# Hackathon Data Dictionary:

- ParentSKU: Parent Stock Keeping Unit.
- Site ID: Location of Warehouse.
- Year: Year on which sales had happened.
- Month: Monthon which sales had happened.
- CategoryNameID: Product Category Name ID.
- Unit: Unit of the product package.
- Qty: Number of products sold.
- Net Sales Calculated: Total amount that should be billed.
- Cash Discount: Per product discount on purchase of bulk.
- Amount to customer: Amount that the customer has been billed.
- MRP: Maximum Retail Price of the product.
- Pack Size: Quantity of per product that has been bought.
- Pack Unit: Unit of the product.
- State: State at which the sales had happened.
- Zone: Zone ID where the sales had happened.
- Master Category: Master Category of the product. Super set of Category Name ID.
- Size: Pack size of the product.
- Rank: Priority of the product to the client. Rank 1 is more important and so on..
- Date: Transaction date.

1. Noise is present in the data, the participant must make a valid judgement on how to deal with it.
2. In real world data, noise varies on a large degree. The given dataset is an actual real world data.
3. Noise mostly gets induced due to the way they collect the data, some human error and most often due to data aggregations while preparing the data.
4. Working on this kind of data would prepare the participants to handle the complicacies that they would face in the future.

5. This a pre-modeling insight extraction competition, so there is no need to use any ML models here.
6. In any company Data Scientists take 1-3 months of duration for Data Analysis itself. Only a small proportion occupies where they'd be using ML models.
7. Due to the insights they were able to capture during the 1-3 months, Data Scientists would know which models meet the Parametric Assumptions of the data. So choosing the right model is not a biggie. That is the reason why analysis is an important aspect.
8. Plots without any meaning are not useful, justify what you're trying to tell through your plots and analysis.
9. Hypothesize your idea and try to find and prove it. This would be the best way to move through the competition.
10. You're free to use any other plotting tools than that you've learned during your course. Just make sure it's a python library.

**GOOD LUCK!**