Improving Term Deposit Subscription Rates through Data-Driven Customer Segmentation and Campaign Optimization

Final report for the BDM capstone Project

Submitted by

Name: Saloni Singh

Roll number: 24DS1000028

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai Tamil Nadu, India, 600036

Table of Contents

List of tables

List of Figures

1. Executive Summary

This project focuses on optimizing term deposit subscription rates for a Portuguese bank using secondary data from a direct marketing campaign. The core objective is to identify actionable customer segments, campaign timings, and strategy enhancements that can improve conversion outcomes through targeted telemarketing.

The dataset used in this project is a subset of the original Bank Marketing Dataset published by the UCI Machine Learning Repository. This particular version (a balanced subset) was obtained from Kaggle, and has 11,162 records and 17 variables such as age, job, education, balance, contact type, call duration, and past campaign outcomes. Key pre-processing steps included label and one-hot encoding, feature engineering (e.g., age groups, call duration buckets), and exploratory visualizations. Methodologies applied spanned descriptive analysis, decision tree classification, logistic regression, and rule-based segmentation.

The analysis revealed several high-performing segments: senior citizens (60+), students, highly educated individuals, and customers with previous positive campaign outcomes. Long call durations and cellular contact types were also linked to high success rates. Conversely, the 30–44 age group, though the largest, showed underperformance due to financial liabilities. Timing analysis identified March, September, October, and December as high-conversion months.

Based on these insights, strategic recommendations include personalized messaging for key segments, adaptive plans for middle-aged customers, data quality improvements, and feedback-based campaign refinement. Seasonality-aware scheduling and KPI-based (Key Performance Indicators) monitoring were proposed to support ongoing optimization. Together, these approaches offer a practical, data-driven roadmap for improving telemarketing ROI and long-term customer engagement.

2. Proof of Originality

This project is based on secondary data. No primary data collection was involved. This dataset is publicly available and used here strictly for academic purposes.
Original Dataset Source: UCI Machine Learning Repository – Bank Marketing Dataset
Original Dataset Link: https://archive.ics.uci.edu/ml/datasets/bank+marketing
Balanced Dataset obtained from Kaggle: Balanced Dataset Link
Analysis Link: Colab Notebook Link
Dataset Description: The dataset contains information collected from a direct marketing campaign of a Portuguese banking institution. The goal was to predict whether a client would subscribe to a term

deposit (deposit = yes/no). It includes 17 variables as described in the next section.

3. Metadata and Descriptive Statistics

Table 1: Metadata

Variable	**Description**	**Data Type**	**Range / Values**
age	Age of the customer	Numeric	18 to 95
job	Type of job	Categorical	‘admin.’, ‘unknown’, ‘unemployed’, ‘management’, ‘retired’, ‘housemaid’, ‘entrepreneur’, ‘student’, ‘blue-collar’, ‘self-employed’, ‘technician’, ‘services’
marital	Marital status	Categorical	‘married’, ‘divorced’, ‘single’
education	Education level	Categorical	‘unknown’, ‘secondary’, ‘primary’, ‘tertiary’
default	Has credit in default?	Binary	‘yes’, ‘no’
balance	Average yearly balance	Numeric	-6847 to 81204 (in Euros)
housing	Has housing loan?	Binary	‘yes’, ‘no’
loan	Has personal loan?	Binary	‘yes’, ‘no’
contact	Communication type	Categorical	‘unknown’, ‘telephone’, ‘cellular’
day	Last contact day	Numeric	1 to 31
month	Last contact month	Categorical	‘jan’, ‘feb’, ..., ‘dec’
duration	Last contact duration	Numeric	2 to 3881 (in seconds)
campaign	Number of contacts during this campaign	Numeric	1 to 63
pdays	Days since last contact	Numeric	-1 to 854 (-1= never contacted)
previous	Number of contacts before this campaign	Numeric	0 to 58
poutcome	Outcome of previous marketing campaign	Categorical	‘unknown’, ‘other’, ‘failure’, ‘success’
deposit	**Target variable**: Subscribed to deposit?	Binary	‘yes’, ‘no’

Descriptive Statistics

The table below summarizes key descriptive statistics for the major numerical variables in the dataset. These statistics provide insights into customer demographics, financial characteristics, and behavioral patterns that can inform marketing strategies and campaign design.

Table 2: Descriptive Statistics

Feature	Mean	Median	Std Dev	Min	Max
Age	41.2	39.0	11.9	18.0	95.0
Balance (€)	1528.5	550.0	3225.4	-6847.0	81204.0
Duration (s)	372.0	255.0	347.1	2.0	3881.0
Campaign	2.5	2.0	2.7	1.0	63.0
Pdays	51.3	-1.0	108.8	-1.0	854.0
Previous	0.8	0.0	2.3	0.0	58.0

These descriptive statistics provide valuable insights into the customer base and engagement history. The average customer is around 41 years old, with most falling within a relatively narrow age range, this indicates a stable middle-aged demographic that can be effectively targeted with tailored financial products. The balance variable shows a highly skewed distribution, with a few clients holding significantly higher balances. This reflects a large variation in financial capacity, reinforcing the need for segmentation in marketing efforts.

Call duration displays a wide spread, with the average call lasting 372 seconds but a median of only 255 seconds. This skew suggests that while most calls are brief, a small number are much longer. These longer durations may be relevant in further stages of analysis. The number of campaign contacts also appears tightly clustered, with most clients contacted just 2–3 times, indicating a generally consistent outreach approach.

4. Analytic Approach and Methodology

4.1 Data Cleaning and Preprocessing

1. Balanced Dataset Selection: The original UCI dataset was highly imbalanced, with only ~11% subscribing to a term deposit, making most patterns biased toward the majority class. To enable fair and reliable analysis, a **rebalanced version** sourced from Kaggle was used. This ensured clearer contrasts between subscriber and non-subscriber segments and improved the quality of insights.

2. Handling "Unknown" Values: The dataset was well-structured with no missing (NaN) values. However, several categorical fields such as job, education, contact, and poutcome included the placeholder "unknown," which represents incomplete information. Instead of removing rows with "unknown" entries, these were retained as a valid category to avoid losing data and analysed as potential indicators of customer behaviour.

3. Data Type Conversion for Regression Analysis:

- Categorical variables like marital, housing, loan, default, and poutcome were encoded using Label Encoding for binary fields and One-Hot Encoding for nominal variables with multiple categories (e.g., job, month).
- Numerical variables such as age, balance, duration, campaign, pdays, and previous were retained in their native formats for statistical analysis and model input.

These steps ensured the dataset remained comprehensive and analysis-ready, without compromising on data integrity or interpretability.

4.2 Telemarketing Campaign Process Mapping

A conceptual business process map was developed to represent a typical telemarketing campaign cycle, informed by the structure of the dataset and common practices in banking-based campaigns. Since the data is secondary and lacks explicit operational documentation, this mapping reflects how such campaigns are generally designed and executed in real-world banking institutions.

