

Bank Marketing EDA

08, September , 2025

Bank Marketing EDA

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.

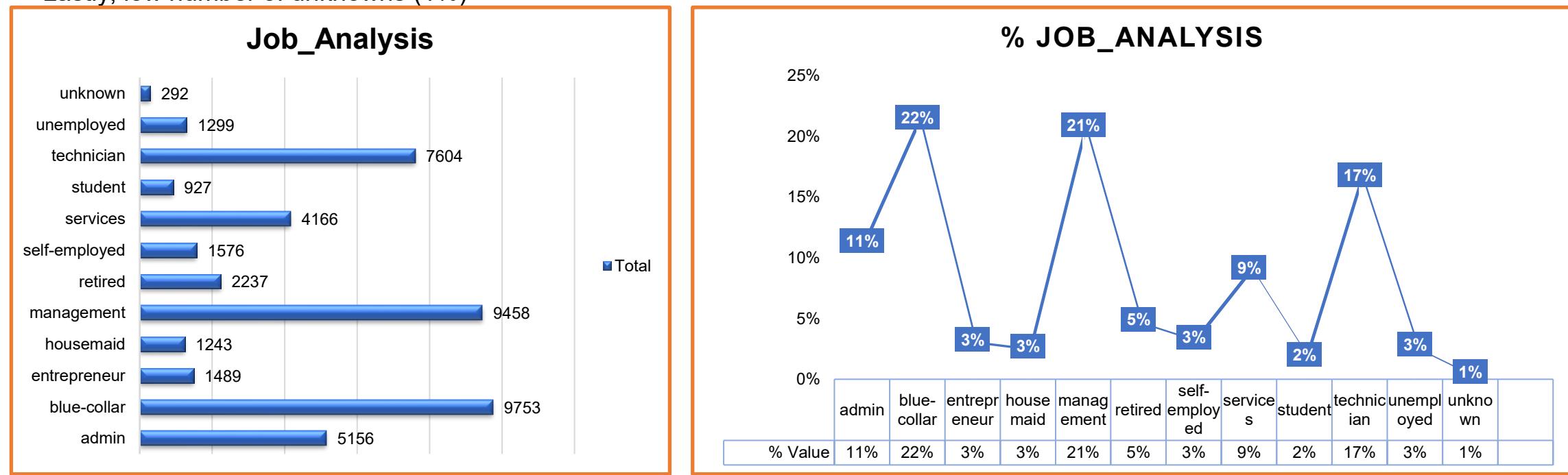
Assignment Objective

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: 1. JOB

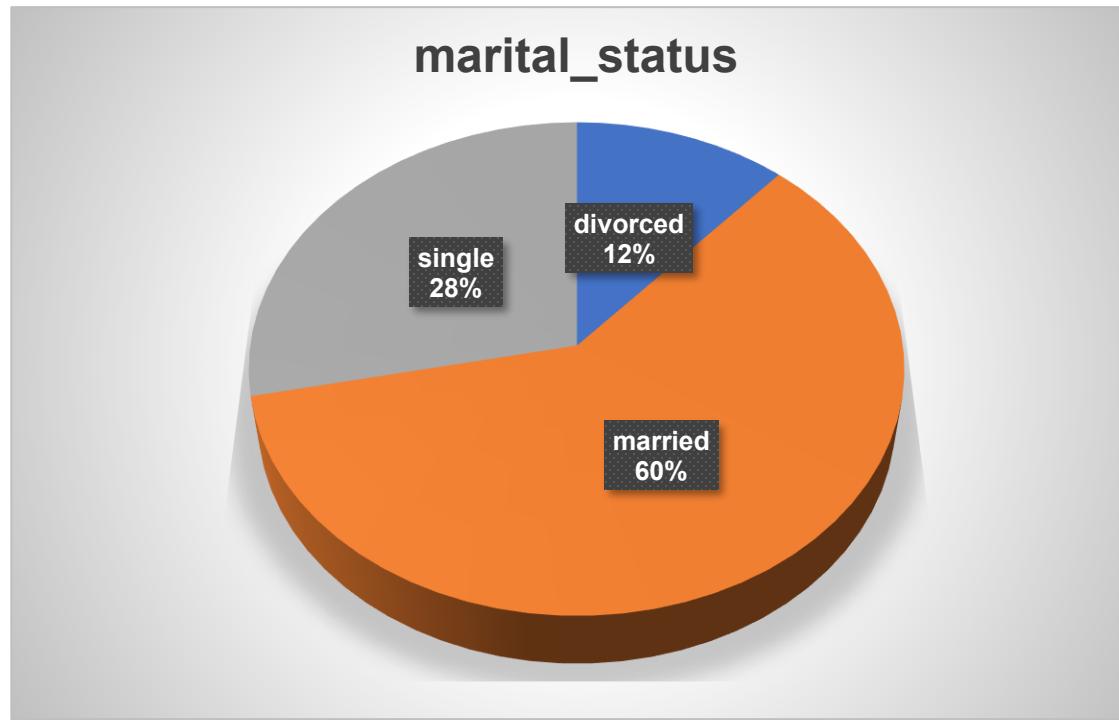
- Blue-collar, management, and technician roles are of the majority: These three account for 60% of all customers.
- Admin and services workers form the next largest chunks (~20% combined)
- Small segments like students, entrepreneurs, housemaids, and unemployed are niche segments: Together, they account for lesser than 15%.
- Lastly, low number of unknowns (1%)



PART I : Univariate Analysis

Variable under consideration: 2. MARITAL STATUS

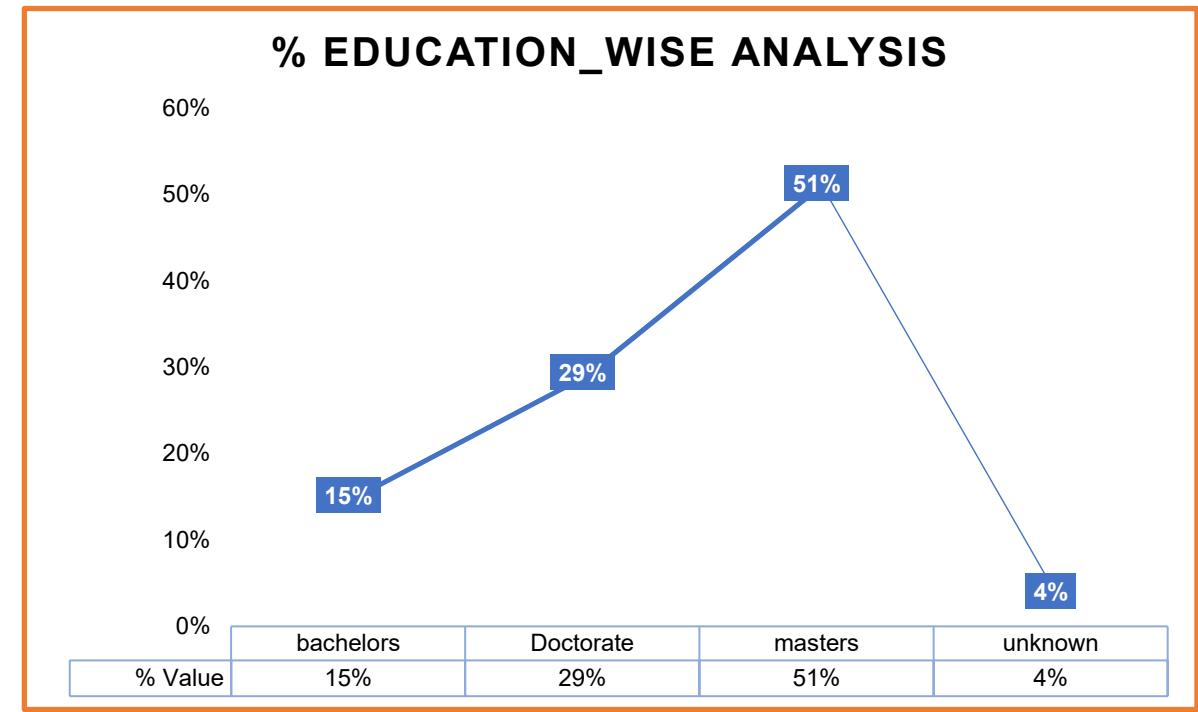
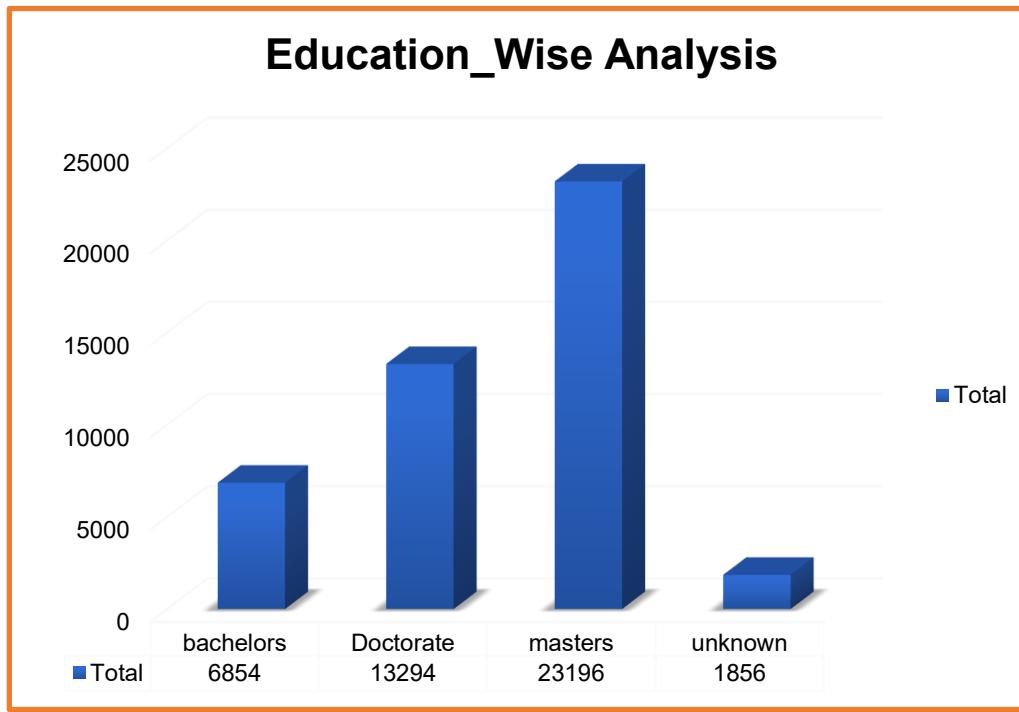
- Majority are married: 60% of customers are married, making them the dominant segment.
- Single customers form a significant minority (28%)
- Divorced customers are the smallest group (12%)



