

ASSIGNMENT GUIDELINES

- Make the changes in the PPT as you solve the parts
- This file contains the template for the **EDA part** of the project.
- Check the instructions added in the note section of every slide for clarity.
- Don't move around any image or text box
- If you require more/lesser elements, be careful when you copy/delete the existing ones.

ASSIGNMENT

Name: Manish Mangaldas Shah

Problem Statement

Bank of Corporate conducted a telemarketing campaign for one of its financial products, 'Term deposits', to build a long-term relationship with the existing customers. Your goal is to identify the target customers for the term deposits from the pool of the bank's existing customers. You should also capture the key driving factors (or driver variables) behind the successful conversion of a customer, i.e., the customer opening a term deposit account with the bank.

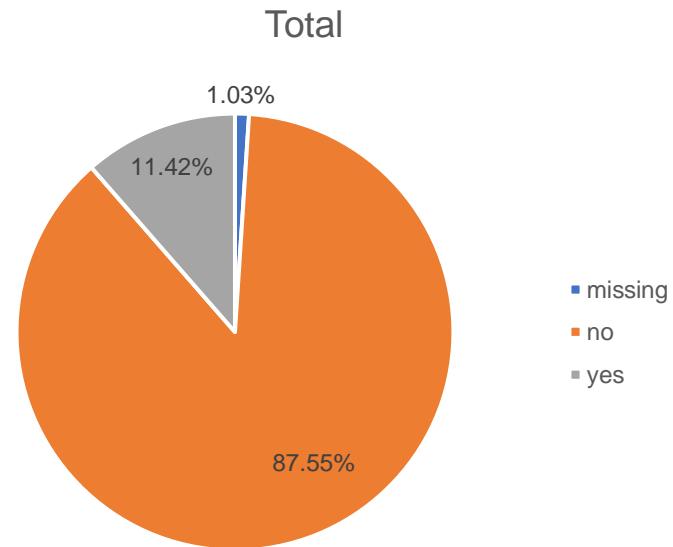
Assignment Objective

To identify the target customers and the driving factors behind a successful customer conversion for future marketing campaigns.

PART I : Univariate Analysis

Variable under consideration: The frequency tables and bar charts detail the distribution of the primary categorical columns, including the target variable, **Subscription**.

Row Labels	Count of Subscription
missing	1.03%
no	87.55%
yes	11.42%
Grand Total	100.00%



PART I : Univariate Analysis

The table below summarizes the central tendency, dispersion, and range for the key numerical columns.

Feature	Count	Mean	Std	Min	25% Quartile	Median (50%)	75% Quartile	Max
Age	45200	40.93	10.59	18	33	39	48	95
Cash_Balance	45200	68062	152306	-400950	3550	22300	71162.5	5.11E+06
pdays	45200	39.74	99.46	-1	-1	-1	-1	854
previous	45200	0.57	2.29	0	0	0	0	275
Call_Duration	45200	258.01	257.83	0	103	180	318	4918

Key Insights from Numerical Data:

Age: The mean age is approximately 41 years, with the majority of customers falling between 33 and 48 years old.

Cash_Balance: The data is highly right-skewed, as the mean (\$68,062) is much higher than the median (\$22,300). This indicates a few customers have very large balances. The minimum value is negative \$(-\\$400,950)\$, which is unusual for a balance and likely represents debt or an overdraft.

