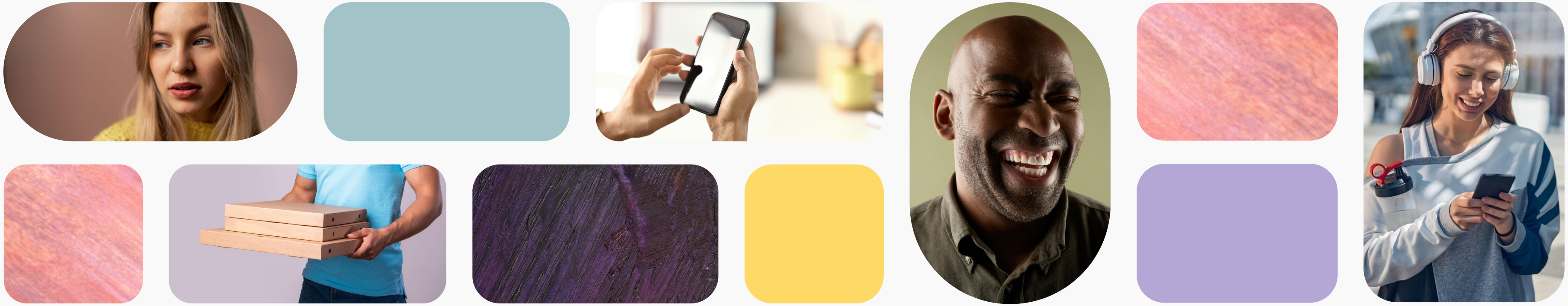
BANK MARKETING - A TERM DEPOSIT

Identify high-value customer groups, design tailored engagement strategies through phone calls, and apply predictive scoring to maximize campaign effectiveness and conversion rates



KEY SECTIONS

1. PROJECT INTRODUCTION

Project Purpose & Key Questions

★ **Project Purpose:** Identify high-value customer groups, design tailored engagement strategies through phone calls, and apply predictive scoring to maximize campaign effectiveness and conversion rates.

★ **Key Questions:**

- Which customer groups represent the high potential value in this campaign?
- How can predictive modeling help estimate individual purchase probabilities and guide resource allocation more efficiently?

Dataset Overview

(After data cleaning was applied)

- Number of rows: 38,234 rows
- Number of columns: 15 columns
  + Customer Demographics & Finances: 6 columns
  + Customer Interaction Data: 8 columns
  + Column “Label”: outcome indicating whether a customer joined the campaign

Label	Count	Percentage
Label_Yes	4257	11.13%
Label_No	33977	88.87%

(The proportion of Yes/No labels in the dataset)

2. DATA EXPLORATION & INSIGHTS

Metrics Analyzed in This Analysis

This analysis is based on three key metrics:

Metric	Description	Formular
Conversion Rate	The effectiveness of each group in terms of conversion performance	$\text{CR} = \frac{Number\ of\ Conversions\ per\ Group}{Total\ of\ Customers\ per\ Group}$
Distribution Rate	The proportion of each group within the total customer population	$\text{Dist} = \frac{Number\ of\ Customers\ per\ Group}{Total\ of\ Customers}$
CR × Dist Index	Each group’s actual contribution to overall performance	$\text{CR x Dist} = \frac{Number\ of\ Conversions\ per\ Group}{Total\ of\ Customers}$

→ **The goal is to identify and prioritize high-potential opportunities:** with both meaningful scale and strong conversion — in order to maximize existing resources and capitalize on proven performance.

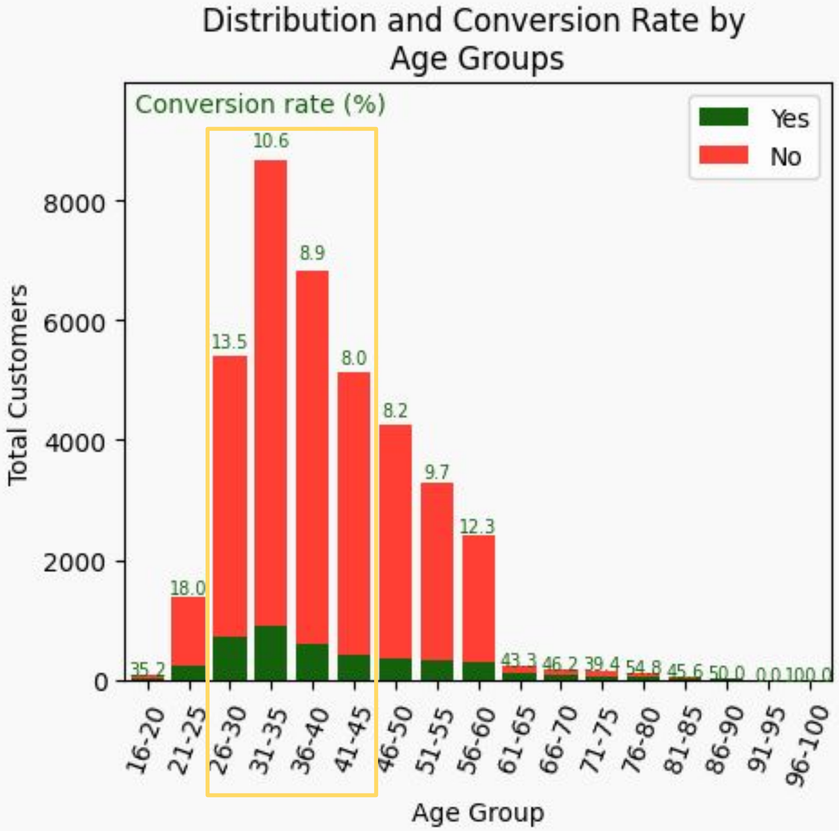
1. Customer Demographics & Finances

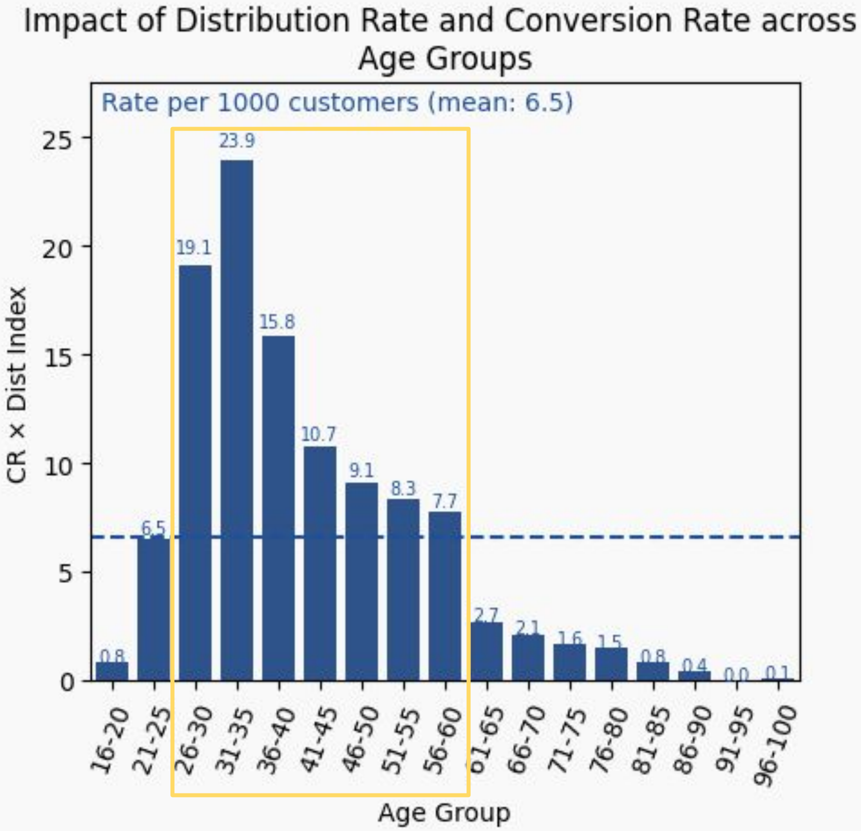
Attribute Information

No.	Columns Name	Type	Description
1	Age Group	Object	Age group in 5-year intervals, from 16 to 100 years old
2	Type of job	Object	'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed'
3	Marital status	Object	marital status ('divorced', 'married', 'single')
4	Education	Object	'illiterate', 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'professional.course', 'university.degree'
5	Housing loan	Object	has housing loan? ('no', 'yes')
6	Personal loan	Object	has personal loan? ('no', 'yes')

Age Group

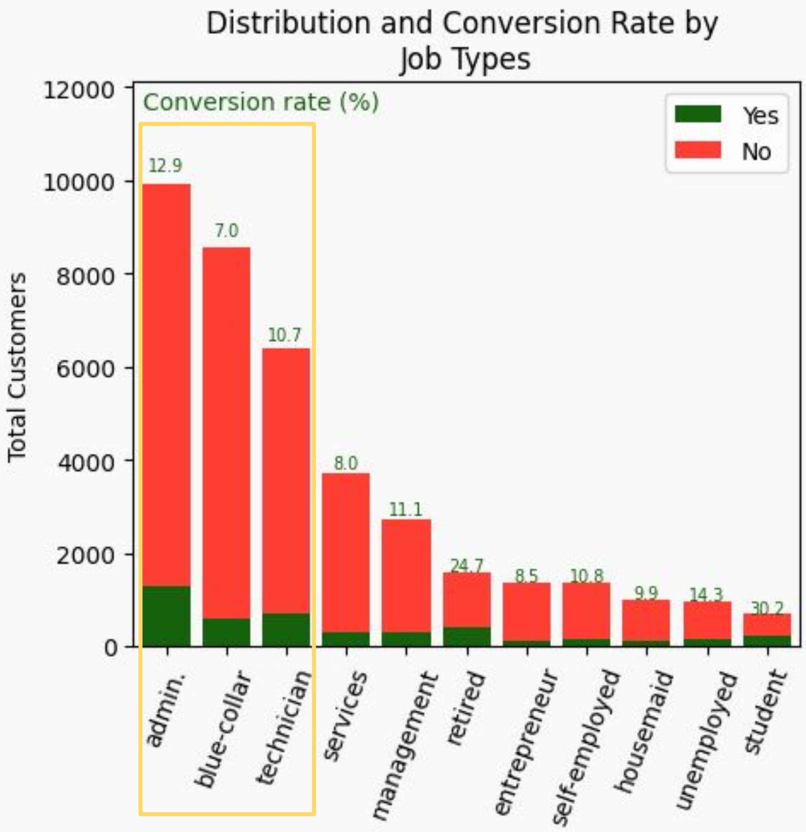
- Strong resources: age 31-35, 36-40, 26-30 (≈ 41-45)
- Potential conversion impact: age 26-60

