Introduction to EDA

Understanding Data through Visualization

While working on the big mart dataset, you have explored all the basic details. You know the shape of the data and the various data types it contains, and you have also explored some statistical information from it. Now, it's time to visualize this data and find more insights to help increase our understanding of this dataset.

There are mainly three ways to explore the dataset using **visualization** tools:

1. **Univariate Analysis**
2. **Bivariate Analysis**
3. **Multivariate Analysis**

You will perform a **Univariate Analysis** of our dataset in the next few questions.

Univariate Analysis

Univariate is a combination of two words where Uni means "**One**" and variate means "**variable**". You will explore each variable and find interesting facts about them individually. The main objective of performing Univariate Analysis is to analyze some patterns like:

1. **Central Tendency (mean, median and mode)**
2. **Standard deviation**
3. **Quartiles (IQR)**
4. **Dispersion (range, variance)**

Types of Graphs for Univariate Analysis

There are mainly two types of data present in any dataset, categorical data and quantitative/numerical data. There are different types of plots you can use to perform univariate analysis.

For Categorical data:

- **Countplot**: Used to show the counts of observations in each categorical bin using bars.
- **Piechart**: Used to show the relative count or frequency of a sample or population.

For Numerical data:

- **Histogram**: Used to analyse and understand the distribution of the sample. It is a value distribution plot of numerical columns.
- **Distplot**: Used to analyse the probability distribution and the skewness.
- **Boxplot**: Used to find out about the outliers present in the feature.