

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: ...SHEMEEMA NAVAS.....

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.

Data Cleaning

1. Replace negative sign in Age column to actual positive value.
2. Delete the rows where there is blank in loan default and subscription column as they are the important columns. 54 records are blank in loan default which account only 0.11 %, so remove those rows. Subscription has 467 records with missing values, so remove those records.
3. Outcome column has 72 rows blank with missing value. Remove those rows.
4. The 412 duplicate records have been found. Removed those duplicate records keeping 44610 unique records.
5. The job column has 'admin' and 'admin.' as two different jobs. So find and Replace the 'admin.' to 'admin' making a unique job type as 'ADMIN'.
6. Standardise the 'call duration' to seconds. Use excel formula
$$=iferror(left(q2,find('hrs'.q2)-1,0)*3600+iferror(mid(q2&"0 min",find("hrs",q2&"0hrs")+5,find("min",q2&"0 min")-find("hrs"q2&"0hrs)-5,0)*60+iferror(mid(q2&"0 secs"find("min"q2"0 mins")+6,find("secs",q2&"0secs")-find("mins"q2&"0 min"),),0)$$
7. Created a combination of marital and education columns as single column using concat function.
8. Created an age-group column using floor(age/10,1). So customers in the age 18-19 will fall in group 1, 20-29 fall in group 2.
9. Derived "month" column from Date using text(date, "mm"). to find out which months has higher subscription rate.
10. Convert the pays column to contacted before or not column to find out the percentage of customers contacted by bank.
11. Create a balance category column from the cash balance column to understand the type of balance for each customer. Customers with balance 0 or less falls in negative balance, customers with balance less than 5000 falls in low balance category whereas customers with balance greater than 50,000 falls in high balance category and in between falls in medium category.
12. Created a duration bin column from duration column to understand where the call falls. calls taken by customers less than 100 seconds falls in short calls, in-between 100-500 falls in medium calls and greater than 501 seconds falls in long calls.

PART I : Univariate Analysis

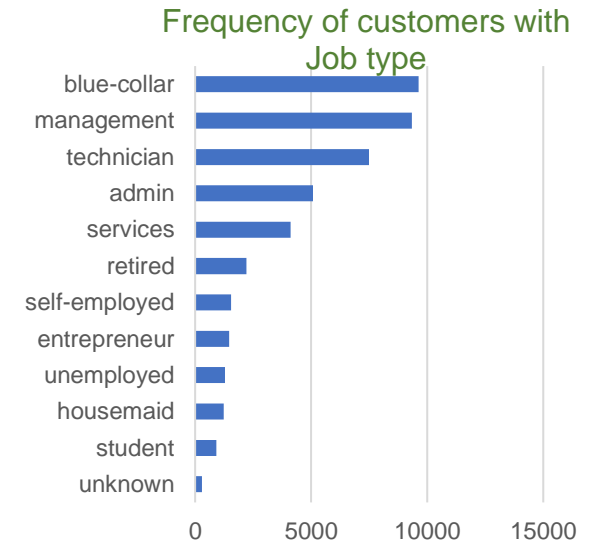
Variable under consideration:

Univariate analysis

Categorical variable

1.Job Type distribution

- The most common job the customers have is *blue-collar job*(9632 customers) and **management**(9335 customers) and least is *students and unknown* with 914 and 288 customers.
- Blue-collar** and **management** professionals form the largest group ,so, focus on them with specific financial products would be beneficial.
- Students and housemaid* are the lower representation, possibly with lower income and less likely to invest in Term deposit.



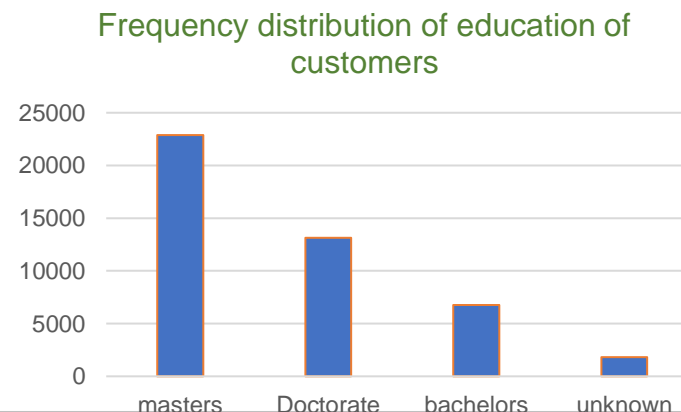
2.Education distribution of customers

- Majority of customers hold **masters degree** (22892 customers) ,followed by customers who are **doctorates**(13132 customers)

•

Highly educated people dominates the customer base ,suggesting they might be more financially aware and like to invest in term deposit.

Education	Count of Education
masters	22892
Doctorate	13132
bachelors	6758
unknown	1827
Grand Total	44609



Job type	Count of Job
unknown	288
student	914
housemaid	1230
unemployed	1282
entrepreneur	1469
self-employed	1556
retired	2211
services	4109
admin	5085
technician	7498
management	9335
blue-collar	9632
Grand Total	44609

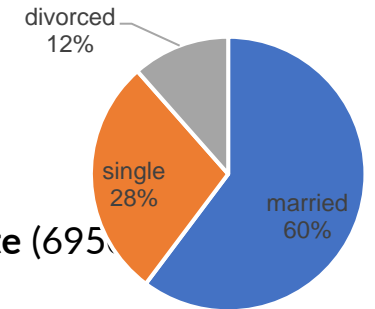
PART I : Univariate Analysis

Variable under consideration:

3.Marital status

- The 60% of customers are married and 28 % of customers are single and only 12% of customers are divorced.