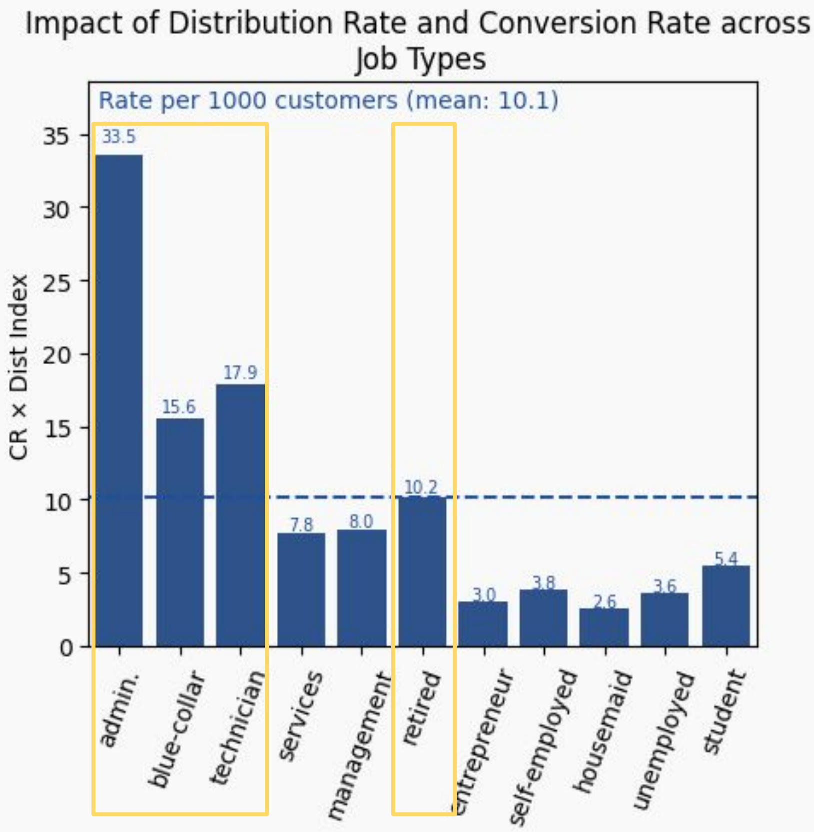




Type of Job

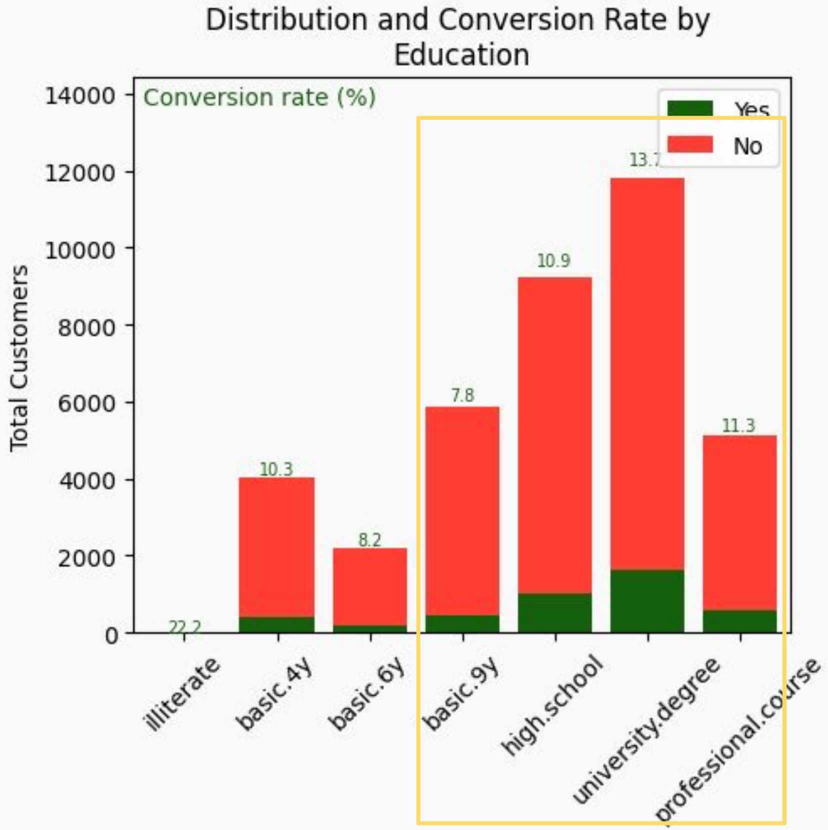
- Strong resources: admin, blue-collar, technician
- Potential conversion impact: admin, technician, blue-collar, retired

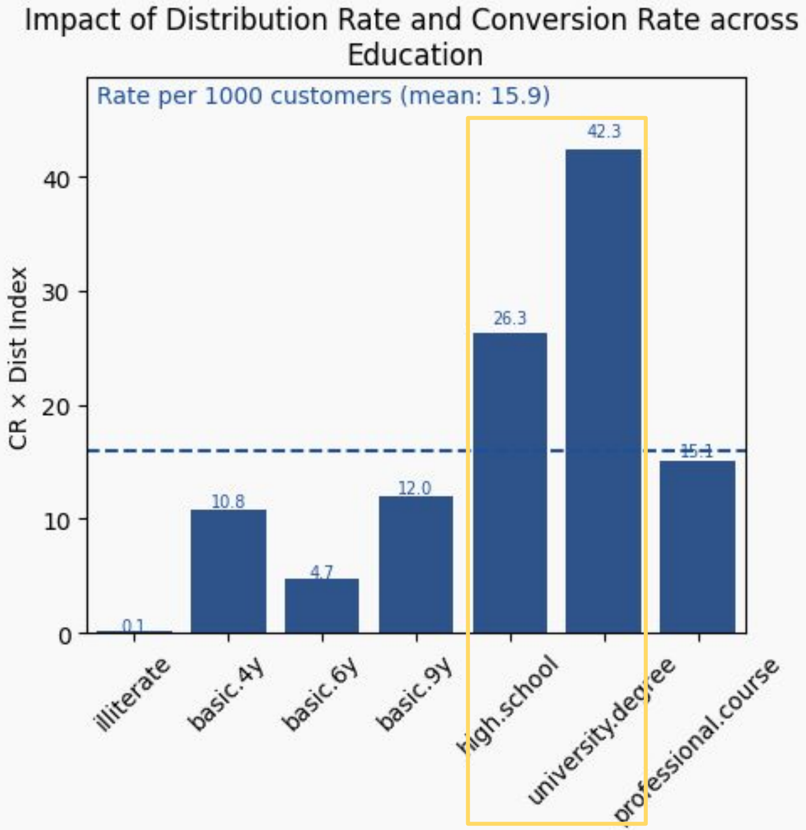




Education

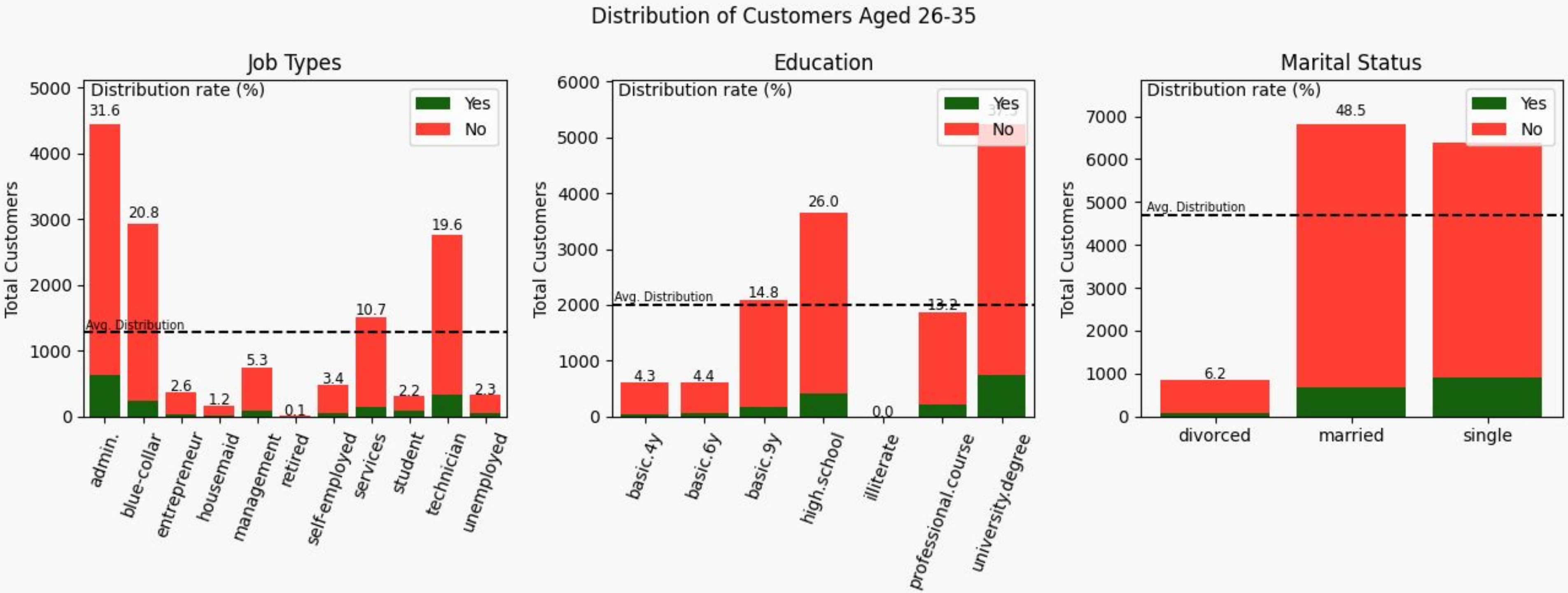
- Strong resources: university, high school, basic.9y ≈ professional
- Potential conversion impact: university, high school





