

Week 1 Assignment:

Q2: If you finally find the dataset of any event/experiment that you've been searching for (say in .csv format), do you think your job of analyzing the data is done, just by performing operation on the computer? Are there any issues you could face?

ANSWER:

If we get the data in a particular format, we can first of all interpret it by visualizations in forms of graphs, histograms, pie charts etc. One important part of interpretation will be the analysis of data in terms of formulating models using techniques like regression but we can face issues if the model is somehow faulty due to limited dataset or has background factors involved which we don't take into consideration say for instance we obtain a dataset which depends on various parameters A,B,C in the correct model and we fail to see dependence of C and the dataset works fine for our model based on A and B then we cannot generalize it to all datasets and we will face issues with it.

So, we also have to do something called Probabilistic Reasoning wherein we can obtain datasets by using the model and see if we can generate a valid dataset using our model.

Hence, only visualization and modelling won't be sufficient but we also have to verify our model and test it repeatedly.