Frequency distribution of marital status of customers

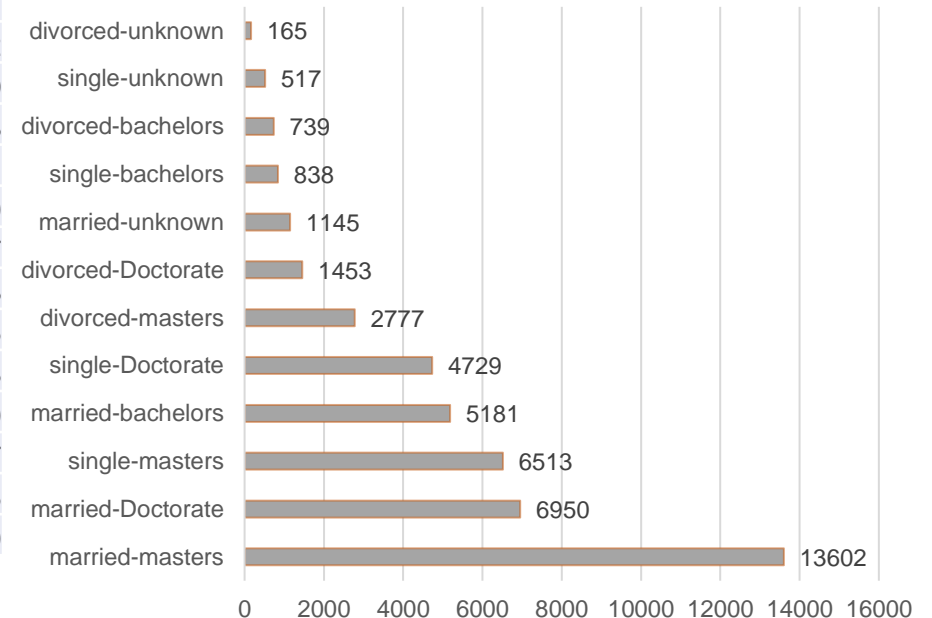


4.marital-Education Combo

- Among these, **married-masters** form the largest customer base with(13602 customers)followed by **married-doctorate** (6950).
- Single-masters**(6513 customers).
- Among **bachelors**, **married** once's are the highest customers base with 5181 customers.
- Divorced and single bachelors** are less likely to invest in term deposit.
- Married bachelors** count up to 5181 customers, which indicates there is a chance for them to Invest in term deposit as customaries marketing plans are required for those category.

Marital-education	Count of Marital-Education
married-masters	13602
married-Doctorate	6950
single-masters	6513
married-bachelors	5181
single-Doctorate	4729
divorced-masters	2777
divorced-Doctorate	1453
married-unknown	1145
single-bachelors	838
divorced-bachelors	739
single-unknown	517
divorced-unknown	165
Grand Total	44609

Frequency distribution of combo



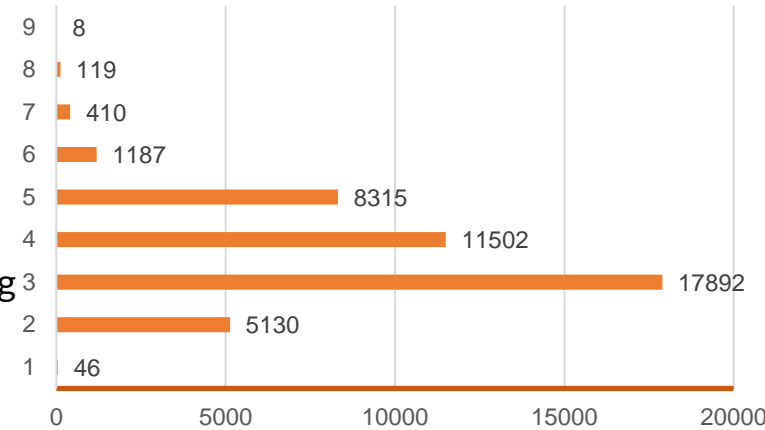
PART I : Univariate Analysis

Variable under consideration:

5. Age group

- Highest customers belongs to **age group 30-39** with 17892 customers followed by customers from **40-49 (11502)**.
- Least customers are from **18-20 and above age group 80+**.
- Most of the customers fall in the *middle aged group* ,indicating prime Earning years.
- Younger customers (group 1 and 2)might not have stable income, making Them less interest to do investment.

Age Group distribution of customers

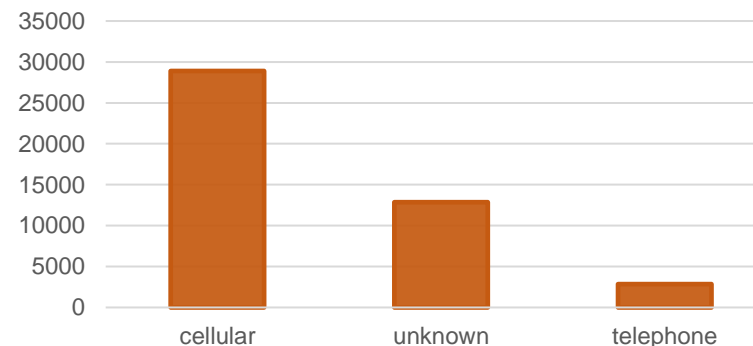


Age group	Count of Age Group
1	46
2	5130
3	17892
4	11502
5	8315
6	1187
7	410
8	119
9	8
Grand Total	44609

6.Contact mode

- Most of the customers are contacted from bank via **cellular** around 64.7% customers contacted ,whereas only 6.4% of **telephone** was Used as method of contact.
- This indicate cellular method of contact are more likely to be effective.

Frequency distribution of customers mode of contact



Contacts	Count of Contact
cellular	28885
unknown	12854
telephone	2870
Grand Total	44609

PART I : Univariate Analysis

Variable under consideration:

7.Loan default

- Around **98.2% of customers have stable credit history** indicates that they have stable or surplus income. Since they are financially disciplined ,they may open to long term investment with the bank if contacted.

