**PROJECT:1**

**Exploratory Data Analysis in R**

**Trainer: Mrs.Indrani Sen**

**Submitted by**

**Dr.Shilpa Joshi**

Exploratory data analysis (**EDA**) the very first step in a data project. We will create a code-template to achieve this with one function.

## Introduction

EDA consists of univariate (1-variable) and bivariate (2-variables) analysis..

* Step 1 - First approach to data
* Step 2 - Analyzing categorical variables
* Step 3 - Analyzing numerical variables
* Step 4 - Analyzing numerical and categorical at the same time

Covering some key points in a basic EDA:

* Data types
* Outliers
* Missing values
* Distributions (numerically and graphically) for both, numerical and categorical variables.

### Type of analysis results

They can be two: informative or operative.

**Informative -** For example plots, or any long variable summary. We cannot filter data from it, but give us a lot of information at once. Most used on the **EDA** stage.

**Operative** - The results can be used to take an action directly on the data workflow (for example, selecting any variables whose percentage of missing values are below 20%). Most used in the **Data Preparation** stage