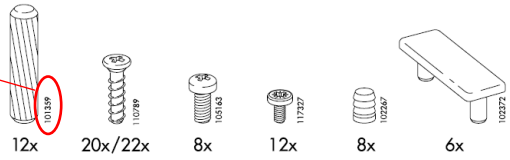
Data Analysis for IKEA Bed Manufacturing

This project explores the use of Python's data carpentry, visualization and analysis tools to investigate a small dataset of parts lists associated with two catagories of IKEA beds:

Part No	Quantity
101359	12
110789	20
105163	8
117327	12
102267	8
102372	6



To do this each student has been allocated two, of the following three, spreadsheets that list the parts uses in IKEA's Single and Guest beds:

- SBmatrixc.csv
- GBmatrixc.csv

The objective is to create a Python notebook that demonstrates your knowledge of:

1. **Data Cleaning**: Write python code that summarizes the size and shape of the data allocated to you and cleans it to remove any inconsistency in the format. These spread sheets have been created from a combination of webpage scrapping, extracting text from pdf of assembly instructions and manual entry. Consequently, there are errors, outliners and inconsistencies in the formatting. You should record any choices or assumption you make (e.g. how to deal with multiple part codes such as '113434/122332') in comments in the Python code and also the markdown cells between the code cells of the notebook. We are looking for code to automate this step, no marks will be given for manual processes of the data with, say, Excell

2. **Data Merge:** Write Python code that combines your datasets and ranks the components by the frequency of their use in the bed designs. Again, there will be choice about how identical and different part numbers are handled. The approach adopted should be described in the comments and markdown cells of the notebook.

3. **Data Analysis:** Write python code that calculates the percentage of components in each bed which are unique? The approach adopted should be described in the comments and markdown cells of the notebook.

4. **Data Discovery:** Apply the PCA *analysis method of your choice* to the combined dataset.

5. **Conclusions:** The notebook should end with a brief discussion and visualization of the results and how they could be used to improve the productivity.