

# **BANK TELEMARKETING CAMPAIGN**

A Project to fulfill the Graded Assignment of Python

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a modern, abstract design. A large red speech bubble is positioned on the left side of the slide, containing the text 'Problem Statement'.

## Problem Statement

To enhance the revenue of Bank by conducting a cost-efficient telemarketing campaign for term deposits among existing customers.

The background of the slide features a series of thin, curved lines in light gray and white, creating a sense of motion and depth. A large, solid red shape, resembling a speech bubble or a stylized 'L' bracket, is positioned on the left side of the slide. The word 'Objective' is written in white, sans-serif font within the red shape.

## Objective

The objective of this project is to conduct an end-to-end Exploratory Data Analysis (EDA) on the campaign dataset, identifying patterns and providing insights to improve the positive response rate.

The analysis involves examining customer demographics, temporal trends, and other factors influencing the success of the campaign, ultimately offering recommendations for targeted improvements in the bank's marketing strategy.



# Software/Tools

- Jupyter Notebook (Python)
- Python
- Numpy
- Matplotlib
- Seaborn
- ScipyStats

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a modern, abstract design.

## Data Description

The given dataset includes the age, salary, balance, marital status, job, education, loan type, loan status, contact period, previous contact details & outcomes and recent responses of the customers of the bank.

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large, solid red speech bubble is centered on the page, pointing downwards.

# Details of Data Analysis

# 1. Understanding the Dataset

In this part, the data has been loaded into the Jupyter file and inspection has been done to clean the data, missing values and outliers has been handled using suitable codes & libraries like missingno, seaborn etc. The titles of the columns (or also called as headers) has also been changed using `iloc` and `reset`. Datatype of Duration column has been modified as a new column `TIME`.

## 2. Descriptive Statistics

Summary statistics have been derived using describe on the relevant columns like age, salary, balance, campaign etc. The distribution of the target variable, indicating responses to the term deposit campaign has been examined which represents that out of the targeted customers of the campaign, the percentage of people responded to Yes.



### 3. Univariate Analysis

The distribution of individual key features like age, balance, call duration etc. has been examined using histogram, density plot and boxplot.

## 4. Bivariate Analysis

Independent variables such as age, salary and Time has been taken with respect to Targeted. At first, jobedu column has been split into two columns, Job and Education. After that the analysis of Salary, Age and Time has been done using boxplot and violin plots. The salary vs targeted plot showed that the people with less salary has responded yes but the difference between both the responses is small. For age and time, the response is almost same.

## 5. Categorical Variables Analysis

[38]:

	1	Job
Edu		
primary	6839	
secondary	23180	
tertiary	13286	
unknown	1856	

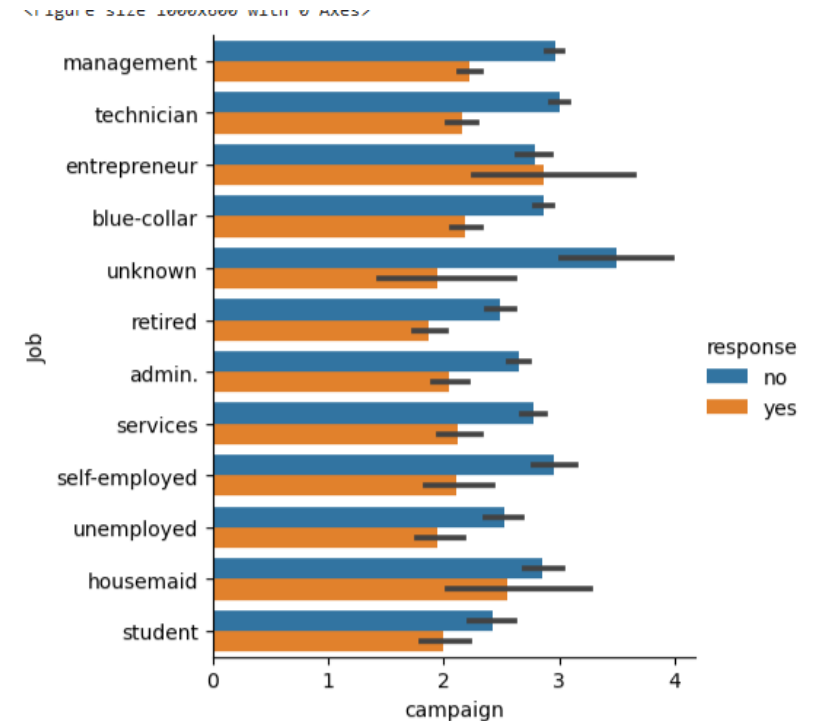
[39]:

	1	Job
marital		
divorced	5198	
married	27185	
single	12778	

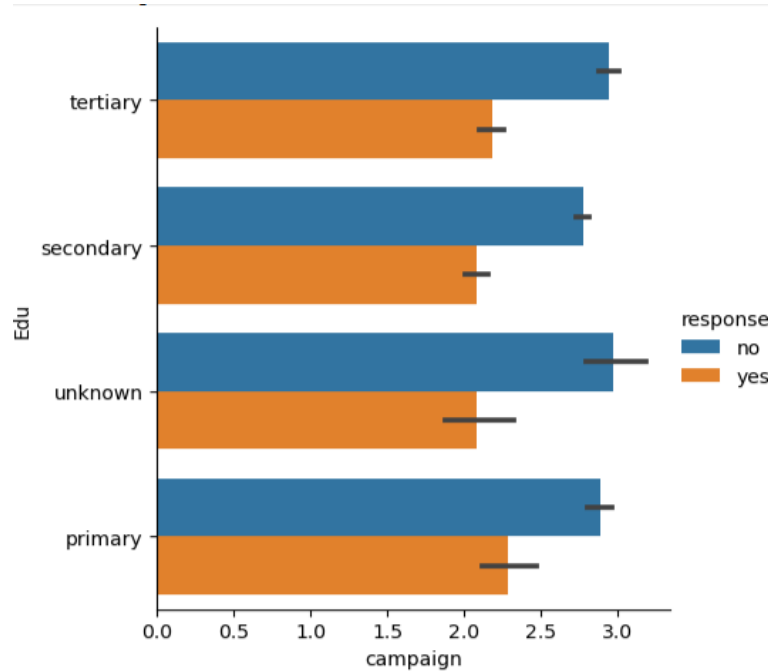
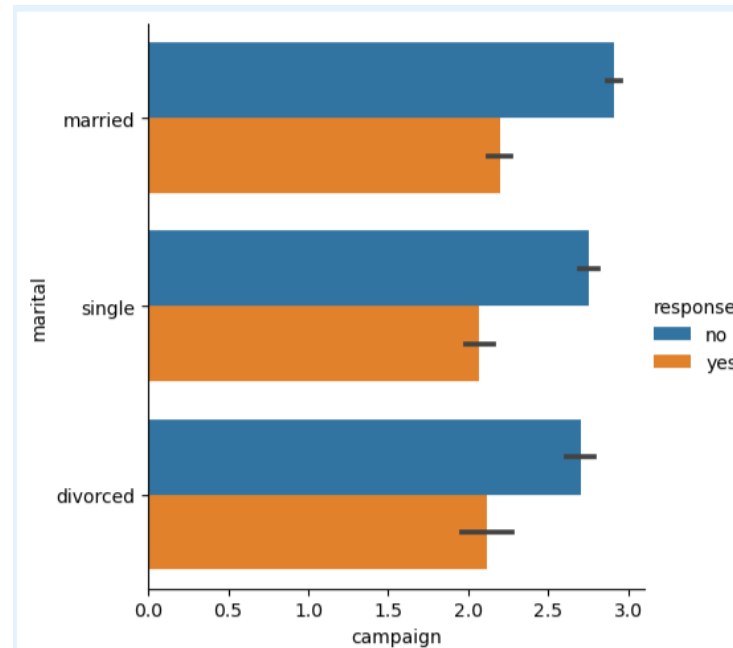
[40]:

	1	Edu
marital		
divorced	5198	
married	27185	
single	12778	

The categorical variable like job, marital and education has been analysed and picture 38, 39 and 40 clearly define the results. The visual representation of impact of categorical variables on campaign can be seen in the pictures below-