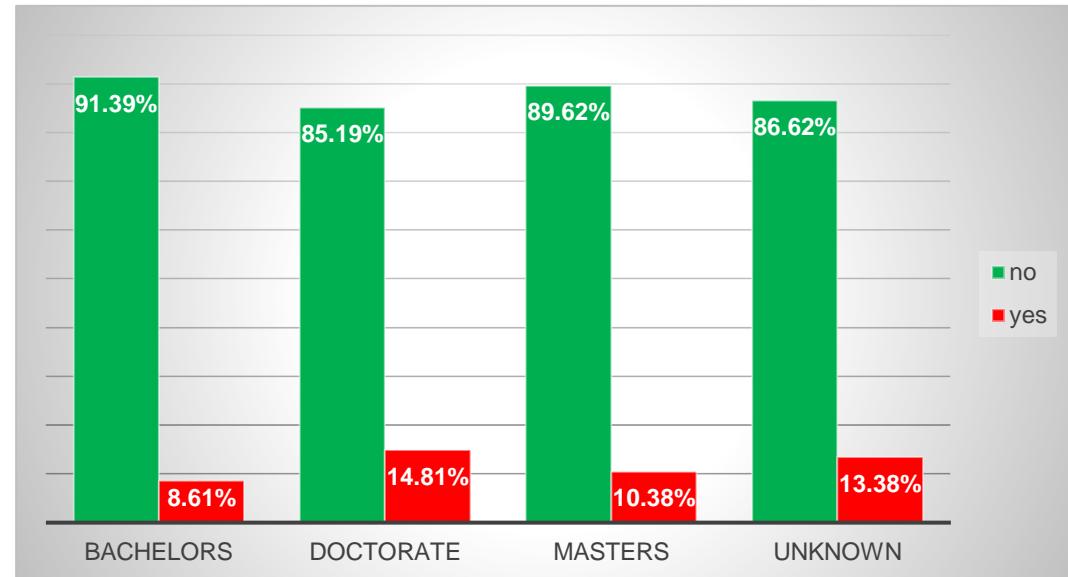
pdays and previous: The median and 75th percentile for pdays is \$-1\$, which is a code indicating that the customer was not previously contacted. This suggests that over 75% of customers were either not part of a previous campaign or were not tracked. The previous contacts column also shows that most customers have had zero previous contacts.

Call_Duration_Sec_: The average call duration is about 258 seconds (approx. 4.3 minutes), with most calls being short (median 180 seconds, or 3 minutes).

PART I : Univariate Analysis

Variable under consideration: Analysis on the basis of Education

Subscription %	Column Labels		
Row Labels	no	yes	Grand Total
Bachelors	91.39%	8.61%	100.00%
Doctorate	85.19%	14.81%	100.00%
Masters	89.62%	10.38%	100.00%
Unknown	86.62%	13.38%	100.00%
Grand Total	88.46%	11.54%	100.00%

