**ERD Companion Commentary**

The purpose of this document is to provide commentary, details, and questions regarding the submitted ERD. Each section will be formatted in the following:

* Name of table
* Table Type: Dimension or Fact
* Sample of one null row in the table
* Sample of one row in the table
* Justification for table and its columns
* Questions or concerns regarding the table

## Dim\_Calendar

* Table Type: Dimension table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date Key | Date | Year | Month Num | Month Name | Day of Week | Quarter Num |
| -1 | 1/1/1995 | 0 | 0 | Not Applicable | Not Applicable | 0 |
| 1 | 1/1/2013 | 2013 | 1 | January | Tuesday | 1 |

* **Justification**: Need to create a calendar table so we can connect target and sales fact tables. I included Month Number and Month Name for flexibility. Being able to calculate sales by day of week and quarter were planned in my data profile.
* **Questions**
  + I am assuming that this table will be created in Snowflake; we will not be uploading a CSV file of dates?

## Dim\_Product

* Table Type: Dimension table



* **Justification**: I want to combine product, product category, and product type CSV files. This should be fairly simple since they have primary keys that can join together. Products will be loaded using the Product\_ID column from Product CSV.
* **Questions**
  + I included Unit of Measure ID, but we do not have a CSV file that expands on this. Is this the best way to address this kind of situation? Should this column we removed?

## Dim\_Branch

* Table Type: Dimension



* **Justification**: I want to combine Store, Customer, and Reseller CSV. They have similar overlapping columns such as Name, Last Name, Phone, and Email but this information does not mean the same for each table (ex. Store’s First Name is the store manager’s name but Customer’s First Name is a customer’s first name). Hence why I want to add a “Branch Type” column that is based off the table this data is taken from (ex. Customer CSV will have Branch Type of Customer). Gender is CHAR(1) to ensure gender format is in a single character format and avoid “F” and “Female” confusion.  
  **(I think this table is the trickiest and I may need to modify it or split store, reseller, and customer tables.)**
* **Questions**
  + I made several assumptions on how I would be able to load and transform my data in Snowflake. Can Natural Keys be varchar? My google research says yes.
  + Same question posed in Dim\_Product; what should I do with Subsegement\_ID?
  + Can I do a Group By to somehow combine these tables?
  + Are there advantages to combining these tables vs. keeping them separate?

## Dim\_Location

* Table Type: Dimension



* **Justification**: I wanted to separate location from Branch information. This is mainly because Targets appear to be location and store specific? I made postal code a VARCHAR because (1) no math will be done with postal codes and (2) I avoided using fixed CHAR so the database can be more flexible.
* **Questions**:
  + None.

## Dim\_Channel

* Table Type: Dimension



* **Justification**: I wanted to split channel so my fact tables can easily be filtered by Channel. My data profile is pretty channel focused as well, so I figured it may make things easier.
* **Questions**:
  + None.

## Fact\_Sales

* Table Type: Fact



* **Justification**: Most of this table was taken from the template. I added a Dim\_Channel\_Key due to my choice of splitting Channel into its own Dimension table. I also changed naming conventions for easier readability.
* **Questions**:
  + None.

## Fact\_Target\_Product

* Table Type: Fact

|  |  |  |
| --- | --- | --- |
| Date Key | Product Key | Target QTY |
| 0 | 0 | 0 |
| 1 | 1 | 750,023 |

* **Justification**: This table will connect with Date and Product dimension table.
* **Questions**:
  + None.

## Fact\_Target\_SRCSales

* Table Type: Fact

|  |  |  |  |
| --- | --- | --- | --- |
| Branch Key | Date Key | Channel Key | Target AMT |
| 0 | 0 | 0 | 0.00 |
| 1 | 1/1/1900 | 1 | 14,760 |

* **Justification**: This table is a little tricky since the original table does not have any keys/ID’s. I assumed we will use Snowflake to clean the tables somehow and be able to link this information to Channel and Branch dimensions.
* **Questions**:
  + Will table cleaning be done in Snowflake?