8.Personal Laon

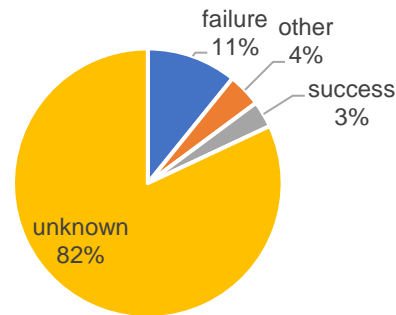
- About **83.9% of customers do not have personal loan** indicating they may have chance for investment.

9.Housing Loan

- More than half of the customers have housing loan ,(55.7%) and 44.3% of customers without housing loan.
- Target the 44.3% of customers for investment as they are more likely to do savings.

housing loan	Count of Housing Loan
yes	24860
no	19749
Grand Total	44609

Previous contact outcome



Loan default	Count of Loan Default
no	43805
yes	804
Grand Total	44609

personal loan	Count of Personal Loan
no	37442
yes	7167
Grand Total	44609

10.Previous contact outcome(poutcome)

- Only 3% of success is due to previous contact, indicating that depending on earlier customer contacts is not a significant predictor of success. There is a need for increased outreach and marketing efforts to target this sizable demographic, as 82% of unknown customers indicate that they have not yet been reached.11% of failures occur due to a lack of interest in the previous offering. Customers' lack of interest in the previously supplied product accounts for a significant percentage of failures.

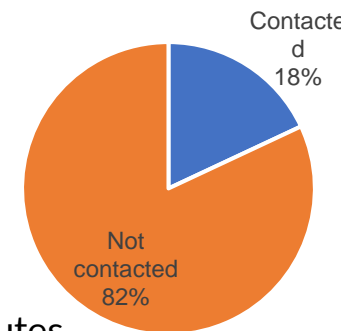
PART I : Univariate Analysis

Variable under consideration:

11. Pdays(contacted customers)

- A vast majority (82%) of the bank's customer base is not being exposed to any marketing messages for the bank's products. This means the bank is potentially missing out on a significant amount of revenue from cross-selling, upselling, and acquiring new customers.

Contact percenatge by bank

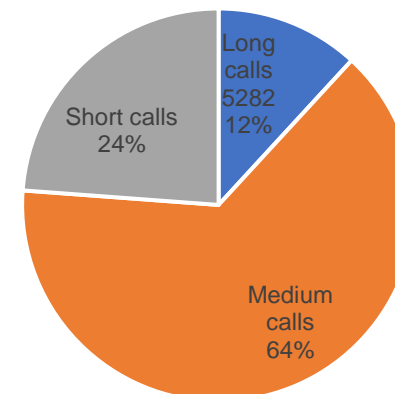


Row Labels	Count of Contacted before
Contacted	8044
Not contacted	36565
Grand Total	44609

12. Call duration

- Around 64% of the customers receives medium calls range from 2 to 8 minutes
- 24% of customers received calls less than 2 minutes and only 12 % Keeps the call longer than 8 minutes.
- This indicates that they show interest but need persuasive arguments and personalized offers to commit.

Call duration bins distribution



Row Labels	Count of duration bin
Long calls	5282
Medium calls	28702
Short calls	10625
Grand Total	44609

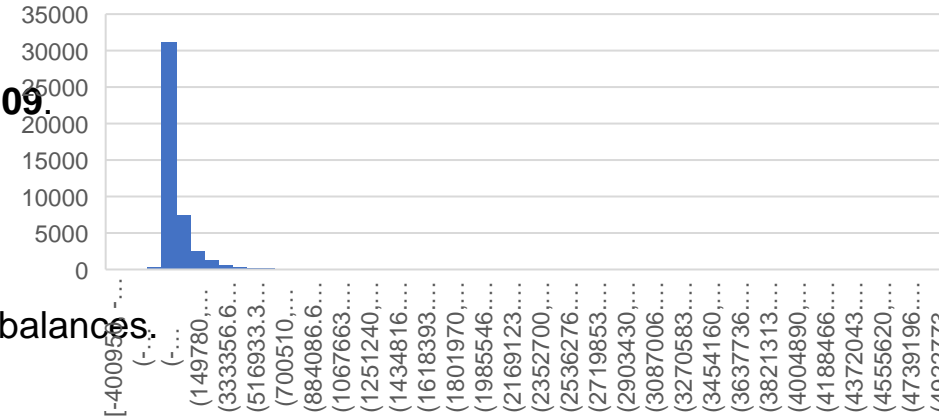
PART I : Univariate Analysis

Variable under consideration:

13. Cash Balance Univariate_numerical_analysis

- The *average* cash balance across all individuals is **68,163.09**.
- The median value is 22,300, meaning that half of the individuals in the dataset have cash balances lower than this amount, and half have higher balances.
- The high **kurtosis** and **skewness**, meaning there are some individuals with extremely large positive or negative balances.

cash balance of customers



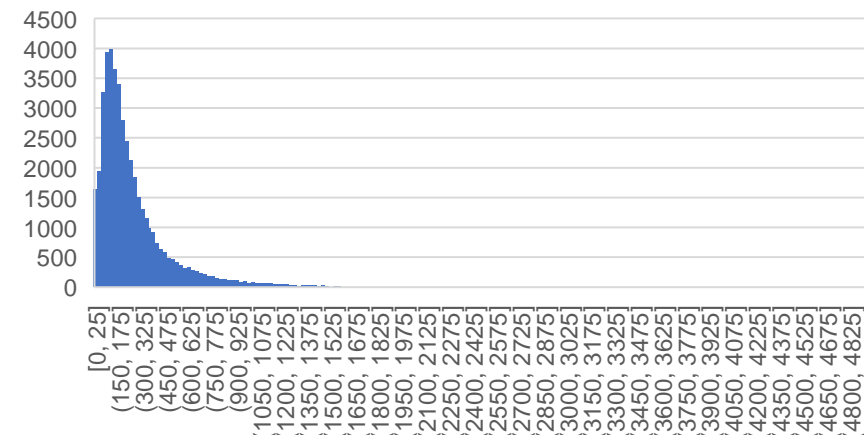
Cash Balance	
Mean	68163.08705
Standard Error	722.5338344
Median	22300
Mode	0
Standard Deviation	152605.2352
Sample Variance	23288357823
Kurtosis	141.0650295
Skewness	8.377417873
Range	5507300
Minimum	-400950
Maximum	5106350
Sum	3040687150
Count	44609