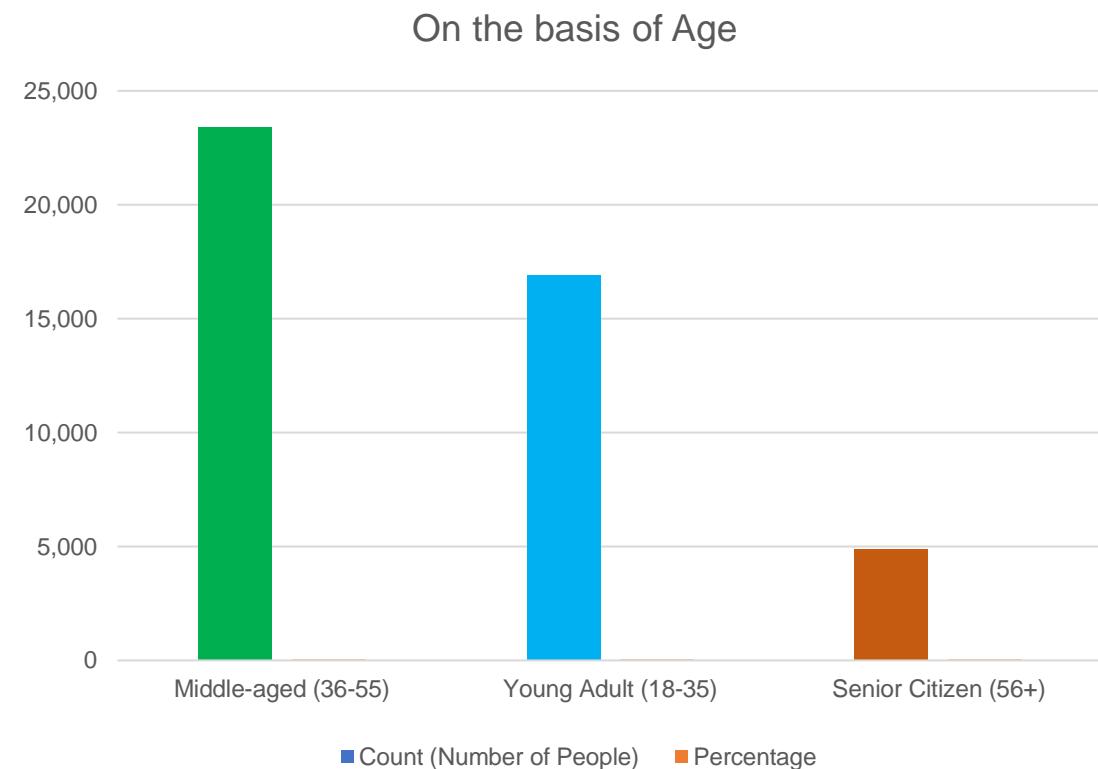


From the above data and charts it is clear that persons having Bachelors degree are less likely to subscribed

PART I : Univariate Analysis

Variable under consideration: On the basis of age factor

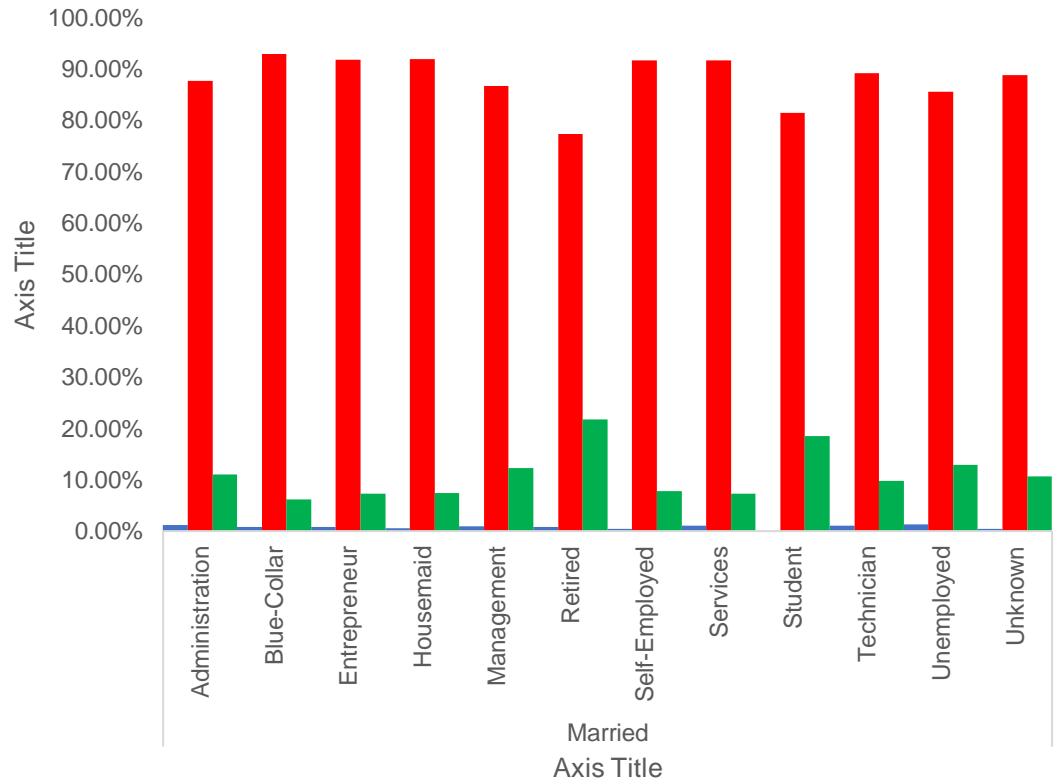
Age Group	Count (Number of People)	Percentage
Middle-aged (36-55)	23,434	51.85%
Young Adult (18-35)	16,894	37.38%
Senior Citizen (56+)	4,872	10.78%



