Download the file bank-full.csv provided in the folder

Data Description:

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.

Domain: Banking

Context:

Leveraging customer information is paramount for most businesses. In the case of a bank, attributes of customers like the ones mentioned below can be crucial in strategizing a marketing campaign when launching a new product.

Learning Outcomes:

- Exploratory Data Analysis
- Preparing the data to train a model
- Training and making predictions using an Ensemble Model
- Comparing model performances

**Objective:**

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

**Steps, deliverables and tasks:**

1. Import the necessary libraries
2. Read the data as a data frame
3. Perform basic EDA which should include the following and print out your insights at every step.
    a. Shape of the data - 2 points
    b. Data type of each attribute - 2 points
    c. Checking the presence of missing values 2 points
    d. 5 point summary of numerical attributes (Statistical Summary) 2 points
    e. Checking the presence of outliers etc 2 points
    f. Data Visualization : Use your learning experience to select the variables to plot with the appropriate charts - 5 points

4. Prepare the data to train a model – check if data types are appropriate, get rid of the missing values etc - 5 points
5. Create the Logistic Regression Model. – 5 points
6. Predict the accuracy of the model - 5 points
7. Verify using some test cases -5 points

References:

- Data analytics use cases in Banking
- Machine Learning for Financial Marketing

Output variable -> y

y -> Whether the client has subscribed a term deposit or not

Binomial ("yes" or "no")

Attribute information For bank dataset

Input variables:

# bank client data:

1 - age (numeric)

2 - job : type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur","student", "blue-collar","self-employed","retired","technician","services")

3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means divorced or widowed)

4 - education (categorical: "unknown","secondary","primary","tertiary")

5 - default: has credit in default? (binary: "yes","no")

6 - balance: average yearly balance, in euros (numeric)

7 - housing: has housing loan? (binary: "yes","no")

8 - loan: has personal loan? (binary: "yes","no")

# related with the last contact of the current campaign:

9 - contact: contact communication type (categorical: "unknown","telephone","cellular")

10 - day: last contact day of the month (numeric)

11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

12 - duration: last contact duration, in seconds (numeric)

# other attributes:

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

14 - 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)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

Output variable (desired target):

17 - y - has the client subscribed a term deposit? (binary: "yes","no")

Submission :

Run each cell of your code.

Download the python file as html and submit the html file.

Submit the python file as well just in case if html is not downloaded correctly.