The campaign process includes the following key stages:

1. **Customer Profiling**: Identifying potential leads using demographic and financial attributes such as age, job, and account balance.
2. **Lead Qualification**: Filtering leads based on previous campaign outcomes, loan status, and contact history to prioritize high-conversion prospects.
3. **Initial Contact**: Executing outreach via telephone or cellular calls and logging variables such as call duration, channel, and date of contact.
4. **Response Evaluation**: Assessing customer engagement using indicators like duration and historical behaviour.
5. **Follow-up Strategy**: Planning re-attempts, escalation, or dropout based on prior response and overall campaign rules.
6. **Final Outcome**: Logging whether the customer subscribed (deposit = yes) or not (deposit = no) which serves as the key target variable for this project.

This process mapping helped us **contextualize dataset variables** within a real-world telemarketing pipeline and **define relevant KPIs** like success rate, contact-to-conversion ratio, and segment-wise performance. This conceptual framework ultimately guided the structure of the entire analysis and ensured that insights generated were not only data-driven but also operationally grounded.

4.3 Data Preparation and Transformation

Several new features were engineered to improve interpretability:

- **age_group:** grouped into 18–29, 30–44, 45–59, 60+ for clearer demographic analysis.
- **duration_bucket:** categorized calls as short (<180s), medium (180–360s), or long (>360s).
- **month_quarter:** grouped campaign months into Q1–Q4 to identify time-based trends. Created new variables like duration_bucket, month_quarter, age_group for easier comparison.

These transformations allowed for group-wise comparisons and helped capture nonlinear **trends and enable interpretable visualizations.**

4.4 Exploratory Data Analysis (EDA)

Methods Used:

- **Descriptive statistics** (mean, median, standard deviation, min, max) were computed for key numeric variables such as balance, duration, and age.
- **Visualizations** included **histograms** to explore the distribution of key continuous variables such as balance and duration. **Grouped bar charts** were used to compare subscription rates across categorical features like job, education, contact type, month, and loan status. **Bubble charts** illustrated temporal conversion patterns by job category and age group. A **line chart** was used in the focused analysis of the 30–59 age group to examine the relationship between financial liabilities (housing and personal loans) and subscription rates.

EDA helped build hypotheses and detect patterns that can guide processing, feature engineering, and model design.

4.5 Customer Segmentation

Customer segmentation was performed using rule-based logic applied across both demographic and behavioural attributes:

- **Demographic**: age_group, job, marital, and education
- **Behavioral**: poutcome, duration_bucket, contact, and campaign frequency

Unlike simple bucketing, segmentation combined multiple features to define practical customer profiles. While many patterns (e.g., poutcome = success, long-duration calls) were already evident through basic visualizations, segmentation was used to **structure and present** these patterns more effectively, aiding communication and decision-making.

This approach mirrors how businesses typically define personas or prioritize leads. These segments are **interpretable, actionable, and form the basis for targeting and KPI comparison**, supporting data-driven campaign refinement and performance tracking.

4.6 Linear Regression Modeling

A linear regression model was developed using scaled numeric and one-hot encoded categorical features. It helped estimate the contribution of each variable toward the likelihood of subscription. Regression quantifies feature influence and works well with transformed numeric data. This model supports strategic campaign planning by identifying which factors have the greatest predictive strength.

The model was used keeping in mind the opportunity for it to be used in future by bank teams to input customer data and generate predictions on subscription likelihood, potentially integrating it into existing CRM lead management systems to support smarter campaign targeting.

4.7 Decision Tree Classification

A decision tree classifier was used to identify how customer attributes interact to determine campaign outcomes. Its visual, rule-based structure made it well-suited for mapping decision paths and uncovering combinations of conditions that lead to subscription or non-subscription.

Decision trees capture non-linear relationships and are highly interpretable, making them useful for campaign managers to derive actionable rules (e.g., "If poutcome = success and duration is long, then high subscription likelihood"; "If contacted multiple times with no prior success, then low likelihood").

Used alongside linear regression, the decision tree provided **complementary insights**:

- Regression quantified **individual feature impact** in a generalized form
- The tree revealed **conditional logic** and **interactions between variables** in a more human-readable way

4.8 Performance Tracking and KPI Development

While formal dashboards or dynamic tracking tools were not developed as part of this project, a set of custom KPIs was conceptually defined to evaluate campaign effectiveness. These include:

1. **Call Success Rate**: Percentage of successful subscriptions from total contacts
   • *Formula:* (Number of customers who subscribed ÷ Total contacted customers) × 100
   • Measures overall campaign effectiveness.

2. **Temporal Conversion Trends**: Subscription rate tracked monthly or quarterly
   • *Formula:* (Deposits in time period ÷ Total contacts in time period) × 100
   • Helps evaluate seasonal performance and optimal outreach periods.
3. **Segment-Wise Conversion Rate**: Subscription rate within defined customer segments
   • *Formula:* (Deposits in segment ÷ Total contacts in segment) × 100
   • Identifies high-performing customer groups for targeted strategies.

These KPIs served as a reference during exploratory analysis and segmentation, helping guide which customer groups or time periods showed relatively higher success rates. In a live business setting, these metrics could be embedded into ongoing campaign monitoring for real-time feedback and optimization.

5. Results and Findings

5.1 Individual Feature Insights

This section analyses each customer attribute individually to understand its standalone influence on term deposit subscription.

Grouped bar charts are used for most categorical variables as they clearly compare subscription rates across categories while also showing the relative size of each group, enabling quick identification of high- and low-performing segments. Bubble charts are employed for variables where both distribution and relative size of customer groups are important (e.g., age distribution, temporal patterns), while histograms with KDE are used for continuous variables to show both frequency and distribution shape. This consistent approach ensures comparability across analyses while choosing the most effective visual form for each type of variable.

