Hand in# 2 feedback for group 11

Very nice document organization 😍

Comment on Revisions (if any)

1. Did the partner group implement suggested revisions from last time?
   a. Yes, except for the one about Dim_Product but then it probably wasn't relevant to the case

2. Are the revised changes explained?
   a. Yes, very well

Comment on ETL

1. Are the steps explained?
   a. Steps and choices from ETL are documented and explained thoroughly, very nice

2. Are the appropriate transformations done?
   a. Yes - the data is cleansed, static Date dimension introduced and appropriate surrogate keys are present

3. Are transformations explained?
   a. Yes, in a lot of detail 👌

4. Are surrogate keys created?
   a. Yes

5. Are the appropriate data types used?
   a. Yes, according to relevancy as well as compatibility to the source system

6. Does the Fact Table store data in lowest granular form?
   a. Yes, OrderLineItem in the EDW sourced from SalesOrderDetails

7. Does the hand in explain choices related to data quality using the data quality terminology?
   a. They explained everything that matters, but didn't insert the actual keywords of terminology. But to be honest, that doesn't really matter - seems rather arbitrary.

Comment on date dimension

1. Does the date dimension contain relevant attributes for analysis? (e.g. week, month, day, weekend etc.)
   a. Yes
2. Does the date dimension use a surrogate key according to Kimball's suggestions (p. 255) and does the group argue for the design choice?

a. Yes, but there is no argument for creating the surrogate key, also no argument for why they added the Date field although we assume it was to be compatible with the source system
b. There is a solid argument on why the dimension is generated as static and was not created until the Load stage of ETL

Comment on documentation

1. Is the SQL code there? Does it contain relevant commenting?
   a. That's a weird thing to ask, but yes, the SQL is there 👀
   b. The comments are present and good, we would like more comments on the updates
   c. There are some queries commented out in several places, was this meant to be deleted before handing in?

2. Are you able to run the SQL code without errors?
   a. Yes, we need to have the AdventureWorks2019 source database to do so, but it is not something you can fix in code

3. Does the hand in include complete source-target mapping?
   a. Yes

4. Is everything documented?
   a. Yes and very cleanly

5. Does the hand in include an activity diagram or a flow diagram showing the steps involved in updating the DWH (or text-alternative)?
   a. The hand in contains an activity diagram

6. Does the diagram explain exactly what has been done (compare it to code)?
   a. Sufficiently so. The general idea is outlined in the activity diagram next to each action and it would be ludicrous to have "exact" documentation

Quality assurance

1. It is important that and ETL procedure does not "lose" data due to misspecification. Do a bit of quality assurance on the partner groups hand in
   a. The data cleansing is thorough

2. Compare number of records in source tables with number of records in corresponding DWH dimension tables. Do they match? If not, why not? (Hint: DimCustomer could be an issue)
   a. They don't match in Product (504 entries in SRC, 295 in the DWH) due to the data cleansing process, could be due to not importing lines that have a null ProductSubcategoryID in the source system

Additional comments

Very well done 🔥 Meticulous documentation and thorough implementation. The remarks are pretty minor and most of them are like "didn't know we have to do that 👀" kind. So yeah, this is splendid. Half of the semester is done! 💪 :D

1. Is there something you particularly liked about the group's implementation?
    a. It works 🔥
    b. The SQL script works splendidly. The only thing is that we needed to restore AdventureWorks2019 source database, but probably that is not something that you can affect through SQL.

2. Is there something that is not explained?
    a. Dim_Employee - JobTitle/EmployeeType - Feels weird to have two types of employee typing, while it's understandable since there can be different work times (full time, part time, etc.), since the case doesn't mention any specific relation regarding this, it was kind of confusing
    b. Dim_Customer - CustomerType: What is an example input of 2 chars?

3. General design
    a. Why is there a Location dimension and then TerritoryId in Order when there's also one in Sale?
    b. Same for Country in Employee and Customer, shouldn't that point to Location dimension instead?
    c. Fact_Sale: Total amount has type numeric(38,6) → having 6 decimal places gives a lot extra 0's on queries. Could this be adding to issues with memory?
    d. What does "created in source system" mean in source target mapping of Product? Do you mean created in the staging?

4. Not sure what that "ran out of memory" means? Did you encounter some performance issues?

5. Line 357 of SQL(below): You don't need this in the code, it's a general case example of date conversion

```sql
DECLARE @Date DATETIME;
SET @Date = getdate()
SELECT CONVERT (CHAR(8), @Date, 112)
select GETDATE()
```