PART I : Univariate Analysis

Variable under consideration: 3. EDUCATION

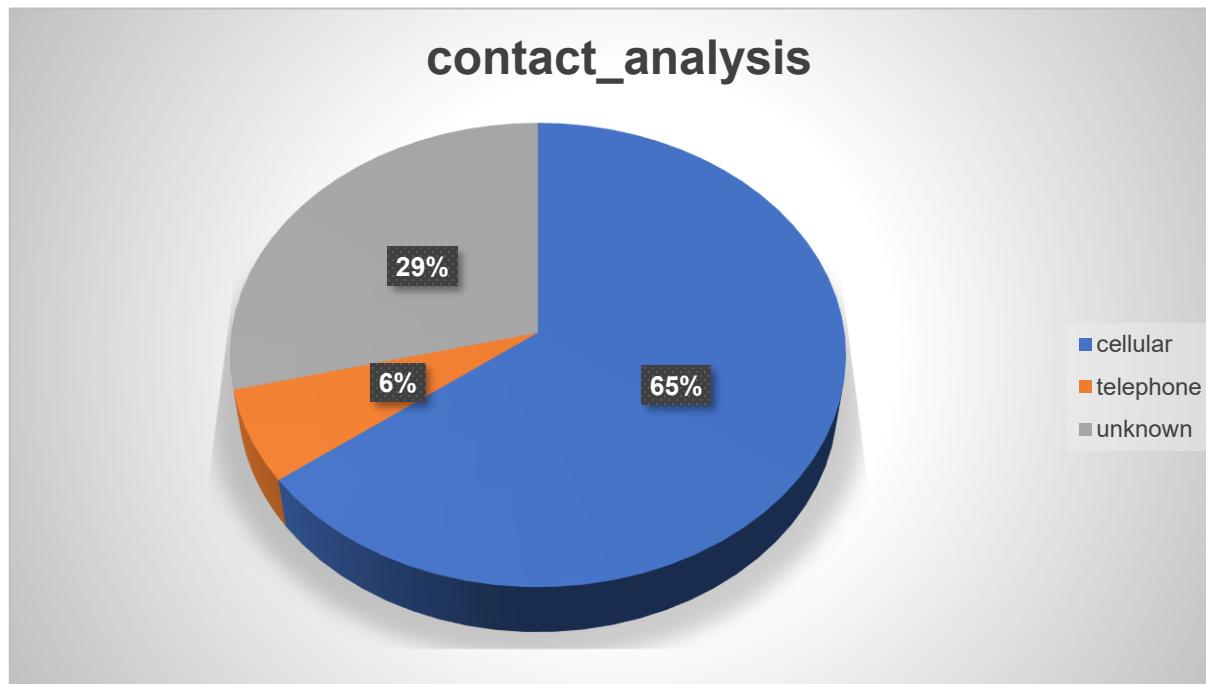
- Master's degree holders dominate: Over half of the customer base (51%) have a master's degree. This denotes a highly educated customer base.
- Doctorate holders form the second-largest group (29%)
- Bachelor's degree holders are a smaller segment (15%)
- Unknown education (4%): Relatively small but still worth acknowledging in campaign design



PART I : Univariate Analysis

Variable under consideration: 4. CONTACT

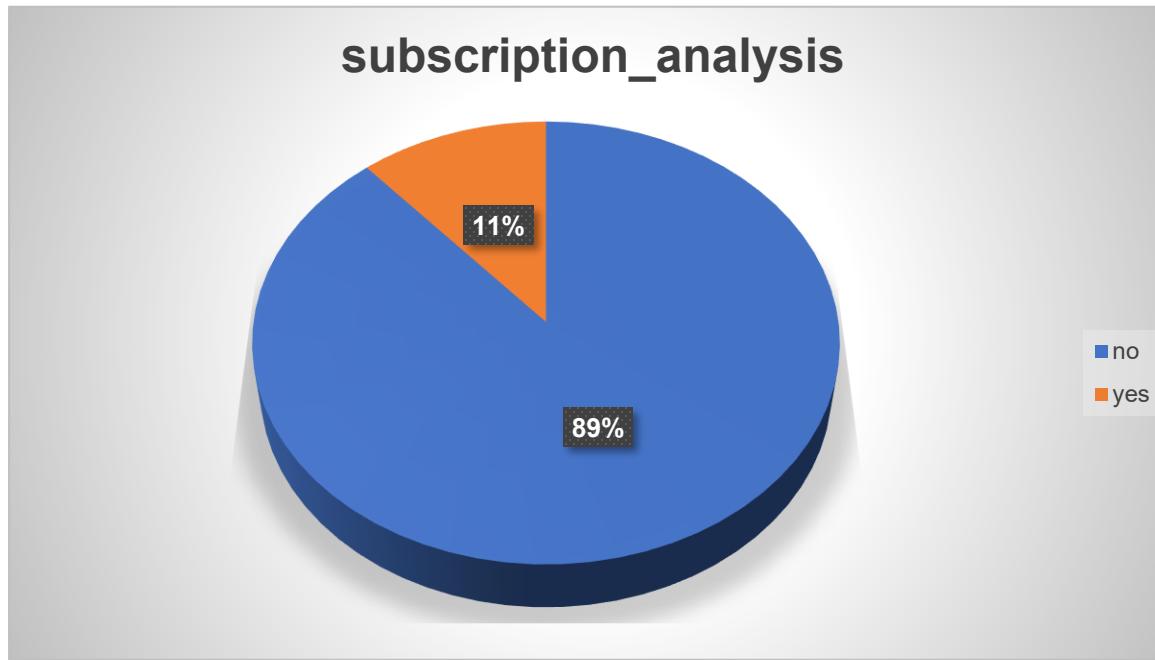
- Cellular is the dominant contact method (65%): This reflects that most customers prefer or are reachable via mobile phones. Suggests convenience.
- Telephone (landline) is rarely used (6%): Could indicate outdated records, or older demographics.
- A significant portion has 'Unknown' as contact (29%): Almost a third of the data has no contact method specified.



PART I : Univariate Analysis

Variable under consideration: 5. SUBSCRIPTION

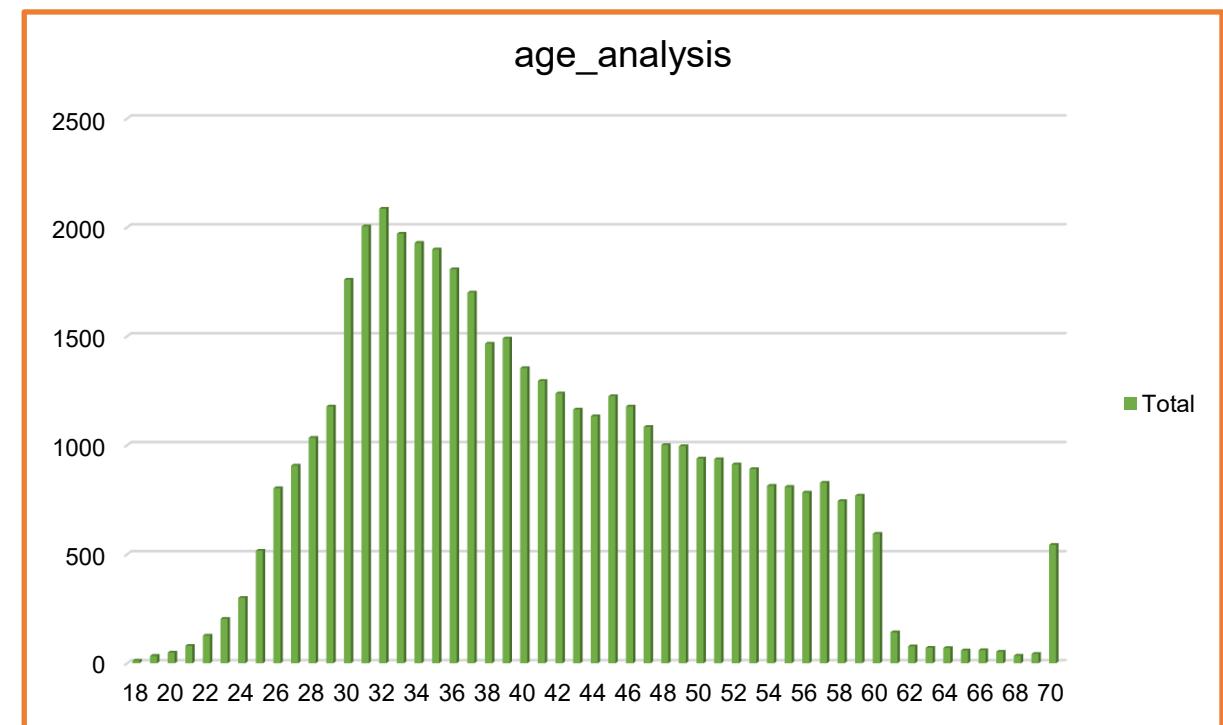
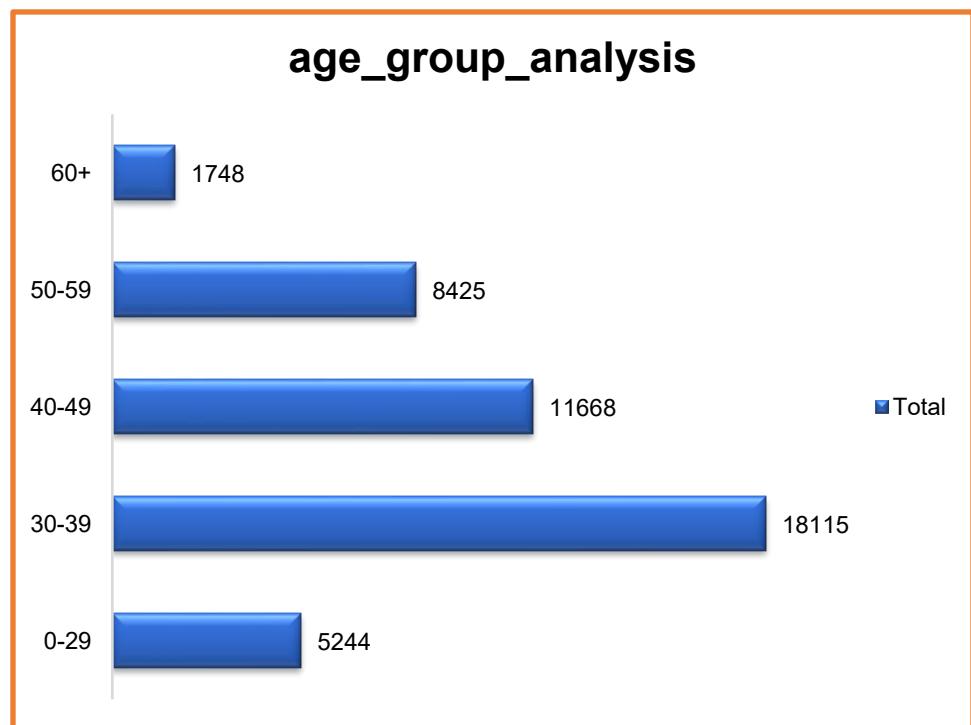
- Low overall conversion rate: Only 11% of customers subscribed to the term deposit.
- A large 89% did not convert, indicating limited campaign effectiveness or lack of customer interest.



