**CHECKPOINT-3**

* One of the major factors for having a decent model is having well-cleaned data so that the model does not get confused.



* Pandas help us with the proper data analysis and data manipulation.

**RESOURCES**

* [Learn Pandas Tutorials (kaggle.com)](https://www.kaggle.com/learn/pandas)
* [10 minutes to pandas — pandas 1.5.2 documentation (pydata.org)](https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html)

**EXERCISE**

Try and explore different pandas’ functions and answer the following problems by analysing the data set given below:

<https://drive.google.com/drive/folders/1hViPFeNvcmi7ZolKFYCeyYeiim4vcm5n?usp=sharing>

QUESTIONS –

1) Reading the above CSV file in google colab using the pandas’ library and converting it into data frames.

2) How many values are missing for each column and try dropping or replacing NULL values using appropriate pandas’ function.

3) Drop the columns which are irrelevant features to feed into our model.

4) Some columns are not statistical data but in the form of strings. It Is difficult for our model to learn when the data is in form of strings. Your task is to encode these strings using proper encoding techniques.