

# Banking Marketing Analysis



Course Name : ALY6000: Data Analysis  
Shamim Sherafati

# Introduction

The dataset contains banking marketing campaign data and we can use it to optimize marketing campaigns to attract more customers to term deposit subscription.

## Business Goals:

1. Analysis of the results of the marketing campaign for each customer and clarification of factors which affect the campaign results. This helps to visualize and understand the marketing campaigns.
2. Finding out customer segments, using data for customers, who subscribed to term deposit. This helps to identify the profile of a customer, who is more likely to acquire the product and develop more targeted marketing campaigns.

# Approach to the problem

## Import data from dataset

Look at the number of rows, look at the missing values, look at dataset columns and their values respective to the campaign outcome.

## Clean the data

Remove irrelevant columns, deal with missing and incorrect values, turn categorical columns into dummy variables.

## Perform high-level analysis

Use Exploratory Data Analysis techniques to obtain the marketing campaign outcomes and to find out answers to various business questions.

# Understanding the Data

## 1. Bank's Client Data:

1. **age**
2. **job** : type of job (categorical: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "selfemployed", "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? (default means the person fail to pay the Minimum Amount Due on the credit card for a few consecutive month)
6. **balance**: average yearly balance, in euros
7. **housing**: has housing loan?
8. **loan**: has personal loan?

# Understanding the Data

## 2. Last contact of the current campaign Data:

- 9. **contact:** contact communication type (categorical: "unknown","telephone","cellular")
- 10. **day:** last contact day of the month
- 11. **month:** last contact month of year (categorical: "jan", ..., "dec")
- 12. **duration:** last contact duration, in seconds

## 3. Other Attributes:

- 13. **campaign:** number of contacts performed during this campaign and for this client
- 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
- 16. **poutcome:** outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

## 4. Output Variable:

- 17. **Deposit:** y - has the client subscribed a term deposit?

# Descriptive Analysis

	age	job	marital	education	default	balance	housing	loan	contact
0	59	admin.	married	secondary	no	2343	yes	no	unknown
1	56	admin.	married	secondary	no	45	no	no	unknown
2	41	technician	married	secondary	no	1270	yes	no	unknown
3	55	services	married	secondary	no	2476	yes	no	unknown
4	54	admin.	married	tertiary	no	184	no	no	unknown

	day	month	duration	campaign	pdays	previous	poutcome	deposit	
0	5	may	1042	1	-1	0	unknown	yes	Table:1
1	5	may	1467	1	-1	0	unknown	yes	
2	5	may	1389	1	-1	0	unknown	yes	
3	5	may	579	1	-1	0	unknown	yes	
4	5	may	673	2	-1	0	unknown	yes	

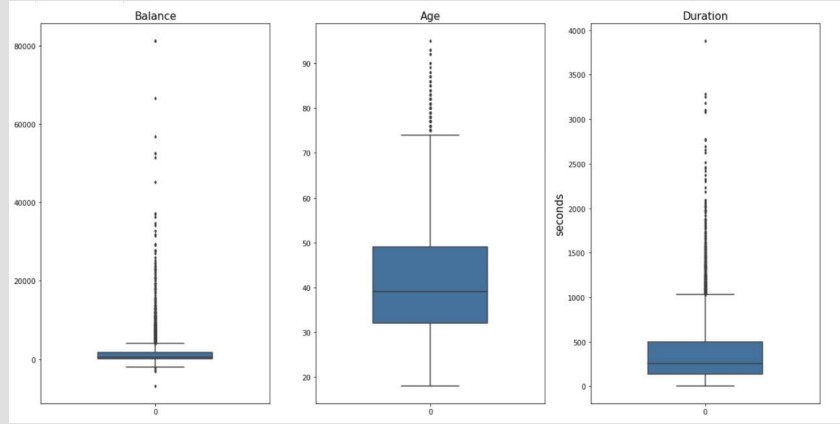
	age	balance	day	duration	campaign
count	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000
mean	41.231948	1528.538524	15.658036	371.993818	2.508421
std	11.913369	3225.413326	8.420740	347.128386	2.722077
min	18.000000	-6847.000000	1.000000	2.000000	1.000000
25%	32.000000	122.000000	8.000000	138.000000	1.000000
50%	39.000000	550.000000	15.000000	255.000000	2.000000
75%	49.000000	1708.000000	22.000000	496.000000	3.000000
max	95.000000	81204.000000	31.000000	3881.000000	63.000000

	pdays	previous
count	11162.000000	11162.000000
mean	51.330407	0.832557
std	108.758282	2.292007
min	-1.000000	0.000000
25%	-1.000000	0.000000
50%	-1.000000	0.000000
75%	20.750000	1.000000
max	854.000000	58.000000