PART I : Univariate Analysis

Variable under consideration: 6. AGE

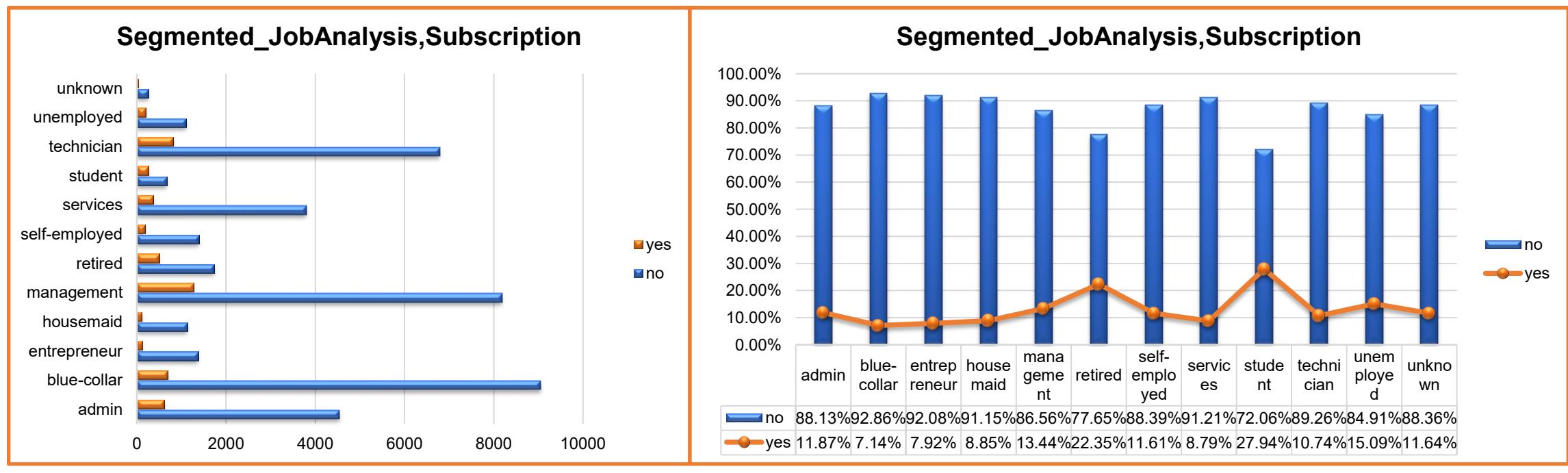
- High engagement from customers aged 30–39: This group makes up 40% of the customer base – the largest segment by far.
- Customers aged 40–49 form the second-largest group (26%) – potential mid-career targets.
- Lower representation from customers under 30 (12%) and over 60 (4%)



PART I : Univariate Analysis

Variable under consideration: 1. JOB [SEGMENTED UNIVARIATE with subscription]

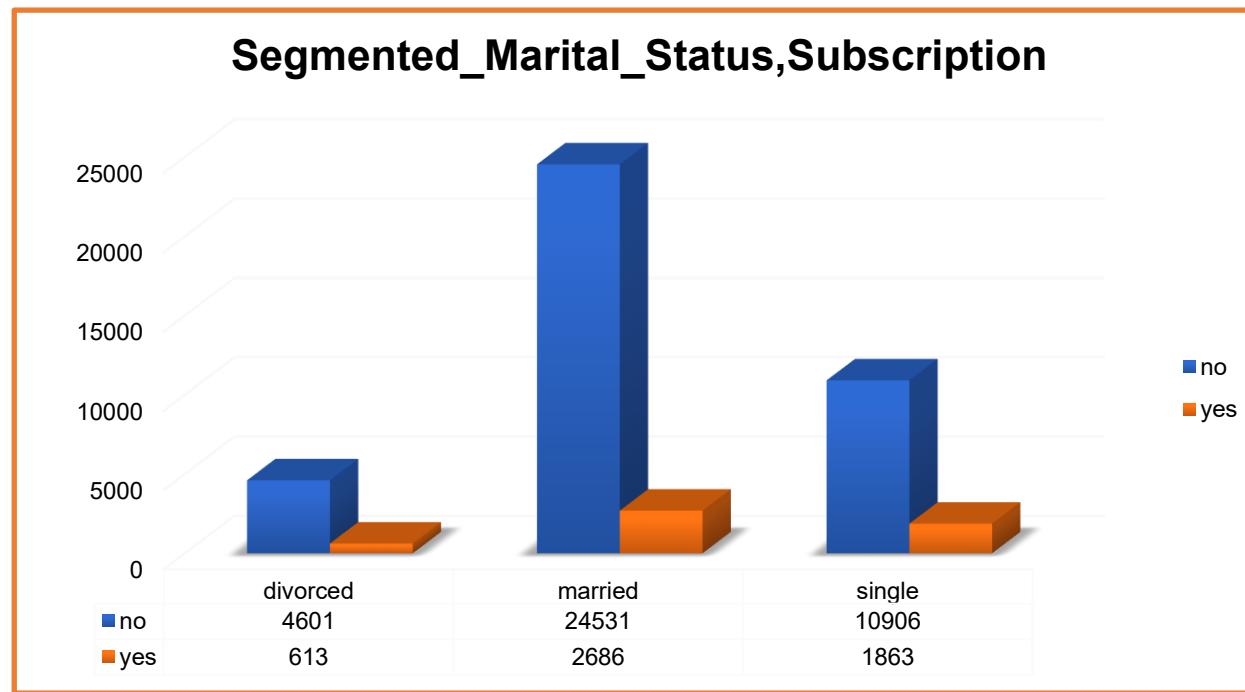
- Students (27.9%) and Retired individuals (22.3%) have the highest subscription rates: These groups may have more time to talk.
- Management (13.4%) and Unemployed (15.1%) also exceed the average conversion rate (11.4%): Management may have disposable income and financial foresight.
- Blue-collar and Entrepreneurial segments have relatively low subscription rates (7–8%): Possibly due to job nature. Need targeted messaging or different communication channels.



PART I : Univariate Analysis

Variable under consideration: 2. MARITAL STATUS [SEGMENTED UNIVARIATE with subscription]

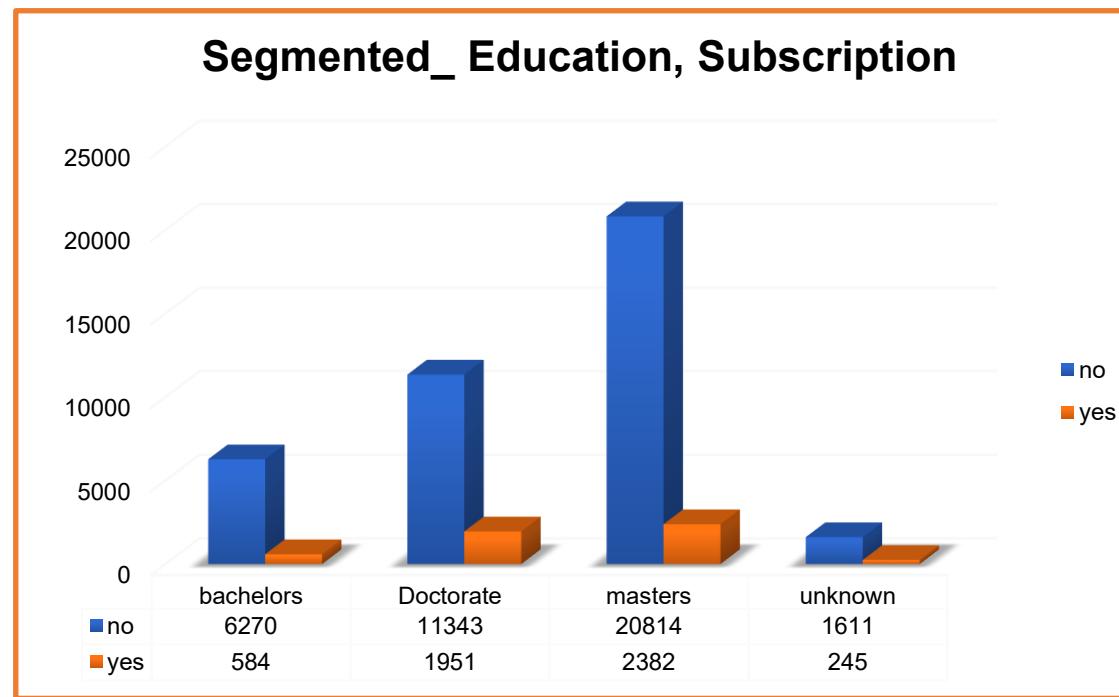
- Single customers show the highest conversion rate (14.6%): They may be more flexible in financial decision-making.
- Married customers have the lowest conversion rate (9.9%), below the overall average (11.4%): Possibly more cautious, or need partner consensus.
- Divorced customers are slightly above average (11.8%)



PART I : Univariate Analysis

Variable under consideration: 3. EDUCATION [SEGMENTED UNIVARIATE with subscription]

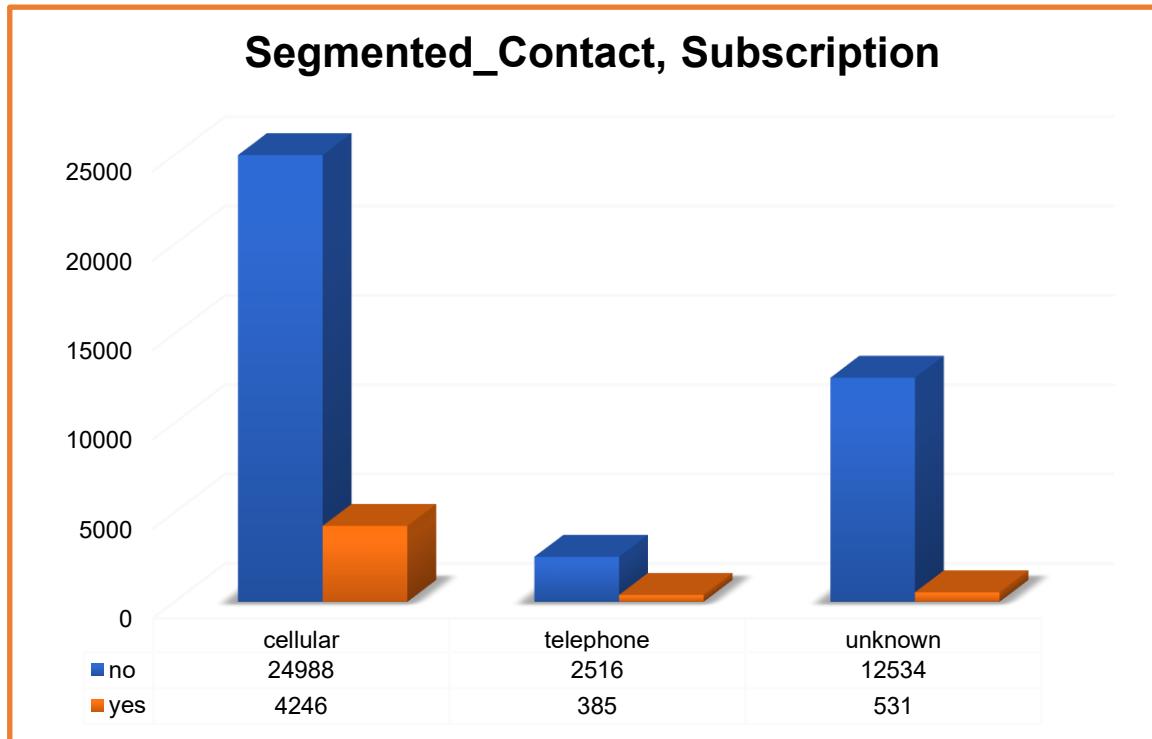
- Doctorate holders have the highest conversion rate (14.7%)
- "Unknown" education group also shows high conversion (13.2%)
- Bachelor's degree holders show the lowest conversion rate (8.5%): Possibly younger or less financially stable compared to those with advanced degrees.
- Masters are near the average with 10.3% conversion.