14. Call duration

- Average duration is 258.08 ,indicating most of the call happens for 4 minute and 18 seconds.
- The median value is 180 seconds (3 minutes),indicates that half of the durations are less than 3minutes,and half are greater, indicating that there may be significant number of shorter durations compared to the average.
- The positive skewness (3.15) indicates that the distribution is right-skewed. This means that most of the durations are relatively short

Duration in sec	
Mean	258.0814
Standard Error	1.22238
Median	180
Mode	124
Standard Deviation	258.1769
Sample Variance	66655.3
Kurtosis	18.21932
Skewness	3.153836
Range	4918
Minimum	0
Maximum	4918
Sum	11512753
Count	44609

Total duration



PART I : Univariate Analysis

Variable under consideration:

Segmented Univariate analysis

1.marital status ,Job vs Duration call

- Single customers had the longest call duration(267 sec),followed by divorced(263 sec)
- Retired ,Self-employed (269 sec) and Blue-collar (263 sec)** workers showed relatively high engagement.
- Management and Entrepreneur** relatively have 4 minutes of call which Ideally makes them all potential customers.
- Students have shorter call duration(246 sec),makes them less interest in investment calls.

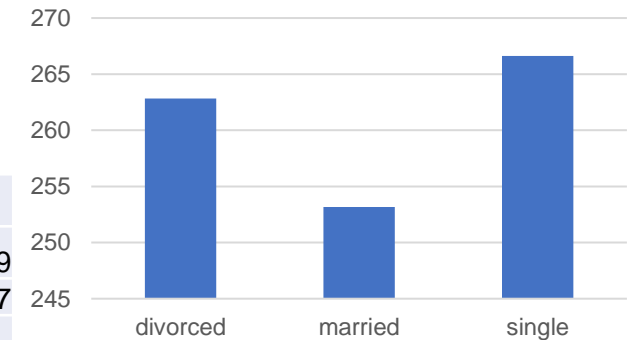
2.Call duration vs contact

- Cellular** calls have an average duration of 263 seconds, making them the most effective mode of communication.
- This suggests that customers are more engaged when contacted via mobile phones, and ideal time is **4 minutes duration call**.

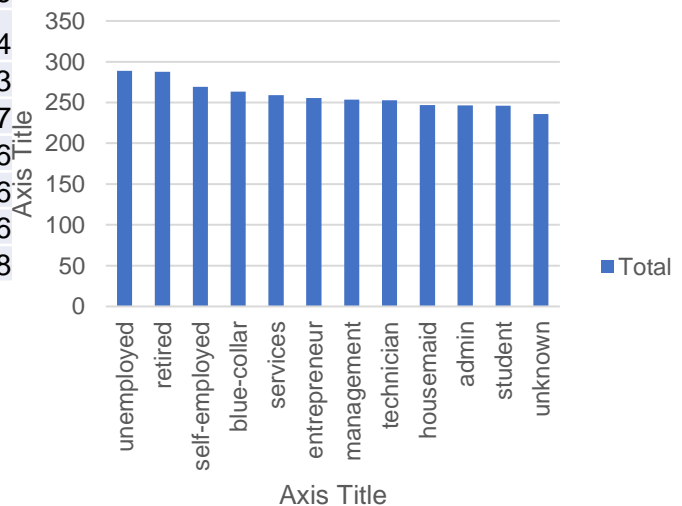
Row Labels	Average of Duration in sec
cellular	263
telephone	234
unknown	253
Grand Total	258

Job	Average call Duration
unemployed	289
retired	287
self-employed	269
blue-collar	263
services	259
entrepreneur	256
management	254
technician	253
housemaid	247
admin	246
student	246
unknown	236
Grand Total	258

Avg call duration by marital status type



Duration cal of each job typeel



PART I : Univariate Analysis

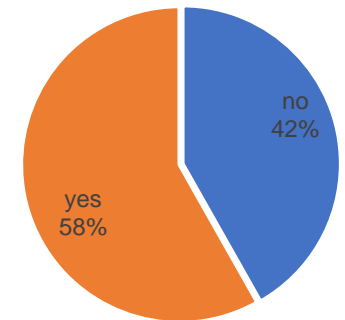
Variables under consideration:

3. Subscription ,age ,Job type vs cash balance

- Customers who opened term deposit has average cash balance of 90754 which makes 16 %more cash balance than customers who does not have deposits.
- Among those **retired** has highest *cash balance*(98714 cash),followed by **unknown**(89266) and **management** (88381 cash balance).
- Doctorate** holders had the highest cash balance (**88,159**), followed by unknown (**76,675**), bachelors (**62,387**), and masters (**57,718**).
- Higher education level correlates with higher cash balances, making them potential high-value customers.
- Customers in **Age 70-89** have the highest average cash balances, making Them potential customers for next marketing campaign.