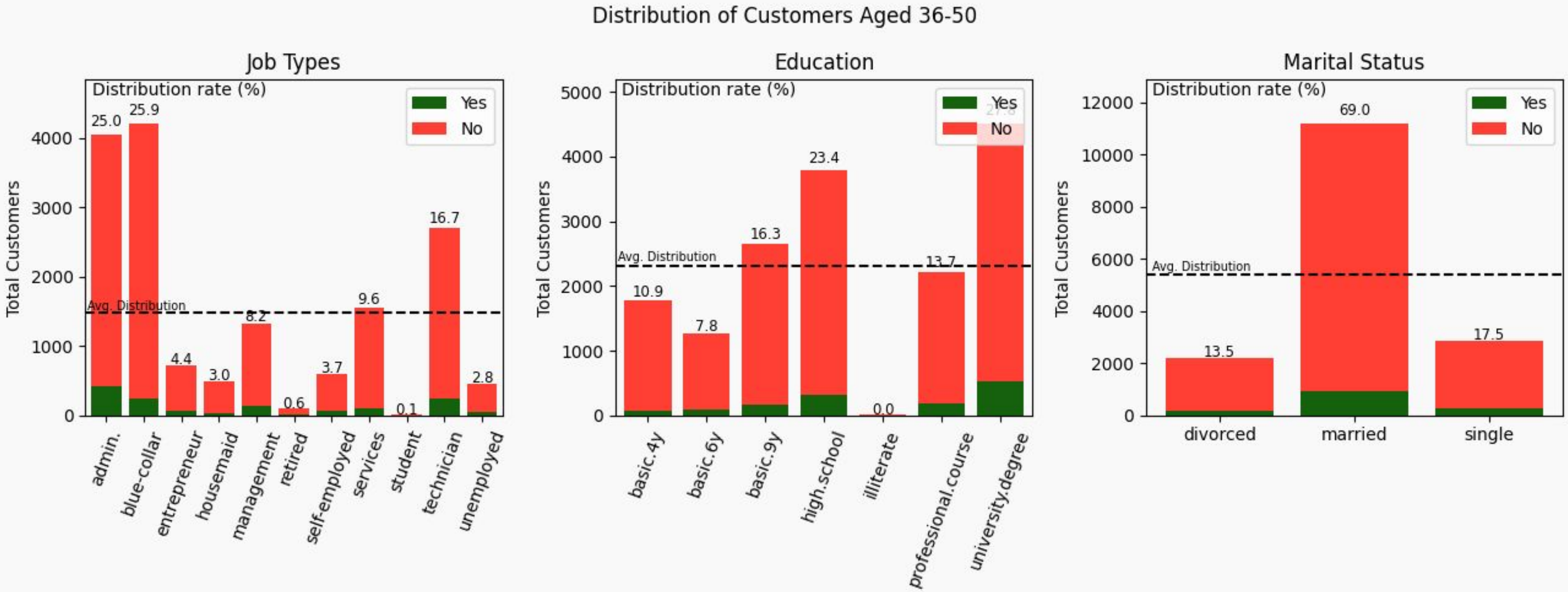
Age 26-35 Individuals

- Job Types: admin, technician, blue-collar, services (high distribution)
- Education: university, high school, basic.9y (high distribution)
- Marital Status: single, married (high distribution)



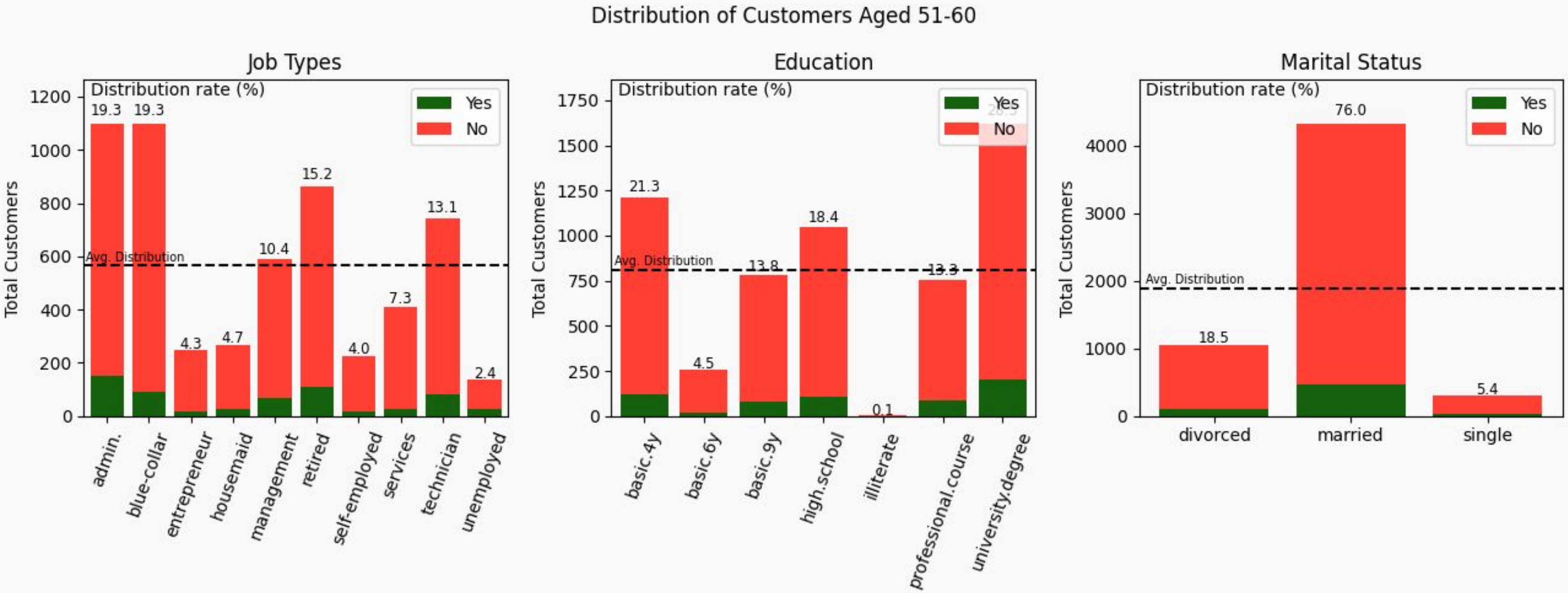
Age 36-50 Individuals

- Job Types: admin, technician, blue-collar, services (high distribution)
- Education: university, high school, basic.9y (high distribution)
- Marital Status: married (high distribution)



Age 51-60 Individuals

- Job Types: admin, blue-collar, retired, technician, management (high distribution)
- Education: university, basic.4y, high school (high distribution)
- Marital Status: married (high distribution)



Data-Driven Insights for Campaign Optimization
(Customer Segmentation)

Based on customer demographics and financial behavior, three key segments were defined:

<table>
<tr><th></th><th>Segment A
Rising Achievers</th><th>Segment B
Established Professionals</th><th>Segment C
Experienced Traditionalists</th></tr>
<tr><td>Age Group</td><td>Age 26–35</td><td>Age 36–50</td><td>Age 51–60</td></tr>
<tr><td>Type of job</td><td colspan="2">admin, technician, blue-collar, services</td><td>admin, blue-collar, retired, technician, management</td></tr>
<tr><td>Marital status</td><td>single, married</td><td colspan="2">married</td></tr>
<tr><td>Education</td><td colspan="2">university, high school, basic.9y</td><td>university, high school, basic.4y</td></tr>
</table>