PART I : Univariate Analysis

Variable under consideration: 4. CONTACT [SEGMENTED UNIVARIATE with subscription]

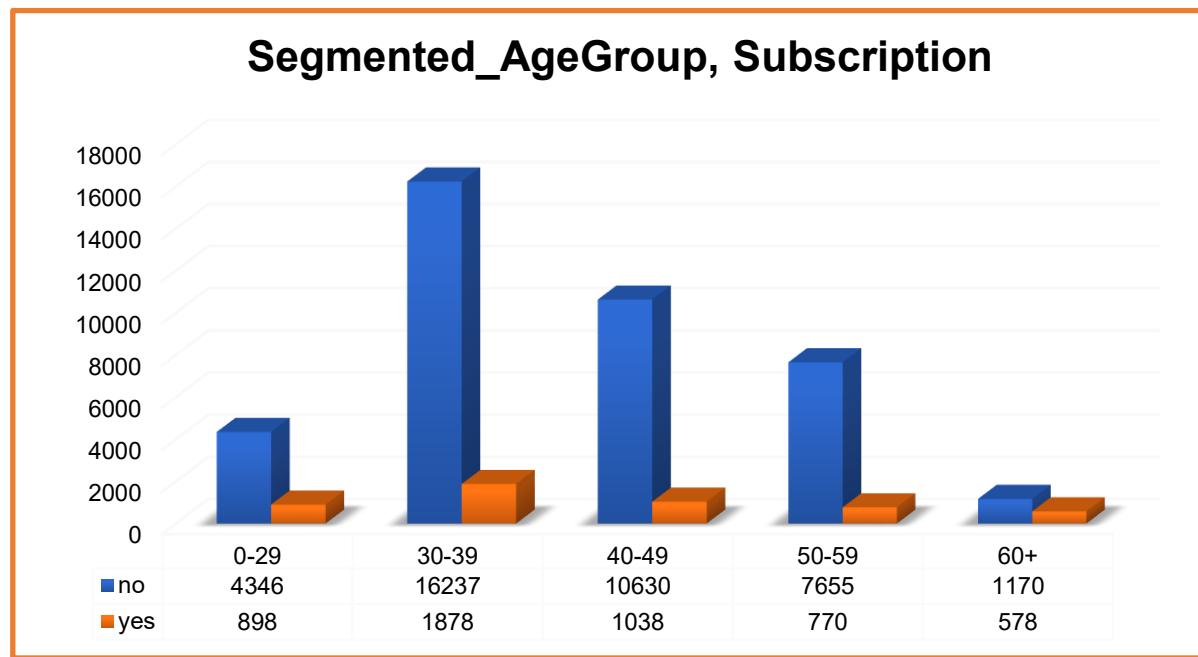
- Cellular contact has the highest success rate (14.5%): Likely due to better availability.
- Telephone performs moderately (13.3%).
- Unknown contact method has the lowest conversion rate (4.1%)



PART I : Univariate Analysis

Variable under consideration: 5. AGE [SEGMENTED UNIVARIATE with subscription]

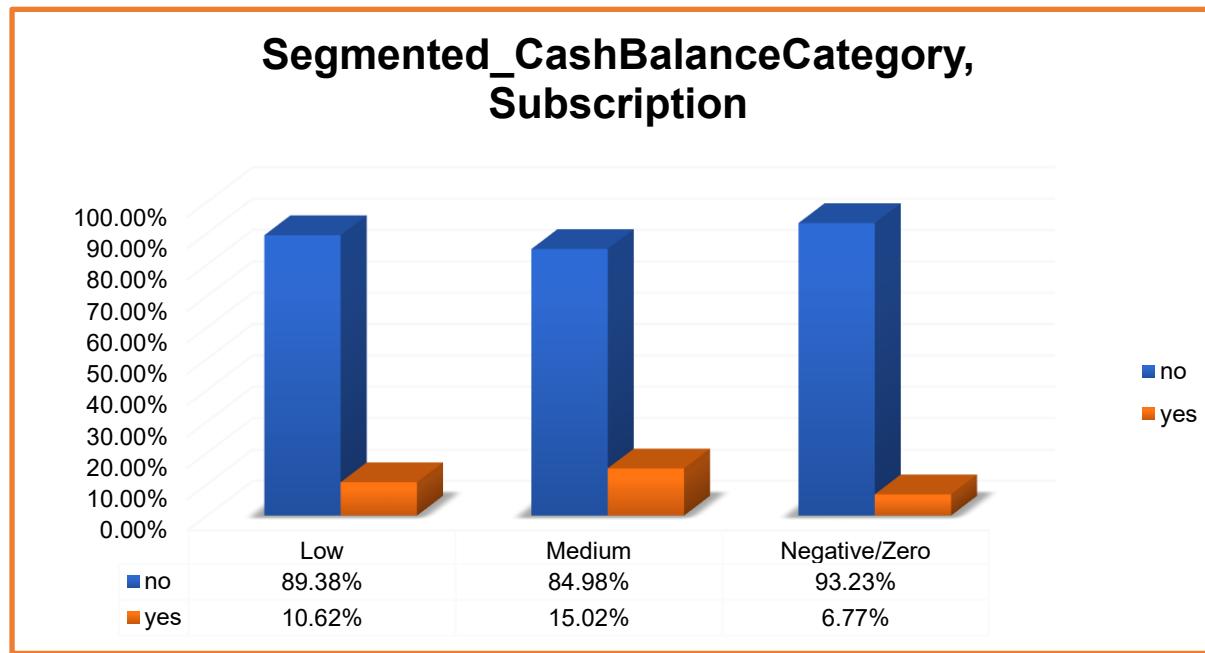
- Customers aged 60+ have the highest subscription rate (33.1%): Possibly because older customers are more interested in secure, long-term investments.
- Young customers (0-29) also show a strong interest (17.1%): This group may be early planners.
- Middle age groups (30-49) show below-average conversion (8.9–10.4%): Possibly due to more financial responsibilities (loans, children).
- The largest age group (30-39) shows only 10.4% conversion despite being 40% of the customer base



PART I : Univariate Analysis

Variable under consideration: 6. CASH BALANCE CATEGORY [SEGMENTED UNIVARIATE with subscription]

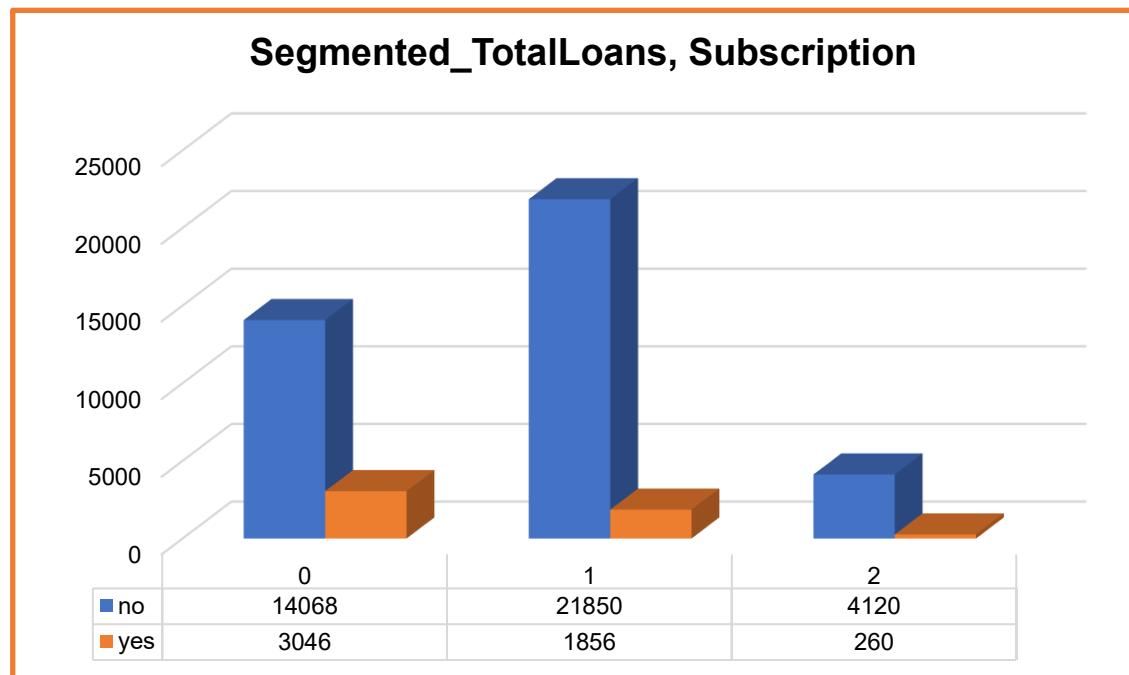
- Medium cash balance customers have the highest subscription rate (15.0%): Likely represent customers with moderate liquidity.
- Low balance customers convert at 10.6%, slightly below average.
- Customers with zero/negative balances are least likely to subscribe (6.8%): Indicates financial instability or lack of disposable funds.



PART I : Univariate Analysis

Variable under consideration: 7. TOTAL LOANS [SEGMENTED UNIVARIATE with subscription]

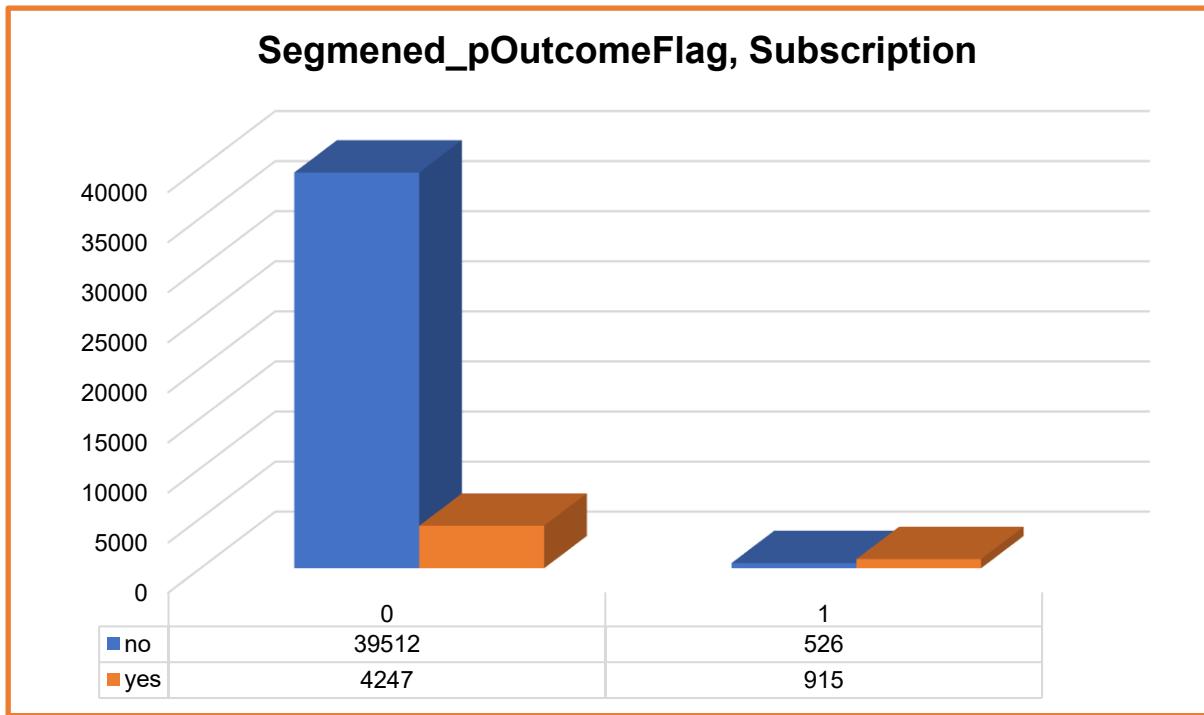
- Customers with no existing loans (Total Loans = 0) have the highest subscription rate (17.8%): They might have less financial burden, making them more likely to invest in term deposits.
- Customers with 1 loan show a much lower subscription rate (7.8%): Existing financial commitments may impact their choice.
- Those with 2 loans convert the least (5.9%): High financial commitments likely reduce their ability or willingness to invest.



