Data Science for Manufacturing Assessment

This course is assessed by two, individual, coursework exercises:
- Assignment 1 (25%): given out Friday 25th of February & submitted on Thursday 17th of March
- Assignment 2 (75%): given out Friday 18th of March & submitted on Thursday 7th of April

This document describes assessment 1, an outline of assessment 2 will be available on Friday 4th March
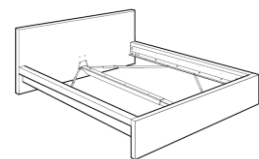
Assignment 1

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

Part No	Quantity
101359	12
110789	20
105163	8
117327	12
102267	8
102372	6
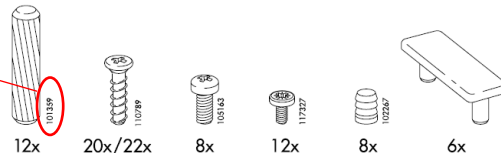
To do this each student has been allocated two, of the following three, spreadsheets that list the parts uses in IKEA's Double, Single and Guest beds:
- DBmatrixc.csv
- SBmatrixc.csv
- GBmatrixc.csv

Assessment 1 Dataset and Analysis Method Assignment

		Datasets				Datasets	
Munzir	**Bakhit**	dbmatrixc	sbmatrixc	**Cristian**	**Ghiorghe**	dbmatrixc	sbmatrixc
Stewart	**Barrows**	dbmatrixc	gbmatrixc	**Kyle**	**Houston**	dbmatrixc	gbmatrixc
Paul	**Bennie**	sbmatrixc	gbmatrixc	**Lisa**	**Hu**	sbmatrixc	gbmatrixc
Daniel	**Brophy**	dbmatrixc	sbmatrixc	**Nicholas**	**Lines**	dbmatrixc	sbmatrixc
Alex	**Campbell**	dbmatrixc	gbmatrixc	**Debaditya**	**Choudhury**	dbmatrixc	gbmatrixc
Gerry	**Morris**	sbmatrixc	gbmatrixc	**Graham**	**Murtagh**	sbmatrixc	gbmatrixc
Michael	**Coleman**	dbmatrixc	sbmatrixc	**Jonathan**	**Nowek**	dbmatrixc	sbmatrixc
Edward	**Curran**	dbmatrixc	gbmatrixc	**Thomas**	**Smith**	dbmatrixc	gbmatrixc
Euan	**Donnelly**	sbmatrixc	gbmatrixc	**Josh**	**Sullivan**	sbmatrixc	gbmatrixc
Aleks	**Ferrari**	dbmatrixc	sbmatrixc	**Mark**	**Wilkie**	dbmatrixc	sbmatrixc
Shi	**Chen**	dbmatrixc	gbmatrixc	**Houbin**	**Wang**	dbmatrixc	gbmatrixc
Shaojiang	**Chen**	sbmatrixc	gbmatrixc	**Chengcheng**	**Wang**	sbmatrixc	gbmatrixc
Shirron	**Ding**	dbmatrixc	sbmatrixc	**Hao**	**Xue**	dbmatrixc	sbmatrixc
Praneeth	**Sarvade**	dbmatrixc	gbmatrixc	**Jing**	**Yu**	dbmatrixc	gbmatrixc
Gavin	**Davis**	sbmatrixc	gbmatrixc				

Your 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

[10 marks]

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.

[10 marks]

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.

[10 marks]

4. **Data Discover**y: Apply the *analysis method of your choice* to the combined dataset. Next week the workshop notebook will give some examples of using python libraries that do cluster analysis[1] and association rule mining[2] to identify families of products in terms of components they share. However, you are free to use any method that provides an insight into the structure of your data. Your notebook should include a number of cells that give a brief overview of the method you are using.

[15 marks]

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

[5 marks]

Notes:

- It is expected that you discuss the assignment with tutors and lectures and ask for help with this exercise. The more detailed the question the easier it is to help.
- Please note that the University has a very strict plagiarism policy and several automated system that check submitted work. Any submitted work must be your own but obviously, it is acceptable to base your code on examples from any source.
- More comments and explanation more mark.
- It is difficult to quantify the size of the assignment in terms of word count, but if a pdf printout of your notebook is more than 20pages long (including tables and plots etc.) please check with one of the staff that what you've done is appropriate.

Submission

You should submit your notebook on the University's "Learn" system (see Assessments). Your submission should take the form of a Python Notebook that is named with the format "Name Assessment1 Partx" (eg "Jean Smith Assessment 1") and has clearly **commented code and cells formatted with markdown**. Please also upload a PDF file of the notebook for review by the course's external examiner. It is acceptable to use functions from pythons Libraries and adapt code examples from the lectures.

You should submit your notebook on Learn (see Assessment section) no later than Thursday the 17th of March

---

[1] https://www.geeksforgeeks.org/multidimensional-data-analysis-in-python/

[2] https://medium.com/@mervetorkan/association-rules-with-python-9158974e761a