Table:2

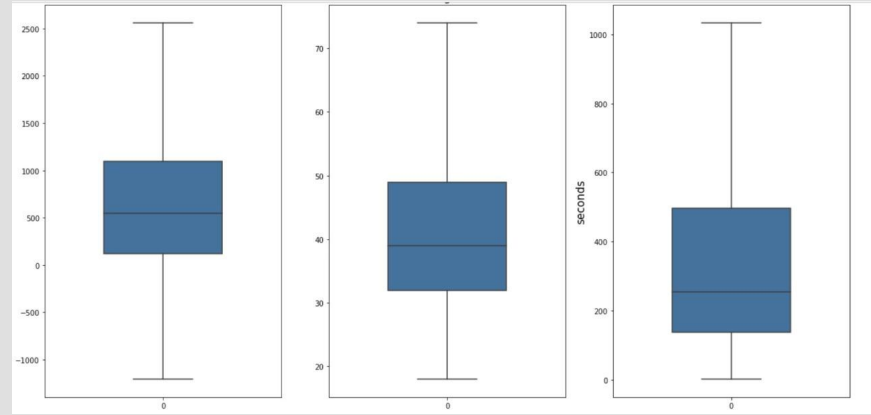
- The dataset includes 11162 rows and 17 columns (attributes).
- Table 1 is the head() of the dataset showing the attributes.
- After cleaning the data, the statistics of the data are as shown in Table 2. We can see that the table shows the mean, standard deviation, minimum and maximum values, and the 25%/50%/75% show their respective percentile values.
- In simple terms, these percentile values mean how many of the values are less than the given percentile.
- From this data, we can see that the average age of people in this dataset is 41 years. The minimum age of a person who participated in this campaign is 18 years old and the maximum age is 95 years old.

# Box Plot Analysis



**Findings:** Box plots are used to show overall patterns of response for a group. They provide a useful way to visualise the range and other characteristics of responses for a large group.

- **Bank Balance:** The majority of the bank balances are close to each other while some are outliers as people have negative bank balances and some have savings present in the bank ranging from -190 to 8000 dollars.
- **Age :** The age of the people contacted in the survey ranges from approx. 16-100 where senior citizens above the age of 80 can be considered outliers. The median age is 40, with 25% being under 32 years old and 75% being under 48 years old.



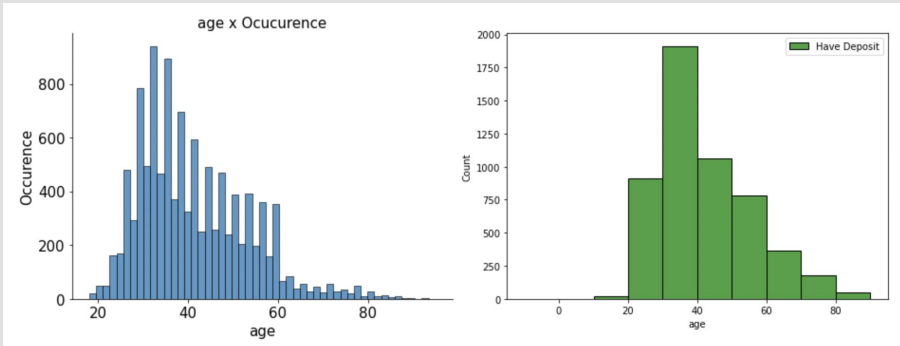
**Duration :** As the last contact information is in seconds. Max duration between contacting a person is an hour and minimum time is few seconds

