

Data Engineer Code Test - SQL

Last Modified: 2023-11-21

Instructions

This task is meant to get an understanding of your SQL skills. Our preference is for you to use the GoogleSQL dialect to complete this task, although not required. We will accept any SQL dialect. The solution should be posted to a public GitHub repository owned by you. Please commit directly to the master branch. Commits can be made as often as needed. When complete, share a link to your repository with your recruiter.

Background

We have a table in a BigQuery dataset that contains coupon redemption data. Every time a coupon is redeemed by a consumer, a record is inserted into this table. Each record is associated with a specific retailer and redemption date. Sometimes data is delayed getting inserted into this table and, on occasion, no records get inserted for a retailer for one or more days.

There is a separate process that runs daily against this table that rolls up the number of coupons redeemed per day, by retailer, for the past 7 days. This data is inserted into the `reporting` dataset in a table called `tblRedemptions-ByDay`. No data is purged from this table.

Data

The following tables and data are available to you. *Optional: If you have access to BigQuery and would like to create physical tables, the syntax is provided at the end of this document.*

tblRetailers

id	retailerName	createDateTime
200	XYZ Store	2020-01-28 11:36:21 UTC
300	ABC Store	2022-05-12 14:27:01 UTC
400	QRS Store	2022-05-12 14:27:01 UTC

tblRedemptions-ByDay

id	retailerId	redemptionDate	redemptionCount	createDateTime
122	200	2023-10-29	2738	2023-11-05 11:00:00 UTC
123	200	2023-10-30	3217	2023-11-05 11:00:00 UTC
124	200	2023-10-31	4193	2023-11-05 11:00:00 UTC
125	200	2023-11-01	2931	2023-11-05 11:00:00 UTC
126	200	2023-11-02	2017	2023-11-05 11:00:00 UTC
127	200	2023-11-03	1936	2023-11-05 11:00:00 UTC
128	200	2023-11-04	2813	2023-11-05 11:00:00 UTC
129	300	2023-10-29	3737	2023-11-05 11:00:00 UTC
130	300	2023-10-30	4216	2023-11-05 11:00:00 UTC
131	300	2023-10-31	5192	2023-11-05 11:00:00 UTC
132	300	2023-11-01	3930	2023-11-05 11:00:00 UTC
133	300	2023-11-03	2935	2023-11-05 11:00:00 UTC
134	300	2023-11-04	5224	2023-11-05 11:00:00 UTC
135	200	2023-10-30	3281	2023-11-06 11:00:00 UTC
136	200	2023-10-31	5162	2023-11-06 11:00:00 UTC
137	200	2023-11-01	2931	2023-11-06 11:00:00 UTC
138	200	2023-11-02	2021	2023-11-06 11:00:00 UTC
139	200	2023-11-03	2007	2023-11-06 11:00:00 UTC
140	200	2023-11-04	2813	2023-11-06 11:00:00 UTC
141	200	2023-11-05	2703	2023-11-06 11:00:00 UTC
142	300	2023-10-30	4274	2023-11-06 11:00:00 UTC
143	300	2023-10-31	5003	2023-11-06 11:00:00 UTC
144	300	2023-11-01	3930	2023-11-06 11:00:00 UTC
145	300	2023-11-03	3810	2023-11-06 11:00:00 UTC
146	300	2023-11-05	3702	2023-11-06 11:00:00 UTC

Task

Write a query to pull back the most recent redemption count, by redemption date, for the date range 2023-10-30 to 2023-11-05, for retailer "ABC Store". Your result should have 2 columns and 7 rows. Provide the query and answer the following questions.

Questions

1. Which date had the least number of redemptions and what was the redemption count?
2. Which date had the most number of redemptions and what was the redemption count?
3. What was the createDateTime for each redemptionCount in questions 1 and 2?
4. Is there another method you can use to get the same answer? In words, describe how you would do this (no need to write a query, unless you'd like to).

Table Syntax (Optional)

If you have access to BigQuery, you can create the tables mentioned above using the below syntax. Replace the `[project]` and `[dataset]` with your own project name and dataset name.

```
CREATE TABLE `[project].[dataset].tblRetailers`  
(  
  id INT64,  
  retailerName STRING,  
  createDateTime TIMESTAMP  
);
```

```
INSERT INTO `[project].[dataset].tblRetailers`  
VALUES (200,'XYZ Store','2020-01-28 11:36:21 UTC'),  
       (300,'ABC Store','2022-05-12 14:27:01 UTC'),  
       (400,'QRS Store','2022-05-12 14:27:01 UTC');
```

```
CREATE TABLE `[project].[dataset].tblRedemptions-ByDay`  
(  
  id INT64,  
  retailerId INT64,  
  redemptionDate DATE,  
  redemptionCount INT64,  
  createDateTime TIMESTAMP  
);
```

```
INSERT INTO `[project].[dataset].tblRedemptions-ByDay`  
VALUES  
(122,200,'2023-10-29',2738,'2023-11-05 11:00:00 UTC'),  
(123,200,'2023-10-30',3217,'2023-11-05 11:00:00 UTC'),
```

(124,200,'2023-10-31',4193,'2023-11-05 11:00:00 UTC'),
(125,200,'2023-11-01',2931,'2023-11-05 11:00:00 UTC'),
(126,200,'2023-11-02',2017,'2023-11-05 11:00:00 UTC'),
(127,200,'2023-11-03',1936,'2023-11-05 11:00:00 UTC'),
(128,200,'2023-11-04',2813,'2023-11-05 11:00:00 UTC'),
(129,300,'2023-10-29',3737,'2023-11-05 11:00:00 UTC'),
(130,300,'2023-10-30',4216,'2023-11-05 11:00:00 UTC'),
(131,300,'2023-10-31',5192,'2023-11-05 11:00:00 UTC'),
(132,300,'2023-11-01',3930,'2023-11-05 11:00:00 UTC'),
(133,300,'2023-11-03',2935,'2023-11-05 11:00:00 UTC'),
(134,300,'2023-11-04',5224,'2023-11-05 11:00:00 UTC'),
(135,200,'2023-10-30',3281,'2023-11-06 11:00:00 UTC'),
(136,200,'2023-10-31',5162,'2023-11-06 11:00:00 UTC'),
(137,200,'2023-11-01',2931,'2023-11-06 11:00:00 UTC'),
(138,200,'2023-11-02',2021,'2023-11-06 11:00:00 UTC'),
(139,200,'2023-11-03',2007,'2023-11-06 11:00:00 UTC'),
(140,200,'2023-11-04',2813,'2023-11-06 11:00:00 UTC'),
(141,200,'2023-11-05',2703,'2023-11-06 11:00:00 UTC'),
(142,300,'2023-10-30',4274,'2023-11-06 11:00:00 UTC'),
(143,300,'2023-10-31',5003,'2023-11-06 11:00:00 UTC'),
(144,300,'2023-11-01',3930,'2023-11-06 11:00:00 UTC'),
(145,300,'2023-11-03',3810,'2023-11-06 11:00:00 UTC'),
(146,300,'2023-11-05',3702,'2023-11-06 11:00:00 UTC');