

# **FINANCIAL EDA REPORT**

NAME: NARAYAN ARAVIND

ROLL: 2404042

BRANCH: ETC

The purpose of this project is to master the exploratory data analysis (EDA) in banking with Pandas framework.

Goals of the Project:

1. Explore a banking dataset with Pandas framework.
2. Build tables.
3. Visualize the dataset with various plot types.

## **Outline**

- Libraries import
- Dataset exploration
- Tables
- Visualization in Pandas

In this project, we will try to give answers to a set of questions that may be relevant when analyzing banking data:

1. What is the share of clients attracted in our source data?
2. What are the mean values of numerical features among the attracted clients?
3. What is the average call duration for the attracted clients?
4. What is the average age among the attracted and unmarried clients?
5. What is the average age and call duration for different types of client employment?

In addition, we will make a visual analysis in order to plan marketing banking campaigns more effectively.

## QUESTIONS:

1. List of 10 clients with the largest number of contacts.

```
#LIST OF CLIENTS WITH MOST CONTACTS
df.sort_values(by = "campaign", ascending = False).head(10)
```

[27] ✓ 0.0s Python

	age	job	marital	education	default	housing	loan	contact	month	day_of_week	...	campaign	pdays	previ
2552	31	services	single	high.school	no	no	no	cellular	jul	thu	...	35	999	
3241	39	services	married	high.school	no	yes	no	cellular	jul	thu	...	29	999	
3564	25	admin.	single	basic.9y	no	no	no	cellular	jul	thu	...	29	999	
56	29	admin.	single	university.degree	no	yes	no	telephone	jun	fri	...	27	999	
2485	41	technician	married	high.school	no	yes	no	telephone	jun	fri	...	24	999	
2988	45	services	married	professional.course	no	yes	no	cellular	jul	mon	...	23	999	
2202	29	technician	married	university.degree	no	no	no	cellular	jul	thu	...	23	999	
3569	31	admin.	single	high.school	no	no	no	telephone	may	thu	...	22	999	
713	43	admin.	married	high.school	no	yes	no	cellular	jul	mon	...	22	999	
886	56	technician	married	university.degree	unknown	no	no	cellular	jul	mon	...	19	999	

10 rows × 21 columns

2. Determine the median age and the number of contacts for different levels of client education.

```
#MEDIAN AGE OF CONTACTS AT DIFFERENT EDUCATION LEVEL
df.pivot_table(
    ["age", "campaign"],
    ["education"],
    aggfunc = ["mean", "count"],
)
```

[28] ✓ 0.0s Python

	mean		count	
	age	campaign	age	campaign
education				
basic.4y	47.657343	2.421911	429	429
basic.6y	40.144737	2.649123	228	228
basic.9y	39.231707	2.348432	574	574
high.school	38.097720	2.630836	921	921
illiterate	42.000000	4.000000	1	1
professional.course	40.207477	2.512150	535	535
university.degree	39.017405	2.583070	1264	1264
unknown	42.826347	2.538922	167	167

3. Output box plot to analyze the client age distribution by their education level.

