Superstore - Challenge

Background - Description of Data Files

The Superstore dataset represents a retail store selling to customers in the US from 2014 to 2017. There are several files – assume that of these are a download of a table as if they were generated by the operational IT system of this retailer. The files are:

- Transaction – this is a fact table. It contains several numerical fields, sales, Quantity and Profit. These are values that we want to group aggregate. It has a lookup field (foreign key) into the four tables below
- Product – a dimension table that lists each product and grouping into categories and subcategories
- Customer – a dimension table that lists each customer and their segment
- Geography – a dimension table that lists the post code and broader geographical attributes (City, State, Region)
- Order – a dimension table that lists each individual order, the dates when it was ordered and shipped and the ship mode (how it was shipped)

There is also a Calendar (date) table that contains all the dates

Instructions

Build a data model in Power BI step by step.

Import each table into the model and then build visuals to show that the model works well. Import the tables in the order they are listed below.

Improve the model after each step: consider relationships, DAX measures and calculations, which fields to hide, hierarchies and so on.

You may want to consider some design aspects such as:

- The pattern of tables that you are building e.g. star schema
- Using the Calendar table rather than the auto – date – time feature
- Placing all measures in a separate Calculations (measure) table

Hint: these are the types of visualisation you may want to experiment with

- bar chart, sales over time, (show Explain Increase)
- table with data bars, sales, profit., margin
- decomposition tree (sales)
- decomposition tree (profit margin) – use conditional formatting
- bar chart by quarter – find where distribution is different
- key influencers
- waterfall

Optional - Advanced Activities

Add time intelligence measures: e.g. YTD Sales

Once the Calendar is imported, the data model becomes a snowflake rather than a star pattern. Is there any benefit of a staying with a star schema design? If so, how would you achieve this?