PART I : Univariate Analysis

Variable under consideration: 8. pOUTCOME FLAG [SEGMENTED UNIVARIATE with subscription]

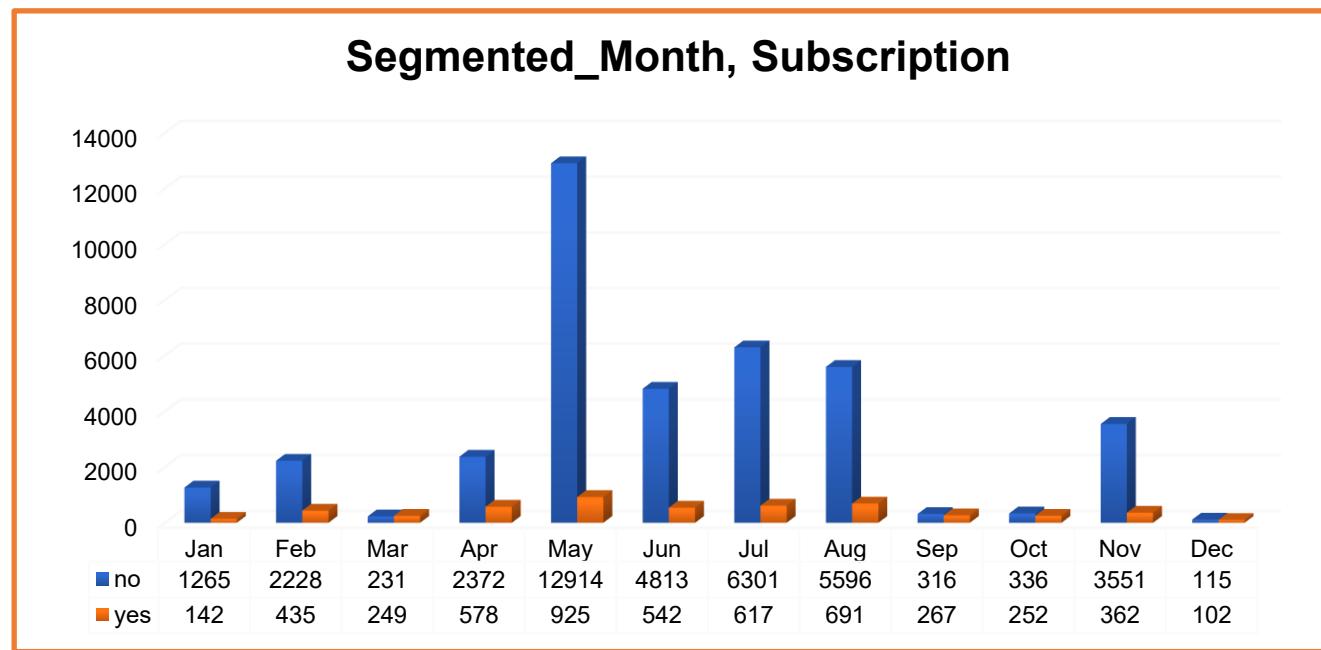
- Customers with a previous successful campaign outcome (`poutcome_flag = 1`) have a significantly higher subscription rate (63.5%):
- Customers without a previous successful contact (`poutcome_flag = 0`) have a much lower subscription rate (9.7%):



PART I : Univariate Analysis

Variable under consideration: 9. Month of Contact [SEGMENTED UNIVARIATE with subscription]

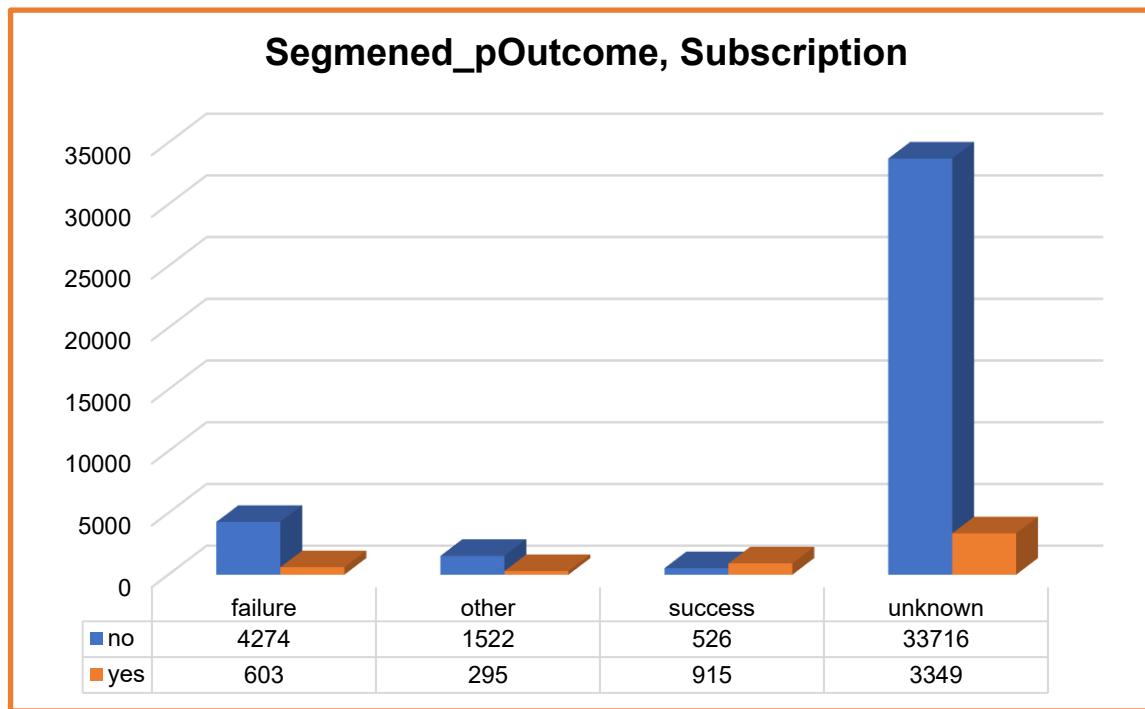
- March, September, October, and December show notably higher subscription rates (~43% to 52%) compared to other months.
- May has a large volume of contacts but a relatively low conversion rate (6.7%), suggesting less effectiveness or possibly less receptive customers during this period.
- Other months like February and April show moderate conversion rates around 16-20%.



PART I : Univariate Analysis

Variable under consideration: 10. pOutcome [SEGMENTED UNIVARIATE with subscription]

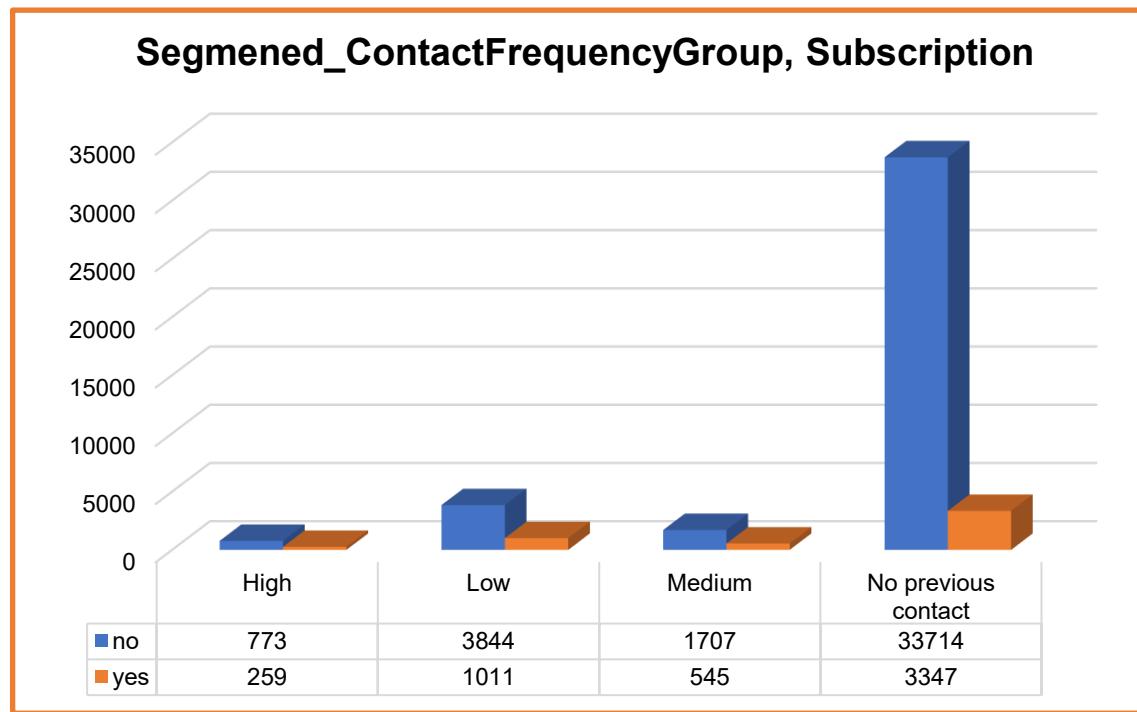
- Customers with previous successful contact (success) have a significantly higher subscription rate (~63.5%), which denotes strong influence from past positive outcomes.
- Those with 'other' previous outcomes have moderate conversion (~16.2%), while 'failure' outcomes convert at around 12.4%.
- The majority of data is unknown with a lower conversion rate (~9%), highlighting the importance of tracking previous contact outcomes.



PART I : Univariate Analysis

Variable under consideration: 11. Contact Frequency [SEGMENTED UNIVARIATE with subscription]

- Customers with high contact frequency have the highest subscription rate (~25.1%), suggesting that more frequent contacts increase the chance of conversion.
- Medium and low contact frequency groups also have relatively good subscription rates (~24.2% and ~20.8% respectively).
- Customers with no previous contact have a much lower conversion rate (~9%), which means prior engagement is critical.

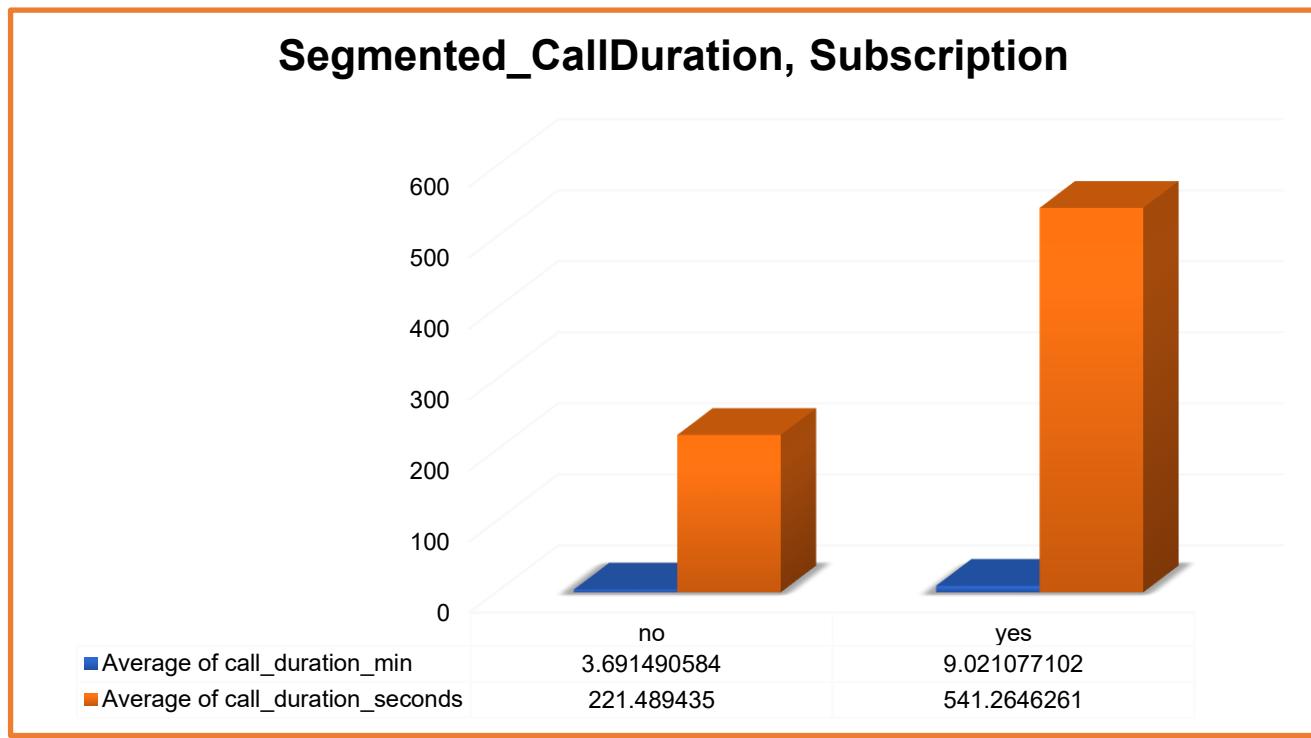


PART I : Univariate Analysis

Variable under consideration: 12. Call Duration[SEGMENTED UNIVARIATE with subscription]

- Customers with high contact frequency have the highest subscription rate (~25.1%), suggesting that more frequent contacts increase the chance of conversion.
- Medium and low contact frequency groups also have relatively good subscription rates (~24.2% and ~20.8% respectively).
- Customers with no previous contact have a much lower conversion rate (~9%), which means prior engagement is critical.

