Introduction

In order to analyze data for increasing the sales of a product or gain of a company, first we have to refine the data in the form that is useful for us to predict the possibility of the different outcomes and for that the data must be structured in such a tidy way that it should be able to facilitate the analysis. Not only is it the first step but at times it is also needed to be repeated as more data is collected or when we are met with a problem of different sort and require more filtering of the given data. It is surprising how little research has been done om how to tidy up your data so in this journal we will cover the topic on how to standardize the data for further analysis.

Literature Review

You can structure the data by representing it in the form of rows and columns so that it is tidy enough to avoid complications that would be faced during analysis. Problems that make a dataset messy are mentioned below:

- Column headers are values, not variable names.
- Multiple variables are stored in one column.
- Variables are stored in both rows and columns.
- Multiple types of observational units are stored in the same table.
- A single observational unit is stored in multiple tables.

Data structured in such a way that it forms perfect relationship between variable and observations is how you know the data is tidy enough to perform tasks on it.

However, there are certain tools that make the data refining worthwhile because there is no point in tidying up the data if it does not make the analysis any easier. These tools are called “Tidy Tools” that take tidy data as an input and gives the desired tidy output. These tools are very useful as the output of one tidy tool could be used as an input for another depending on the data required.

The analysis of data is divided into 3 important components that are as follows:

- Manipulation.
- Visualization.
- Modelling.

Conclusion

This summary covers one of the many tasks performed by every data scientist when it comes to analyzing the given data for the betterment and increasing the gain of a company, that is “Data Cleaning/Tidying”. It is surprising how important it is to tidy up the data yet it is such an uncommon subject. Hopefully there will be more improvements made to clean the data so that we are enabled to reduce the friction of getting data in the form that is useful.