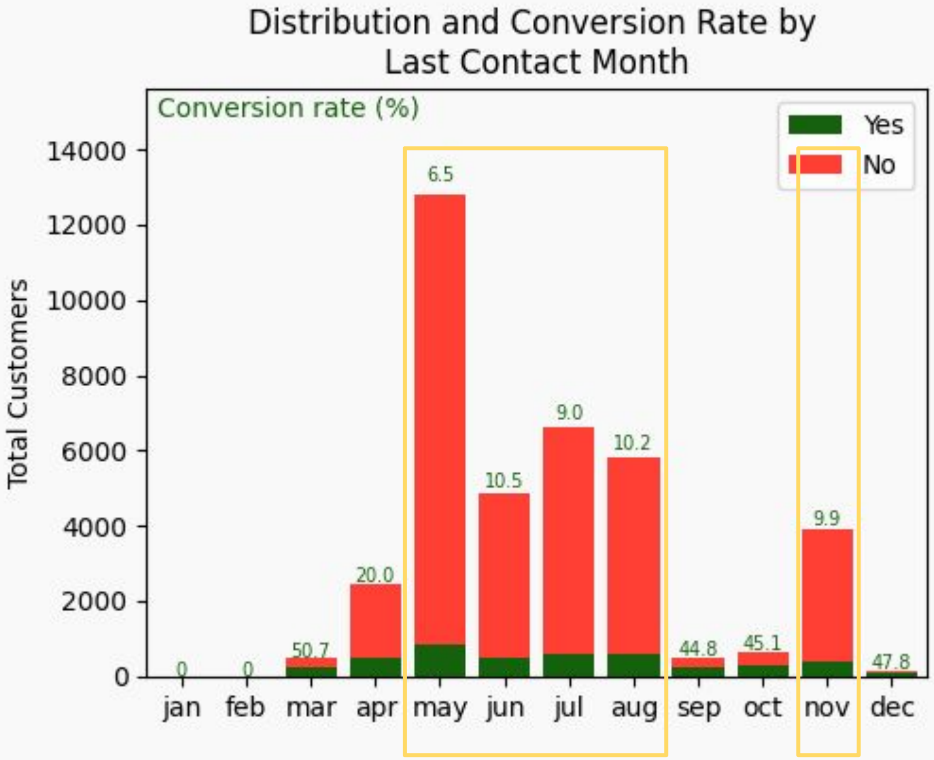
2. Customer Interaction Data

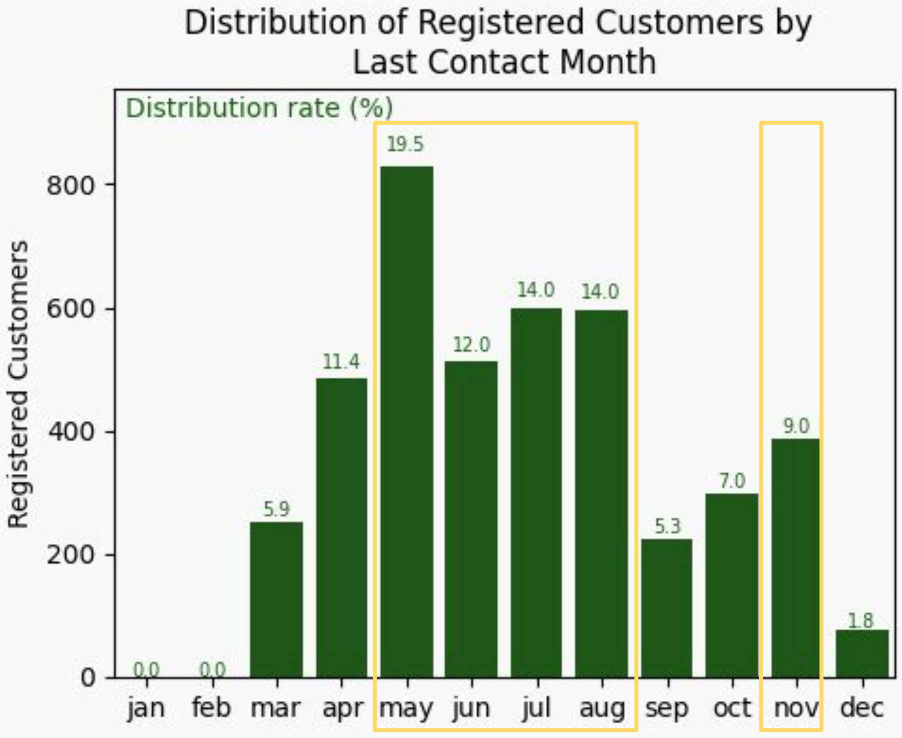
Attribute Information

No.	Columns Name	Type	Description
1	Latest month	Object	Last contact month of year ('jan', 'feb', 'mar',…, 'nov', 'dec')
2	Latest day	Object	Last contact day of the week ('mon','tue','wed','thu','fri')
3	Number of contacts	Int64	Number of contacts performed during this campaign
4	Latest duration	Int64	Last contact duration, in seconds
5	Contact status	Int64	Customers who were never contacted (‘1’, ‘0’)
6	Passed days	Int64	Number of days that passed by after the client was last contacted from a previous campaign
7	Previous campaign's number	Int64	Number of contacts performed before this campaign
8	Previous campaign outcome	Object	Outcome of the previous campaign ('failure', 'nonexistent', 'success')

Latest month

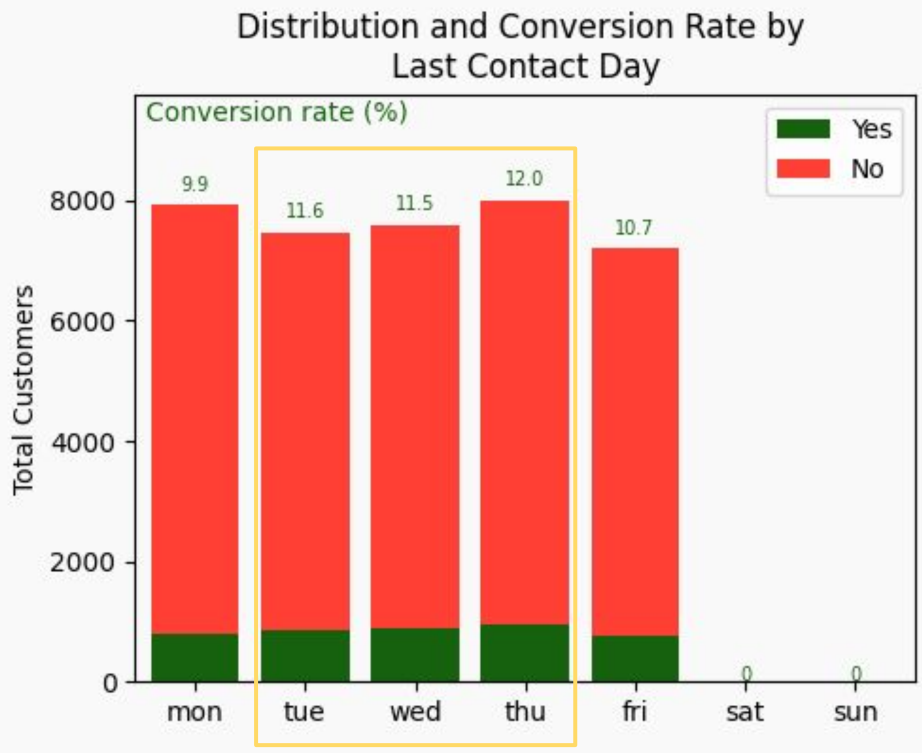
- High-traffic months with low conversion rate: 5, 6, 7, 8, 11
- Low-traffic months with high conversion rate: 3, 4, 9, 10