Customers who subscribed stayed on the call for about 9 minutes (541 seconds) on average. Customers who did not subscribe stayed for about 3.7 minutes (221 seconds) on average. This is a significant difference — subscribers talked more than twice as long.

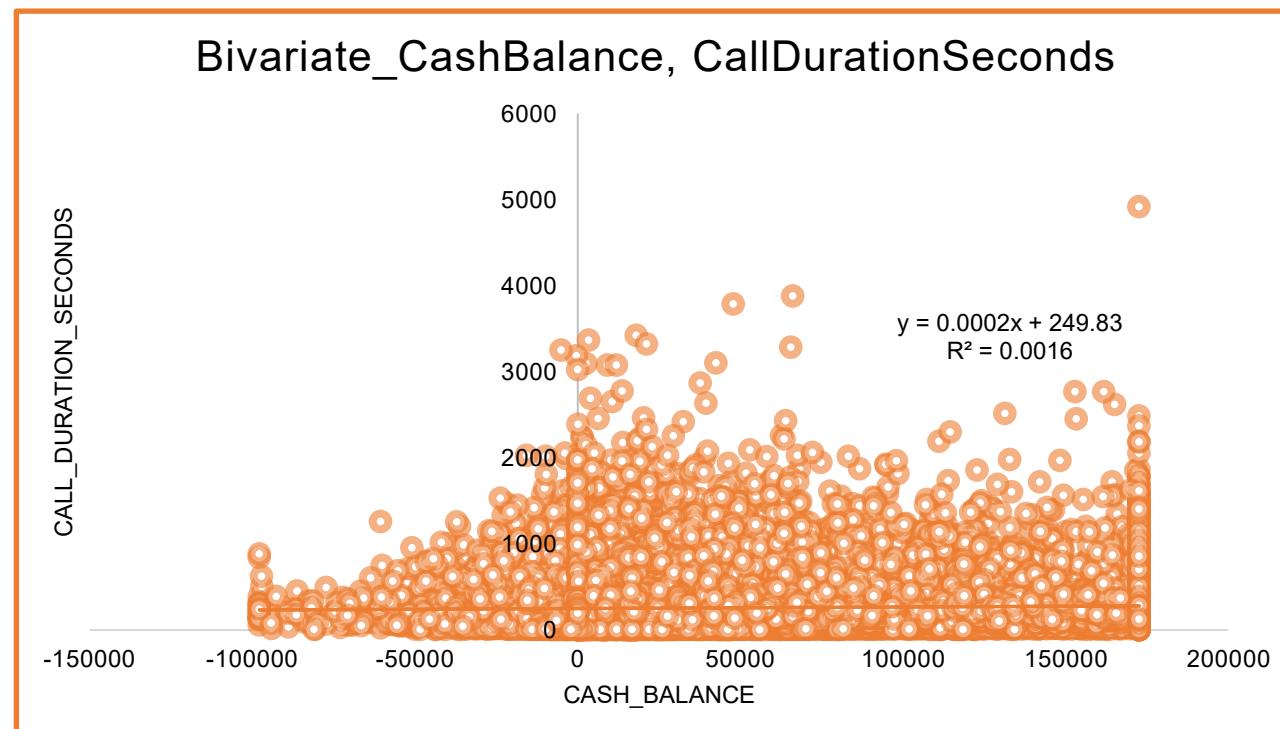


PART II : Bivariate Analysis

Variables under consideration: 1. Cash Balance and Call Duration in Seconds

- The correlation between cash balance and call duration is very weak and positive (around 4%), indicating almost no linear relationship.
- The very low R^2 value (0.0016) confirms that cash balance explains less than 0.2% of the variance in call duration.
- In practical terms, cash balance and call duration are almost independent in this dataset.

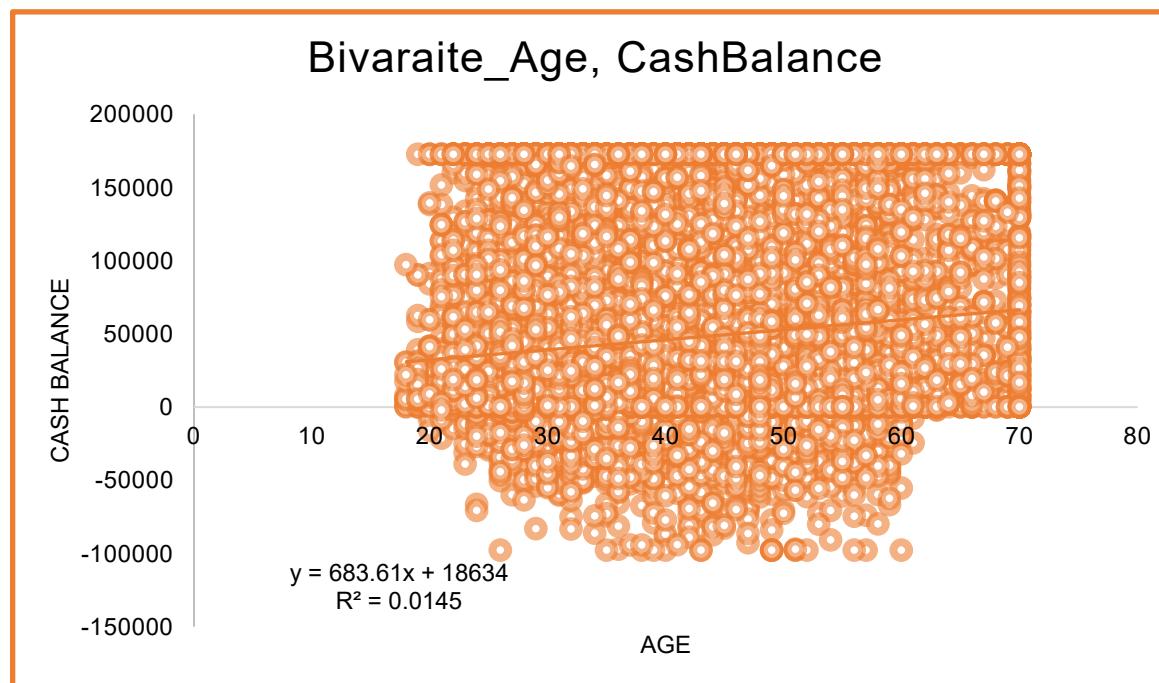
**R^2 is low:
No linear relationship**



PART II : Bivariate Analysis

Variables under consideration: 2. Age and Cash Balance

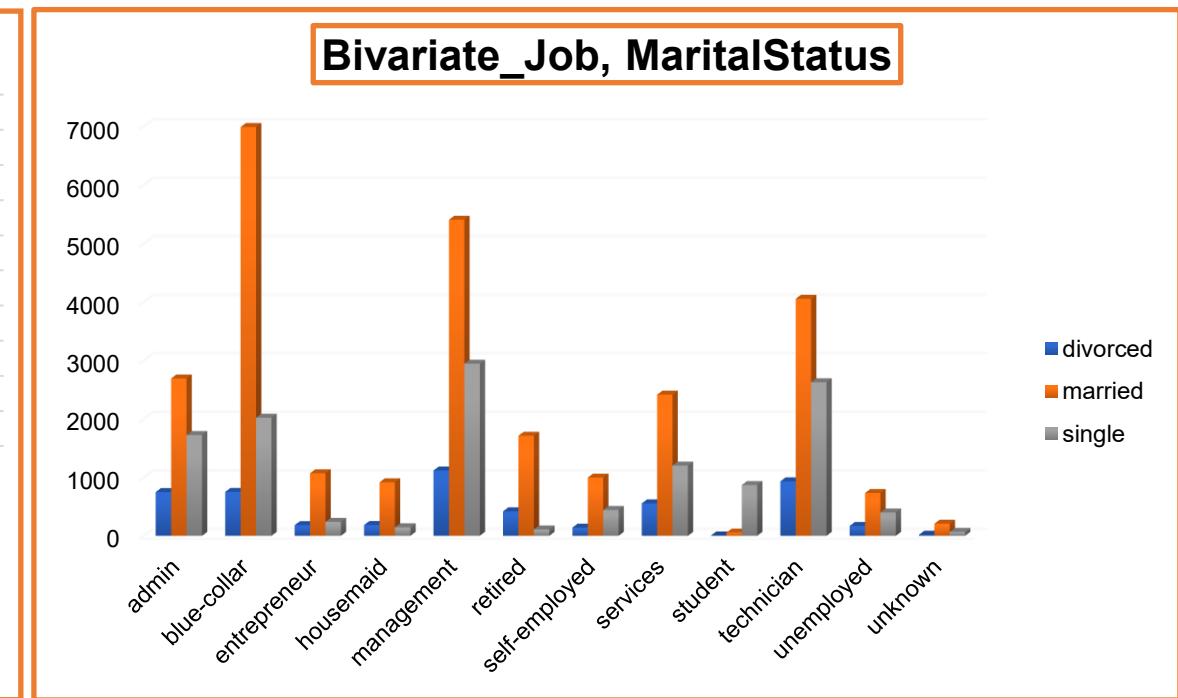
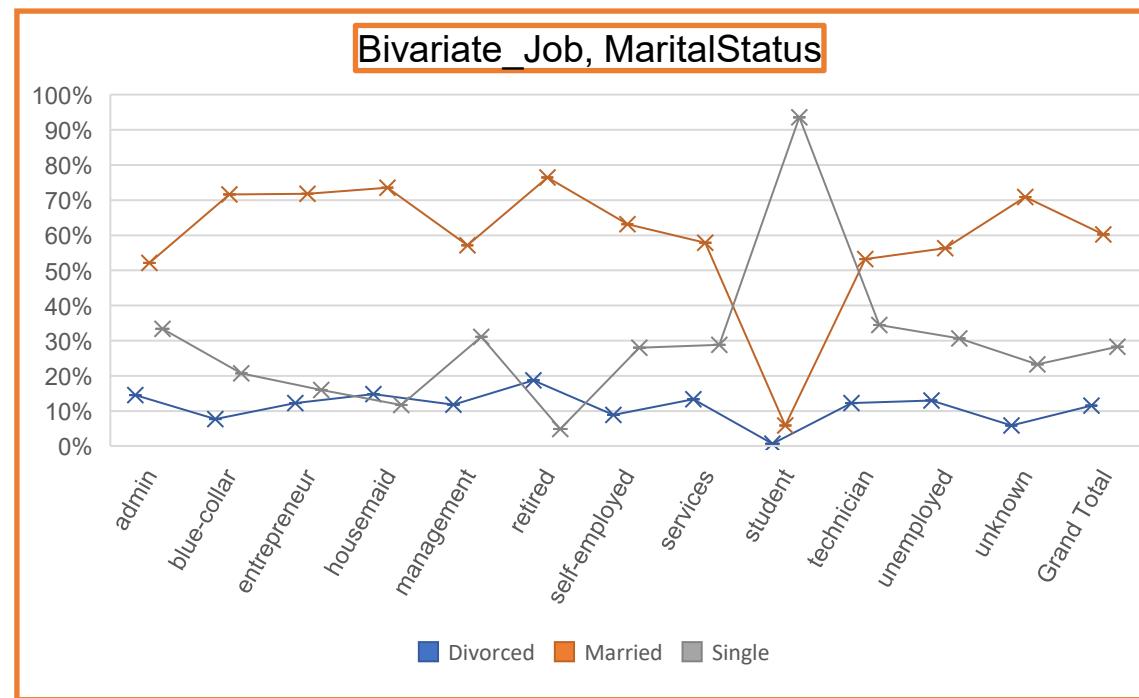
- The correlation between age and cash balance is weak but positive (~12%), indicating a slight tendency for cash balance to increase as age increases.
- The R² value (0.0145) means that age explains only about 1.45% of the variation in cash balance.



