

Quiz "Advanced SQL"

Komplexpraktikum DB-Anwendungsentwicklung

Wintersemester 2024/25

Hinweise

Nutzen Sie zur Lösung der Aufgaben den vollständigen Stats-Datensatz¹. Im OPAL-Kurs befindet sich ein neuer Test². Geben Sie dort Ihre Antworten ein! *Kursiv* markierte Textteile beziehen sich auf Tabellen bzw. Spalten aus dem Stats-Schema. Mit *Tag* ist also bspw. die Tabelle *tags* gemeint, mit Tag hingegen eine Zeitdauer von 24h.

Tipp: Eine Übersicht über das Stats-Schema gibt es hier:

<https://relational-data.org/assets/img/datasets-generated/stats.svg>

1 Antwort-Qualität

Bestimmen Sie zu allen *User* die durchschnittliche Bewertung ihrer Kommentare. Accounts mit einer überdurchschnittlichen Bewertung sollen als *Good poster* markiert werden, Accounts mit einer unterdurchschnittlichen Bewertung als *Troll*. Liefern Sie zu jedem User den Accountnamen, die durchschnittliche Bewertung, sowie die Markierung.

2 Schnelle Antworten

Welche fünf *Posts* wurden am schnellsten (korrekt) beantwortet und was waren die Antworten? Bei welchen *Posts* dauerte es am längsten, bis eine korrekte Antwort abgegeben wurden? Wie viel Zeit ist pro *Tag* durchschnittlich vergangen, bis eine korrekte Antwort vorlag?

Tipp: Stelle Sie sicher, dass Fehler im Datensatz passend gehandhabt werden.

¹<https://db4701.inf.tu-dresden.de:8443/index.php/s/JSeYtLidKkaERGR>

²<https://bildungsportal.sachsen.de/opal/auth/RepositoryEntry/45990117386/CourseNode/1731987343077782007>

3 Post-Aktivität

Wie viel Zeit in Tagen vergeht durchschnittlich zwischen zwei Antworten auf einen Post?

Tipp: Verwenden Sie eine Window-Funktion.

4 Wichtige Posts

Erstellen Sie einen neuen View *top_posts*. Er soll für jeden *Tag* die 10 Posts mit dem höchsten *Score* beinhalten. Jeder Eintrag soll die folgenden Informationen enthalten:

- ID des Posts
- Titel des Posts
- Datum der letzten Aktivität
- Zugehöriger *Tag*
- Den Text der drei Antworten mit dem höchsten Score

Tipp: Verwenden Sie ein Array, um die Antworten zu speichern.

5 Eigenwerbung

Auf welche Domains verlinken *User* am häufigsten in ihren Profilen?

Tipp: Verwenden Sie einen regulären Ausdruck, um die Domains zu extrahieren. Mit `'\1'`, `'\2'`, usw. können Sie jeweils auf die *n*-te Gruppe im Ausdruck zugreifen.

6 Top User

Gibt es einen Zusammenhang zwischen der Anzahl an Profilaufrufen oder der Anzahl an Upvotes und der durchschnittlichen Antwortzeit? D.h., haben *User*, die besonders oft Upvotes erhalten, oder deren Profile besonders häufig aufgerufen werden, besonders kurze bzw. lange Antwortzeiten?

Zur Beantwortung können wir folgende Vereinfachungen treffen: um zu bestimmen, ob ein User besonders viele Views/Upvotes erhält, kann das 90te Perzentil herangezogen werden. D.h., alle User, die mehr Views/Upvotes als 90% der anderen User erhalten, fallen in die jeweilige Kategorie.

Tipp: Verwenden Sie ein **GROUPING SET**, um die unterschiedlichen User-Klassen zu vergleichen. Strukturieren Sie die Anfrage durch Verwendung von Common Table Expressions und berechnen sie die Ergebnismenge in einzelnen Teilschritten.