**Chart Type:** We have used **Boxplot** to show the range of balance, age and last contact duration.

**Attributes Behaviour**

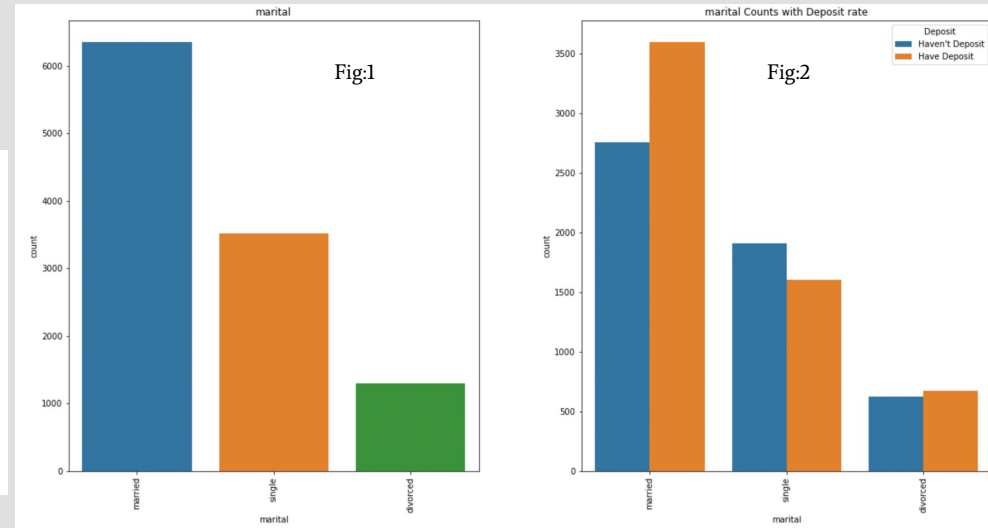


# What is the majority age group that made deposit ?



**Findings:** The majority of customers were in their **30s and 40s**. The distribution of age is fairly normal with a smaller standard deviation. The subscription rate is also more in the range of 30 to 40 years old.

**Chart Type:** We used a **histogram** for plotting the age distribution. The histogram clearly compares subscription to non-subscription for each age group.

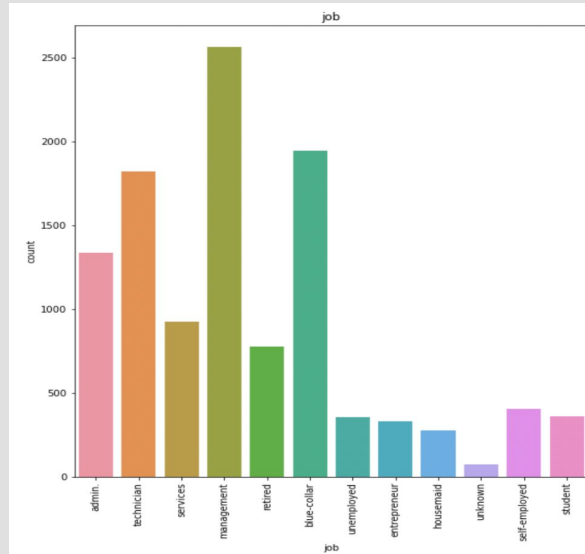


**Findings:** From the Bar Graph(Fig. 1), we can see term deposits based on their **marital status**. We understand that the **married** clients have made **more deposits** than other groups.

The Fig 2 Bar Graph shows that among individual groups, **married and single people have a higher rate** of term deposits compared to **Divorced People**.

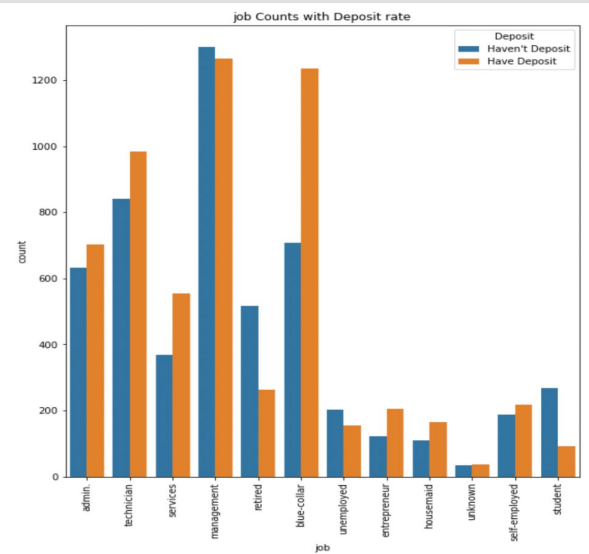
**Chart Type:** We have used **Bar Graphs** for showing that how many people made the term deposit based on their marital status.. It's very clear through a bar graph to compare the categorical data.

# What is the trend among different professions that made deposit ?



**Findings:** From the above (Job vs Count Bar Graph), we can see that the clients who are “**Management**”, “**Blue-color**” and “**Technician**” have made more term deposits compared to any other **profession**.