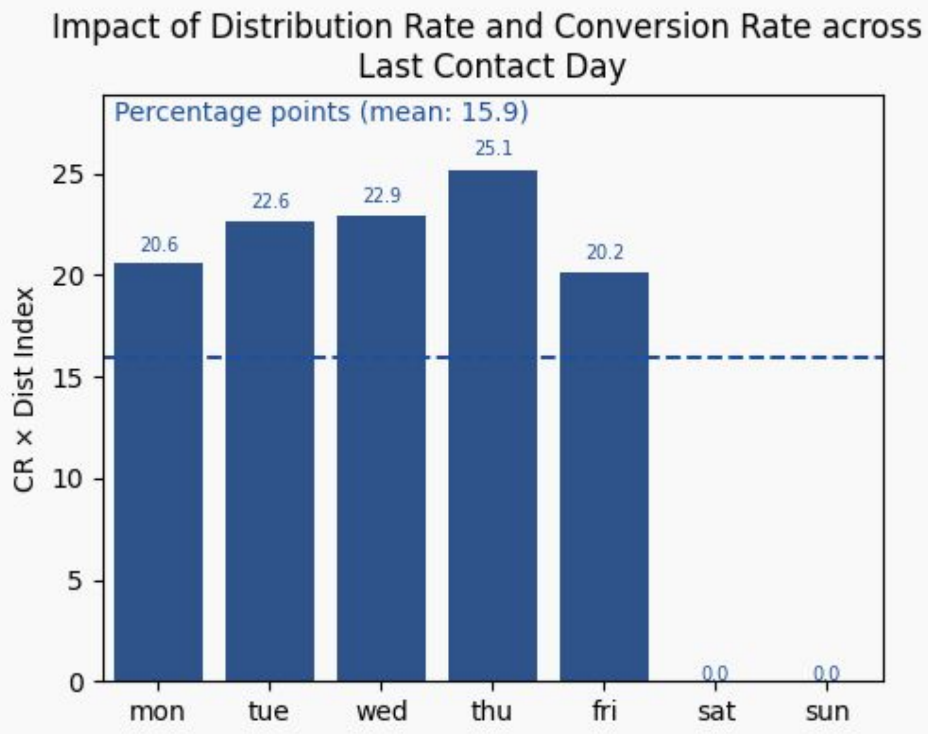




Latest day

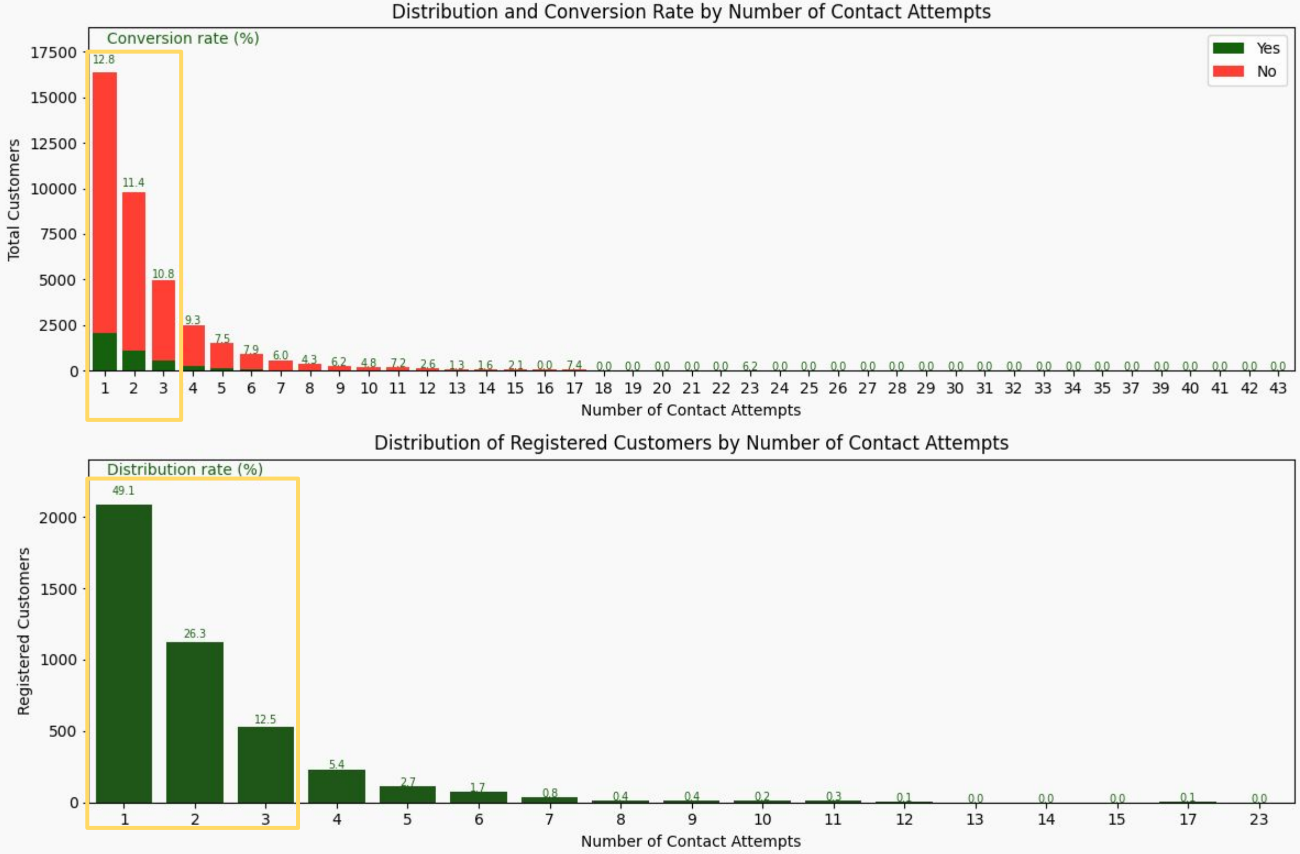
- Most effective contact days: Tuesday, Wednesday, especially Thursday





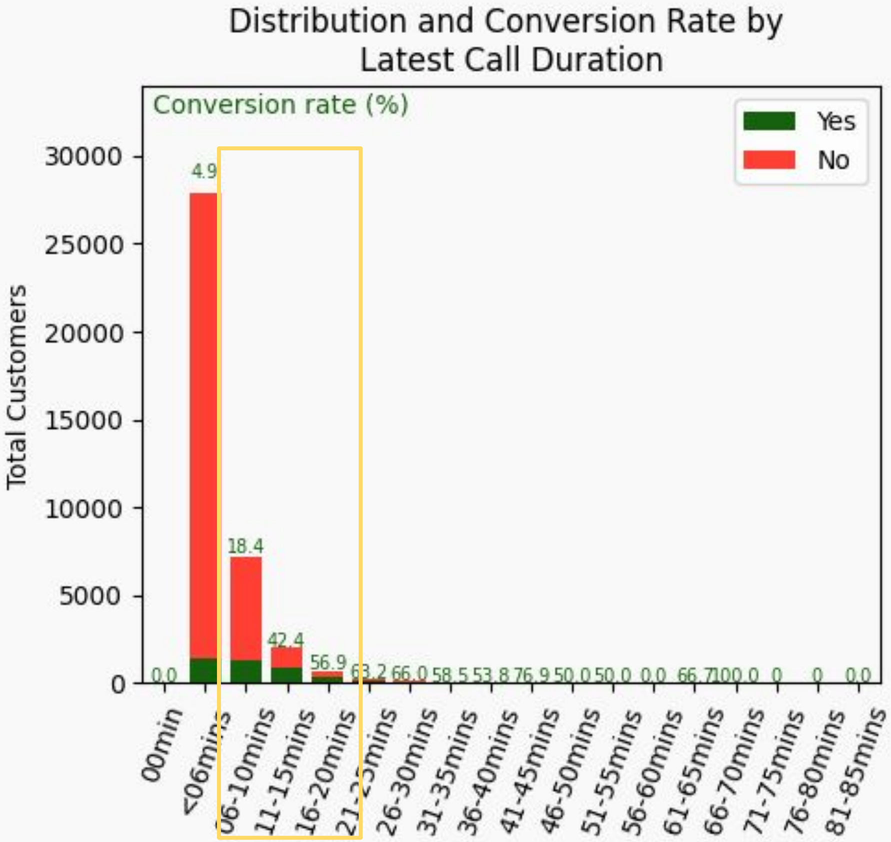
Number of contacts

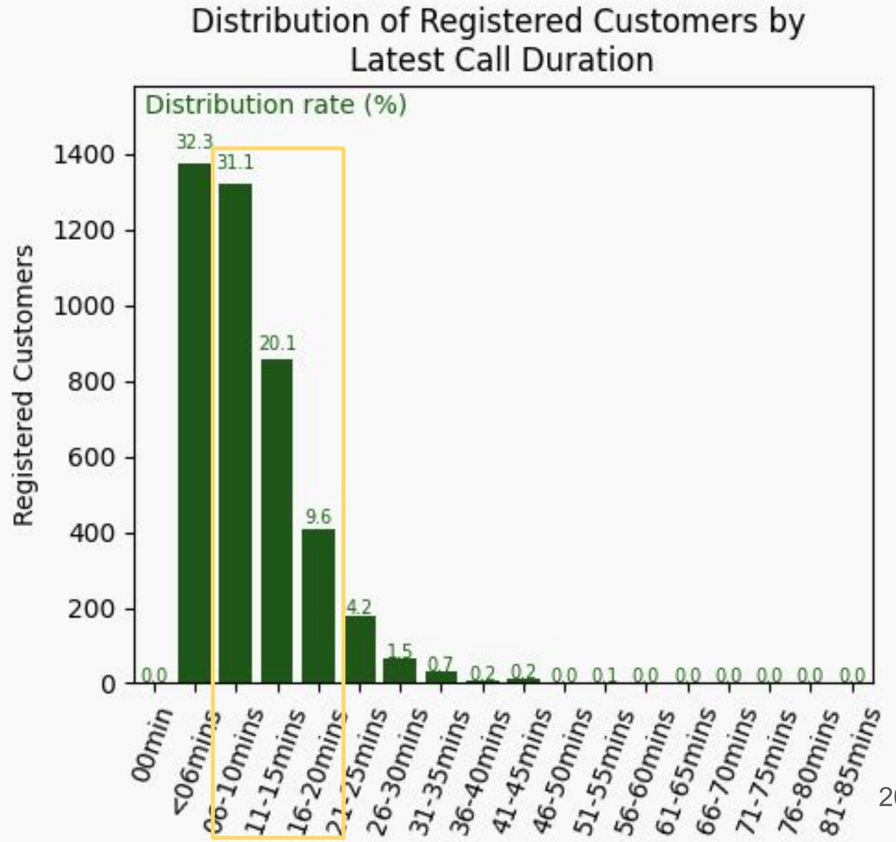
- Effective number of contact attempts: 1-3 times (effective conversion)
- Avoid exceeding 3 calls to prevent resource waste (except for complex cases)



Latest duration

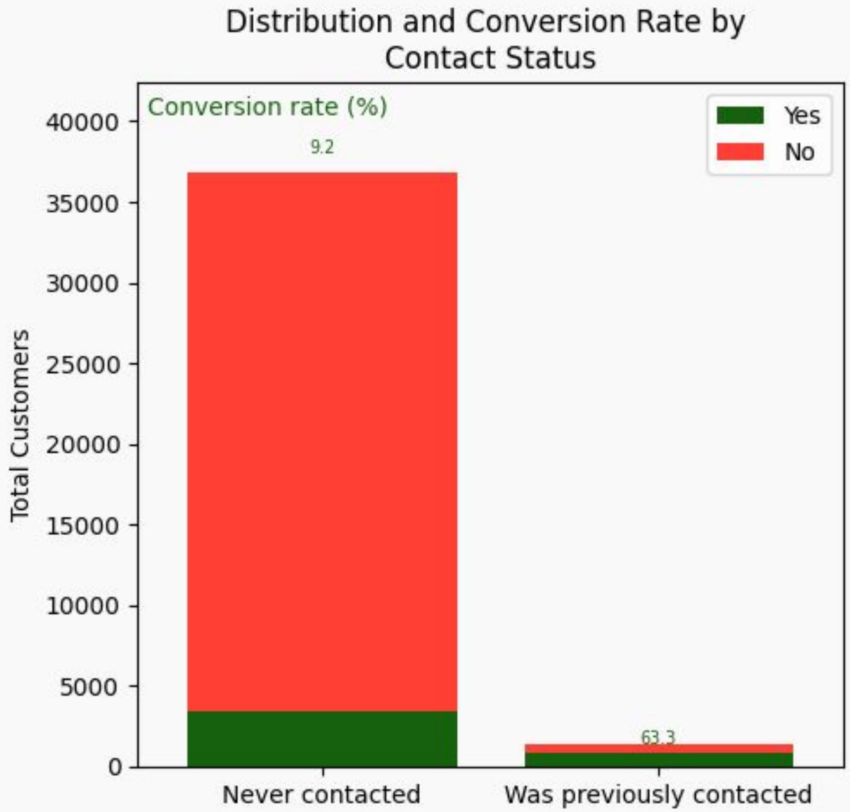
- Focus on effective call duration: 6-20 minutes (effective conversion)
- Shorter calls 1–5 minutes: for high conversion potential cases
- Longer calls 21–45 minutes: for demanding cases; avoid exceeding to stay efficient

