



# Data Engineering Interview - Practice Test

In this task we will be using kaggle public dataset **[superstore\_data retail sales]**

1. Download the dataset using the link below to your local disk, you may need to register for kaggle login before downloading

<https://www.kaggle.com/jr2ngb/superstore-data>

2. Define list of Entities in a your Data Model
3. Try to build a multidimensional schema data model (tables) using open source MySQL database or any DB of your favorite
4. Apply any data cleansing/ transformation when needed
5. Define relationships Primary, Foreign keys or constraints in your data modelling
6. Write SQL Query statements, using new data model to answer the questions below:

- **Qry 1:**

Total sales items by month for the last 4 years



- **Qry 2:**

Top 5 products sold during January period, along with Category and sub- Category names

- **Qry 3:**

Total quantity sales for Weekends by Region by Week

**7. Discuss the issues/disadvantages with this dimensional model**