1. Write a SQL code to retrieve the current lifecycle stage of users and the duration in days that took for them to be in the current lifecycle stage, considering the below sample table where users go through three stages before onboarding, with stage 1 as the starting point and stage 3 as the final stage.

|  |  |  |
| --- | --- | --- |
| date | user\_name | lifecycle\_stage |
| 10-06-2022 | user 1 | LF\_1 |
| 15-06-2022 | user 2 | LF\_1 |
| 21-06-2022 | user 2 | LF\_2 |
| 15-06-2022 | user 1 | LF\_2 |
| 30-06-2022 | user 6 | LF\_3 |
| 13-06-2022 | user 4 | LF\_1 |
| 18-06-2022 | user 5 | LF\_1 |
| 19-06-2022 | user 5 | LF\_2 |
| 30-06-2022 | user 2 | LF\_3 |
| 20-06-2022 | user 6 | LF\_1 |
| 22-06-2022 | user 6 | LF\_2 |
| 13-06-2022 | user 3 | LF\_1 |

1. From the “Sales” data below, write a SQL query to find the days when the sales were higher than the previous day.

|  |  |  |
| --- | --- | --- |
| ID (int) | SaleDate (date) | SaleAmount (Float) |
| 1 | 2021-06-04 | 23000 |
| 2 | 2021-06-05 | 13000 |
| 3 | 2021-06-05 | 11000 |
| 4 | 2021-06-07 | 25000 |
| 5 | 2021-06-08 | 26000 |
| 6 | 2021-06-09 | 24500 |
| 4 | 07-06-2021 | 25000 |
| 5 | 2021-06-08 | 26000 |
| 6 | 09-06-2021 | 24500 |

1. From the “Contact” table below, write a SQL query to delete the duplicates. Also write a query to display invalid emails.

|  |  |
| --- | --- |
| ID (int) | Email (string) |
| 1 | abc@xyz.com |
| 2 | def@x.com |
| 3 | hij@y.com |
| 4 | jkl@dcom |
| 1 | a@b.com |
| 6 | b@d.com |
| 7 | abc123 |

1. Consider the sample data below that shows the activity of users. You are requested to write SQL query to find the users who got a “credit” for their streak completion. Here, a continuous activity of four weeks is considered as a streak completion (i.e., a user must be active at least once in a week for 4 consecutive weeks). Output the below mentioned columns in your query.

Output columns :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| user\_id | activity\_date | activity\_week | streak\_number | credited? |

Here

* streak\_number – streak count of users runs from 1 to 4.
* credited? – At the fourth streak\_number or fourth week of activity user gets a credit. A binary column (0/1)

Sample data :

|  |  |  |
| --- | --- | --- |
| user\_id | activity\_date | activity\_week |
| U1 | 27-12-2022 | 01-01-2023 |
| U2 | 05-01-2023 | 08-01-2023 |
| U2 | 10-01-2023 | 15-01-2023 |
| U1 | 08-01-2023 | 08-01-2023 |
| U1 | 09-01-2023 | 15-01-2023 |
| U3 | 07-02-2023 | 12-02-2023 |
| U4 | 04-01-2023 | 08-01-2023 |
| U3 | 19-02-2023 | 19-02-2023 |
| U4 | 20-01-2023 | 22-01-2023 |
| U3 | 01-03-2023 | 05-03-2023 |
| U3 | 05-03-2023 | 05-03-2023 |
| U2 | 21-01-2023 | 22-01-2023 |
| U1 | 30-01-2023 | 05-02-2023 |
| U2 | 22-01-2023 | 22-01-2023 |
| U1 | 29-12-2022 | 01-01-2023 |
| U3 | 16-03-2023 | 19-03-2023 |
| U1 | 19-01-2023 | 22-01-2023 |

1. From the sample data provided, make a week wise report of win back users. Users who joined back in 7- 60 days after churning are called win back users. Also, get any possible insights that you can make from the data.(use python to complete this task)