subscription	Average of Cash Balance
no	65215
yes	90754
Grand Total	68163.08705

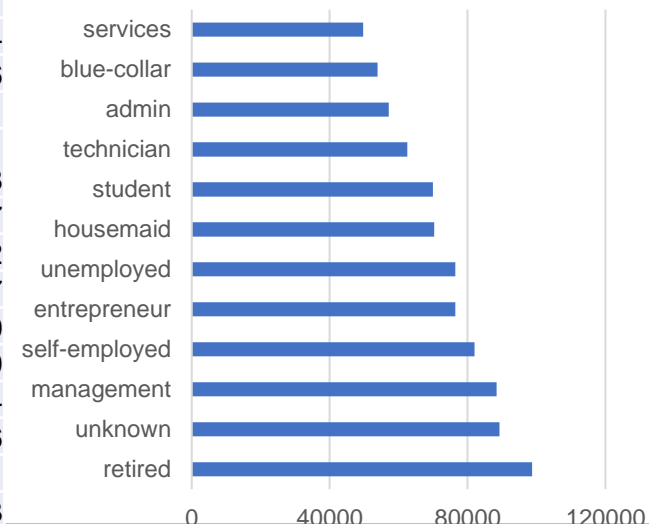
mean cash baalnce vs subscription



Row Labels	Average of Cash Balance	Min of Cash Balance	Max of Cash Balance
bachelors	62387	-130200	3559400
Doctorate	88159	-342350	5106350
masters	57718	-400950	4060200
unknown	76675	-72250	3217150
Grand Total	68163	-400950	5106350

Job type	Average of Cash Balance
retired	98714
unknown	89266
management	88381
self-employed	82058
entrepreneur	76487
unemployed	76452
housemaid	70317
student	69990
technician	62509
admin	57184
blue-collar	53916
services	49741
Grand Total	68163

avg cash balance with different jobs



Age group	Average of Cash Balance	Min of Cash Balance
1	25126	0
2	49665	-400950
3	61419	-152900
4	67136	-342350
5	84560	-165650
6 Total	111213	-202850
7	142846	0
8	164913	0
9	40963	0
Grand Total	68163	-400950

PART I : Univariate Analysis

Variables under consideration:

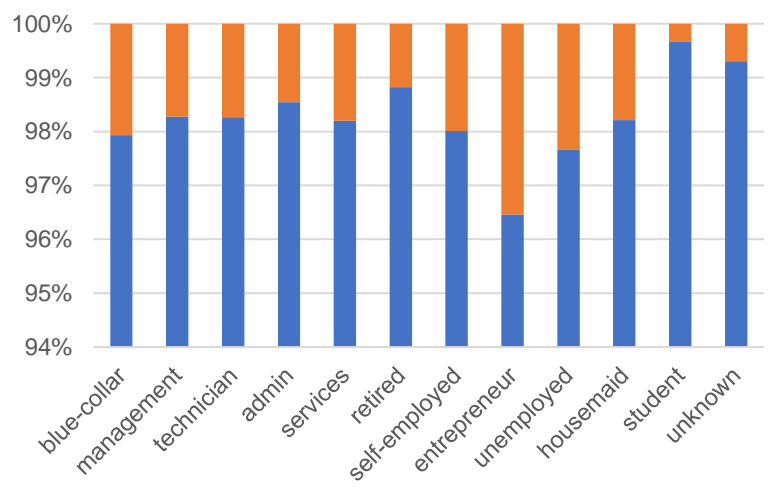
- Customers in the age group of **30-49** has average cash balance of 64,277, have moderate financial stability, which indicates they are the looking for secure investment options like term deposits.
- Some individuals in these groups have negative balances (Min Cash Balance: -152,900 for Age Group 3, -342,350 for Age Group 4), Offer flexible deposit plans with lower initial investment for those still building their savings.

4.Education ,Job type vs loan default

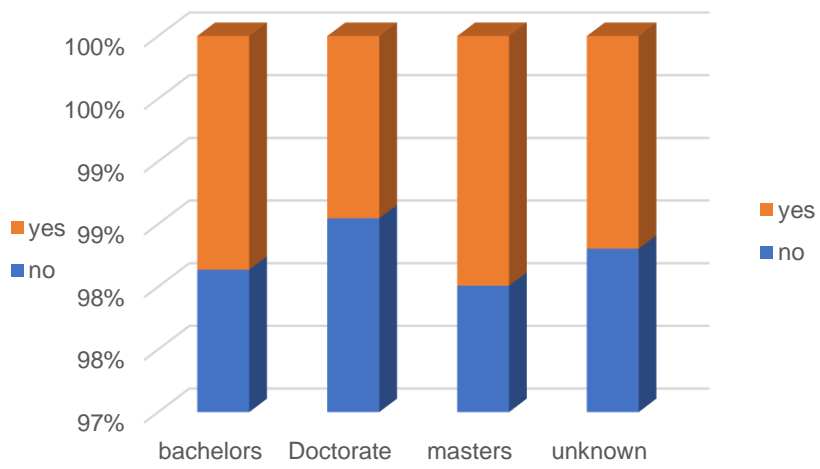
- Customers with a **Doctorate (98.5% no defaults)** and **Bachelors (98.1% no defaults)** have the lowest loan default rates.
- Masters Degree Holders Have the Highest Defaults456 defaults out of 22,892 (2.0%), which is the highest default rate.

- Blue-collar (9433 no defaulters, 21.5%)** and **Management (9174 no defaulters, 20.9%)**
- have the lowest number of loan defaulters.

Job type vs loan default



education vs loan default



Count of Loan Default	loan default		Grand Total
	no	yes	
education			
bachelors	6632	126	6758
Doctorate	12941	191	13132
masters	22436	456	22892
unknown	1796	31	1827
Grand Total	43805	804	44609

PART I : Univariate Analysis

Variables under consideration:

- Retired customers are a great target for **fixed-income investment options**.
- **Technician: 7,368 individuals** (16.8%) A strong pool, but may need mid-range investment plan
- **Focus on Management, Technicians, and Blue-collar workers** as they form the largest financially stable groups.
- **Avoid aggressive marketing to students and housemaids**, as they have lower savings potential.

PART II : Bivariate Analysis

Variables under consideration:

Co-relation

The correlation of all variable with respect to Subscription rate.

