





1. Determine the number of active ads on the site per day (use generate\_series).

Postgres Query:

--Postgres Query:

SELECT c."activedate" as "ActiveDate",COUNT(a."id") as "#ActiveAds"

FROM public."Ads" a

LEFT OUTER JOIN LATERAL

(SELECT generate\_series(MIN("create"),DATE(now()),'1 day') as ActiveDate FROM public."Ads") c

ON true

WHERE c."activedate" >= a."publish" AND c."activedate" <= a."delete"

GROUP BY c."activedate"

ORDER BY c."activedate";



SOL Server Query:

DECLARE @StartDate DATETIME = (SELECT MIN(create) as SDate FROM dbo.Ads)

DECLARE @EndDate DATETIME = CONVERT(DATE,GETDATE())

;WITH Generate\_series\_cte

AS

(

SELECT @StartDate AS [ActiveDate]

UNION ALL

SELECT DATEADD(dd, 1, [ActiveDate]) FROM Generate\_series\_cte

WHERE DATEADD(dd, 1, [ActiveDate]) <= @EndDate

)

SELECT [ActiveDate], COUNT(id) as [#ActiveAds]

FROM Generate\_series\_cte

CROSS JOIN dbo.Ads

WHERE [ActiveDate] >= CONVERT(DATE,publish) AND [ActiveDate] <= CONVERT(DATE,delete)

GROUP BY [ActiveDate]

ORDER BY [ActiveDate]

OPTION (MAXRECURSION 0)



2. Determine the YoY growth of Beef Cattle ads in county Tipperary from 2016 to 2017 (use lag).

SELECT \* INTO temporary table YearMonth

FROM (

SELECT generate\_series(cast(MIN(EXTRACT(MONTH FROM "Ads"."publish")) as integer),cast(MAX(EXTRACT(MONTH FROM "Ads"."publish")) as integer),1) as "Month"

FROM public."Ads"

WHERE EXTRACT(year FROM "Ads"."publish") IN (2016,2017)

) as a

LEFT OUTER JOIN LATERAL

(

SELECT generate\_series(cast(MIN(EXTRACT(YEAR FROM "Ads"."publish")) as integer),cast(MAX(EXTRACT(YEAR FROM "Ads"."publish")) as integer),1) as "Year"

FROM public."Ads"

WHERE EXTRACT(year FROM "Ads"."publish") IN (2016,2017)

) b ON true

ORDER BY "b"."Year", "a"."Month"

SELECT "Year", "Month",count("a"."id") "TotalAds",

--,lag(count("a"."id")) OVER( partition by "Month" ORDER BY "Month","Year" ) "PYAds"

(CASE

WHEN cast(count("a"."id") - lag(count("a"."id")) OVER( partition by "Month" ORDER BY "Month","Year") as integer) IS NULL THEN 0

ELSE cast(count("a"."id") - lag(count("a"."id")) OVER( partition by "Month" ORDER BY "Month","Year") as integer)

END) "YoY Growth"

FROM "yearmonth"

LEFT JOIN

(

SELECT "Ads".\* FROM public."Ads"

LEFT JOIN public."category" ON category."id" = "Ads"."category\_id"

LEFT JOIN public."region" ON region."id" = "Ads"."region\_id"

WHERE category."subsubcategory" = 'Beef Cattle' AND region."county"= 'Tipperary' AND EXTRACT(YEAR FROM "Ads"."publish") IN (2017,2016)

) a ON EXTRACT(MONTH FROM "a"."publish") = "Month" AND EXTRACT(YEAR FROM "a"."publish") = "Year"

GROUP BY "Month","Year"

Order BY "Month" ,"Year"

Donedeal Sample Data Integrity Issues:

1. Missing (Null) values for columns:

price

model

year

color

mileage

mileageType

engine\_cc

body\_type

fuel\_type

transmission

doors

2. Inappropriate values in the “Year” column e.g (0,132,141,142 and 200)

3. Inconsistent values in the “Model” column (numeric values like 308 etc)

4. Inconsistent values in the “Body\_type” column e.g please select…

5. Inconsistent values in the “Doors” column e.g. 0

6. Inconsistent values in the “Engine\_cc” column e.g. 0

7. Inconsistent values in the “Mileage\_Type” column e.g. (km vs Kilometres, 136, 26500 etc)

8. Inconsistent values in the “Make” column e.g. numeric values and date

9. Issues with table structure for table “car\_ads”:

Column “year” should be a numeric field instead of varchar

Column “mileage” should be a numeric field instead of varchar

Column “engine\_cc” should be a numeric field instead of varchar

Column “doors” should be a numeric field instead of varchar

Inconsistent column name: mileageType, should be mileage\_type