Recitation 6-Nov-2023

PostgreSQL crosstab(), Google BigQuery

Files:

GitHub repo: https://github.com/acatlin/sql
Deck: recitation_crosstab.pdf

crosstab() hands-on lab

response_id [PK] integer	office text	flavor text
1	Uptown	Chocolate
2	Midtown	Chocolate
3	Downtown	Strawberry
4	Uptown	Chocolate
5	Midtown	Chocolate
6	Downtown	Strawberry
7	Uptown	Chocolate
8	Midtown	Chocolate
9	Downtown	Strawberry
10	Downtown	Chocolate
11	Midtown	Chocolate
12	Downtown	Chocolate

office text	chocolate bigint	strawberry bigint	vanilla bigint
Downtown	23	32	19
Midtown	41	[null]	23
Uptown	22	17	23

Example from: *Practical SQL, 2nd Edition*

crosstab() demonstration

station_name [PK] text	observation_date [PK] date	max_temp integer	min_temp integer
CHICAGO NO...	2016-01-01	31	20
CHICAGO NO...	2016-01-02	34	23
CHICAGO NO...	2016-01-03	32	26
CHICAGO NO...	2016-01-04	32	27
CHICAGO NO...	2016-01-05	34	20
CHICAGO NO...	2016-01-06	38	27
CHICAGO NO...	2016-01-07	40	34
CHICAGO NO...	2016-01-08	45	36
CHICAGO NO...	2016-01-09	44	27
CHICAGO NO...	2016-01-10	28	4
CHICAGO NO...	2016-01-11	25	2
CHICAGO NO...	2016-01-12	22	6

station	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
CHICAGO NORTHERLY ISLAND IL US	34	36	46	50	66	77	81	80	77	65	57	35
SEATTLE BOEING FIELD WA US	50	54	56	64	66	71	76	77	69	62	55	42
WAIKIKI 717.2 HI US	83	84	84	86	87	87	88	87	87	86	84	82

Example from: *Practical SQL, 2nd Edition*

Writing a crosstab() query

1. Enable crosstab() function
2. Load data (in long format) into SQL table
3. Write query that returns three columns:
   a. first column names rows,
   b. second column names columns,
   c. third column is cell values
4. Write query that returns set of category names for columns
5. Write crosstab() function with two queries above (#3 and #4) and AS clause that specifies crosstab's output columns in the order generated by the query (#4b).

crosstab() demonstration

Files:

GitHub repo: https://github.com/acatlin/sql
Data: temperature_readings.csv
Code: median_high_temps_by_city.sql

station_name [PK] text	observation_date [PK] date	max_temp integer	min_temp integer
CHICAGO NO...	2016-01-01	31	20
CHICAGO NO...	2016-01-02	34	23
CHICAGO NO...	2016-01-03	32	26
CHICAGO NO...	2016-01-04	32	27
CHICAGO NO...	2016-01-05	34	20
CHICAGO NO...	2016-01-06	38	27
CHICAGO NO...	2016-01-07	40	34
CHICAGO NO...	2016-01-08	45	36
CHICAGO NO...	2016-01-09	44	27
CHICAGO NO...	2016-01-10	28	4
CHICAGO NO...	2016-01-11	25	2
CHICAGO NO...	2016-01-12	22	6

station	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
CHICAGO NORTHERLY ISLAND IL US	34	36	46	50	66	77	81	80	77	65	57	35
SEATTLE BOEING FIELD WA US	50	54	56	64	66	71	76	77	69	62	55	42
WAIKIKI 717.2 HI US	83	84	84	86	87	87	88	87	87	86	84	82

Example from: *Practical SQL, 2nd Edition*

crosstab() hands-on lab

response_id [PK] integer	office text	flavor text
1	Uptown	Chocolate
2	Midtown	Chocolate
3	Downtown	Strawberry
4	Uptown	Chocolate
5	Midtown	Chocolate
6	Downtown	Strawberry
7	Uptown	Chocolate
8	Midtown	Chocolate
9	Downtown	Strawberry
10	Downtown	Chocolate
11	Midtown	Chocolate
12	Downtown	Chocolate

office text	chocolate bigint	strawberry bigint	vanilla bigint
Downtown	23	32	19
Midtown	41	[null]	23
Uptown	22	17	23

Files:

GitHub repo: https://github.com/acatlin/sql
Data: ice_cream_survey.csv
Starter Code: ice_cream.sql

Example from: *Practical SQL, 2nd Edition*

Developing expertise with crosstab() queries

Try and find (or create) a dataset in long format that would be interesting to convert to wide format using PostgreSQL's crosstab() function.

Consider showing off your work at the beginning of our next recitation session!

BigQuery Beachhead

1. Set up a Sandbox environment
   - https://www.youtube.com/watch?v=StEuT-pntZQ
2. Follow a short two-part example from YouTube
   - https://www.youtube.com/watch?v=Davhwj_8b8Q
   - https://www.youtube.com/watch?v=Abzj-Vyhi74
3. Create your own BigQuery example, and consider showing it at the beginning of our next recitation session!