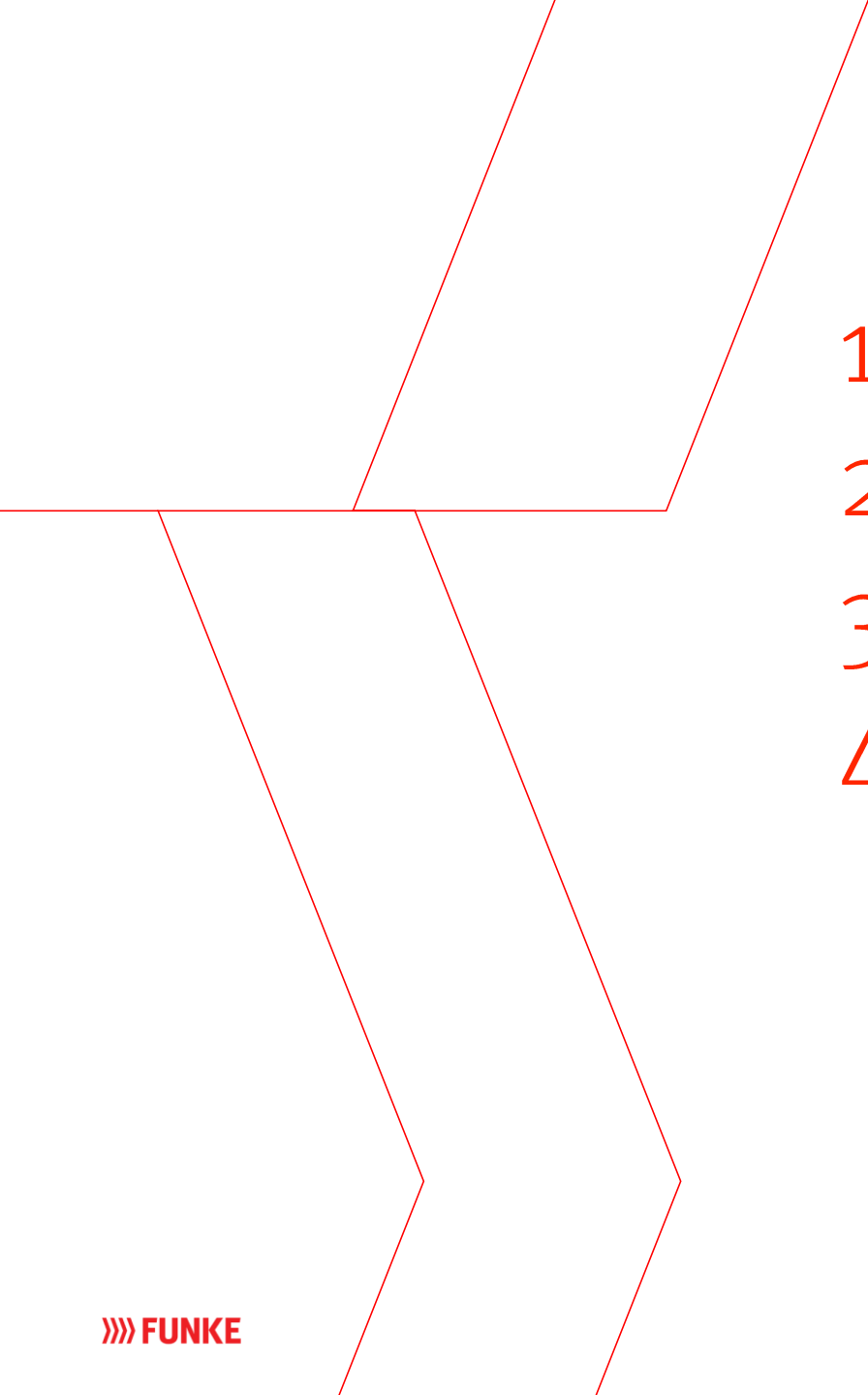


Data Challenge

Data Product & Editorial



- 
1. Allgemeine Information
 2. Set Up
 3. Aufgaben
 4. Ausblick

Allgemeine Information

- Hiermit erhältst du unsere Data Challenge, diese besteht aus den Aufgaben (dieses PDF) und zwei CSV-Dateien mit Daten. Für die ersten beiden Fragen ist die Datei "dataset_web_usage" relevant. Die Datei "dataset_app_usage" ist nur für Aufgabe 3 relevant.
- Dein Ansatz und die Kommunikation der Ergebnisse sind wichtiger als absolute Korrektheit.
- Du kannst das Mittel zur Datenanalyse (z.B. SQL, Python oder Excel) und die Präsentationsform frei wählen.
- Bei Rückfragen melde dich gerne per E-Mail. Wir wünschen dir viel Spaß und Erfolg!

Set Up

Du arbeitest für eine regionale Mediengruppe als Analytics Engineer. In einem interdisziplinären Meeting mit der Redaktion kommt die Frage auf, ob man Inhalte gezielter nach Tageszeit oder Wochentag ausspielen soll. Diese Frage nimmst du mit und wirfst einen Blick in die Daten...

Aufgaben

1. Was fällt dir beim Leseverhalten am Wochenende (Samstag / Sonntag) auf? Wie würdest du dir das erklären?
2. Zu welcher Tageszeit sind die meisten User auf der Seite? Kommen die gleichen User, die morgens die Seite besuchen, auch abends nochmal wieder?

Aufgaben

- 3.1 Anhand des Datasets zum Web Usage: Zu welcher Schicht im Datenmodellierungsprozess passt es deiner Meinung nach am besten und warum? Welche Anpassungen würdest du vorschlagen, um es besser an diese Schicht anzupassen?
- 3.2 Das Dataset wird in Zukunft weiter anwachsen. Wie würdest du deine Transformationen optimieren um die Performance aufrechtzuerhalten?