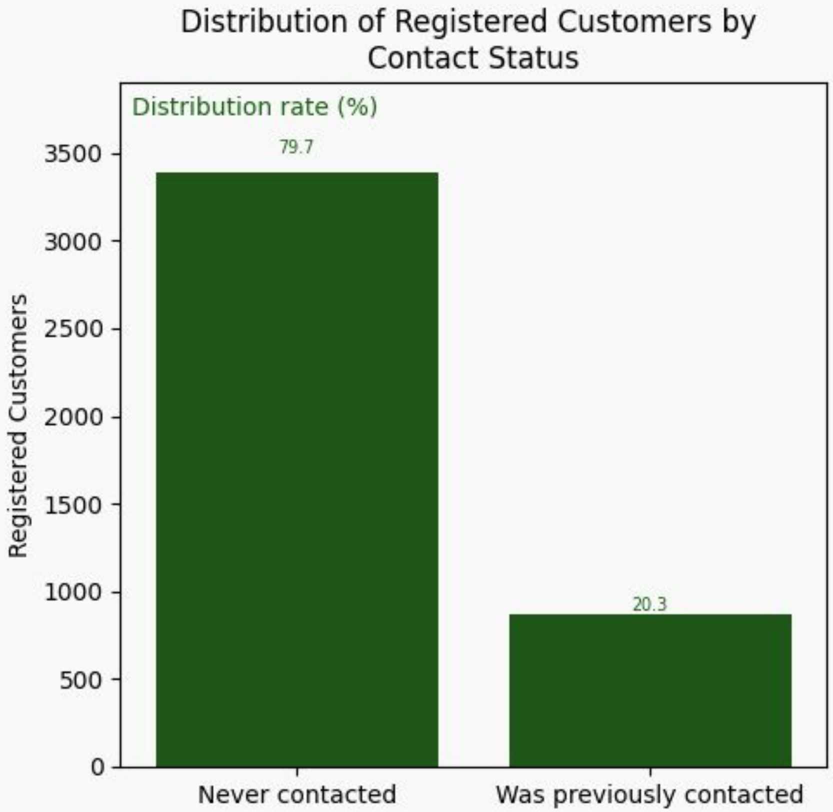




Contact status

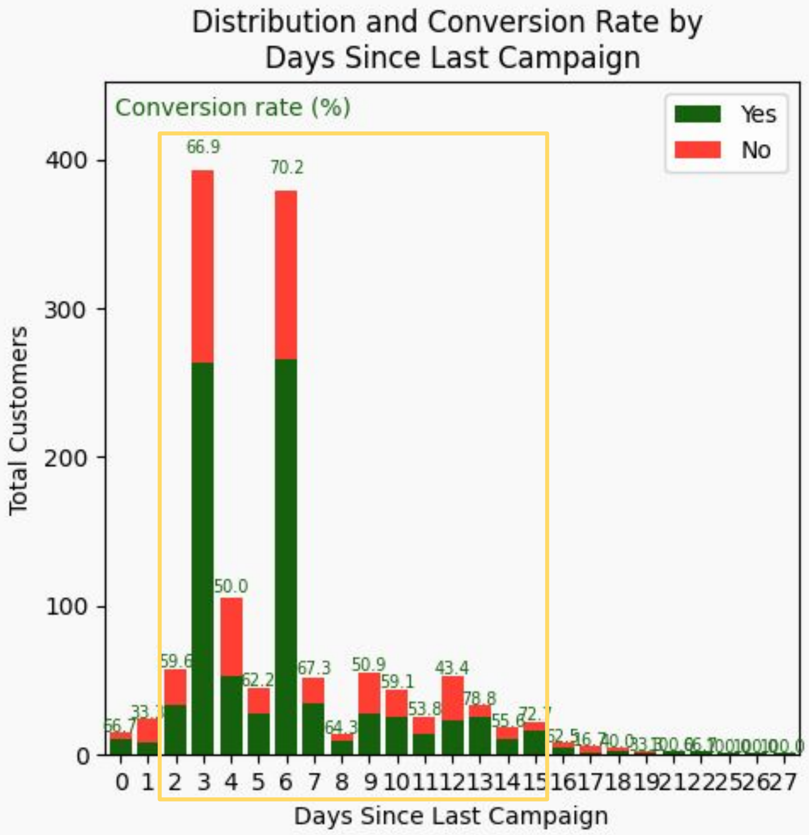
- Effective conversion: contacted
- Strong resources: never contacted

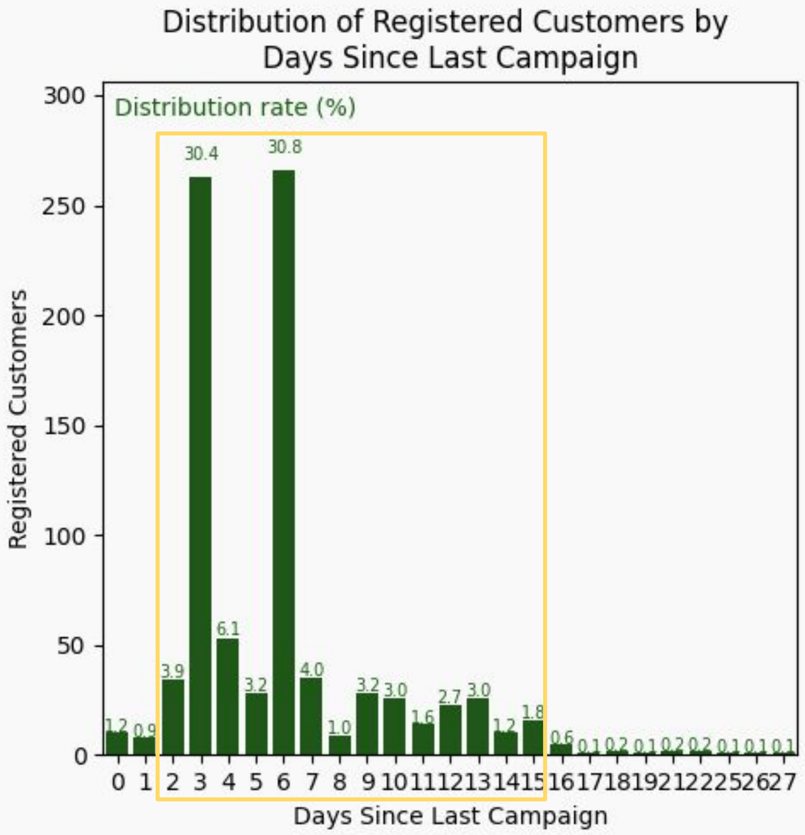




Passed days

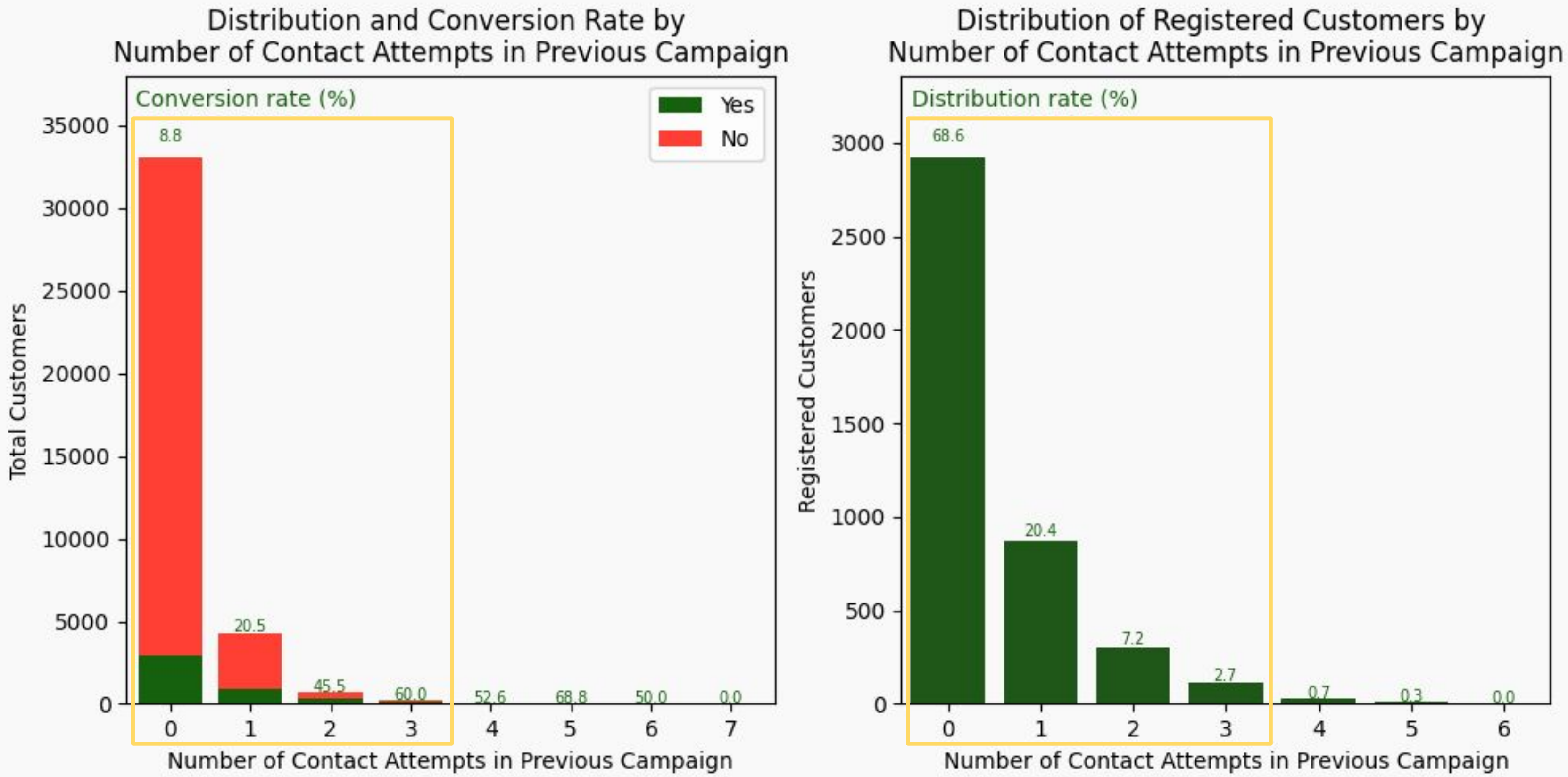
- Optimize interaction recency: within 2-15 days (effective conversion)
- Reallocate customers for efficient, avoid clustering on Days 3 & 6





