BANK MARKETING (CAMPAIGN)

WEEK 8

Group Name	Tentacool
Name	Murat Kıran
Email	murattkiran@gmail.com
Country	Türkiye
Specialization	Data Science

1. Problem description:

ABC Bank is preparing to launch a term deposit product and seeks to develop a predictive model to determine if customers will purchase the product based on their past interactions with the bank or other financial institutions.

2. Data understanding:

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Input Variables:

1. **age (numeric):** Represents the age of the bank client.
2. **Job (categorical):** Describes the type of job the client has. Categories include 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', and 'unknown'.
3. **marital (categorical):** Indicates the marital status of the client. Categories include 'divorced', 'married', 'single', and 'unknown'.
4. **education (categorical):** Represents the educational background of the client. Categories include 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', and 'unknown'.
5. **default (categorical):** Specifies whether the client has credit in default. Categories are 'no', 'yes', and 'unknown'.
6. **housing (categorical):** Indicates whether the client has a housing loan. Categories include 'no', 'yes', and 'unknown'.

7. **loan (categorical):** Specifies whether the client has a personal loan. Categories are 'no', 'yes', and 'unknown'.

8. **contact (categorical):** Describes the communication type for the last contact. Categories are 'cellular' and 'telephone'.

9. **month (categorical):** Represents the last contact month of the year. Months are denoted as 'jan', 'feb', 'mar', ..., 'nov', 'dec'.

10. **day_of_week (categorical):** Indicates the last contact day of the week. Days are 'mon', 'tue', 'wed', 'thu', and 'fri'.

11. **duration (numeric):** Represents the duration of the last contact in seconds. Note: This attribute highly affects the output target.

12. **campaign (numeric):** Indicates the number of contacts performed during this campaign for this client, including the last contact.

13. **pdays (numeric):** Represents the number of days that passed since the client was last contacted from a previous campaign. '999' means the client was not previously contacted.

14. **previous (numeric):** Represents the number of contacts performed before this campaign for this client.

15. **poutcome (categorical):** Represents the outcome of the previous marketing campaign. Categories are 'failure', 'nonexistent', and 'success'.

Social and Economic Context Attributes:

16. **emp.var.rate (numeric):** Represents the employment variation rate, a quarterly indicator.

17. **cons.price.idx (numeric):** Represents the consumer price index, a monthly indicator.

18. **cons.conf.idx (numeric):** Represents the consumer confidence index, a monthly indicator.

19. **euribor3m (numeric):** Represents the Euribor 3-month rate, a daily indicator.

20. **nr.employed (numeric):** Represents the number of employees, a quarterly indicator.

Output Variable:

21. **y (binary):** The desired target variable indicating whether the client has subscribed to a term deposit. Categories are 'yes' and 'no'.

3. Data types:

```
###################### Shape ######################
Rows: 41188
Columns: 21

###################### Types ######################
age                int64
job               object
marital           object
education         object
default           object
housing           object
loan              object
contact           object
month             object
day_of_week       object
duration           int64
campaign           int64
pdays              int64
previous           int64
poutcome          object
emp.var.rate     float64
cons.price.idx   float64
cons.conf.idx    float64
euribor3m        float64
nr.employed      float64
y                 object
dtype: object
```

4. Problems in the data and approaches:

There are unknown values for many variables in the dataset. Variables with unknown values include 'education,' 'job,' 'housing,' 'loan,' 'default,' and 'marital'. Logical assignments will be made for variables where it is necessary to replace 'unknown' values. Additionally, outlier values exist in some variables in the dataset. Approaches to address these outliers will be implemented based on the utilized models.