5.1.1 Job Type vs. Subscription Outcome

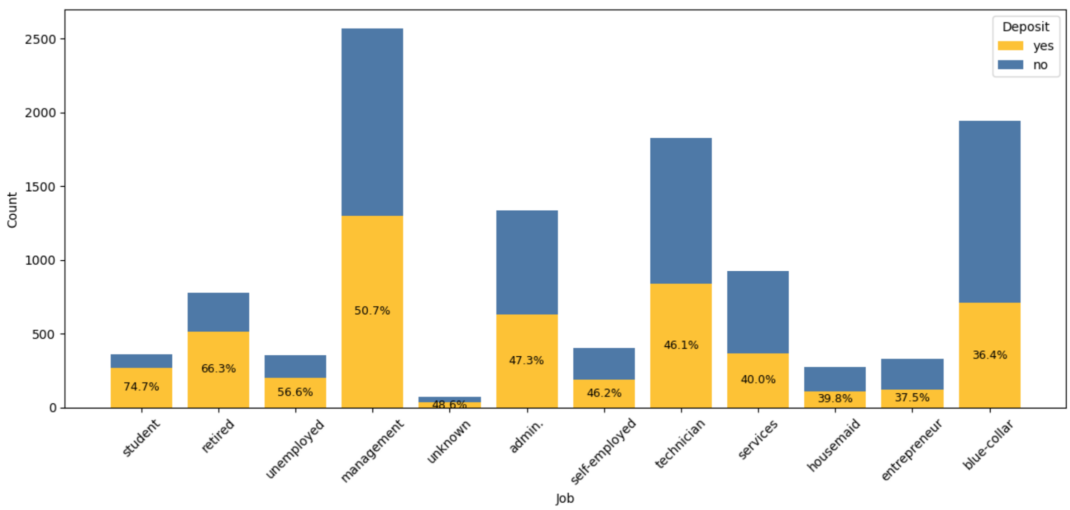

Fig.1 Job Categories vs. Subscription Outcome

A grouped bar chart was chosen as it effectively displays both the distribution of job categories and their corresponding subscription outcomes side by side.

Analysis of customer occupation reveals marked variation in subscription behavior. Customers categorized as **retired** and **students** exhibit the **highest subscription rates**, despite forming smaller portions of the total customer base. These groups are either in the post-retirement phase with stable savings or at the beginning of their financial journey and are therefore more inclined to commit to low-risk savings options.

Conversely, customers in **blue-collar**, **services**, **housemaid** and **entrepreneur** categories show **significantly lower subscription rates**. These segments may be more sensitive to liquidity needs or have limited disposable income. Mid-performing groups include **technicians**, **admin**, **management** and **self-employed individuals**.

5.1.2 Education Level vs. Subscription Outcome

Subscription likelihood increases with the customer's level of education. Those with **tertiary education** (i.e., university-level) display the **strongest conversion rates**, possibly due to greater financial literacy and long-term planning behavior. Customers with only **primary education** or whose education status is **unknown** show the lowest propensity to subscribe**.** Those with **secondary education** are moderate performers.

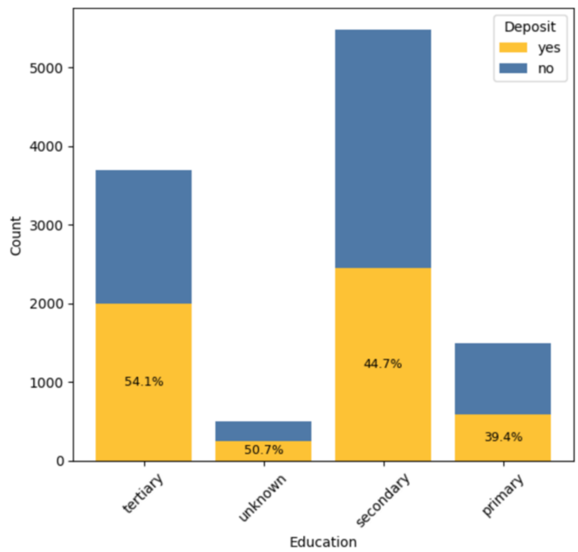

Fig.2 Education Level vs. Subscription Outcome

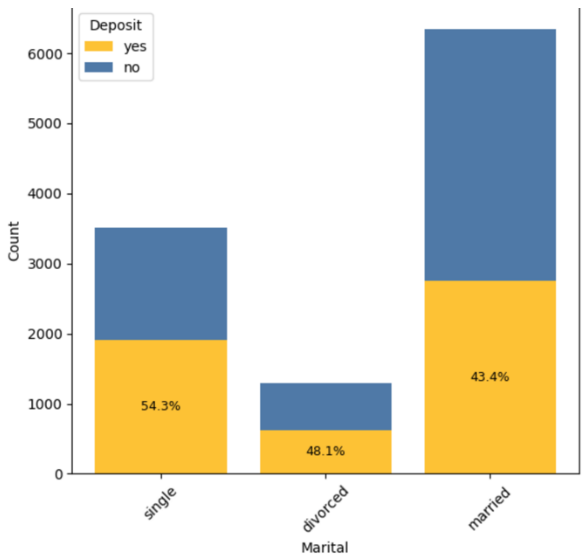

Fig.3 Marital Status vs. Subscription Outcome

5.1.3 Marital Status vs. Subscription Outcome

Subscription success differs across relationship statuses. **Single** customers are significantly more likely to subscribe than their **married** or **divorced** counterparts. While **married customers** form the largest population share, their lower conversion suggests higher financial responsibilities, risk aversion, or joint decision-making delays.

5.1.4 Contact Type vs. Subscription Outcome

The method of communication substantially affects campaign results. Contacts made via **cellular phones** yield **much higher subscription rates** than those made over **landlines**. This is likely due to mobile users being more reachable, responsive, and digitally engaged, leading to better engagement during the call.

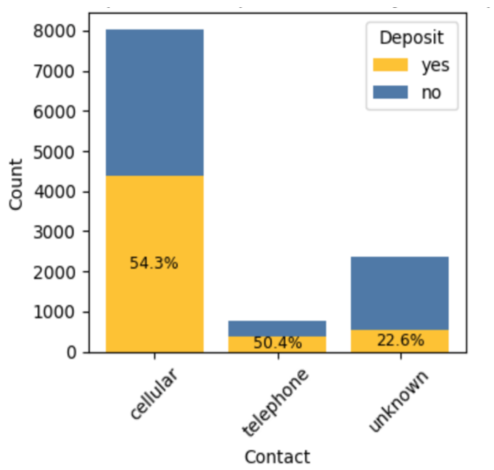

Fig. 4 Contact Type vs. Subscription Outcome

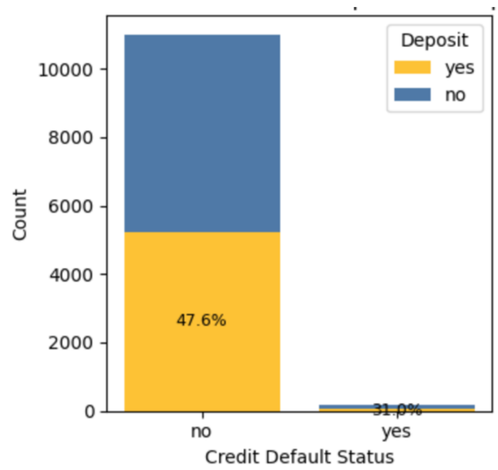

Fig. 5 Credit Default Status vs. Subscription Outcome

5.1.5 Loan / Housing / Credit Default vs. Subscription Outcome

Three financial risk indicators were analyzed: Customers with an existing **personal loan** or **housing loan** were less likely to subscribe. Customers marked as having a **credit default** history had the **lowest conversion rates** overall. These trends suggest that individuals already carrying financial burdens are cautious about locking funds in fixed-term products.

