Data Science: Cleansing And Visualization For Beginners

This article is an outcome of a short project undertaken by collaborators, persevering in discovering the ropes of Data Science. The names of all the collaborators are mentioned at the end of this article.

Data science is a fascinating discipline that is both artistic and scientific at the same time. A project journey in Data Science involves extracting and gathering insightful knowledge from data that can either be structured or unstructured. The entire tour commences with data gathering and ends with exploring the data entirely for deriving business value, during which many procedures are applied systematically. Broadly speaking, the cleansing of the data, selecting the right algorithm to use on the data, and finally devising a machine learning function is the objective in this journey. The machine learning function derived is the outcome of this art that would solve the business problems creatively.

This article focuses exclusively on the Data analysis, cleansing, exploration, and imputation of data. I describe the steps that we undertook in this journey, forming the crux of this article.

Step 1. Loading the data set

We were handed over e-commerce data to load.

<<TBD>>

Step 2. Interpreting and transforming the data set

This step involved loading the data set and trying to understand what the data set contains. In a real-world scenario, the data information that one starts with

could be either raw or unsuitable for Machine Learning purposes. We had to transform the incoming data suitably.

<<TBD>>

Step 3. Cleansing and Imputing the data:

We looked into invalid values in the data set ignorer to impute the dataset with clean data in this step.

<<TBD>>

In our case, we introduced errors purposefully to understand how we can impute the data.

<<TBD>>

Step 4. Exploring and Analysing the data: A cleaned up and structured data is suitable for analyzing and finding exemplars using visualization.

Conclusion

<<TBD>>