MINI PROJECT- Python for Data Science

PROBLEM STATEMENT:

The problem is that the Bank Marketing campaigns of a Portuguese banking institution need to identify the factors that cause the customers to tend to take the subscription, as well as Bank Marketing campaigns of a Portuguese banking institution need to identify the reasons behind the customer which make them not take the subscription.

Aim:

You aim to Determinate/Analysis factors for the subscription and nonsubscription. Using the ITP and NPV techniques find the below questions.

DATASET DESCRIPTION:

Bank Marketing: The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact with the same client was required to assess if the product (bank term deposit) would be (or not) subscribed.

DATA DICTIONARY:

1 - age (numeric)

2 - job : type of job (categorical:
"admin.","unknown","unemployed","management","housemaid","entrepreneur",
"student","bluecollar","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 a housing loan? (binary: "yes","no")

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

- related to 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 the last contact

14 - P-days: 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 to a term deposit? (binary: "yes","no")

QUESTIONS TO SOLVE:

Check Point 1:

1. Import data set

2. Make the data proper to make use of data for analysis

   A. Identify the Features data types before entering into the analysis

   B. .Convert the datatypes which are wrongly identified according to the business(domain). Kindly use the User Defined function and loop to convert the data types once.

   C. Find and Remove missing if any. Use visualization to find the missing values or Use general method to find the missing values.

   D. Find duplicates (if necessary)

Check Point 2:

3. Find the average balance of the customer who belongs to the subscribed customer and non-subscribed customer and also use a related plot to show them in visualization.

4. Use a pie plot to find the distribution(frequency) of the education. Make sure to add labels and show the percentage of each education distribution.

5. Create a function that should be able to create a new feature(Variable) called season using the month column.

6. Use the count plot with a variable that you created in the above question and also the Y variable to find the class distribution.

7. Use the Pdays feature and find does it cause any effect on the subscription of the term using the bar plot.

8. Replace the -1 as nan values for the P-days store.

9. Once you are done with question number 8, do the same analysis as question number 7. And observe the difference between question number 7 and question number 9.

10. Does the customer take the term subscription who has less than 0 balance? Hint: Use any kind of plot which would you the related information to this question.

11. Use Pivot table to find the maximum balance for each type of job.

12. Use the Age, balance, and Y column to plot the scatter plot and find what kind of relationship Age and balance had, and See the points which belong 0 and 1 class and how they are distributed.

13. Use the violin plot and also the box plot to find the distribution of the balance for each class of the Y column. And try to tell why we have a Violin plot and Box plot both rather than one.

14. Use a pie plot to know the Proportion(distribution) of the defaulters and non-defaulters.
Note: Try to explore more parameters that are there in the pie-plot method.

15. Use Box plot and strip plot to know the distribution of the Pdays with respect to Y classes and differentiate both plots.