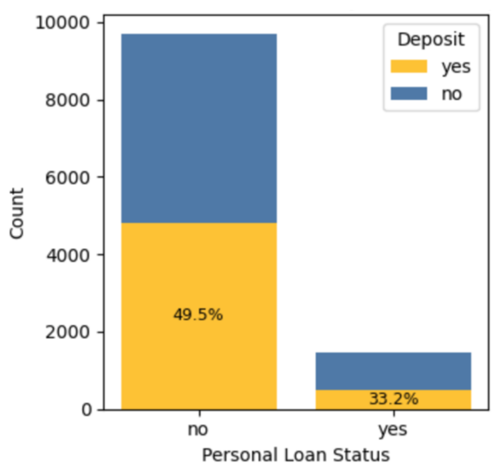

Fig. 6 Personal Loan vs. Subscription Outcome

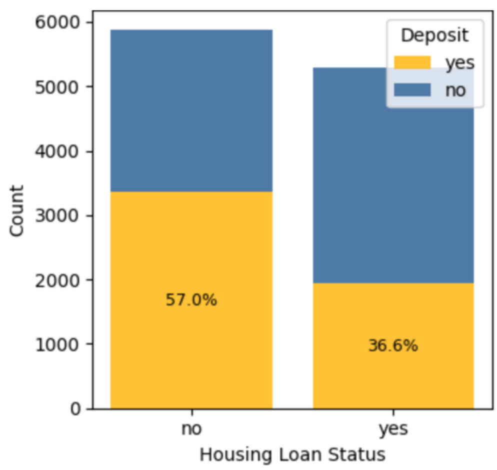

Fig. 7 Housing Loan Status vs. Subscription Outcome

5.1.6 Age Distribution vs. Subscription Outcome

This bubble chart used here is ideal for age, a continuous variable, as it captures nuanced variations in conversion rates across the entire age range. The use of bubble size to represent customer count adds a second layer of insight, helping visualize both distribution and performance in a single view, unlike bar charts, which are more suited to discrete categories.

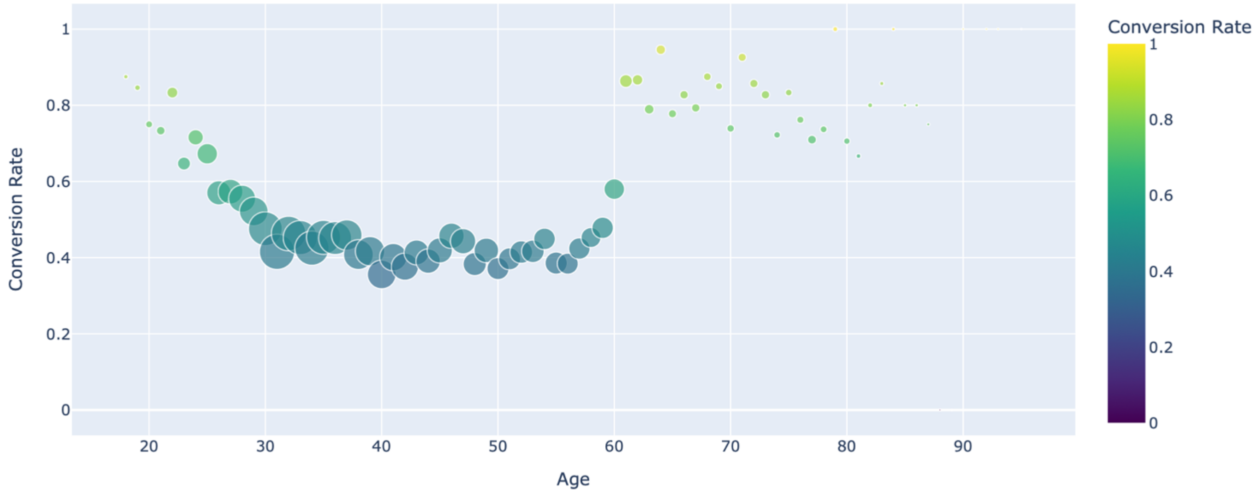

Fig. 8 Age Distribution vs. Subscription Outcome (Bubble Chart)

We see that most customers fall in the **30–45 age range**, making it the core demographic. However, **younger customers (18–30)** and **older customers (60+)** show **higher conversion rates** than this middle band. Younger customers may be new earners, open to financial planning and saving habits. Older customers, especially retirees, may be shifting toward conservative, stable financial instruments.

5.1.7 Balance Distribution vs. Subscription Outcome

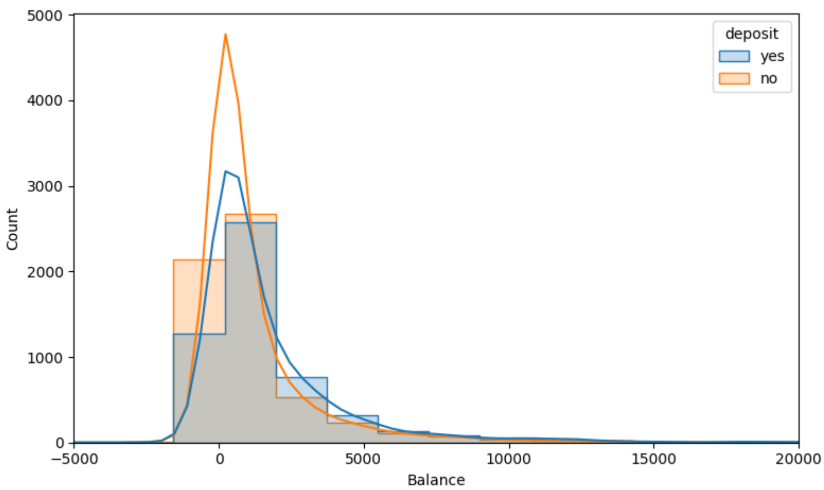

Fig. 9 Balance Distribution vs. Subscription Outcome (Histogram + KDE)

We used a histogram with overlaid KDE (Kernel Density Estimation) to visualize this because it clearly illustrates the frequency of balance values as well as the smoothed shape of the distribution.
Analysis of account balance reveals a **right-skewed distribution** for both subscribing and non-subscribing customers, with a majority having relatively low balances. However, the KDE curves show that customers who subscribed to the term deposit tend to have slightly higher balances on average, with a longer tail toward the higher end.

