#find the most common hashtag pairs in the data

WITH cte AS (

SELECT tweet\_date, tweet\_Location\_logged, LTRIM(RTRIM(hashtag\_list)) as hashtag\_list,

LTRIM(RTRIM(SUBSTRING(hashtag\_list, 1, charindex(',', hashtag\_list + ',') - 1))) AS hashtag

FROM [dbo].[hashtag\_profile (1)]

UNION ALL

SELECT tweet\_date, tweet\_location\_logged, LTRIM(RTRIM(SUBSTRING(hashtag\_list, charindex(',', hashtag\_list) + 1, LEN(hashtag\_list)))),

LTRIM(RTRIM(SUBSTRING(hashtag\_list, charindex(',', hashtag\_list) + 1, charindex(',', hashtag\_list + ',', charindex(',', hashtag\_list) + 1) - charindex(',', hashtag\_list) - 1)))

FROM cte

WHERE charindex(',', hashtag\_list) > 0

)

SELECT CONCAT(LOWER(c1.hashtag), ' & ', LOWER(c2.hashtag)) as hashtag\_pairs, COUNT(\*) as count

FROM cte c1

JOIN cte c2 ON c1.tweet\_date = c2.tweet\_date

AND c1.tweet\_location\_logged = c2.tweet\_location\_logged

WHERE c1.hashtag < c2.hashtag

GROUP BY CONCAT(LOWER(c1.hashtag), ' & ', LOWER(c2.hashtag))

ORDER BY count DESC

#top 3 hashtags from each countries

WITH cte AS (

SELECT tweet\_date, LTRIM(RTRIM(tweet\_Location\_logged)) as tweet\_Location\_logged, LTRIM(RTRIM(hashtag\_list)) as hashtag\_list,

LTRIM(RTRIM(SUBSTRING(hashtag\_list, 1, charindex(',', hashtag\_list + ',') - 1))) AS hashtag

FROM [dbo].[hashtag\_profile (1)]

UNION ALL

SELECT tweet\_date, LTRIM(RTRIM(tweet\_location\_logged)), LTRIM(RTRIM(SUBSTRING(hashtag\_list, charindex(',', hashtag\_list) + 1, LEN(hashtag\_list)))),

LTRIM(RTRIM(SUBSTRING(hashtag\_list, charindex(',', hashtag\_list) + 1, charindex(',', hashtag\_list + ',', charindex(',', hashtag\_list) + 1) - charindex(',', hashtag\_list) - 1)))

FROM cte

WHERE charindex(',', hashtag\_list) > 0

),

cte1 AS (

SELECT tweet\_Location\_logged, hashtag, COUNT(hashtag) as hashtag\_count, ROW\_NUMBER() OVER (PARTITION BY tweet\_Location\_logged ORDER BY COUNT(hashtag) DESC) as rn

FROM cte

WHERE tweet\_Location\_logged != 'Unknown'

GROUP BY tweet\_Location\_logged, hashtag

)

SELECT tweet\_Location\_logged, hashtag, hashtag\_count

FROM cte1

WHERE rn < = 3

ORDER BY tweet\_Location\_logged , hashtag\_count desc;

/\*\*\*\*\*\* Script for SelectTopNRows command from SSMS \*\*\*\*\*\*/

SELECT \*

FROM [NORTHWND].[dbo].[hashtag\_profile (1)]

--count hashtags by month

SELECT MONTH(tweet\_date) as month, hashtag\_list, COUNT(\*) as hashtag\_count

FROM [dbo].[hashtag\_profile (1)]

GROUP BY MONTH(tweet\_date), hashtag\_list

order by hashtag\_count desc, month

--top 3 hash tags by month

WITH CTE AS (

SELECT MONTH (tweet\_date) AS month, hashtag\_list, COUNT(\*) AS hashtag\_count,

ROW\_NUMBER() OVER (PARTITION BY month(tweet\_date) ORDER BY COUNT(\*) DESC) AS rn

FROM [dbo].[hashtag\_profile (1)]

GROUP BY month(tweet\_date), hashtag\_list

)

SELECT month, hashtag\_list, hashtag\_count

FROM CTE

WHERE rn <= 3;

--count hashtags by year

SELECT year(tweet\_date) as year, hashtag\_list, COUNT(\*) as hashtag\_count

FROM [dbo].[hashtag\_profile (1)]

GROUP BY year(tweet\_date), hashtag\_list

order by year desc, hashtag\_count desc

---top 5 hashtags by year

WITH CTE AS (

SELECT YEAR(tweet\_date) AS year, hashtag\_list, COUNT(\*) AS hashtag\_count,

ROW\_NUMBER() OVER (PARTITION BY YEAR(tweet\_date) ORDER BY COUNT(\*) DESC) AS rn

FROM [dbo].[hashtag\_profile (1)]

GROUP BY YEAR(tweet\_date), hashtag\_list

)

SELECT year, hashtag\_list, hashtag\_count

FROM CTE

WHERE rn <= 5;

/\*\*\*\*\*\* Script for SelectTopNRows command from SSMS \*\*\*\*\*\*/

SELECT \*

FROM [NORTHWND].[dbo].[full\_data\_for\_sql (1)]

ALTER TABLE [NORTHWND].[dbo].[full\_data\_for\_sql (1)]

ALTER COLUMN [engagement\_rate] float;

UPDATE [NORTHWND].[dbo].[full\_data\_for\_sql (1)]

