Re: Working with Data Assignment 2 TU060

Lucas Rizzo <Lucas.Rizzo@TUDublin.ie>

Fri 31/12/2021 12:55

To: C06543201 Rachel Donnelly <C06543201@mytudublin.ie>

Hi Rachel,

I answer inline.

You seem to be almost there. Well done!

Regards,
Lucas.

From: C06543201 Rachel Donnelly <C06543201@mytudublin.ie>
Sent: Thursday 30 December 2021 14:08
To: Lucas Rizzo <Lucas.Rizzo@TUDublin.ie>
Subject: Working with Data Assignment 2 TU060

Hi Lucas,

Hope you're well.

I've been working away on this assignment, and although I've already spent a lot of time on it I'm afraid I'm not getting on too great with it... I just wanted to check with you that I am actually doing the right thing and proceeding through it in the correct way. Is the process for completing the assignment:

Section 1 - Seems all good!

1. Design star schema. Justify design decisions in report. State the business questions you anticipate answering with this design.
2. Import all 8 data files into SQL Developer using the Data Import wizard, WITHOUT changing them in any way before import. Take screenshots of this process and add to the report.
3. Create ERD diagram of star schema in Data Modeler and generate DDL script. Include ERD diagram in report.
4. Use automatically generated DDL script to create a new empty set of tables in SQL Developer in the desired shape of star schema.
5. Create an SQL script to populate the new tables with the data from the imported tables, WITHOUT changing the imported files in any way.
6. Add the DDL script and the transformation script together into one script file that is to be submitted along with the report. Should this be a .txt file? DDL and SQL scripts are both SQL code, not sure why you need to use a .txt file. The idea is that you (Lucas) will be able to import the data files just once and then generate each student's star schema just by running their scripts once without needing to adjust anything? Yes, that is the idea.
7. Just in case the script doesn't work, provide screenshots in the report of what the tables look like such as this one:

	PHONE_NUMBER	CALL_TIME	DURATION	CONNECTION_ID	CALL_TYPE	CALL_RATE	COST
1	022 640 7415	26-JAN-21 11:00:00.000000000	363.5955278	5b2bf121-de36-4ac5-b308-eef87cb0a9fc	peak	0.12	0.73
2	022 640 7415	04-MAR-21 11:00:00.000000000	468.2356927	2afade06-7ae5-4f2b-a25a-b368a9812de2	peak	0.12	0.94
3	022 640 7415	25-MAR-21 17:00:00.000000000	412.4217463	1b1f76f4-585a-4af6-be88-a8557c100f85	roaming	0.57	3.92
4	022 640 7415	11-MAR-21 13:00:00.000000000	51.7193495	f7dab53d-5bab-4f0d-8f5d-2c807a1cf65a	peak	0.12	0.1
5	01 637 6101	11-JAN-21 13:00:00.000000000	901.4575955	99049c7a-d3bb-4d84-8b4d-cff871e3b714	peak	0.09	1.35
6	01 637 6101	12-MAR-21 05:00:00.000000000	439.0758634	61855ad4-fa6c-44c7-b8c3-7e0d6cabcee9	roaming	0.57	4.17
7	01 637 6101	27-FEB-21 13:00:00.000000000	669.3785716	9f7995d7-57db-4e31-84cb-312743b99c9a	peak	0.09	1
8	01 637 6101	28-MAR-21 17:00:00.000000000	811.7916428	4e221165-cab7-40b4-892b-e1a85039c59e	peak	0.09	1.22
9	01 637 6101	06-FEB-21 18:00:00.000000000	666.8633995	8289af29-34db-44c1-aaa0-7fc4170c12d8	off_peak	0.09	1
10	01 637 6101	12-FEB-21 13:00:00.000000000	712.4501546	50e3313e-8252-42b1-b664-23a6f4bd0d93	international	0.26	3.09

Section 2:

1. Write a series of SQL queries to answer the questions posed in Section 1.
2. A separate script file does not need to be provided for this part. Just write the SQL queries in the report, followed by a screenshot of the query output? That is fine.
3. Am I correct in thinking that at least some of these queries should be done using PL/SQL with complex logic like loops, procedure and cursors? Or is just normal SQL fine for this section? They don't need to be done using PL/SQL, but if you think you need PL/SQL because you are thinking on more complex queries that could be a good addition.
4. How many queries should there be, or how long should this section be. Will 10-15 query result screenshots be enough to include in the report, if the queries are complex enough? 10-15 should be more than enough if the queries make sense, and are not too similar to each other.

Section 3:

1. Design a new fact table. Justify design decisions in the report.
2. Write an SQL script to create this table (or automatically generate it in Data Modeler).
3. Once again, populate the table using data from the original import files, WITHOUT changing them. At this stage, is it also possible to populate the new fact table using the data in the star schema tables created in Section 1?
   Yes. In fact it is better if you populate the fact table using the data in the star schema tables.
4. Now split this new fact table into training and test sets, creating a new table or view in order to do so.
5. Implement ML models to predict churn
6. Write PL/SQL script to evaluate models and present results to user.
7. Put screenshot of these results in report and briefly discuss.
8. Put the SQL code for creating the new fact table, splitting it, running the ML models and evaluating them all into one new SQL script file to be submitted along with the report. Again, you should just be able to import the original csv files into SQL Developer, run the two script files provided in the correct order and the student's entire assignment should be created instantly without any need for changes by you. We should not need to submit any data files at all, just the report and 2 SQL scripts, is that correct? Yes, correct.

I already have my star schema implemented in SQL, but the problem is I did it by editing the original import files, which I now think was incorrect. Is this right?
Yes, I wouldn't expect you to edit the original import files to create the star schema. But also I don't know which changes you made and why. There are different star schemas that could be correct, so if you can justify well the changes in the original files MAYBE it won't be a problem. If you don't want to risk, use the original files as they are given.

Thanks in advance for your help with this Lucas!

Best regards,

Rachel