5.1.8 Call Duration vs. Subscription

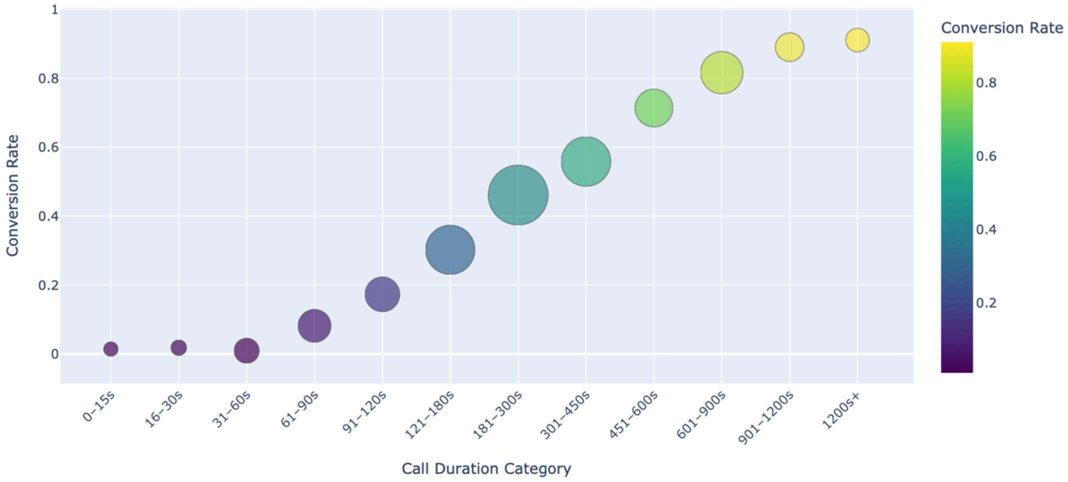

Fig. 10 Call Duration vs. Subscription Outcome

Call duration exhibits a direct positive correlation with campaign success. Most successful conversions occurred during **medium to long calls**, while **short calls** typically ended in non-subscription.
This suggests that longer calls allow agents to build rapport, clarify doubts, and communicate product value more effectively.

5.1.10 Pdays (Days Since Last Contact) vs. Subscription

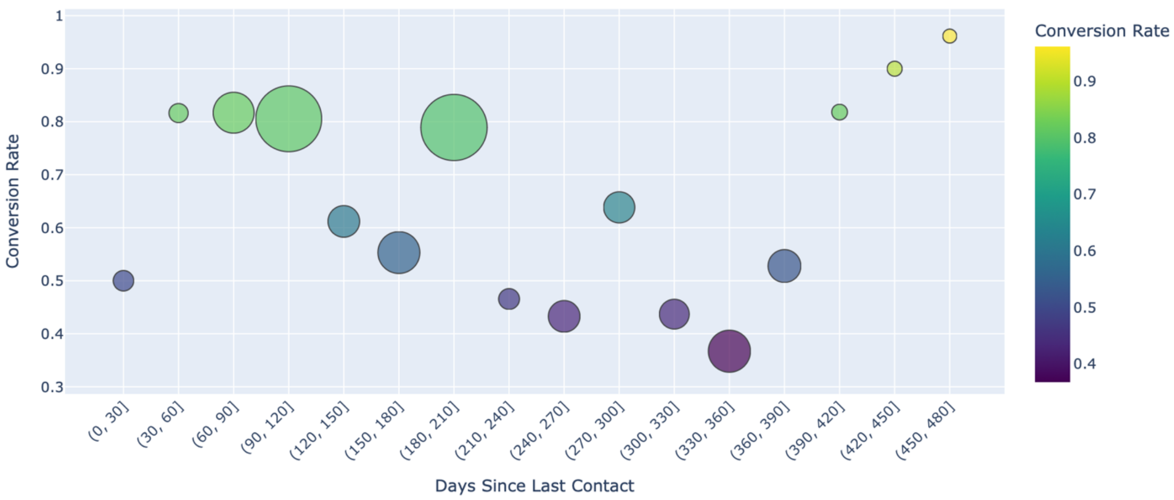

Fig. 11 Number of Days since Last Contact vs. Subscription Outcome

The bubble chart illustrates conversion rates across different intervals of days since last contact, with bubble size representing customer volume. It offers a clear view of both success rates and scale, which bar or dot plots cannot convey simultaneously.

The variable pdays primarily contains the value **-1**, denoting no previous contact (removed from visualization to see clearer trends in non-negative data) . **Spikes around 90 and 180 days** show that follow-ups are likely happening at **fixed intervals**, aligning with campaign cycles. At these intervals **conversion rates are higher**, hinting at **higher effectiveness** of delayed or planned follow-ups. Despite the lower customer count in extreme bins, such as the **450–480** day group, the **high conversion rates** suggest that carefully re-engaging long-dormant leads may be highly fruitful.

5.1.9 Number of Contacts (During Campaign) vs. Subscription

Most conversions occur within **2 to 3 contact attempts**. Beyond this point, conversion rates drop, indicating **diminishing returns** and likely customer resistance or fatigue.

5.1.10 Conversion Rate by Month with Campaign Volume Context

This chart compares the **conversion rate** across each month of the year, with bubble sizes representing the **number of customers contacted**. It helps evaluate not just when customers are more responsive, but also how efficiently campaign efforts are distributed.

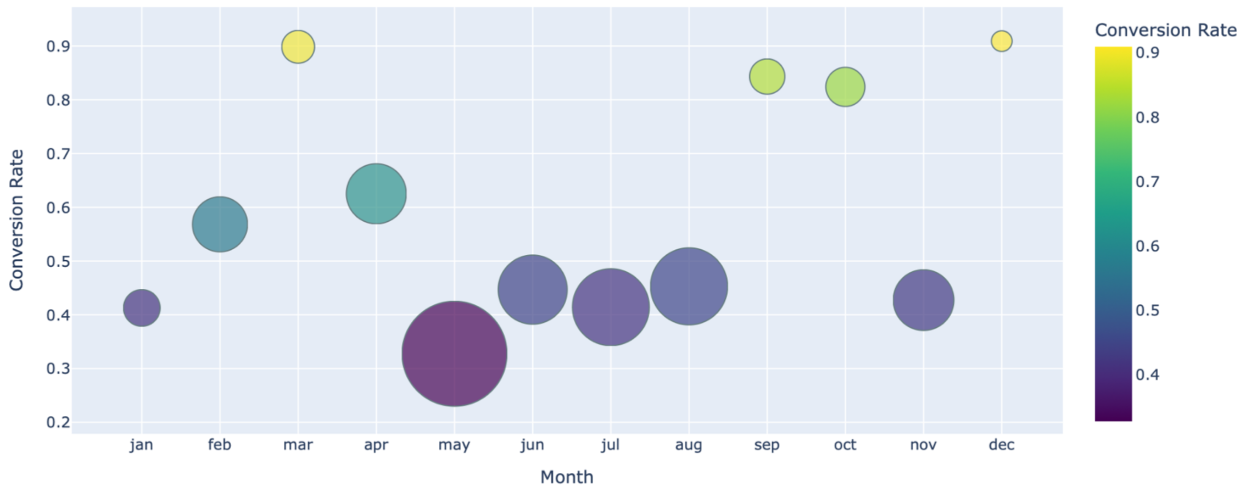

Fig. 12 Conversion Rate by Month with Campaign Volume

December had the **highest conversion rate (~91%)**, but was also the **least contacted month** , suggesting a highly receptive but underutilized window. Similarly, **March (89.9%)**, **September (84.3%)**, and **October (82.4%)** achieved excellent conversion outcomes, each with relatively low to moderate contact volumes. These months stand out as **high-return periods** that may benefit from scaled-up efforts.

