Documentation

This solution tackles the problem of calculating the cancellation rate for unbanned users by structuring data analysis into several critical steps. First, it involves filtering the dataset to only include trips where both the client and driver were not banned, which is essential for maintaining data integrity and meeting the problem's specific requirements. Users are divided by roles as either "client" or "driver," and any user with a banned status is excluded from the relevant joins. This allows the program to filter the `Trips` data accurately, ensuring that only unbanned interactions are considered when calculating statistics.

After filtering, the solution focuses on the date range of interest ("2013-10-01" to "2013-10-03"), during which the cancellation rate will be calculated. This step is crucial to ensure the analysis adheres to the specified timeframe, which allows for targeted results and ensures the output's relevance. Only trips occurring within these dates are considered, and those falling outside this range are automatically excluded, simplifying the subsequent analysis steps.

For each day within this timeframe, the solution calculates two metrics: the total number of unbanned requests and the number of canceled requests among them. The cancellation conditions are derived from the `status` field in the `Trips` table, where trips can be categorized as "completed," "cancelled_by_driver," or "cancelled_by_client." Canceled requests are thus those with a status indicating cancellation by either the driver or the client. These metrics are grouped by day, creating a summarized view that captures the daily cancellation dynamics of the taxi service.

The next step involves calculating the cancellation rate, which is determined by dividing the number of canceled requests by the total number of requests for each day. This metric is computed using precise rounding to two decimal places, as specified in the problem. Rounding to two decimal places ensures clarity and ease of understanding in the output, making it more presentable and professionally formatted. This cancellation rate calculation is vital for understanding the proportion of failed requests due to cancellations among unbanned users, offering insights into user behavior and service reliability on a day-to-day basis.

Finally, the resulting dataset is presented in a clear and structured format that matches the required output schema, with columns representing the day and its corresponding cancellation rate. By focusing on unbanned user data and analyzing it within a precise timeframe, this solution effectively provides a snapshot of the service's performance over specific dates. This format also supports extensibility, allowing additional date ranges or user criteria to be analyzed with minimal modification, making the solution versatile for broader business intelligence purposes.