PART II : Bivariate Analysis

Variables under consideration: On the basis of Job and Marital Status

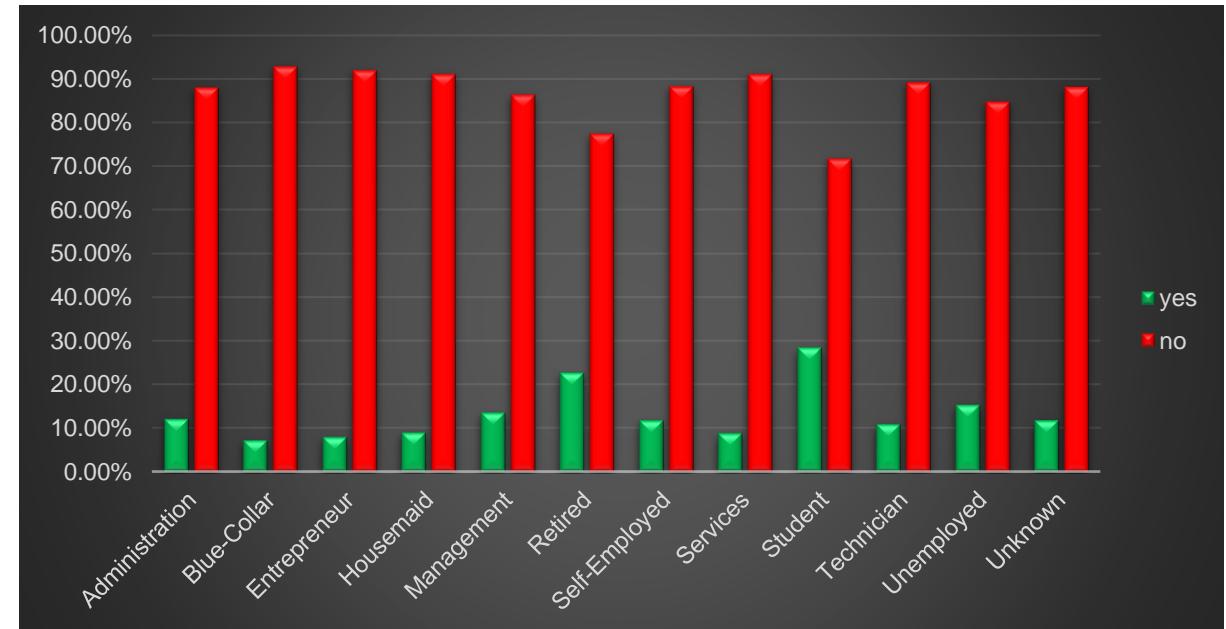
Subscription %	Column Labels			Grand Total
Row Labels	missing	no	yes	100.00%
Married	0.97%	89.16%	9.87%	100.00%
Administration	1.23%	87.72%	11.05%	100.00%
Blue-Collar	0.89%	92.96%	6.16%	100.00%
Entrepreneur	0.84%	91.86%	7.30%	100.00%
Housemaid	0.55%	92.01%	7.44%	100.00%
Management	1.02%	86.69%	12.29%	100.00%
Retired	0.82%	77.37%	21.81%	100.00%
Self-Employed	0.50%	91.66%	7.84%	100.00%
Services	1.04%	91.67%	7.30%	100.00%
Student	0.00%	81.48%	18.52%	100.00%
Technician	1.09%	89.16%	9.75%	100.00%
Unemployed	1.37%	85.66%	12.98%	100.00%
Unknown	0.48%	88.89%	10.63%	100.00%
Grand Total	0.97%	89.16%	9.87%	100.00%