In contrast, **May** saw the **highest campaign volume** (2,824 contacts), but delivered a low conversion rate (~32.8%), indicating inefficiency or campaign fatigue. **July and November**, despite significant contact volumes, also yielded modest results (~41–43%), suggesting they may not justify continued investment at current levels.

5.2 Predictive Modeling Insights: Regression and Decision Tree

5.2.1 Logistic Regression Feature Influence

To better quantify which features influence deposit subscription, a logistic regression model was fitted on this balanced dataset using scaled numerical variables and one-hot encoded categorical features. The model provides interpretable coefficients, indicating the direction and strength of each feature's impact. Logistic regression achieved reasonably high accuracy for both classes. The model reliably predicts both subscription and non-subscription outcomes.

Positive Influencers: poutcome_success had the strongest positive coefficient (~2.18), reaffirming that customers who responded positively to past campaigns are far more likely to convert again. **duration** (longer calls), **month_mar**, and **month_dec** also positively influenced conversions, suggesting both communication quality and campaign timing matter.

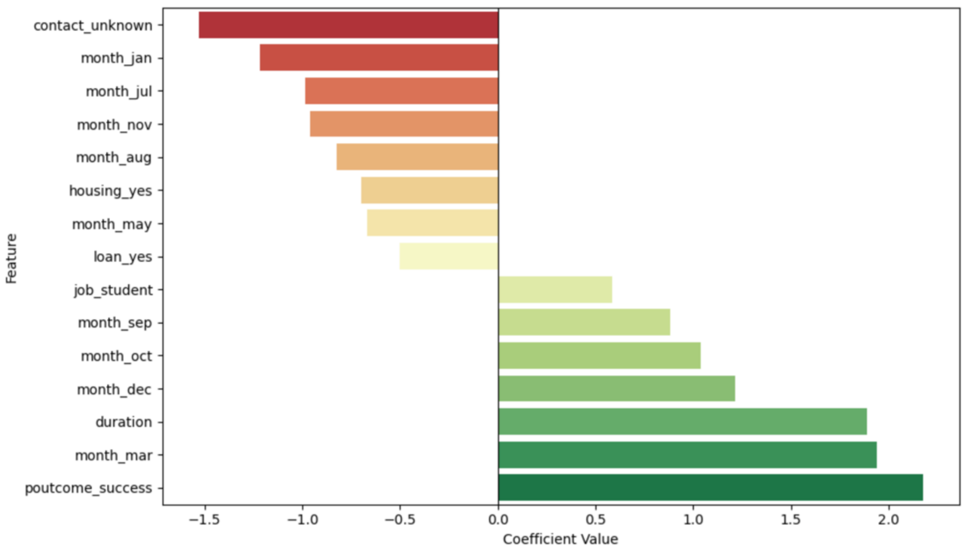

Fig. 13 Logistic Regression Coefficients – Feature Impact

Negative Influencers: contact_unknown was the most detrimental, with a strong negative coefficient (~-1.57), showing that insufficient contact information heavily reduces conversion chances. Months like **January**, **July**, and **November**, along with existing **loans** and **housing loans**, also showed a negative association. These insights reinforce earlier exploratory findings while **quantifying the relative strength** of each variable.

5.2.2 Decision Tree Classification and Strategic Rules

A decision tree classifier was trained to identify actionable patterns in how combinations of customer attributes influence subscription outcomes.

Key Rules and Branch Patterns:

1. **Call Duration** was the primary splitter, longer calls consistently led to higher conversions.
2. **Previous Outcome** : Customers who had previously converted remained strong leads.
3. **Contact type**: **cellular** contact paths often led to "yes," while "unknown" or "telephone" led to dead ends.
4. **High-performing segments** emerged around: Older customers (60+), Long-duration calls, Prior success in campaigns, Educated clients

Dead-end segments were clearly visible:

- Medium calls + contact = unknown + age = 30–44 = consistently poor results
- Overlap of low-importance features often led to zero-conversion paths

5.2.3 Top Customer Segments

To complement the tree-based findings, we compiled a table that highlights the top customer segments (with significant number of customers) ranked by conversion rate. These serve as ready-to-use targeting profiles for future campaigns. As an example the top 10 are mentioned below:

Table 3: Top 10 customer segments

Combination	Segment 1	Segment 2	Conversion Rate (%)
Age Group + Quarter	Under 30	Q4	86.7
Duration Bucket + Age Group	60+	Long	84.9
Age Group + Education Level	60+	High	82.7
Age Group + Quarter	60+	Q4	81.5
Job Group + Age Group	White-collar	60+	81.5
Age Group + Contact	60+	Cellular	80.3
Age Group + Quarter	60+	Q1	80.1
Job Group + Duration Bucket	Others	Long	79.6
Duration Bucket + Contact	Long	Telephone	79.0
Age Group + Education Level	60+	Medium	78.1

Section 5.3: Segmented and Special-focus Comparison Insights

5.3.1 Temporal Conversion Patterns by Customer Profile

This section investigates how conversion behavior varies over time across different customer groups, using segmented bubble charts by **job category** and **age group**. While individual feature analysis in Section 5.1 established which groups tend to perform well or poorly overall, the current analysis introduces a **temporal dimension**, allowing us to evaluate the **month-wise consistency and variability** of these trends.