1.*Age* (0.0244) indicates that older customers are likely to invest in term deposit.

2.*Loan default* (-0.0230) has a negative correlation indicates that if loan default rate decrease further ,more customers are likely to invest in term deposit.

	Age	Age Group	Loan default	Housing loan	personal loan	Cash Balance	pdays	previous	Duration in sec	Subscribed or not
Age	1									
Age Group	0.963721735	1								
Loan default	-0.017644966	-0.017353433	1							
Housing loan	-0.185437423	-0.176411891	-0.0064654	1						
personal loan	-0.015945804	-0.017106663	0.07655108	0.040179	1					
Cash Balance	0.096251957	0.094075385	-0.0666516	-0.06834	-0.08395	1				
pdays	-0.026177445	-0.023927599	-0.0292326	0.129348	-0.02105	0.002421	1			
previous	-0.001301029	-0.000804507	-0.0208188	0.048313	-0.00916	0.019947	0.544041	1		
Duration in sec	-0.00570257	-0.007709202	-0.0096481	0.005945	-0.01235	0.020994	-0.00391	-0.00283	1	
Subscribed or not	0.02448117	0.02301722	-0.0209969	-0.13742	-0.06751	0.053482	0.099218	0.107534	0.396287	1

PART II : Bivariate Analysis

Variables under consideration:

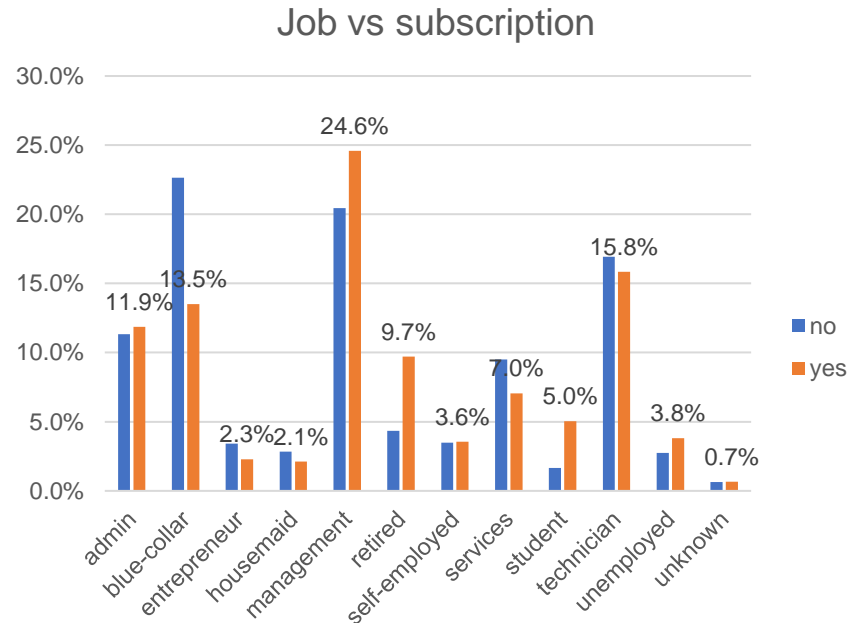
3. A small negative correlation (-0.0675) between *personal loans* and term deposit investment suggests a weak inverse relationship. This indicates that individuals with personal loans are slightly less likely to invest in term deposits. The same trend is observed for *housing loans*.
4. A weak positive correlation (0.054) exists between *cash balance* and term deposit subscription, indicating a slight direct relationship. Customers with higher cash balances are marginally more likely to subscribe to term deposits. Consequently, focusing marketing efforts on high-balance customers may yield increased subscription rates.
5. Positive correlation of *duration of call* (0.3963) indicates that longer the calls, more chance to convert the call to subscription. This suggests that effective communication or persuasion during the call plays a major role.
6. *Pdays* with (0.099) with week positive correlation indicates that most recent past contacts are likely to convert to subscription. As the days increase chance of converting the call to successful deposit decreases.
7. Only 18% of customers were contacted by the bank. Despite a weak positive correlation (0.10) between contact and conversion, this low contact rate suggests a significant opportunity to increase customer engagement and potential investment by expanding communication efforts.

PART II : Bivariate Analysis

Variables under consideration:

1. Job vs subscription rate

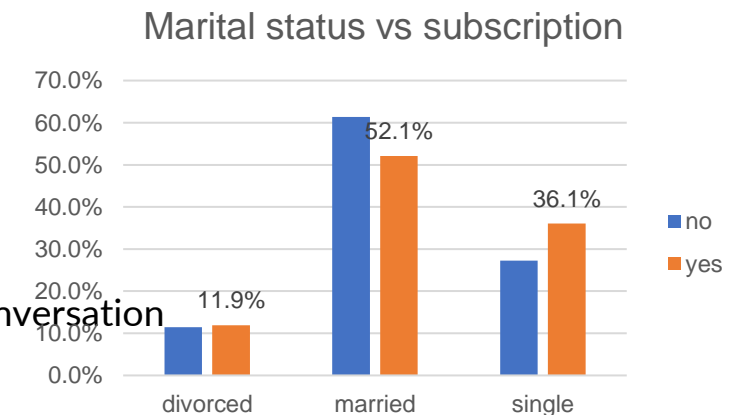
- Management (24.6%)** has the highest term deposit subscription rate, followed by **technicians (15.8%)** and **blue-collar workers (13.5%)**.
- Blue-collar subscriptions are low relative to their customer base, indicating untapped potential.
- Admin customers have an even split in subscription rates, suggesting a strong possibility for increased conversions with focused marketing.



job vs subscription	Subscription	
	no	yes
admin	11.3%	11.9%
blue-collar	22.6%	13.5%
entrepreneur	3.4%	2.3%
housemaid	2.8%	2.1%
management	20.4%	24.6%
retired	4.3%	9.7%
self-employed	3.5%	3.6%
services	9.5%	7.0%
student	1.7%	5.0%
technician	16.9%	15.8%
unemployed	2.8%	3.8%
unknown	0.6%	0.7%