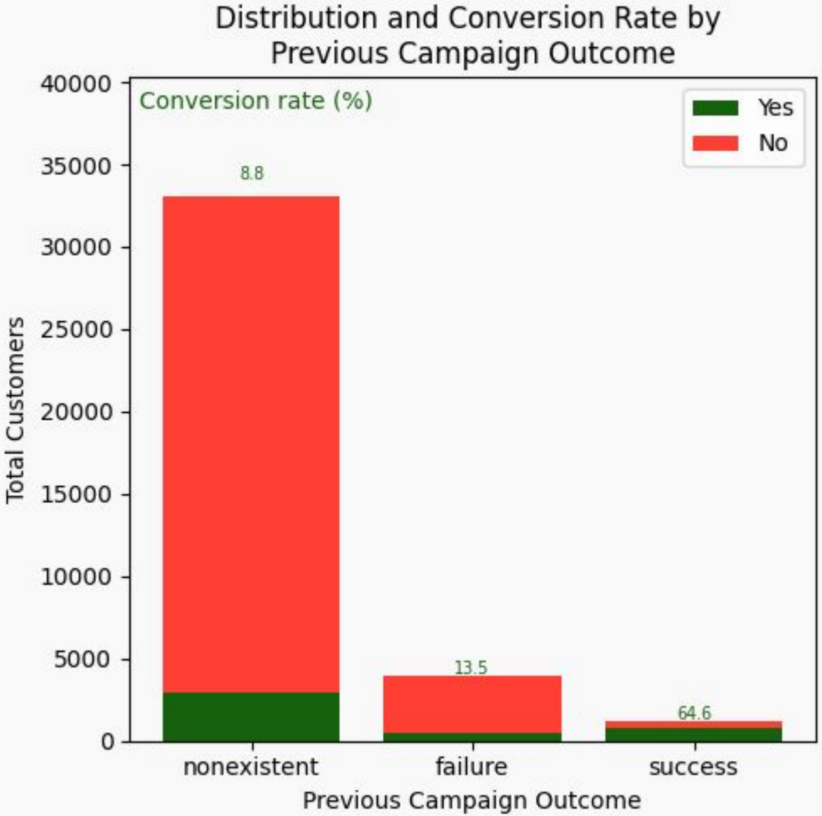
Previous campaign's number

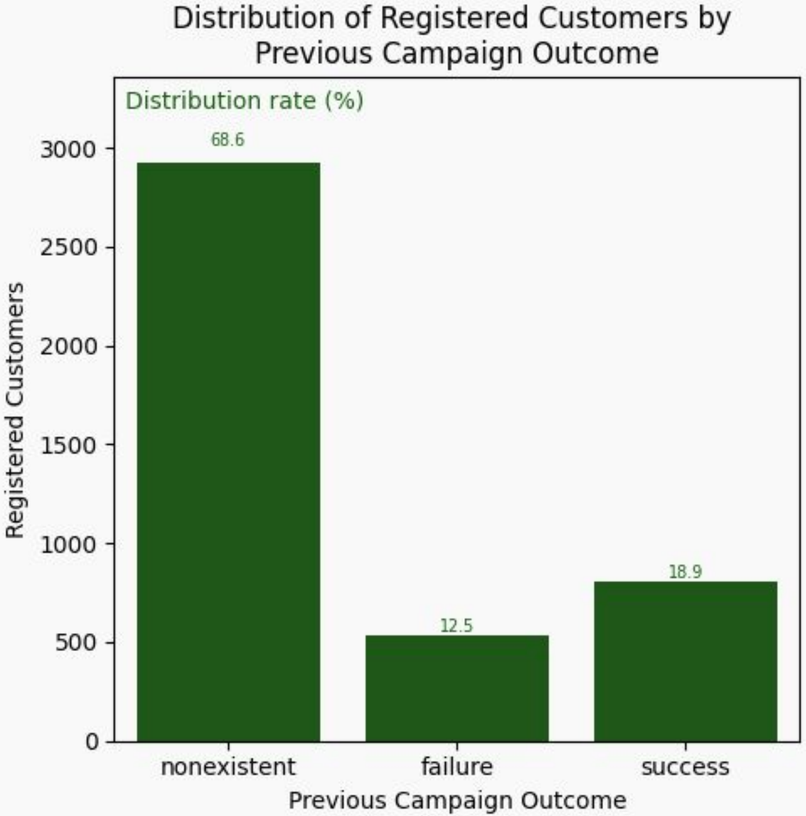
- Customers with 1–3 times in previous: focus on contacting (effective conversion)
- New leads: pre-qualify carefully and invest in trust-building first-touch content (strong resources)



Previous campaign outcome

- Previously converted leads: high-priority (high conversion rates)
- New leads: screen carefully, invest in first-touch content (strong resources)





Data-Driven Insights for Campaign Optimization
(The Optimal Customer Interaction through Phone Call)

Customer interaction analysis revealed clear engagement patterns, allowing us to categorize customers into high, moderate, and low engagement types for tailored strategy design:

	Recommendation
Contact Month	- Distribute campaign efforts more evenly throughout the year - Avoid overload between May to August
Contact Day	- Optimize outreach on Tue–Thu (with focus on Thursday) - Avoid low-efficiency days: Monday & weekends
Number of Contact Attempts	- Optimizing content, timing, and approach in the first 1–3 contact attempts - Consider a stop rule after 4 contacts unless there's positive engagement
Days since Last Campaign	- Focus contact within 2-15 days post-campaign - Can use gentle reminders

Data-Driven Insights for Campaign Optimization

(The Optimal Customer Interaction through Phone Call)

	Recommendation
Call Duration	- Optimal effectiveness at 6–20 mins - Can extend call (20+ mins) to build trust and clarity - Avoid exceeding 45 minutes
Number of Contact Attempts (Previous Campaign)	- Customers with 1–3 prior contacts: prioritize and personalized messaging that reinforces previously discussed benefits - New leads: pre-screen carefully, invest in strong first-touch marketing (introductory content and trust-building)
Previous Campaign Outcome	- Previously converted: prioritize, reconnect early - New leads: pre-screen carefully, invest in strong first-touch marketing