PART II : Bivariate Analysis

Variables under consideration: 3. Job, Marital Status

- Most jobs are in a majority married group, e.g., Blue-collar (71.6%), Entrepreneur (71.8%), and Retired (76.4%).
- Students are mostly single (93.6%), which is typical.
- Retired customers mostly married (76.4%), very few single (4.9%).
- Management and Admin have more singles compared to other jobs (31.1% and 33.4% respectively), but married still lead.
- Jobs like Technician and Services have a relatively higher share of singles (~28-35%).

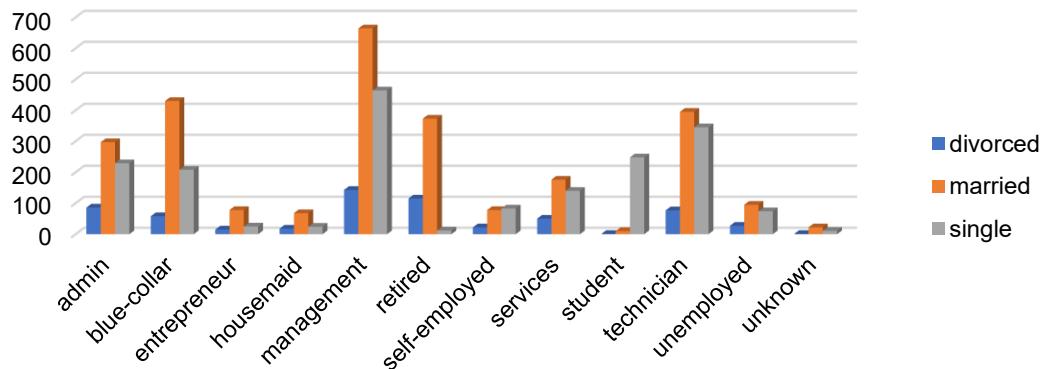


PART II : Bivariate Analysis

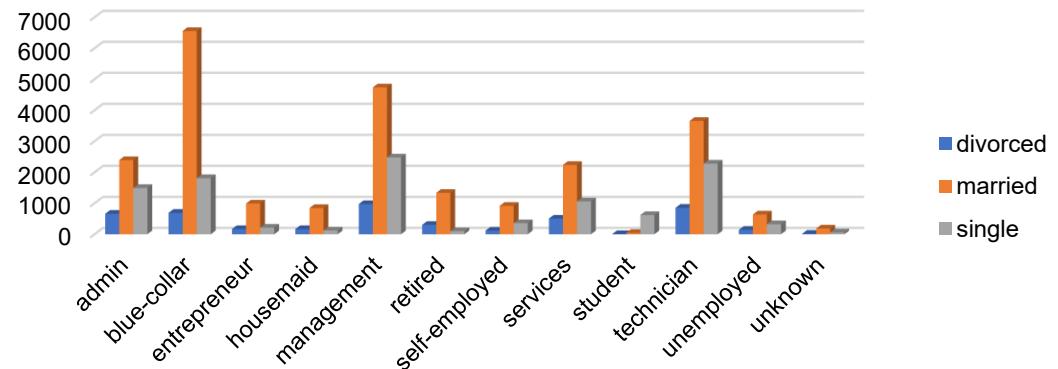
Variables under consideration: 4. Job, Marital Status [Subscription as Filter]

- Single customers generally have a higher proportion of "yes" subscriptions compared to "no" — especially in technical, blue-collar, and self-employed roles.
- Married customers dominate non-subscribers in all jobs, showing a potential saturation or resistance.
- The largest shift from no to yes appears for:
 - Blue-collar singles (+9.9%)
 - Technician singles (+8.6%)
 - Management singles (+6.2%)

Bivariate_Job, Marital Status
with
Subscription as YES



Bivariate_Job, Marital Status
with
Subscription as NO

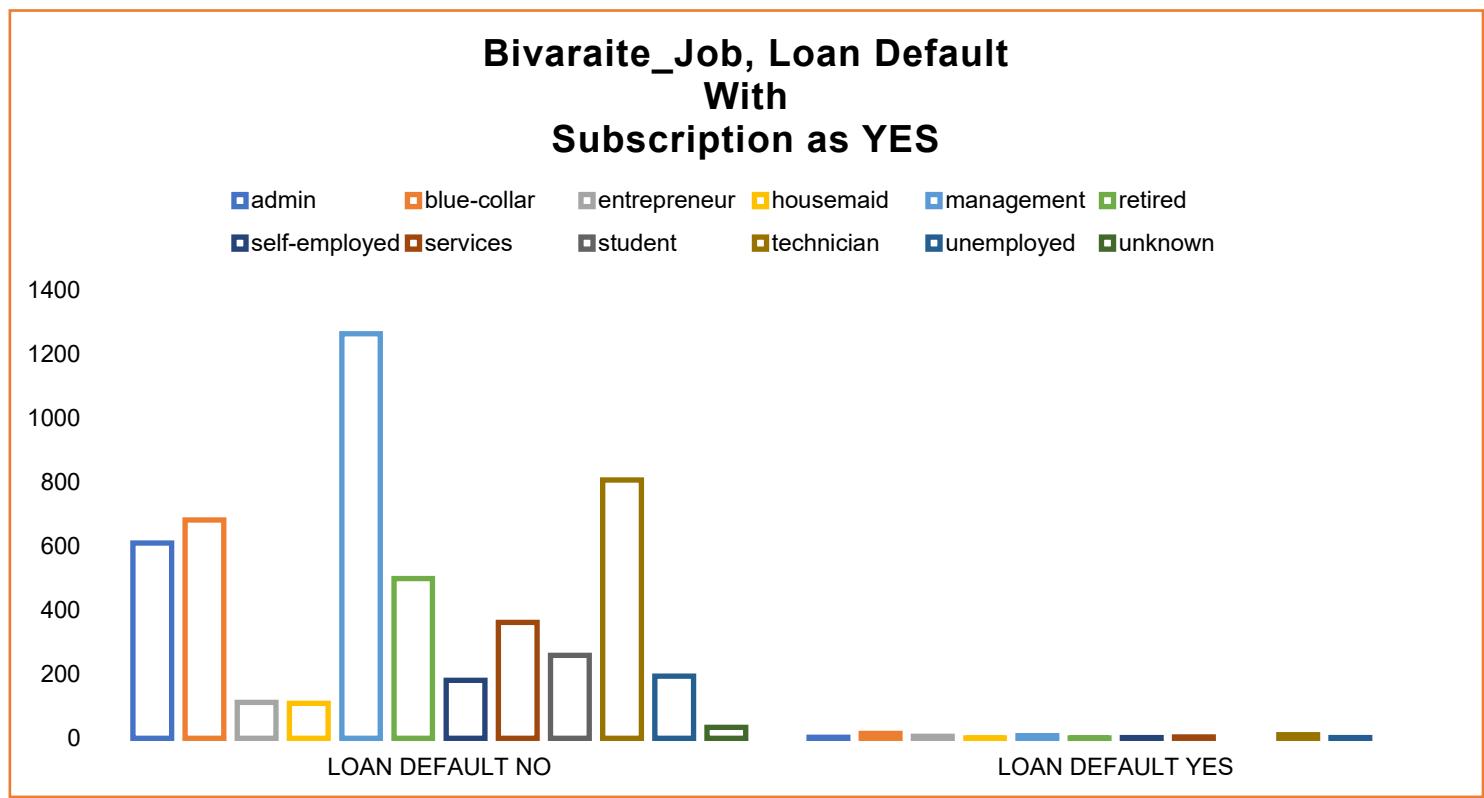


PART II : Bivariate Analysis

Variables under consideration: 5. Job, Loan Default [Subscription as Filter]

- Entrepreneurs stand out: Highest default rate (5.1%) even among subscribers.
- Most other jobs have low loan default among subscribers (<1.5%).
- Students and Unknowns have 0% loan default among subscribers.

JOB	LOAN DEFAULT NO	LOAN DEFAULT YES
admin	609	3
blue-collar	681	15
entrepreneur	112	6
housemaid	109	1
management	1263	8
retired	499	1
self-employed	181	2
services	362	4
student	259	
technician	806	11
unemployed	194	2
unknown	34	

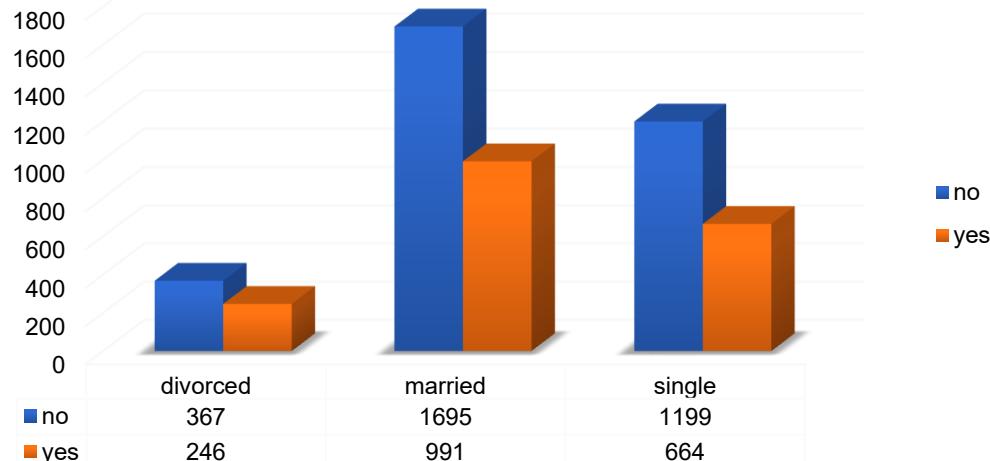


PART II : Bivariate Analysis

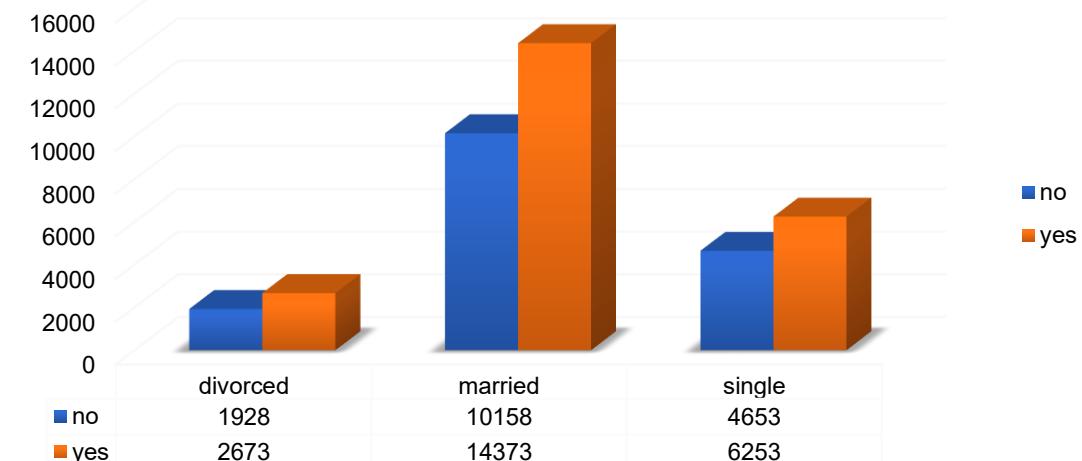
Variables under consideration: 6. Marital Status, Housing Loan [Subscription as Filter]

