**Beginner SQL Project Questions for ACLED Dataset answer with your validated and running SQL statements**

**Level 1**

1. **Select all records from the dataset**

SELECT \*

FROM TBL\_ACLED\_FINAL;

1. **Select the first 10 records**

SELECT \*

FROM TBL\_ACLED\_FINAL

FETCH FIRST 10 ROWS ONLY;

1. **Count the total number of events**

SELECT *COUNT* ( EVENT\_TYPE) AS NUMBER\_OF\_EVENTS

FROM TBL\_ACLED\_FINAL;

1. **Find the earliest and latest event dates≠**

SELECT

*MIN*(EVENT\_DATE) AS EARLIEST\_EVENT\_DATE,

*MAX*(EVENT\_DATE) AS LATEST\_EVENT\_DATE

FROM TBL\_ACLED\_FINAL;

1. **List unique event types**

SELECT DISTINCT EVENT\_TYPE

FROM TBL\_ACLED\_FINAL;

1. **Find the total number** of events for each year

SELECT

YEAR,

*COUNT*(EVENT\_TYPE) AS NUMBER\_OF\_EVENTS

FROM TBL\_ACLED\_FINAL

GROUP BY YEAR;

1. **Find the number of events for each disorder type**

SELECT

DISORDER\_TYPE,

*COUNT*(EVENT\_TYPE) AS NUMBER\_OF\_EVENTS

FROM TBL\_ACLED\_FINAL

GROUP BY DISORDER\_TYPE;

1. **Select events with more than 10 fatalities**

SELECT \*

FROM TBL\_ACLED\_FINAL

WHERE FATALITIES > 10;

1. **Find the average number of fatalities per event type**

SELECT

DISTINCT EVENT\_TYPE,

*AVG*(FATALITIES) AS AVG\_FATALITIES

FROM TBL\_ACLED\_FINAL

GROUP BY EVENT\_TYPE;

1. **Find events with missing latitude or longitude values**

SELECT \*

FROM TBL\_ACLED\_FINAL

WHERE LATITUDE IS NULL OR LONGITUDE IS NULL;

**Level 2**

1. **Find the top 5 most common actors involved in events**

SELECT

ACTOR1,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY ACTOR1

ORDER BY EVENT\_COUNT DESC

FETCH FIRST 5 ROWS ONLY;

SELECT ACTOR1

FROM (SELECT

ACTOR1,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY ACTOR1

ORDER BY EVENT\_COUNT DESC

FETCH FIRST 5 ROWS ONLY)

WITH TopActors AS (

SELECT

ACTOR1,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY ACTOR1

ORDER BY EVENT\_COUNT DESC

FETCH FIRST 5 ROWS ONLY

)

SELECT ACTOR1

FROM TopActors;

1. **Calculate the total number of events and total fatalities for each country**

SELECT

COUNTRY,

*COUNT*(\*) AS NUMBER\_OF\_EVENTS,

*SUM*(FATALITIES) AS TOTAL\_FATALITIES

FROM TBL\_ACLED\_FINAL

GROUP BY COUNTRY;

1. **Find events that occurred between two specific dates (one month apart) - Should we create bins? ; if you use specific dates to get the count won’t the data contain duplicate values?**

SELECT

*TO\_CHAR*(EVENT\_DATE, 'YEAR'||' MONTH'),

*COUNT*(EVENT\_ID\_CNTY)

FROM TBL\_ACLED\_FINAL

GROUP BY EVENT\_DATE, *TO\_CHAR*(EVENT\_DATE, 'YEAR'||' MONTH');

SELECT \*

FROM TBL\_ACLED\_FINAL

WHERE EVENT\_DATE BETWEEN '2024-03-01' AND '2024-04-01'

1. **Identify the events with the highest number of fatalities in each year**

WITH EVENT\_COUNT AS (

SELECT

YEAR,

EVENT\_TYPE,

*SUM*(FATALITIES) AS TOTAL\_FATALITIES,

*DENSE\_RANK*() OVER(PARTITION BY YEAR ORDER BY YEAR, *SUM*(FATALITIES) DESC) AS EVENT\_RANK

FROM TBL\_ACLED\_FINAL

GROUP BY YEAR, EVENT\_TYPE

ORDER BY YEAR, TOTAL\_FATALITIES DESC)

SELECT

YEAR,

EVENT\_TYPE,

TOTAL\_FATALITIES

FROM EVENT\_COUNT

WHERE EVENT\_RANK = 1;

1. **List events by geographical precision and count them**

SELECT

EVENT\_TYPE,

GEO\_PRECISION,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY GEO\_PRECISION, EVENT\_TYPE;

SELECT

GEO\_PRECISION,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY GEO\_PRECISION;

1. **Find the most common sub-event type for each event type**

WITH POPULAR\_EVENT AS (

SELECT

*DENSE\_RANK*() OVER(PARTITION BY EVENT\_TYPE ORDER BY *COUNT*(\*) DESC) AS POPULAR\_SUB\_EVENT,

EVENT\_TYPE,

SUB\_EVENT\_TYPE,

*COUNT*(\*) AS EVENT\_COUNT

FROM TBL\_ACLED\_FINAL

GROUP BY EVENT\_TYPE, SUB\_EVENT\_TYPE

ORDER BY EVENT\_TYPE, EVENT\_COUNT DESC)

SELECT

EVENT\_TYPE,

SUB\_EVENT\_TYPE

FROM POPULAR\_EVENT

WHERE POPULAR\_SUB\_EVENT = 1;

1. **Calculate the percentage of events that are peaceful protests ( USE CTE)**

SELECT SUB\_EVENT\_TYPE,

*COUNT*(\*)/ (SELECT *COUNT* (\*)

FROM TBL\_ACLED\_FINAL) AS PERCENTAGE\_OF\_EVENTS

FROM TBL\_ACLED\_FINAL

WHERE SUB\_EVENT\_TYPE = 'Peaceful protest'

GROUP BY SUB\_EVENT\_TYPE;

1. **Identify events that involved more than one actor - ( NULL VALUES)**

SELECT

EVENT\_TYPE,

*COUNT*(DISTINCT ACTOR1)

FROM TBL\_ACLED\_FINAL

GROUP BY EVENT\_TYPE

HAVING *COUNT* (DISTINCT ACTOR1) > 1

ORDER BY EVENT\_TYPE;

1. **Find the total number of events reported by each source**

SELECT

SOURCE,

*COUNT*(\*) AS COUNT\_EVENT

FROM TBL\_ACLED\_FINAL

GROUP BY SOURCE;

1. **Identify regions with the highest average fatalities per event**

WITH FATALITIES AS (

SELECT

REGION,

EVENT\_TYPE,

*AVG*(FATALITIES) AS AVG\_FATALITIES,

*DENSE\_RANK*() OVER(PARTITION BY EVENT\_TYPE ORDER BY EVENT\_TYPE, *AVG*(FATALITIES) DESC) AS RANK\_FATALITIES

FROM TBL\_ACLED\_FINAL

GROUP BY REGION, EVENT\_TYPE

ORDER BY EVENT\_TYPE,AVG\_FATALITIES DESC)

SELECT \*

FROM FATALITIES

WHERE RANK\_FATALITIES =1;