Case Study: Analyzing Rides Data in SQLite

Nwe Thiri May

thiri.jewel.14@gmail.com +4917684619518 / Berlin, Germany/
https://www.linkedin.com/in/nwe-thiri-may-403589174 https://github.com/May-NweThiri/

Task 1: Determine the Month with the Highest Number of Rides and the Month with the Lowest Number of Rides.

Answer:

The month with Hghest number of Rides is November 2022 (2022-11).
The month with Hghest number of Rides is September 2024 (2024-09).

Explanation:
To identify the months with the highest and lowest number of rides, I used the following SQL queries**:**

Query to find the month with the highest number of rides:

```
SELECT strftime('%Y-%m', date) AS month, COUNT(*) AS ride_count
FROM rides
GROUP BY strftime('%Y-%m', date)
ORDER BY ride_count DESC
LIMIT 1;
```

```
1 -- Task 1: Task 1: Determine the Month with the Highest Number of Rides
2 SELECT strftime('%Y-%m', date) AS month, COUNT(*) AS ride_count
3 FROM rides
4 GROUP BY strftime('%Y-%m', date)
5 ORDER BY ride_count DESC
6 LIMIT 1;
7
8
```

month	ride_count
2022-11	53

Query to find the month with the lowest number of rides:

```
SELECT strftime('%Y-%m', date) AS month, COUNT(*) AS ride_count
FROM rides
GROUP BY strftime('%Y-%m', date)
ORDER BY ride_count ASC
LIMIT 1;
```

```
1 -- Task 1: Task 1: Determine the Month with the lowest Number of Rides
2 SELECT strftime('%Y-%m', date) AS month, COUNT(*) AS ride_count
3 FROM rides
4 GROUP BY strftime('%Y-%m', date)
5 ORDER BY ride_count ASC
6 LIMIT 1;
7
8
```

month	ride_count
2024-09	11

Task 2: Who do you Consider to be the Best Performing Driver(s), and Why?

Answer:

The best performing driver is Nymphadora Tonks. This conclusion is based on the following criteria:

1. Highest Average Rating: Nymphadora Tonks has the highest average rating among all drivers.
2. Total Rides: Nymphadora Tonks completed a total of 61 rides.
3. Minimal Late Rides: The driver has a lower percentage of late rides compared to others.

Criteria for Best Performing Driver:

1. **Average Rating**: Higher average ratings indicate better performance as rated by passengers.
2. **Total Number of Rides**: A higher number of completed rides reflects experience and reliability.
3. **Percentage of Late Rides**: Fewer late rides demonstrate punctuality and reliability.

Data Preparation:

Initially, I checked for any null values in data . I found some null values in rating , I updated to 0 to avoid skewing the analysis:

Query to check the null values:

SELECT * FROM drivers
WHERE id is **NULL**
OR full_name is **NULL**
OR country_code is **NULL**;

```
1 SELECT *
2 FROM drivers
3 WHERE id IS Null
4 OR full_name IS NULL
5 OR country_code IS NULL;
6
```

Query to upade null values in rating table :

UPDATE rides **SET** rating = 0
WHERE rating **IS NULL**;

Check the null values again with count function:

```
1 SELECT COUNT(*)
2 FROM rides
3 WHERE rating IS NULL;
```

COUNT(*)
0

Query to Find the Best Performing Driver:

```
SELECT d.full_name,
COUNT(r.id) AS total_rides,
AVG(r.rating) AS avg_rating,                                    -- rating average
SUM(r.is_driver_late) * 1.0 / COUNT(r.id) AS total_late_rides    -- in percentage
FROM drivers d JOIN rides r ON d.id = r.driver_id               -- Join drivers with rides
GROUP BY d.id                                                   -- Group by driver
ORDER BY avg_rating DESC, total_rides DESC, total_late_rides ASC;
```

The list of the driver acorrding to the rating , total rides and total_late ride :

full_name	total_rides	avg_rating	total_late_rides
Nymphadora Tonks	61	2.9344262295081966	0.47540983606557374
Bellatrix Lestrange	57	2.8596491228070176	0.5263157894736842
Ron Weasley	66	2.696969696969697	0.5303030303030303
Cho Chang	58	2.6551724137931036	0.39655172413793105
Luna Lovegood	66	2.6515151515151514	0.5151515151515151
Draco Malfoy	76	2.6447368421052633	0.5263157894736842
Lily Evans	91	2.6263736263736264	0.5274725274725275
Hermione Granger	66	2.621212121212121	0.48484848484848486
Sirius Black	83	2.5542168674698793	0.5180722891566265

The result of the query will list drivers according to their average rating, total number of rides, and percentage of late rides. **Nymphadora Tonks** ranks **highest** based on these criteria.