**Welch’s business analytics project – initial steps**

Our goal will eventually be forecasting units shipped at the customer – product level. Our initial steps will involve data cleaning and exploratory data analysis (EDA). Here are some concrete initial steps. Each team must do at least these and is encouraged to expand on these.

1. Data cleaning
   1. Clean up the INVENTORY\_ITEM\_DESC column (remove unnecessary parts/ separate into size and product name)
   2. Remove unnecessary parts of the EBS\_CUSTOMER\_NUMBER\_DESC column
   3. Add a date column with the format ‘mdy’. The current date column contains “Fiscal year – week”. The Welch’s fiscal year begins on September 1. So, “2019 – 1” corresponds with the week of “9/1/2019”.
2. Data exploration
   1. Who are the top 10 customers? Do they change over the years (2019 – 2022)?
   2. What proportion of units are accounted for by the top 10 customers in each year?
   3. What are the top 10 products? Do they change over the years (2019 – 2022)?
   4. What proportion of units are accounted for by the top 10 products in each year?
   5. Time series of units shipped by consolidated major group (CONSOLIDATED\_MAJOR\_GROUP)
   6. Look at the overall time series by week over the entire period. What trends do you see? Any possible pandemic effects?