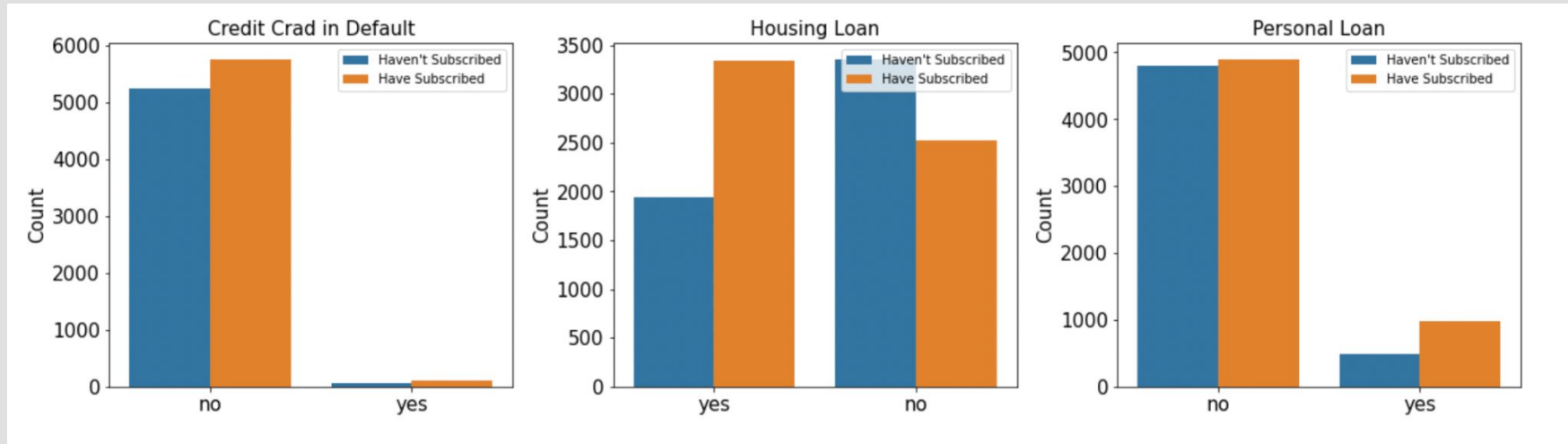
**Chart Type:** We used a **Bar Graph** for showing that how many people made the term deposit based on their profession. It's very clear through a bar graph to compare the categorical data.



**Findings:** From the above (Multiple Bar Graph), we can see that for each profession, how many people have made the term deposit or not. We can see that the **highest acceptance rate** is from “**Blue-color**” profession which is **about 50%**.

**Chart Type:** We used a **Multiple Bar Graph** to show the deposit rate of each profession. It's very easy to showcase whether a person made the term deposit or not for each profession in a single graph using a multiple bar plot.

# What can we say about people who take different Loans ?



**Findings:** We also discovered information about loan segments. As a result, those with **Credit Card defaults** (users who failed to pay) are very few, and their subscriber - to-term-deposit ratio is nearly the same. But for those who **don't have CC defaults**, more people **make term deposits**.

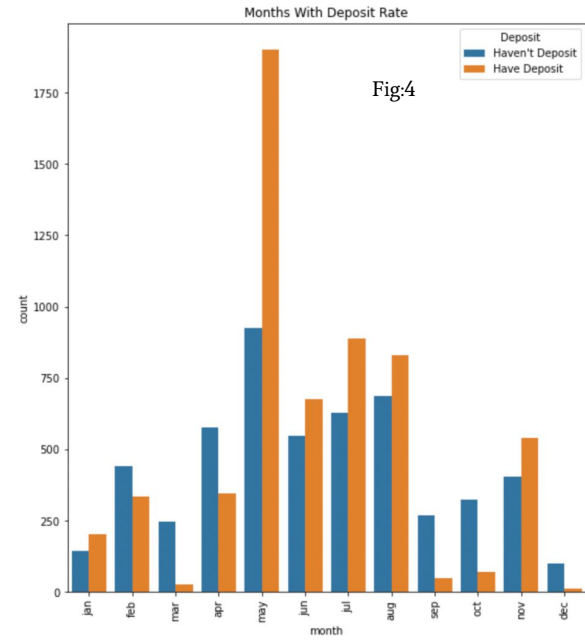
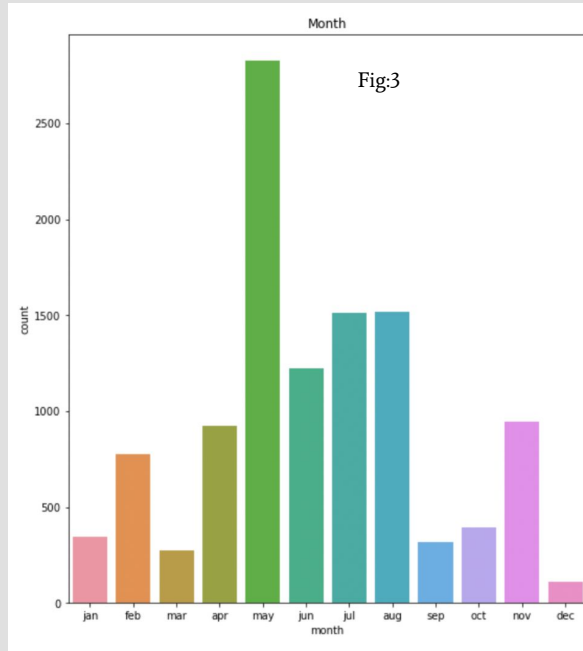
50% of the total people in this campaign have **housing loans**. Of those **having housing loans**, almost **60%** of them have **subscribed** to a term deposit. The majority of those once **not having housing loans didn't subscribe** to the term deposits.