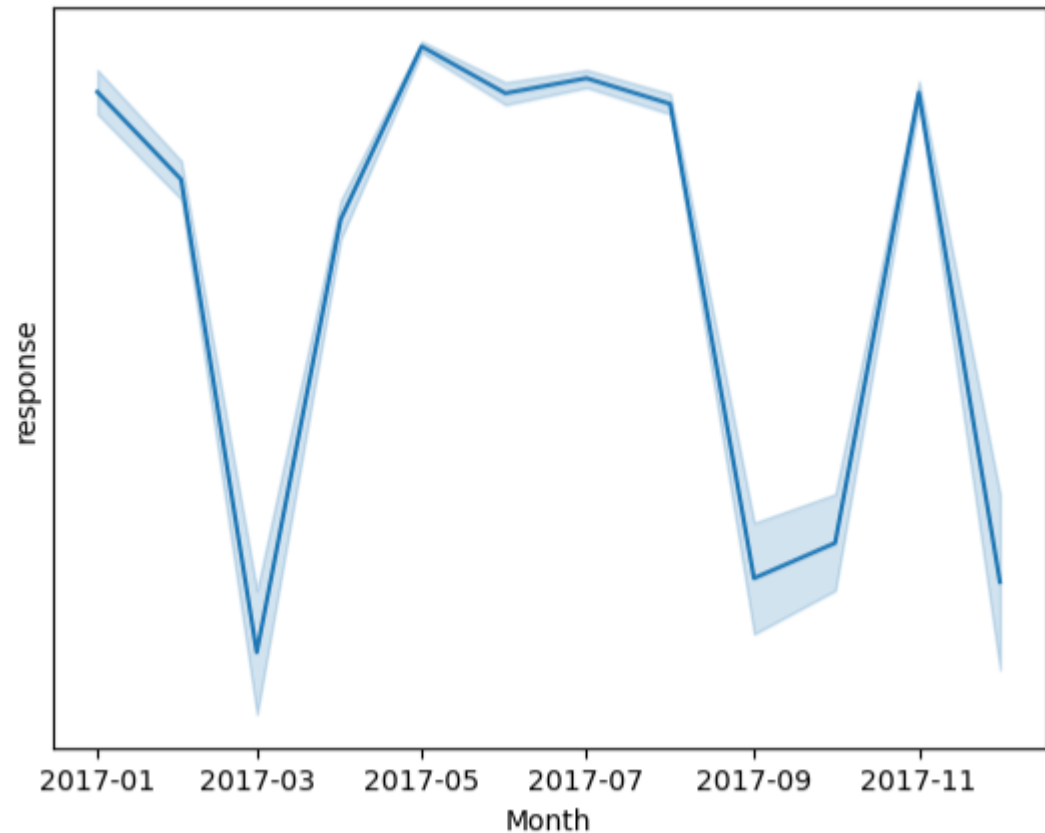


## 5. (continued)



## 6. Temporal Analysis

To investigate the Temporal patterns, a new column Month was made with date type of datetime. It represented that May has the highest number of positive responses followed June, July and November. The count of the responses can be seen in the ipynb file.



## 7. Feature Engineering

A new column for age groups has been made. Age and Salary have been divided into 4 individual groups. The responses for all can be seen in the picture. Focus is required on High and Very Pricely groups of all age groups.

18-30	Low	no	2033
		yes	432
	Medium	no	2008
		yes	363
	High	no	728
		yes	92
	v. Pricely	no	1092
		yes	246
31-45	Low	no	6302
		yes	524
	Medium	no	7378
		yes	851
	High	no	2091
		yes	197
	v. Pricely	no	5500
		yes	761
46-58	Low	no	3230
		yes	264
	Medium	no	3886
		yes	430
	High	no	921
		yes	73
	v. Pricely	no	2620
		yes	328
58+	Low	no	343
		yes	60
	Medium	no	1147
		yes	527
	High	no	39
		yes	7