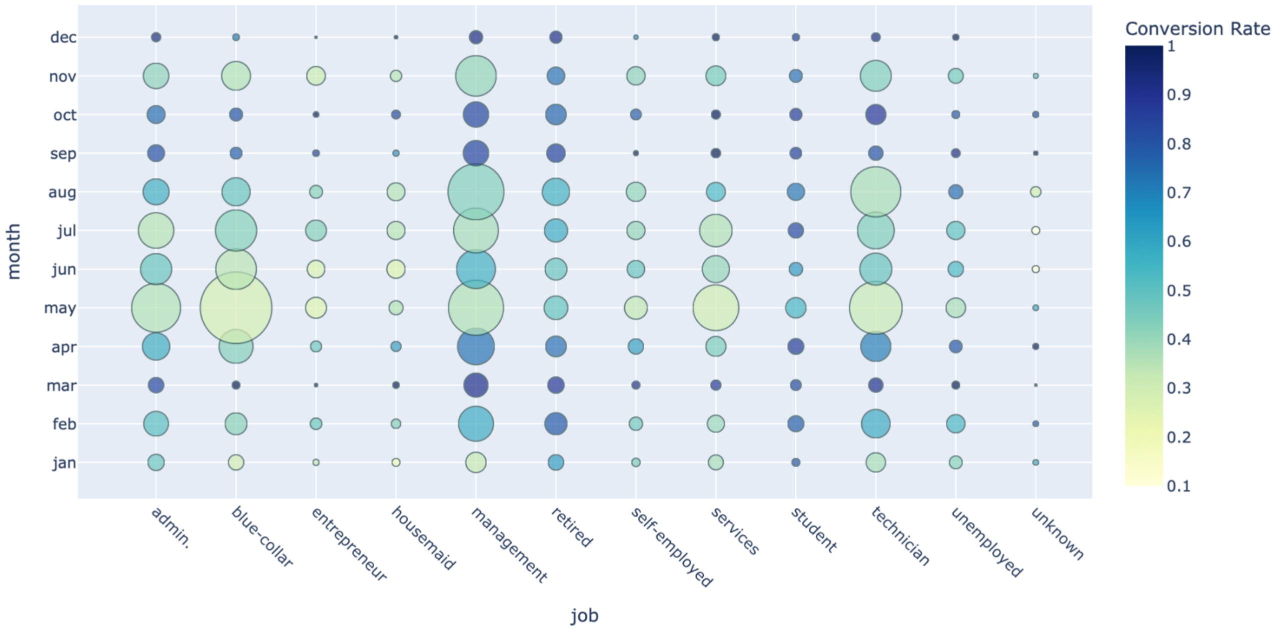

Fig. 14 Temporal Conversion Patterns by Job Category (Bubble Chart)

The job-based segmentation reveals that the **student**, **retired** and **management** categories are not only high-performing in aggregate, but also demonstrate **stable conversion rates across most months**, including lower-performing periods such as May and November. This consistency suggests strong receptiveness that is **largely independent of campaign timing**. In contrast, segments such as **entrepreneur**, **housemaid**, and **unknown** occupations remain underperforming across the entire year, reinforcing their limited value as target groups.

Among more variable segments, **blue-collar** and **services** categories exhibit **month-specific surges** (e.g., May or July) but lack sustained performance, indicating potential sensitivity to campaign timing or external seasonal factors. This pattern of **volatility**, while not visible in the overall job-based conversion rates, is captured clearly in the time-based view.

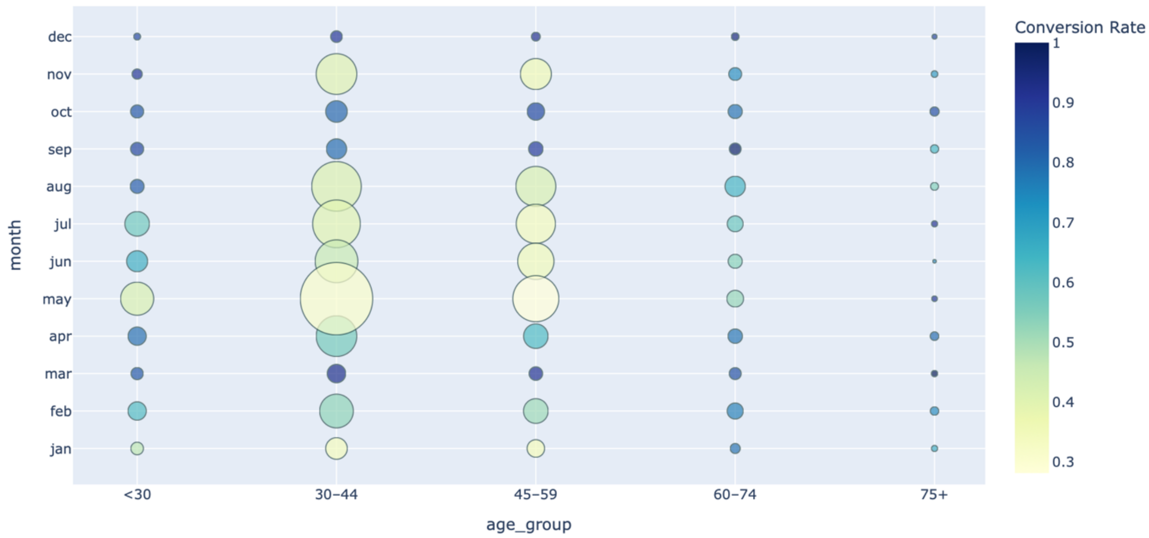

Fig. 15 Temporal Conversion Patterns by Age Group

The age-group analysis similarly highlights segment-specific temporal behavior. Customers aged **60 and above** maintain **consistently high conversion rates** throughout the year, reinforcing their position as **reliably responsive** despite lower contact volumes. The **30–44** age group, which receives the highest number of contacts, displays **notable month-to-month fluctuations**, with stronger results in March and September and reduced performance in May and June. Such instability indicates a **need for precise timing** when targeting this segment. The **under-30** group, by comparison, shows **moderate but stable** performance, with less pronounced peaks or drops.

In summary, this analysis extends prior findings by highlighting **which segments are consistently high-performing**, which are **highly timing-sensitive**, and which **underperform across the board**. These distinctions are critical for optimizing campaign timing and prioritizing stable versus variable audience groups.

5.3.2 Focused Analysis: Understanding Underperformance in the 30–44 Age Group

The 30–44 age group constitutes the **largest demographic segment in the dataset**, comprising approximately **39% of the total contacted customers**. Visual exploration showed that the 30–44 segment was **heavily targeted** across months but displayed **low conversion rates** compared to both younger and older cohorts. Further, decision tree analysis identified this group as a **frequent terminal node** in low-conversion branches. These observations warranted focused examination. To better understand this underperformance, a targeted analysis was conducted on the group's **financial burden**, through the lens of **housing and personal loan exposure**.

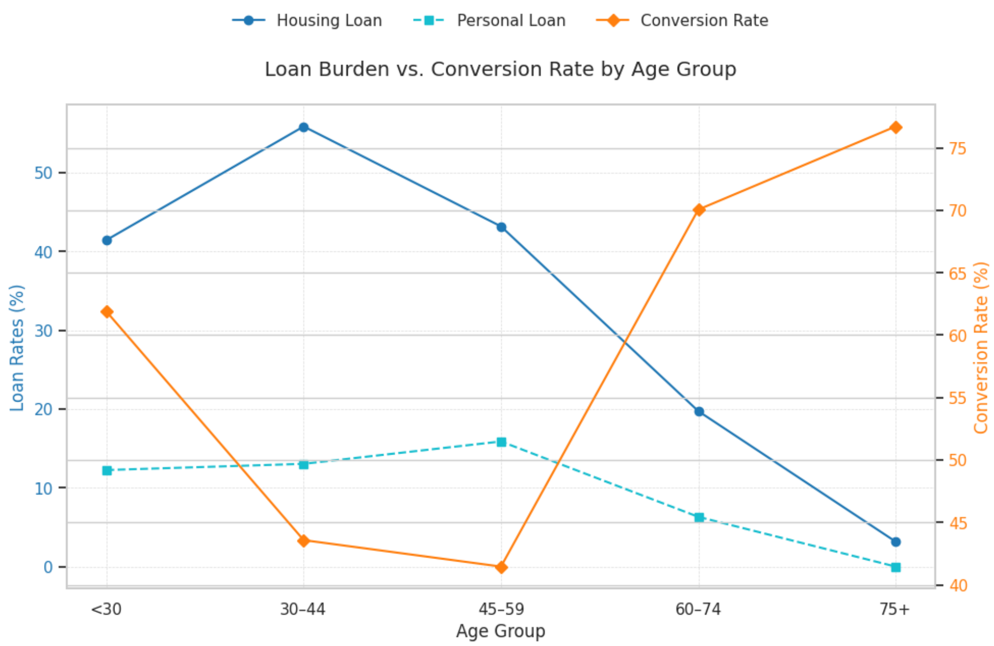

Fig. 16 Focused Analysis – Financial Liabilities across Age Group vs Conversion Rate