SET Polarity = ABS(Polarity) \* -1

WHERE Polarity < 0;

SELECT COUNT(\*)

FROM [NORTHWND].[dbo].[full\_data\_for\_sql (1)]

WHERE tweet\_hashtags like '%palestine%';

SELECT COUNT(DISTINCT Tweet\_User\_TwitterHandle) as 'Distinct users',

avg(tweet\_retweetCount) AS 'avg retweets' ,

avg(tweet\_replyCount) AS 'avg replies',

avg(tweet\_likeCount) AS 'avg likes',

avg(tweet\_quoteCount) AS 'avg quoted',

avg(tweet\_user\_followersCount) AS 'avg user\_followers ',

avg(tweet\_user\_friendsCount) AS ' avg user\_friends',

CASE WHEN avg(Polarity) > 0 THEN 'Positive' ELSE 'Negative' END AS sentiment

FROM [dbo].[full\_data\_for\_sql (1)]

WHERE tweet\_hashtags like '%israel%';

SELECT [tweet\_date\_MonthDay] ,avg(engagement\_rate)\*100 'Avg Engagement rate'

FROM [dbo].[full\_data\_for\_sql (1)]

WHERE tweet\_hashtags like '%,israel' or tweet\_hashtags like 'israel,%'

GROUP BY [tweet\_date\_MonthDay]

-- Most tags related to israel

WITH hashtags AS (

SELECT

SUBSTRING(hashtag\_list, 1, CHARINDEX(',', hashtag\_list + ',') - 1) AS hashtags,

SUBSTRING(hashtag\_list, CHARINDEX(',', hashtag\_list + ',') + 1, LEN(hashtag\_list)) AS remaining\_hashtags

FROM

[dbo].[hashtag\_profile (1)]

WHERE

hashtag\_list LIKE '%Israel%'

UNION ALL

SELECT

SUBSTRING(remaining\_hashtags, 1, CHARINDEX(',', remaining\_hashtags + ',') - 1),

SUBSTRING(remaining\_hashtags, CHARINDEX(',', remaining\_hashtags + ',') + 1, LEN(remaining\_hashtags))

FROM

hashtags

WHERE

remaining\_hashtags > ''

)

SELECT

hashtags,

COUNT(\*) AS count

FROM

hashtags

GROUP BY

hashtags

ORDER BY

count DESC;

-- most common tags with thw word israel in it

WITH hashtags AS (

SELECT

LOWER(RTRIM(LTRIM(SUBSTRING(hashtag\_list, 1, CHARINDEX(',', hashtag\_list + ',') - 1)))) AS hashtags,

LOWER(RTRIM(LTRIM(SUBSTRING(hashtag\_list, CHARINDEX(',', hashtag\_list + ',') + 1, LEN(hashtag\_list))))) AS remaining\_hashtags

FROM

[dbo].[hashtag\_profile (1)]

WHERE

CHARINDEX(',', hashtag\_list) > 0

UNION ALL

SELECT

LOWER(RTRIM(LTRIM(SUBSTRING(remaining\_hashtags, 1, CHARINDEX(',', remaining\_hashtags + ',') - 1)))),

LOWER(RTRIM(LTRIM(SUBSTRING(remaining\_hashtags, CHARINDEX(',', remaining\_hashtags + ',') + 1, LEN(remaining\_hashtags)))))

FROM

hashtags

WHERE

CHARINDEX(',', remaining\_hashtags) > 0

)

SELECT

hashtags,

COUNT(\*) AS count

FROM

hashtags

WHERE

hashtags LIKE '%israel%'

GROUP BY

hashtags

ORDER BY

count DESC;

WITH hashtags AS (

SELECT

tweet\_Location\_logged,

LOWER(RTRIM(LTRIM(SUBSTRING(hashtag\_list, 1, CHARINDEX(',', hashtag\_list + ',') - 1)))) AS hashtags,

LOWER(RTRIM(LTRIM(SUBSTRING(hashtag\_list, CHARINDEX(',', hashtag\_list + ',') + 1, LEN(hashtag\_list))))) AS remaining\_hashtags

FROM

[dbo].[hashtag\_profile (1)]

WHERE

tweet\_Location\_logged != 'Unknown' AND

CHARINDEX(',', hashtag\_list) > 0

UNION ALL

SELECT

tweet\_Location\_logged,

LOWER(RTRIM(LTRIM(SUBSTRING(remaining\_hashtags, 1, CHARINDEX(',', remaining\_hashtags + ',') - 1)))),

LOWER(RTRIM(LTRIM(SUBSTRING(remaining\_hashtags, CHARINDEX(',', remaining\_hashtags + ',') + 1, LEN(remaining\_hashtags)))))

FROM

hashtags

WHERE

CHARINDEX(',', remaining\_hashtags) > 0

),

agg AS (

SELECT

tweet\_Location\_logged,

hashtags,

COUNT(\*) AS count

FROM

hashtags

GROUP BY

tweet\_Location\_logged,

hashtags

)

SELECT

tweet\_Location\_logged,

hashtags,

count

FROM

(

SELECT

tweet\_Location\_logged,

hashtags,

count,

ROW\_NUMBER() OVER (PARTITION BY tweet\_Location\_logged ORDER BY count DESC) AS rn

FROM

agg

) t

WHERE

rn <= 3

ORDER BY

tweet\_Location\_logged,

count DESC