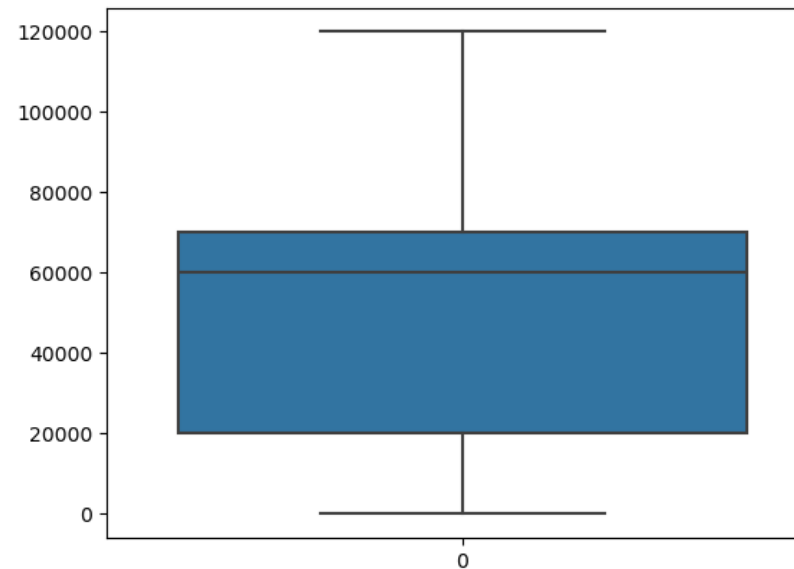
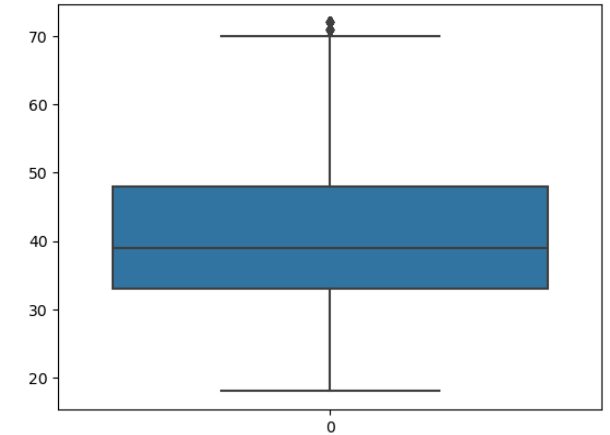
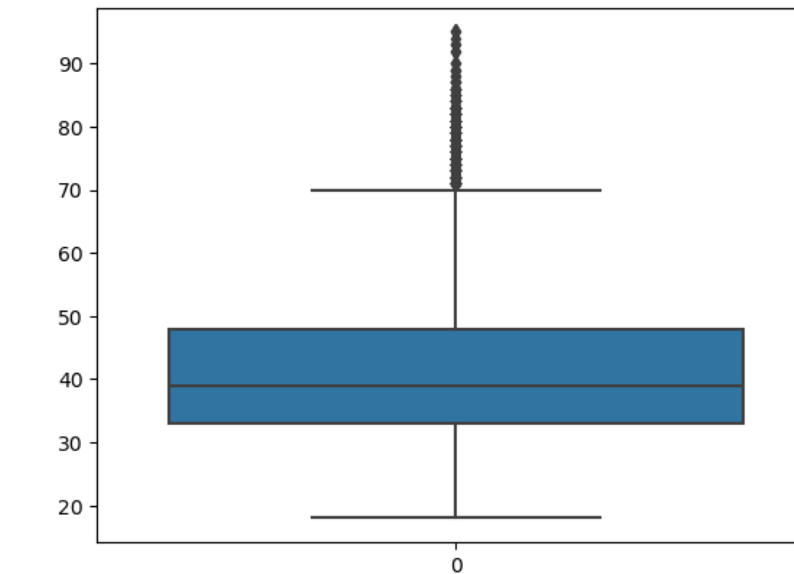
## 8. Correlation Analysis

The correlation between independent variables is as-

	1	age	salary	campaign	TIME
1					
age	1.000000	0.024513	0.004814	-0.004392	
salary	0.024513	1.000000	0.015010	-0.009988	
campaign	0.004814	0.015010	1.000000	-0.084379	
TIME	-0.004392	-0.009988	-0.084379	1.000000	

For evaluation of correlated features influencing target variable can be seen in ipynb.

## 9. Outlier Detection and Handling



```
# Removing Outliers of Salary  
bank_data4['salary'].skew()
```

0.13770754746825994

As the SKEWNESS of SALARY is 0.1377, which is lying between -0.5 to 0.5, shows that it is Approximately Symmetric.

