

Data Team Coding Challenge

Assessment Criteria

Technical

- Assess candidate's ability to write complex SQL statements
- Assess candidate's analytical and logical skill

Documentation

- Assess whether code was documented to enhance readability
- Assess whether the result is well elaborated and conveyed

Coding Challenge

Question 1

This is a snapshot of a pageview activity log table

ID	User_ID	Page_ID	Visit_Date	Visit_Time
1	1	54	2018-01-01	11:54:34
2	1	55	2018-01-01	11:55:10
3	1	56	2018-01-02	13:11:12
4	1	55	2018-01-02	17:10:08
5	1	56	2018-01-02	17:12:45
6	2	55	2018-01-01	10:25:18
7	2	55	2018-01-01	17:30:12
8	2	55	2018-01-01	17:45:57
9	3	54	2018-01-02	00:00:12
10	3	56	2018-01-02	00:03:22
11	3	55	2018-01-02	01:20:11
12	3	56	2018-01-02	01:40:09

Write an SQL statement to find the total number of user sessions each page has each day.
(A user session is defined as continuous activity on a site where each activity is within 10 mins of each other.)

Link to the pageviews table csv file:

<https://s3-ap-southeast-1.amazonaws.com/ms-data-coding-challenge/SamplePageviews.csv>

Expected result:

Page_ID	Visit_Date	Total_User_Sessions
54	2018-01-01	1
54	2018-01-02	1
55	2018-01-01	4
55	2018-01-02	2
56	2018-01-02	4

Question 2

This is a link to a sample order dataset:

<https://s3-ap-southeast-1.amazonaws.com/ms-data-coding-challenge/SampleOrders.csv>

There are some queries from the sales team you need to solve. You can defined your own metrics and tell your data story if they didn't specify. SQL is preferred but not necessary for these question.

1. Which month has the highest sales? Is there any seasonality effect?

2. Which product is the recent best seller?

Hint: There is no `right` answer for this question, please feel free to make your own interpretation and give suggestion

3. Is there any group of products which are often bought together?

Hint: You can define you own metric, however, please explain the rationale behind

4. Is there any other insight you can get to help to improve sales number?

Hint: It is okay to not having any significant insight, but you have to state the hypothesis and verify

5. Based on the data we have, what kind of BI dashboards you would build in order to help the sales team monitoring the performance?

Submission Instructions

- 1) Create a public repository in any popular version control service (Github / Bitbucket etc)
- 2) Upload your answer scripts and packages to the public repo
- 3) Create a readme file that describes which scripts contain the answers to each question in the coding challenge
- 4) Share the link to the repo with us