PART II : Bivariate Analysis

Variables under consideration: Bivariate Analysis on the basis of Jobs

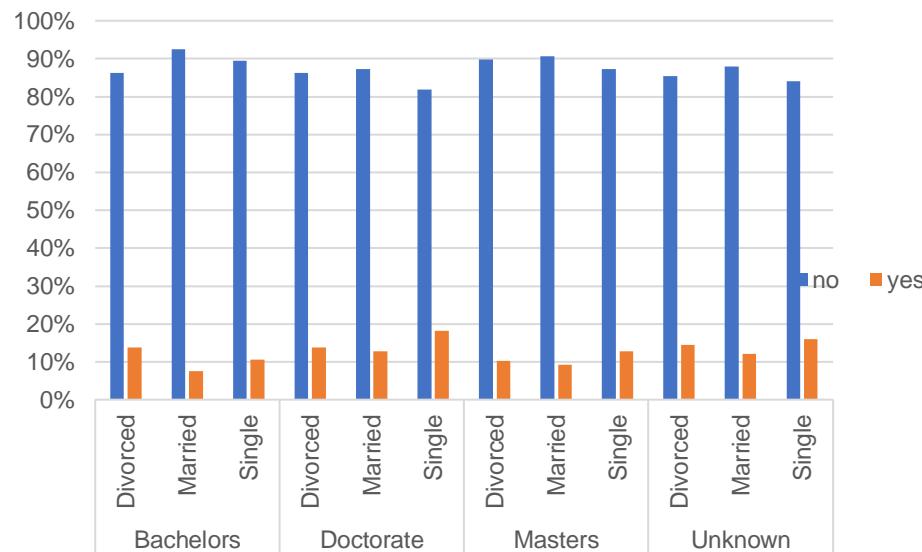
Subscription %	Column Labels		
Row Labels	yes	no	Grand Total
Administration	12.01%	87.99%	100.00%
Blue-Collar	7.20%	92.80%	100.00%
Entrepreneur	8.01%	91.99%	100.00%
Housemaid	8.91%	91.09%	100.00%
Management	13.58%	86.42%	100.00%
Retired	22.55%	77.45%	100.00%
Self-Employed	11.71%	88.29%	100.00%
Services	8.88%	91.12%	100.00%
Student	28.24%	71.76%	100.00%
Technician	10.87%	89.13%	100.00%
Unemployed	15.24%	84.76%	100.00%
Unknown	11.81%	88.19%	100.00%
Grand Total	11.54%	88.46%	100.00%



PART II : Bivariate Analysis

Variables under consideration: Education and Marital Status

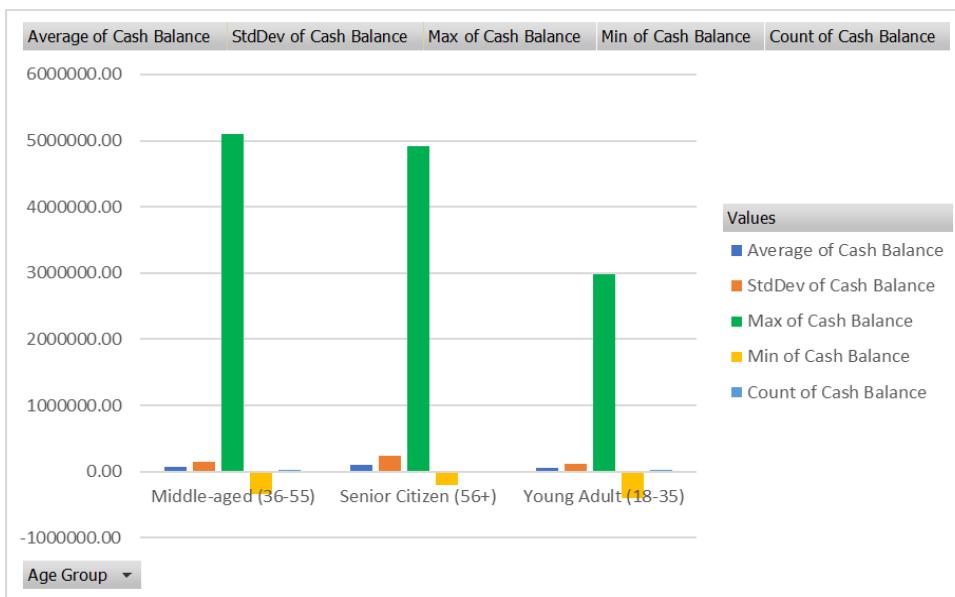
% of Subscription	Column Labels	
Row Labels	yes	Grand Total
Bachelors	91.39%	8.61% 100.00%
Divorced	86.22%	13.78% 100.00%
Married	92.45%	7.55% 100.00%
Single	89.40%	10.60% 100.00%
Doctorate	85.19%	14.81% 100.00%
Divorced	86.20%	13.80% 100.00%
Married	87.25%	12.75% 100.00%
Single	81.84%	18.16% 100.00%
Masters	89.62%	10.38% 100.00%
Divorced	89.74%	10.26% 100.00%
Married	90.71%	9.29% 100.00%
Single	87.29%	12.71% 100.00%
Unknown	86.62%	13.38% 100.00%
Divorced	85.45%	14.55% 100.00%
Married	87.95%	12.05% 100.00%
Single	84.07%	15.93% 100.00%
Grand Total	88.46%	11.54% 100.00%