1 Age

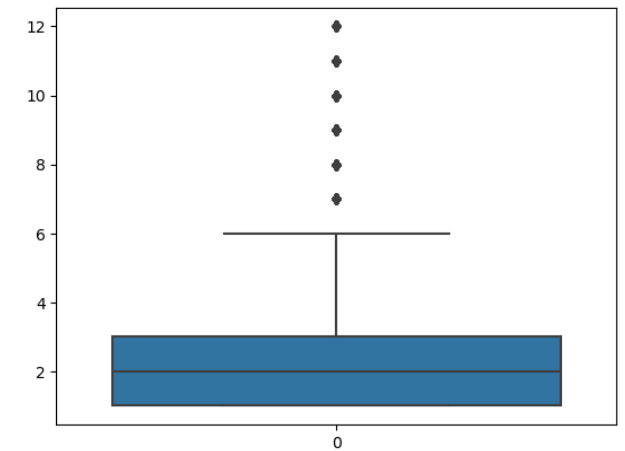
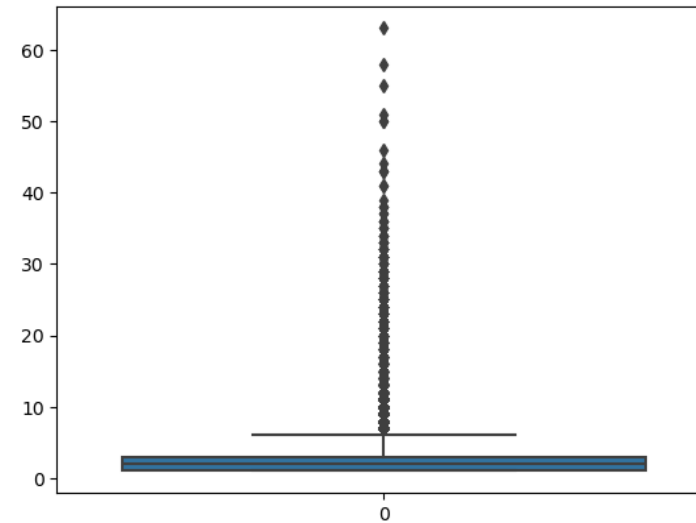
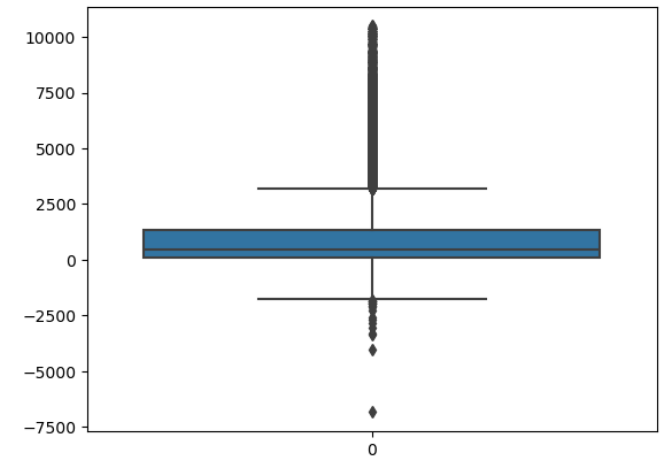
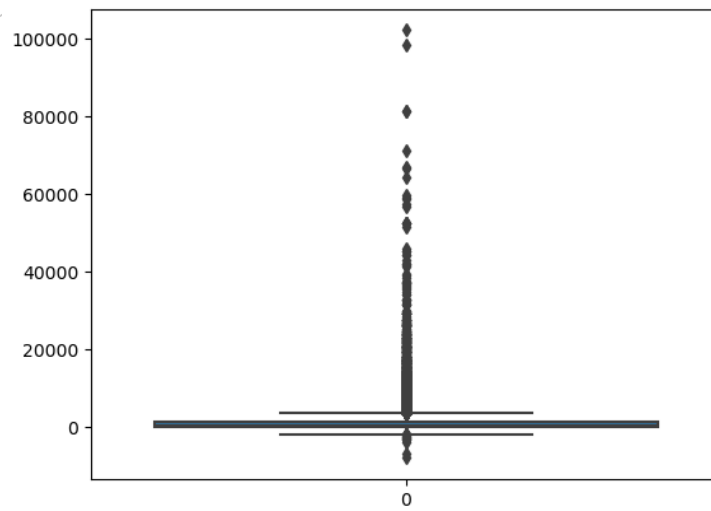
2 Salary

On Left- Outliers

On Right- Results after removing Outliers



## 9. (Continued...)



1 Balance

2 Campaign

On Left- Outliers

On Right- Results after removing Outliers



*Thank You*

By-  
Vattandeep Kaur  
ABADS, Batch - 12