**Date : 12th July, 2024**

**Assignment- 1 : Excel**

A bank's **telemarketing business problem** is given to you to understand the *types of customers which should be targeted* for future marketing campaigns. The dataset contains demographic and personal information of about 45000 people and their response to a telemarketing campaign, i.e. whether they responded positively or negatively to the campaign.

Some of the attributes are described here:

**Age:** This is the numerical age of the customer. It can be used to understand the target audience and tailor marketing campaigns accordingly. For example, younger customers might be more receptive to online banking solutions, while older customers might prefer traditional in-person banking services.

**Job:** This is the type of job the customer holds (e.g., "admin," "unemployed," "entrepreneur"). It can be used to assess a customer's financial stability and risk profile. Customers with stable jobs might be more likely to take on loans or invest in financial products.

**Marital Status:** This indicates whether the customer is married, single, divorced, etc. It can be a factor in a customer's financial needs and spending habits. Married couples, for instance, might be more likely to seek joint accounts or mortgages.

**Education:** This refers to the highest level of education attained by the customer (e.g., "primary," "secondary," "tertiary"). Education level can be correlated with income and financial literacy, potentially influencing a customer's receptiveness to complex financial products.

**Balance:** This represents the average yearly balance maintained by the customer in their account(s). It's an indicator of a customer's financial resources and potential for investment or lending opportunities.

**Housing Loan:** This is a binary variable indicating whether the customer has a housing loan with the bank. It can be used to understand a customer's debt level and potential financial obligations.

**Contact:** This specifies the communication channel used to contact the customer during a marketing campaign (e.g., "telephone," "cellular," "unknown"). Understanding the most effective channels can help optimize future campaigns.

**Day of Last Contact:** This indicates the day of the month when the customer was last contacted for the marketing campaign. It can be used to analyze if specific days are more effective for outreach.

**Duration of Last Contact:** This represents the length of the last marketing call with the customer, measured in seconds. It might be helpful in understanding customer engagement and potential areas for improvement in communication strategies.

**Campaign:** number of contacts performed during this campaign and for this client (numeric, includes last contact)

**pdays** : number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)

**previous** : number of contacts performed before this campaign and for this client

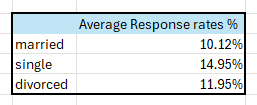
**poutcome** : outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

y: has the client subscribed a term deposit? ( 0 no, y yes)

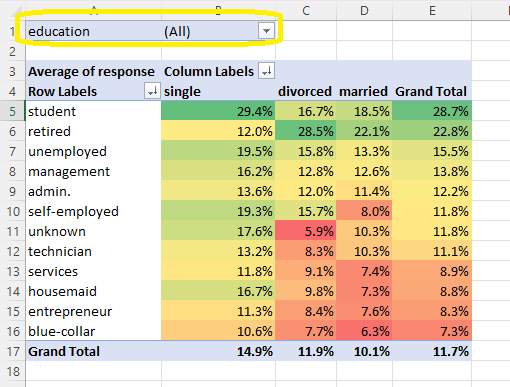
Source: <https://archive.ics.uci.edu/dataset/222/bank+marketing>

As a part of *acquisition analytics* where companies use past data to predict which customers are likely to buy certain products, let us try and answer the following questions:

1. What is the range of age. Calculate mean, median and mode of age.
2. Create a new column called age group that has values 0,1,2,3,4 etc. A person with an age falling between 0-10 would be in the age group 0, 10-20 in 1 , 20-30 in 2 etc. Explore round(), rounddown(), roundup()
3. Replace responses to 0 and 1s instead of no and yes. The use of binary values (especially 0 & 1)  makes the computation of average easier.
4. Calculate the average response rates for the different marital status categories.



1. Create a new attribute (marital-education) that takes the combination of marital status and education . How many such unique combinations we have? Which combination gives us the highest response rate? Also the values in column should be saved in a way that it takes first 3 characters from marital status and first three characters from education and combines both with “-“ as delimiter. (married-tertiary should be MAR-TER)
2. Create a new column called “eligible” to check for the eligibility for loan application. A person is eligible for a loan only if his age is between 20 and 60, including boundary values. What % of people are eligible for loan
3. Represent the age group and % response rates as a bar chart
4. Create the following pivot table and conditionally format. The table shows response rates for the different job roles and marital status.



1. Does the response rates change with the education?
2. Based on your analysis of the data, can you suggest which group of customers are responding positively to the marketing campaign?