2. Marital Status vs subscription rate

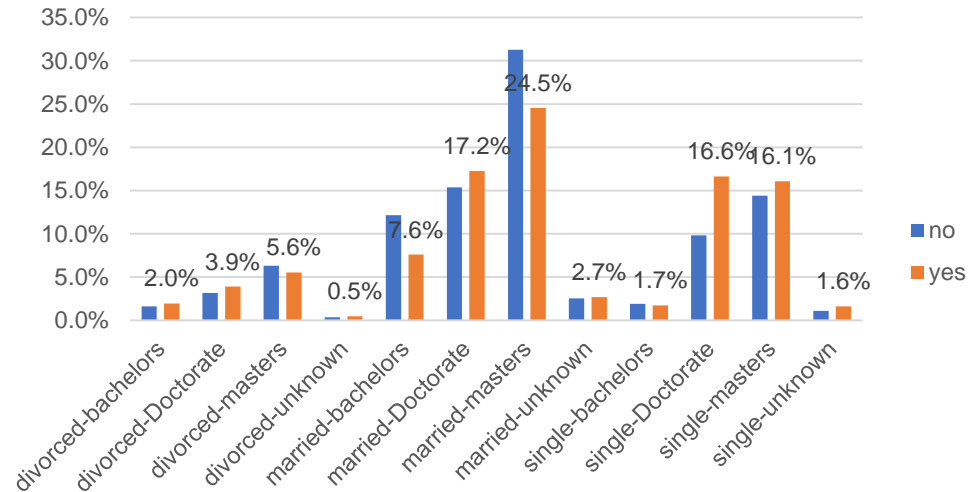
- Married people around(52.1%) have highest subscription rate followed by singles with 36%.
- High educated groups of people are more likely to subscribe ;**Married-masters(24.5%)** and **Married-doctorate (17%)** and single-doctorate are the group of customers with high subscription rate.
- Focus on single -doctorates and masters customers as they are more likely to convert the conversation to successful deposit.



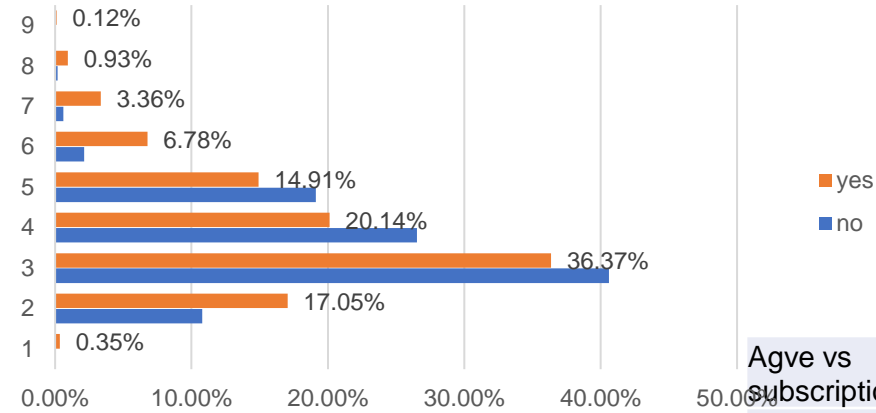
PART II : Bivariate Analysis

Variables under consideration:

Marital-Education vs subscription



Age group vs subscription



3.Age vs subscription rate

- The **30-39 age group** demonstrates the highest term deposit subscription rate at 36.37%. This age range, along with the **40-49** (20%) and 20-29 (17.05%) groups, constitutes the largest segment of subscribers.
- The 30-49 age group exhibits a significant 'no subscription' rate. Therefore, targeted and effective marketing strategies are crucial to convert this demographic into term deposit subscribers.

Age vs subscription	Subscription		
Age group	no	yes	Grand Total
1	0.07%	0.35%	0.10%
2	10.78%	17.05%	11.50%
3	40.60%	36.37%	40.11%
4	26.52%	20.14%	25.78%
5	19.13%	14.91%	18.64%
6	2.12%	6.78%	2.66%
7	0.60%	3.36%	0.92%
8	0.18%	0.93%	0.27%
9	0.01%	0.12%	0.02%
Grand Total	100.00%	100.00%	100.00%

PART II : Bivariate Analysis

Variables under consideration:

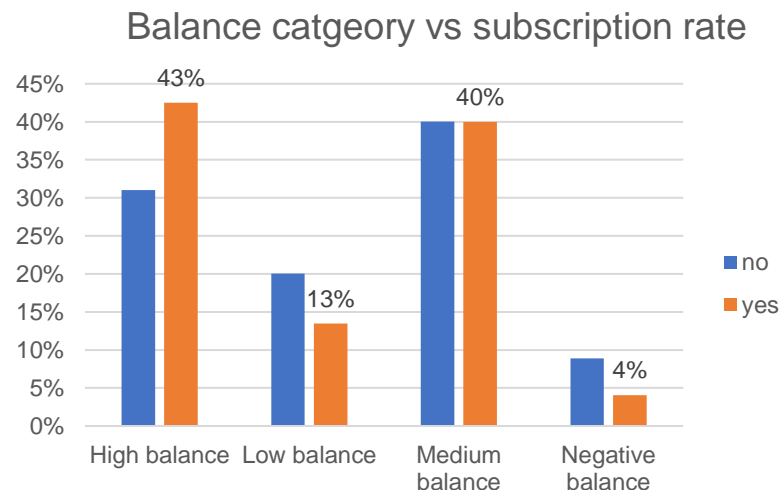
4.Loan Default: The data suggests a potential counterintuitive trend: customers with no outstanding loan defaults exhibit a lower propensity for term deposit subscription. Further analysis is required to validate this finding and understand the underlying dynamics. There appears to be a weak or negligible direct correlation between loan default status and term deposit subscription.