The results indicated that individuals aged 30–44 have the **highest incidence of housing loans**, with over **55%** carrying active obligations. Personal loan prevalence is also elevated in this group relative to others. In contrast, their conversion rate is among the **lowest (~43%)**, diverging sharply from the 60+ segment, where debt levels are minimal and conversion exceeds **75%**. Trends similar to the 30-44 age group were noticed in the 45-59 age group. These findings suggest a potential link between **active financial liabilities** and **reduced campaign responsiveness** for long-term savings products in this age group.

6. Interpretation of Results and Recommendations

6.1 Leverage High-Performing Segments to Maximize Early Returns

Analysis across job type, education level, age, and call duration revealed consistently high conversion rates among certain customer groups. These include:

- **Senior Customers (Age 60+)**
- **Retired and Student Profiles**
- **Highly Educated Individuals (Tertiary Education)**
- **Customers with Previous Campaign Success**
- **Long Call Duration (>360s)**

These segments not only convert at high rates but also exhibit **month-on-month consistency**. Targeting them with personalized outreach, premium deposit schemes, or loyalty-based incentives can generate reliable returns with minimal risk.

6.2 Segment-Specific Messaging and Call Strategies

The decision tree analysis revealed that certain combinations of features consistently resulted in poor outcomes - particularly medium-duration calls with unknown contact type and customers aged 30–44. While these combinations represent low-yield paths, not all zero-conversion segments are inherently non-viable.

Recommended Strategies:

- **Avoid Dead-End Combinations**: Deprioritize efforts when multiple low-performing features overlap, unless mitigated by other strong predictors.
- **Design Segment-Specific Scripts**: Tailor communication for key customer profiles. Say,
    - Customers aged 45–59 with medium-duration calls may benefit from early trust-building techniques and clearly stated product benefits**.**
    - Short-duration calls should be reframed to immediately address customer objections or schedule follow-up discussions.
- **A/B Test in Borderline Segments**: Run pilot tests in zero-conversion but large-volume segments (e.g., 30–44 with cellular contact and medium-duration calls) to assess whether **message framing or call tone** can improve engagement.

6.3 Strategic Engagement of the 30–59 Age Segment

The age group between 30 and 59 years comprises the largest share of the customer base, accounting for over 60% of total contacts. However, this segment consistently displayed lower-than-average conversion rates. Detailed analysis reveals that this cohort carries significant financial liabilities, which may restrict their willingness to commit to long-term savings products. Nonetheless, the segment's size, financial activity, and middle-income status render it too valuable to deprioritize.

Recommended Strategies:

- **Introduce Flexible Deposit Plans**: Position term deposits as adaptable products that offer low entry barriers, partial withdrawal options, and periodic top-ups.
- **Develop Loan-Linked Deposit Products**: Create offerings that integrate better returns for customers already servicing loans, thereby turning existing liabilities into touchpoints for engagement.
- **Promote Tax-Saving Instruments**: Emphasize tax benefits of term deposits

- **Enable Digital Micro-Savings**: Encourage adoption of automated savings tools linked to salary accounts for small but consistent contributions.
- **Utilize Goal-Based Campaigns**: Frame deposits around personal milestones (child's future, home upgrade, vacation plans) and provide planning tools such as calculators or bundled consultation

6.4 Timing Optimization: Data-Driven Campaign Scheduling

The analysis of conversion rates by month revealed significant temporal variation, offering clear guidance for scheduling future outreach efforts.

Key Observations and Recommendations:

- **Prioritize High-Conversion Months**: Months such as **March, September, October, and December** consistently demonstrated high subscription rates (~82–91%) despite moderate outreach volumes. These months represent high-return opportunities and should receive greater campaign focus and budget allocation.
- **Reallocate Effort from Underperforming Months**: **May** had the highest campaign outreach but yielded the lowest conversion rate (~33%). Similarly, **June and July** showed suboptimal performance despite moderate to high effort. These months should be reconsidered for campaign volume unless strategic changes are made.
- **Test and Learn from Success Cases**: Controlled experiments should be conducted in high-performing months to identify underlying success factors - whether behavioural, seasonal, or messaging-related - and replicate those conditions in adjacent months.

Aligning campaign timing with demonstrated seasonal responsiveness will not only improve effectiveness but also reduce wasted effort and customer fatigue.

6.5 Data Quality Enhancement for Improved Targeting

A substantial number of records in the dataset contain "unknown" values in key fields such as **contact type**, **education level**, and **previous campaign outcome** (poutcome). These records correlate strongly with low or zero conversions, making them unreliable for precision targeting. Improving data completeness will directly enhance model performance and campaign segmentation accuracy.

Recommendations:

- **Enforce Mandatory Data Capture**: Future campaigns should ensure complete logging of customer attributes through better CRM integration or agent prompts.
- **Investigate Unknown Data Sources**: Determine whether "unknown" entries arise from specific lead acquisition channels (e.g., third-party vendors) and apply stricter vetting protocols for those channels.
- **Segment and Reclassify**: Where possible, impute or cluster "unknown" entries based on available attributes to reclassify them into usable categories.

6.6 Continuous Improvement Through Feedback and Performance Monitoring

To sustain and enhance campaign effectiveness over time, banks should implement dynamic feedback and evaluation mechanisms.

Recommendations**:**

- **Integrate predictive modeling into CRM systems**: Deploy the linear regression model within the bank's CRM to score incoming leads. This would allow campaign teams to focus on high-probability customers and continuously refine targeting based on live data.
- **Integrate Feedback Loops:** Capture customer feedback and agent input through post-call surveys or CRM notes to iteratively refine targeting strategies and messaging tone.
- **Monitor Uplift with KPIs:** Regularly track key performance indicators such as contact-to-conversion ratios, monthly campaign efficiency, and segment-wise conversion trends to identify patterns and guide mid-cycle adjustments.
- **Establish Learning Systems:** Create a structured process to document what worked (or didn't) in each cycle, and use these learnings to guide future campaign design, script updates, and model tuning.