Almost **90%** of users **didn't have Personal Loans**, and the ratio of subscribers and non-subscribers to term deposits is almost the same. Of those **having a personal loan**, more people **subscribed** to term deposits.

**So we can conclude that people who are not on loans show more interest in term deposits.**

**Chart Type:** We have used **Bar Graphs** to show how many people made the term deposit based on their Loan Status.

# What analysis can be made based on months ?



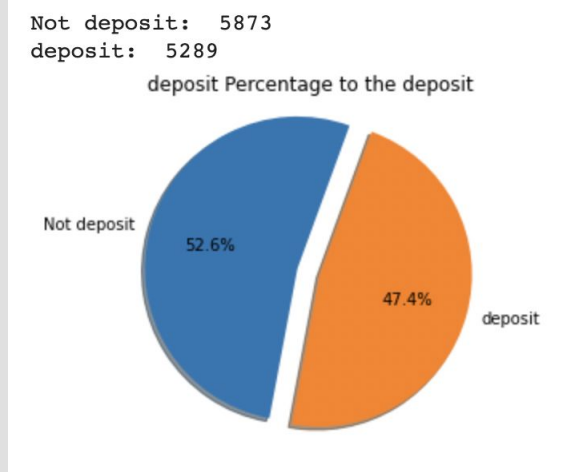
**Findings: (Fig 3).** The bank contacted most clients between May and August. The highest contact rate is around **30%**, which happened in **May**, while the contact rate is very low in March, September, October, and December.

**(Fig 4)** The highest subscription rate occurred in **May**, which is over **50%**. This was followed by subscription rates in June, July, and August.

**From this, we come to the conclusion that if the contact rate is higher, a greater number of subscriptions will take place.**

For **example, in the month** of May, as the contact rate was higher, the subscription was also higher.

# Overall Result of the campaign



**Findings:** According to the Pie chart, nearly 50% of all people contacted by banks during this campaign accepted term deposits.

**Reasons for this results:** This success was achieved because of targeting the right age group, the right profession, and making the most calls in a particular set of months (May-August). We also saw that people who are not taking loans tend to accept the term deposits.

**Conclusion:** We can improve the results by focusing on the following points:

1. Target the right age group (in their **30's and 40's**) more.
2. We can see that people having the **marital status "Single"** and **"Married"** invest more and have more focus on them compared to Divorced.
3. People having the **professions of "Management"** and **"Blue-color"** invest more in deposits so we can target such professions more.
4. People who **do not have loans or credit card defaults** tend to take in term deposits.
5. Term deposit subscribers tend to **have higher balances and age values**.

## Improvisations in the dataset:

1. If we even include the **salaries of the people**, it can help to determine their ability to take the policy.
2. We can include the **timestamp of the call** as well to determine at what time of the day people are interested in talking and buying the term deposit.
3. We can also check if the person has **previous investments** to know about his interests in the deposits.

# References

1. Analytics University. (2016, February 18). Boxplots in R | Creating Box and Whisker plots in R [Video]. YouTube  
<https://www.youtube.com/watch?v=KQfym04mIDY>
2. *Bank Marketing Campaign || Opening a Term Deposit*. (2019, March 16). Kaggle.  
<https://www.kaggle.com/code/janiobachmann/bank-marketing-campaign-opening-a-term-deposit>
3. Bluman, A. (2017). *Elementary Statistics: A Step By Step Approach* (10th ed.). McGraw Hill.
4. Chad C. Williams. (2021, February 9). *Get R Done | R Stats Tutorials: Professional Grouped Bar Plot (w/ ggplot)* [Video]. YouTube. <https://www.youtube.com/watch?v=-bPjGgD5bZQ>
5. Kabacoff, R. (2015). *R in Action: Data Analysis and Graphics with R* (2nd ed.). Manning.
6. Loo, M. van der, Jonge, E. de, van der Loo, M., & de Jonge, E. (2018). *Statistical Data Cleaning with Applications in R*. Wiley.