5.Housing Loans: 37% of customers with housing loans have subscribed to term deposits, while 58% have not. This indicates a mixed response, suggesting that the presence of a housing loan does not uniformly correlate with increased investment in term deposits.

6.Personal Loans: The term deposit subscription rate among customers with personal loans is notably low, at only 9%. This suggests a potential inverse relationship between personal loan acquisition and term deposit investment.

7.Cash balance vs subscription rate

- Customers with cash balances exceeding £5,000, encompassing both **high and medium balance segments**, demonstrate a higher propensity to subscribe to term deposit accounts



loan default vs subs	subscription		
Loan default	no	yes	Grand Total
no	98%	99%	98%
yes	2%	1%	2%
Grand Total	100%	100%	100%

housing loan vs subscri	subscription		
Housing loan	no	yes	Grand Total
no	42%	63%	44%
yes	58%	37%	56%
Grand Total	100%	100%	100%

personal loan vs subscri	subscription		
personal loan	no	yes	Grand Total
no	83%	91%	84%
yes	17%	9%	16%
Grand Total	100.00%	100.00%	100.00%

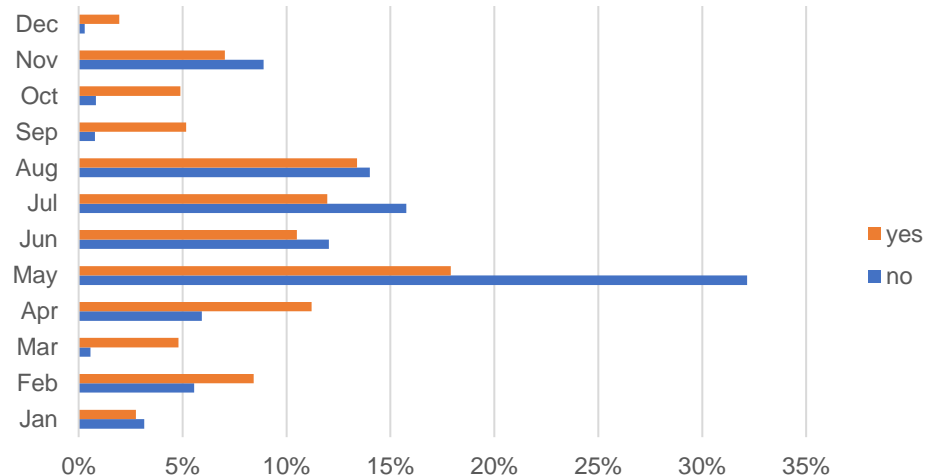
PART II : Bivariate Analysis

Variables under consideration:

8. Mode of contact ,duration vs subscription rate

- **Cellular calls** demonstrate a high conversion rate to term deposit subscriptions, with 82% success.
- The most effective call duration falls within **the medium range of 2 to 8 minutes**.
- Calls shorter than 2 minutes are less likely to result in subscriptions, while calls **exceeding 8 minutes show a higher propensity** for conversion
- The conversion rate is high on May (18%) and Followed by August and July.

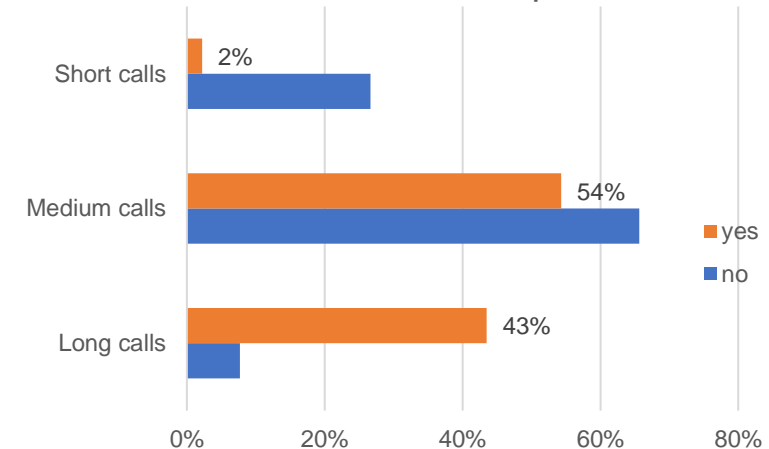
Months vs subscription rate



mode vs subscription		Subscription		Grand Total
Mode of contact		no	yes	
cellular		62%	82%	65%
telephone		6%	7%	6%
unknown		31%	10%	29%
Grand Total		100%	100%	100%

duration vs subscription		subscription rate		Grand Total
Duration bin		no	yes	
Long calls		8%	43%	12%
Medium calls		66%	54%	64%
Short calls		27%	2%	24%
Grand Total		100%	100%	100%

Duration vs subscription rate



PART III: Major insights

Target High-Potential Customer Segments

- Focus on Management, blue-collar and technicians as they represent the largest stable income groups.
- Focus on high educated group with married-doctorate and masters and even single doctorate who are in the age range of 30-49 age group.
- customers in the 70-89 age group have the highest average cash balances, making them potential targets.
- Target customers with high cash balance, those customers are in the range of retirement. so focus more marketing strategies to convert such customers as they prefer one to one contact rather than call.
- Offer flexible deposit plans with lower initial investment for those still building their savings, as some individuals have medium balance.

Optimise communication channel:

- Utilize cellular calls as the primary contact method, as they have proven to be the most effective .
- Aim for medium-length calls (2-8 minutes) for optimal engagement and conversion.
- Increase outreach efforts, as a significant portion (82%) of the customer base has not been contacted. As Previous contact outcomes are not strong predictors of success, so more communication initiatives are crucial.
- Recent past contacts are more likely to subscribe so increase the frequency of customer contact.