- 3.3 Du erhältst einen weiteren Datensatz aus einem anderen System, das App-Daten enthält. (Dataset_App_Usage) Dieser hat einige Unterschiede zu den Web-Daten. Wie würdest du die Daten transformieren und modellieren, um sie in unsere bestehende Struktur zu integrieren? Auf welche Herausforderungen würdest du dabei achten?

Datenerklärungen (dataset_web_usage)

- Portal: eine Website von FUNKE
- Timestamp: Datum und Uhrzeit (Unix)
- Device: Gerät, mit dem die Seite besucht wurde (Mobile für Mobilgerät, Desktop für PC/Laptop)
- Visits: Anzahl der Besuche auf dem Portal
- Pageviews: Einzelne Seiten- oder Artikelaufrufe auf dem Portal
- Visit_Duration: Dauer eines Besuchs in Sekunden

Datenerklärungen (dataset_app_usage)

- User_id: eindeutiger Bezeichner für einen in-App Nutzer
- App_version: Die Version der App, die der Nutzer beim Zugriff verwendet hat (z. B. v1.0.0)
- Os_type: Betriebssystem des Geräts, z. B. iOS oder Android
- Session_duration: Dauer der App-Nutzung in Sekunden während einer Session
- Num_screens: Anzahl der besuchten Screens in einer Session
- Crashes: Gibt an, ob die App während der Session abgestürzt ist (0 = nein, 1 = ja)
- Device_model: Modell des verwendeten Geräts (z. B. „iPhone 14“, „Samsung Galaxy S21“)
- Is_premium_user: Gibt an, ob der Nutzer ein Premium-Abonnement besitzt (True / False)
- Push_notification_enabled: Gibt an, ob der Nutzer Push-Benachrichtigungen aktiviert hat
- Last_access: Datum des letzten App-Zugriffs (YYYY-MM-DD)

Ausblick

- Es gibt keine Frist für die Challenge, sobald du fertig bist kannst du uns dein Ergebnis per E-Mail senden.
- Im anschließenden Team-Interview hast du die Chance nochmal über dein Vorgehen zu sprechen, ggf. stellen wir dazu Rückfragen.