


bite-sized.sql

SQL: 3 EASY BINNING QUERIES

1. CREATE n BINS

NTILE(X) OVER (ORDER BY Y) creates X bins on column Y with an equal (or as close to equal as possible) number of rows in each bin



```
WITH Bins
AS
(
  SELECT
    YearlyIncome,
    NTILE(10) OVER (ORDER BY YearlyIncome) AS Bin
  FROM DimCustomer
)
SELECT
  MIN(YearlyIncome) AS MinYearlyIncome,
  MAX(YearlyIncome) AS MaxYearlyIncome,
  COUNT(*) AS records
FROM Bins
GROUP BY Bin
ORDER BY Bin;
```

	MinYearlyIncome	MaxYearlyIncome	records
1	10000.00	20000.00	1849
2	20000.00	30000.00	1849
3	30000.00	40000.00	1849
4	40000.00	40000.00	1849
5	40000.00	60000.00	1848
6	60000.00	60000.00	1848
7	60000.00	70000.00	1848
8	70000.00	80000.00	1848
9	80000.00	100000.00	1848
10	100000.00	170000.00	1848

2. CREATE BINS WITH n ROWS IN EACH

Use a sub-query to calculate the **NTILE** argument as **COUNT(*) / X** where X is the number of records you want in each bin

```
WITH Bins
AS
(
  SELECT
    YearlyIncome,
    NTILE(
      (SELECT COUNT(*) / 1200 FROM DimCustomer)
    ) OVER (ORDER BY YearlyIncome) AS Bin
  FROM DimCustomer
)
SELECT
  MIN(YearlyIncome) AS MinYearlyIncome,
  MAX(YearlyIncome) AS MaxYearlyIncome,
  COUNT(*) AS records
FROM Bins
GROUP BY Bin
ORDER BY Bin;
```

	MinYearlyIncome	MaxYearlyIncome	records
1	10000.00	20000.00	1233
2	20000.00	20000.00	1233
3	20000.00	30000.00	1233
4	30000.00	30000.00	1233
5	30000.00	40000.00	1232

...


12	70000.00	80000.00	1232
13	80000.00	90000.00	1232
14	90000.00	110000.00	1232
15	110000.00	170000.00	1232

3. CREATE BINS WITH EQUAL RANGE

Calculate the range for a column X using **MAX(X) OVER () - MIN(X) OVER ()**. Divide the range by the # of bins you want. This is the bin range.

Divide the column by the bin range to get the decimal bin value. Use **CEILING** to get the integer bin.

```
WITH Bins AS (  
    SELECT  
        YearlyIncome,  
        CEILING(  
            YearlyIncome /  
            ((MAX(YearlyIncome) OVER ()  
              - MIN(YearlyIncome) OVER ())) / 10)  
        ) AS Bin  
    FROM DimCustomer  
)  
SELECT  
    MIN(YearlyIncome) AS MinYearlyIncome,  
    MAX(YearlyIncome) AS MaxYearlyIncome,  
    COUNT(*) AS records  
FROM Bins  
GROUP BY Bin  
ORDER BY Bin;
```



	MinYearlyIncome	MaxYearlyIncome	records
1	10000.00	10000.00	1155
2	20000.00	30000.00	4054
3	40000.00	40000.00	2747
4	50000.00	60000.00	3797
5	70000.00	80000.00	3691
6	90000.00	90000.00	842
7	100000.00	110000.00	1045
8	120000.00	120000.00	332
9	130000.00	130000.00	512
10	150000.00	160000.00	197
11	170000.00	170000.00	112