**Reading workbooks/datasets**

>Transactions <- read\_excel("C:/Internships Taken/KPMG/Data Quality Assessment/task 1/KPMG\_VI\_New\_raw\_data\_update\_final.xlsx", sheet = "Transactions")

// count of missing values for each worksheet

sum(is.na(Transactions))

// # find location of missing values column wise from each worksheet

**which(is.na(Transactions$transaction\_id))**

**// at once find missing values in worksheet for all columns**

* sapply(Transactions, function(Transactions) which(is.na(Transactions)))
* or sapply(Transactions, function(x) sum(is.na(x)))

#### **Dealing With Missing Values**

Transactions$online\_order = ifelse(is.na(Transactions$online\_order),ave(Transactions$online\_order, FUN = function(x) mean(x, na.rm = 'TRUE')),Transactions$online\_order)

Same for other columns with empty values

#### **Dealing With Categorical Data**

Transactions$brand=as.factor(Transactions$brand)

levels(Transactions$brand)

unique(Transactions$product\_id)

Azzuming product id belongs to brand