- Housing loan prevalence is significantly lower among subscribers (36.8%) than non-subscribers (58.2%).
- Across all marital statuses:
 - The housing loan percentage drops by ~20 percentage points when comparing subscribers to non-subscribers.
 - Divorced subscribers have the highest housing loan rate (40.1%), but still far lower than the 58.1% among divorced non-subscribers

Bivariate_MaritalStatus, Housing Loan with Subscription as YES



Bivariate_MaritalStatus, Housing Loan with Subscription as NO

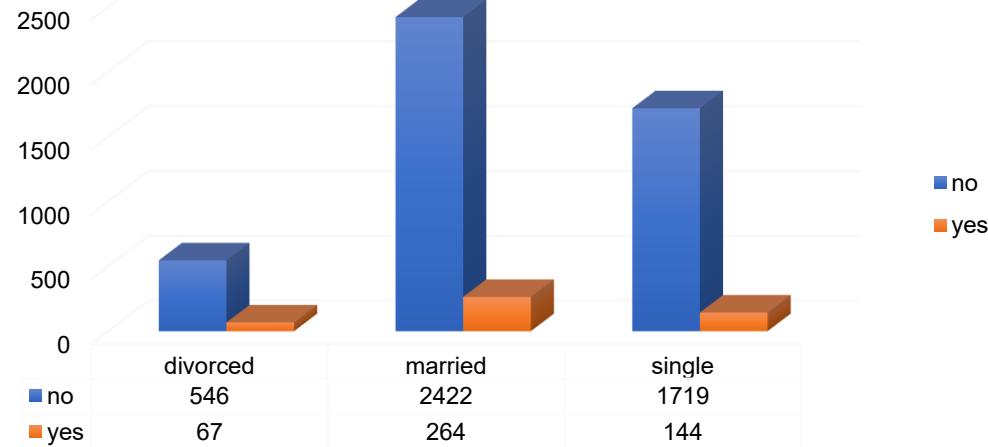


PART II : Bivariate Analysis

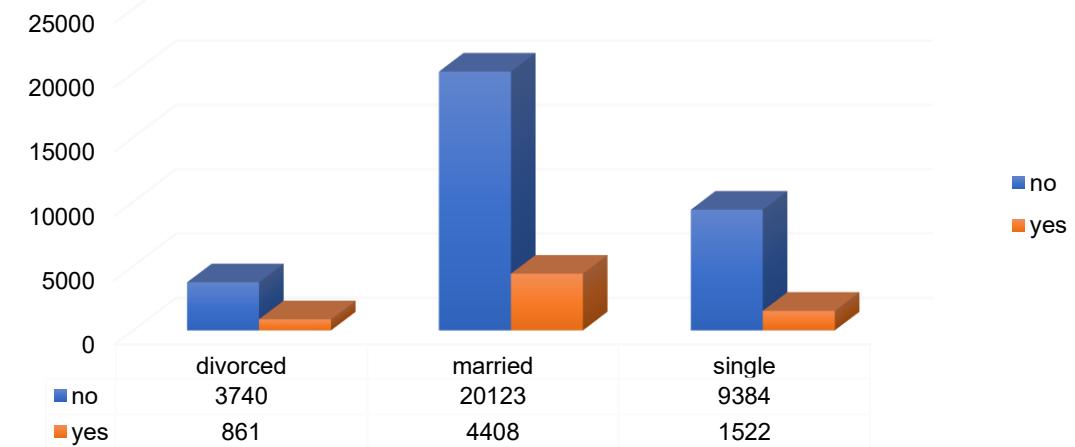
Variables under consideration: 7. Marital Status, Personal Loan [Subscription as Filter]

- Likelihood: Subscribers are less likely to have personal loans (only ~9.2% have one) compared to non-subscribers (17.0% have one).
- Among all marital groups, singles show the lowest personal loan uptake—both for subscribers (7.7%) and non-subscribers (14%).

Bivariate_MaritalStatus, Personal Loan with Subscription as YES



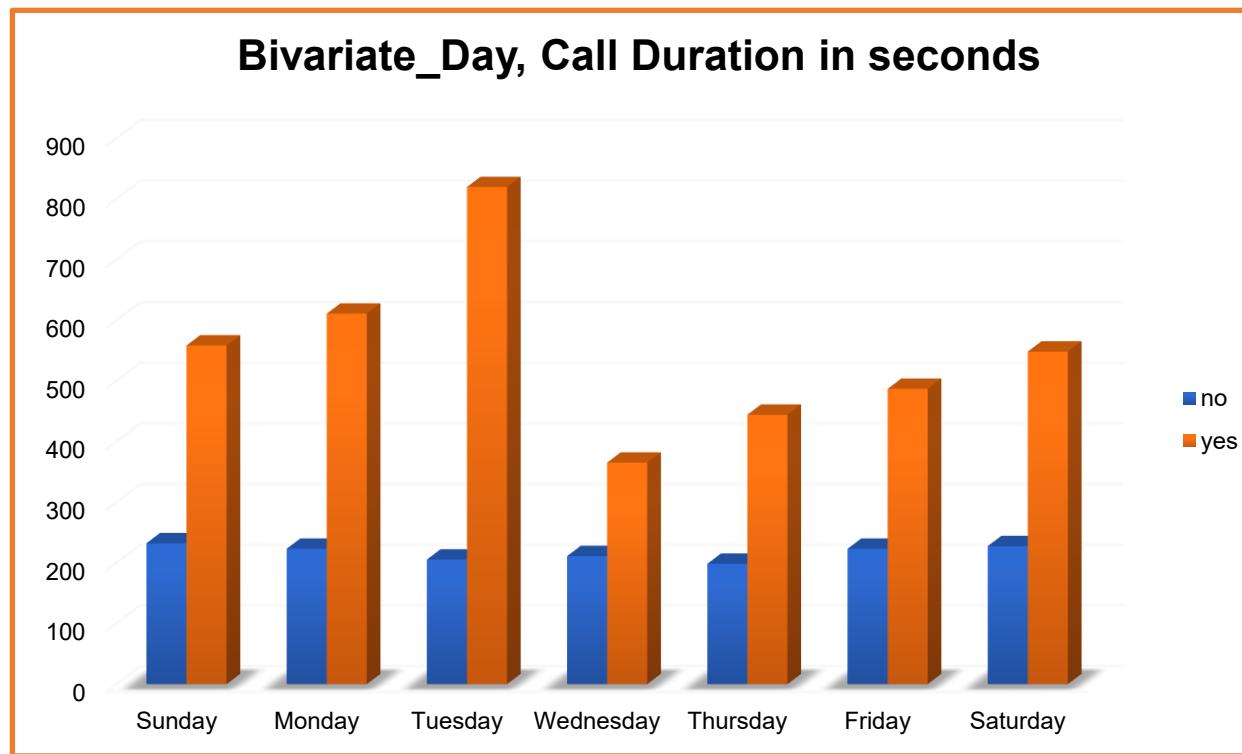
Bivariate_MaritalStatus, Personal Loan with Subscription as NO



PART II : Bivariate Analysis

Variables under consideration: 8. Day, Call Duration in Seconds [Subscription as Filter]

- Subscribers are consistently found to have longer call durations across all days of the week.
- The largest difference is observed on Tuesdays, where subscribers' calls are nearly 4 times longer than non-subscribers (819s vs 205s).
- Shortest calls for subscribers are on Wednesdays, but still, they are longer than non-subscribers.



PART III: Major insights

- **Month Matters :** Subscriptions peak in spring/early summer (Mar–May). However, winter (Dec–Jan) shows lowest conversions. Conversion fluctuates with season.
- **Focus marketing during high-performing months.**
- **Past Outcome signifies Future Likelihood:** “Success” in past means >50% subscription now. “Failure” or “Other” outcomes denotes poor performance. “Non-contacted” indicates untapped but colder segment.
- **Retarget “success” leads; reframe pitch for others.**
- **Loan Load Hurts Subscription:** Only 36.8% of subscribers have housing loans vs 58.2% non-subscribers. Personal loans also lower in subscribers (~9% vs 17%). Singles take fewer loans and subscribe more.
- **Push to low-debt customers; avoid high-loan segments.**
- **Loan Defaults mean Risk:** Highest defaults seen in Entrepreneurs (3.6%) and Unemployed (2.3%). On the contrary, lowest seen in Students (0.3%), Retired (1.2%) — also higher subscribers. Defaults negatively linked to subscription.
- **Filter high-risk job groups before campaign push.**
- **Call Duration signals Engagement:**
Subscribers talk longer, especially Tuesdays (4× longer). Longer calls happen early in the week.
 - **Prioritize long calls and early-week outreach.**

Variable	Influence on Subscription	Actionability	Targeting Usefulness
Poutcome	High	High	High
Call Duration	High	High	High
Month	High	Medium	High
Housing Loan	Medium	High	High
Personal Loan	Medium	Medium	Medium
Loan Default	Medium	Low	Medium

PART III: Major insights

- **Segment by job + marital + loan for better ROI:** Married customers dominate both yes/no subscriptions. Single students, technicians, self-employed convert well (e.g. Students: 95.9%). Retired customers show high subscription and low default — ideal low-risk, high-value segment.
- **Target students (flexibility), retired (stability), technicians/self-employed;**
- **Loan holders vs Job:** Loan holders (housing/personal) are less likely to subscribe across all marital groups. Divorced customers and loan holders present a lower ROI due to low conversions and higher risk.
- **Skip risky combos (loan + high-default jobs)**
- **Cash Balance vs Call Duration:** Extremely weak correlation ($r \approx +0.04$, $R^2 = 0.0016$) found. Essentially independent. Cash Balance doesn't affect engagement. It's not a driver for call duration or conversions.
- **Age vs Cash Balance:** Weak but positive correlation ($r \approx +0.12$, $R^2 = 0.0145$). Older customers tend to have higher balances. Age alone isn't predictive, but older customers have slightly higher financial value. Use age + other variables (like job, marital status) to segment high-balance, low-risk customers.

Job	Subscription Rate (%)	Default Rate (%)
Student	95.9	0.3
Retired	74.6	1.2
Technician	42.2	1.8
Self-employed	45.4	2
Management	38.2	1.5
Blue-collar	27.8	1.7
Entrepreneur	28.2	5.1
Unemployed	20.5	2.3

PART III: Major insights

Driver	What Works	What Doesn't	Action
Job + Marital	Single Students (96%), Retired (married), Technicians & Self-employed (single)	Entrepreneurs, Divorced + Loans	Target high-converting, low-risk combos
Loan Status	No housing/personal loan = higher conversion	Loan holders convert less (esp. divorced)	✗ Filter out or de-prioritize loan holders
Past Outcome	"Success" = >50% likely to subscribe	"Failure" has lowest conversion	Focus on warm leads from past campaigns
Call Timing	Early-week (Mon/Tue) = longer calls → higher success	Short calls (<3 mins), late-week less effective	Schedule key calls early in the week
Month (Seasonality)	March–June → high conversions	Dec–Jan → low performance	Boost budget/calls in peak months
Default Risk	Retired, Students = low risk	Entrepreneurs, Unemployed = high default	🚫 Avoid high-risk job groups
Age & Cash Balance	Older customers → slightly higher cash	Weak predictors individually	Combine with job + marital for targeting
Cash vs Call Duration	No real link → cash ≠ call length	Irrelevant for engagement	✗ Don't use for campaign planning