PART II : Bivariate Analysis

Variables under consideration: The Bivariate Analysis of Age group Vs Cash balance reveals distinct financial relationship across different states of life. We used descriptive statistics and a box plot (excluding extreme outliers) to understand the distribution of Cash_Balance within each age segment.

Row Labels	Average of Cash Balance	StdDev of Cash Balance	Max of Cash Balance	Min of Cash Balance	Count of Cash Balance
Middle-aged (36-55)	68895.88	148523.33	5106350	-342350	23434
Senior Citizen (56+)	104169.08	235306.11	4920850	-202850	4872
Young Adult (18-35)	56492.49	122437.30	2982450	-400950	16894
Grand Total	68061.989	152305.925	5106350	-400950	45200



PART III: Major insights

1. Campaign Baseline and Data Imbalance  The overall campaign performance was poor, creating a significant challenge for predictive modeling:
 - Low Conversion Rate: Only 11.42% of all customers subscribed to the term deposit (yes).
 - Data Imbalance: The data is highly imbalanced, with nearly 88% of customers responding negatively (no). This confirms the necessity of targeted marketing, as blanket campaigns are inefficient.
2. High-Value Customer Segments (Focus Areas)
 - The bivariate analysis (Job \times Marital Status \rightarrow Subscription) revealed the specific demographic segments with the highest conversion rates, suggesting where marketing spend should be prioritized.

Segment Rank	Marital Status	Job	Subscription Rate (%)
1 (Top Target)	Single	Student	28.94%
2	Divorced	Retired	27.85%
3	Married	Retired	21.99%

3. Low-Value Customer Segments (Avoid/Re-evaluate)

The analysis also highlighted segments that are least likely to convert, indicating areas where current marketing resources should be reduced or fundamentally changed.

- Least Receptive Segment: Married Blue-Collar customers show the lowest conversion rate at just \$6.21\%\$.
- General Non-Performers: Entrepreneurs (\$8.01\%\$) and Blue-Collar workers overall (\$7.20\%\$) are significantly below the average conversion rate.

4. Financial and Age Distribution Insights

The analysis of Cash Balance and Age provides crucial context about the customer base:

- Wealth Increases with Age: Both the mean and median Cash_Balance steadily increase across age groups. Senior Citizens (56+) hold the highest average balance (\$\approx \\$104\$K), but also show the largest variability (standard deviation), indicating a wide disparity in wealth within this older segment.
- Core Customer Base: The Middle-aged (36-55) group is the largest demographic segment, comprising \$51.85\%\$ of the customer base. While this group has a high number of contacts, the marketing strategy must be optimized, as their conversion rate is mixed based on their job/marital status.