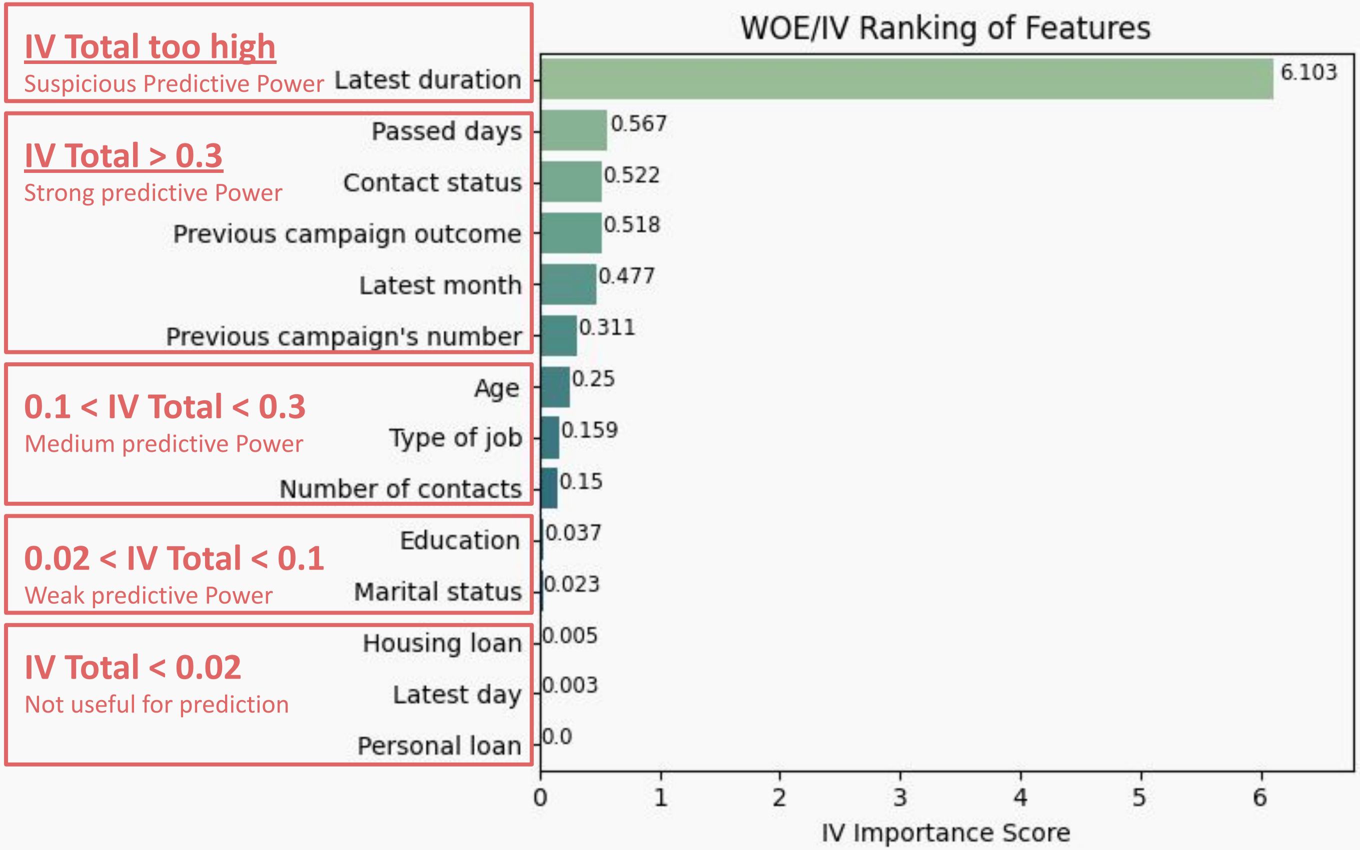
3. MODELING & EVALUATION

Data Preprocessing

- **Train–Validation–Test Split**
- **Feature selection (Woe & IV):**

Latest Call Duration: this feature poses a data leakage risk in real-world prediction scenarios prior to the call, because as longer calls may indicate higher customer interest.

→ Remove from training features



Data Transformation

(For Logistic Regression)

- **Remove columns:** 'Latest duration', 'Latest day', 'Personal loan', 'Housing loan' → Prevent data leakage and reduce redundant features
- **WOE Encoding:** for categorical variables → Converts categories into numerical values that reflect predictive power while preserving interpretability
- **Split into X (features) & y (target):** prepare data for training and evaluation in a structured way
- **StandardScaler:** standardize features for models like Logistic Regression that rely on gradient-based optimization

Data Transformation

(For Decision Tree)

- **Remove columns:** 'Latest duration' → Prevent data leakage
- **One-hot Encoding:** for categorical variables → Converts categories into binary numeric columns that preserve all information without imposing order
- **Split into X (features) & y (target):** prepare data for training and evaluation in a structured way

Model Performance
(Logistic Regression)

- The Logistic Regression model achieved a **reasonable Accuracy** (81.5%)
- Perform well in identifying non-buyers
- Ability to correctly predict actual buyers was quite limited (**Precision: 30%**)
- Ability to capture true buyers was better (**Recall: 52%**)
- The model did not achieve strong balance between precision and recall (**F1-score: 38%**)

→ Suitable for filtering out customers who are highly unlikely to make a purchase

```
--- Logistic Regression - Training Results ---
accuracy score: 0.8150
[[8685 1514]
 [ 608  664]]
              precision    recall  f1-score   support

         0.0       0.93      0.85      0.89     10199
         1.0       0.30      0.52      0.38      1272

    accuracy                           0.82     11471
   macro avg       0.62      0.69      0.64     11471
weighted avg       0.86      0.82      0.84     11471
```

(WOE Encoding and StandardScaler Model trained with class_weight='balanced')

Model Performance
(Decision Tree)

- The Decision Tree model delivered a **higher Accuracy** of approximately 86%
- Perform well for non-buyers
- Capability in correctly identifying buyers was not bad (**Precision: 38%**)
- Ability to capture all actual buyers was slightly better (**Recall: 43%**)
- Moderately improved balance between predicting correctly and minimizing missed opportunities (**F1-score: 40%**)

→ More appropriate for identifying potential customers to target for engagement

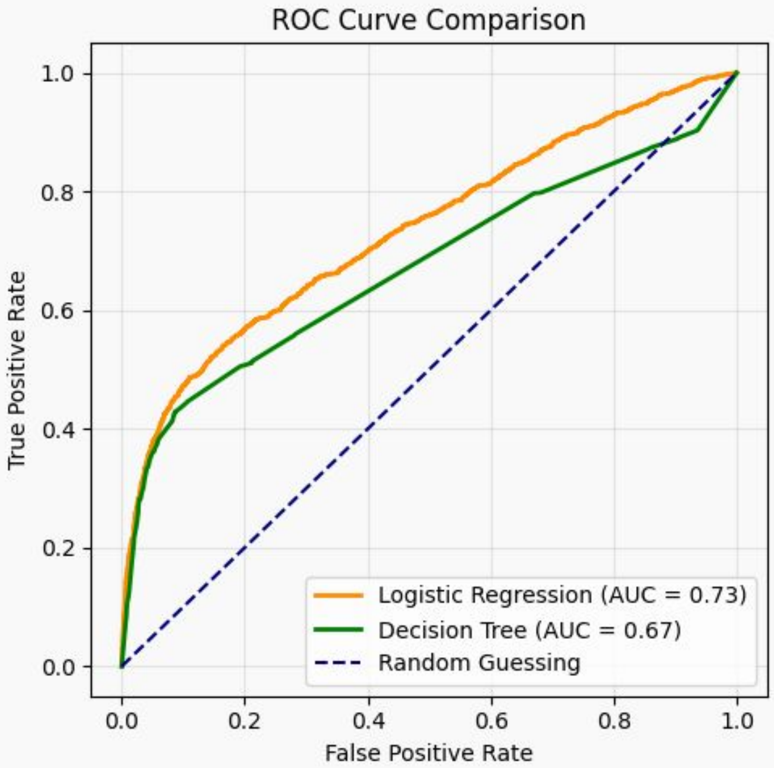
```
--- Decision Tree - Training Results ---
accuracy score: 0.8596
[[9318  881]
 [ 730  542]]
              precision    recall  f1-score   support

         0.0       0.93      0.91      0.92     10199
         1.0       0.38      0.43      0.40      1272

    accuracy                           0.86     11471
   macro avg       0.65      0.67      0.66     11471
weighted avg       0.87      0.86      0.86     11471
```

(One-hot Encoding Model trained with class_weight='balanced', criterion='entropy', max_depth=12)

Model Evaluation

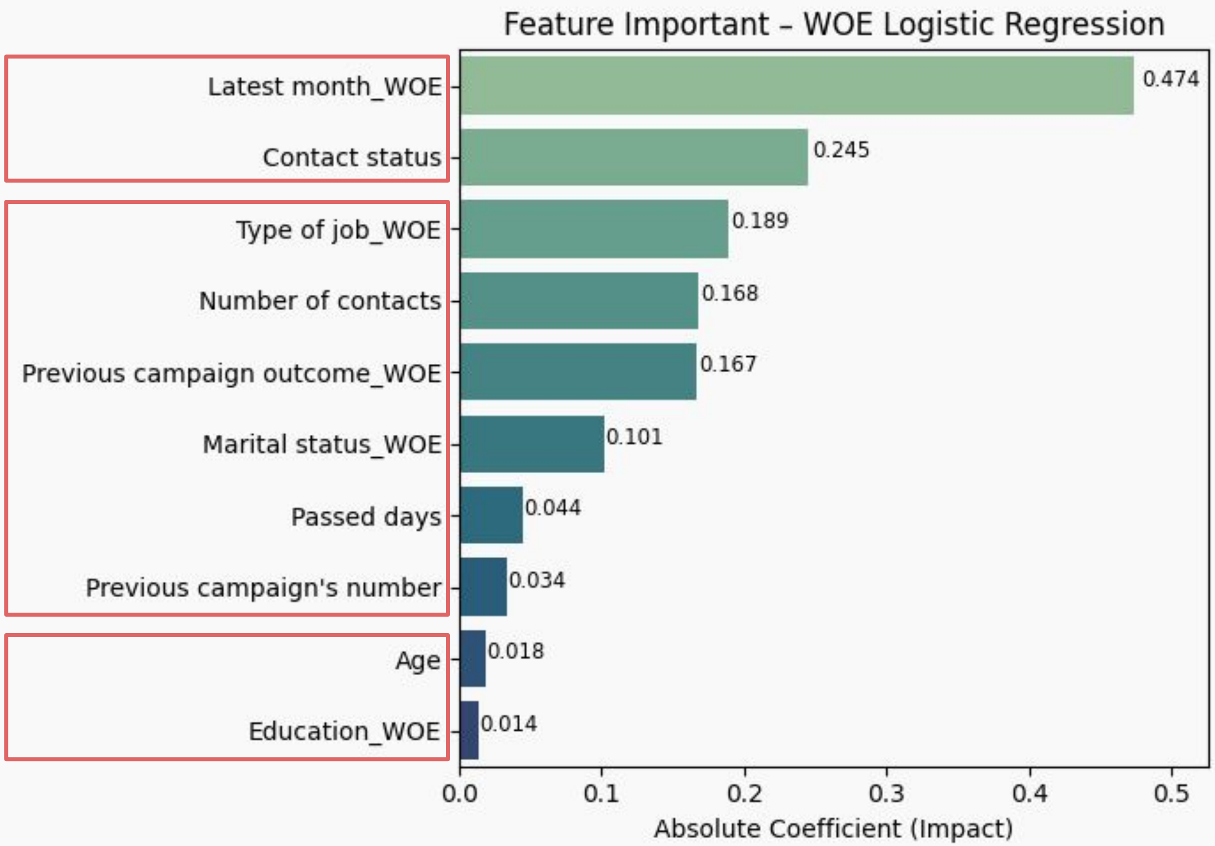


Metric	Logistic Regression	Decision Tree
Accuracy	81.5%	**85.96%** ✅
Precision	0.305	**0.381** ✅
Recall	**0.522** ✅	0.426
F1-score	0.385	**0.402** ✅
AUC	**0.734** ✅	0.672

→ Logistic Regression shows better separation ability.

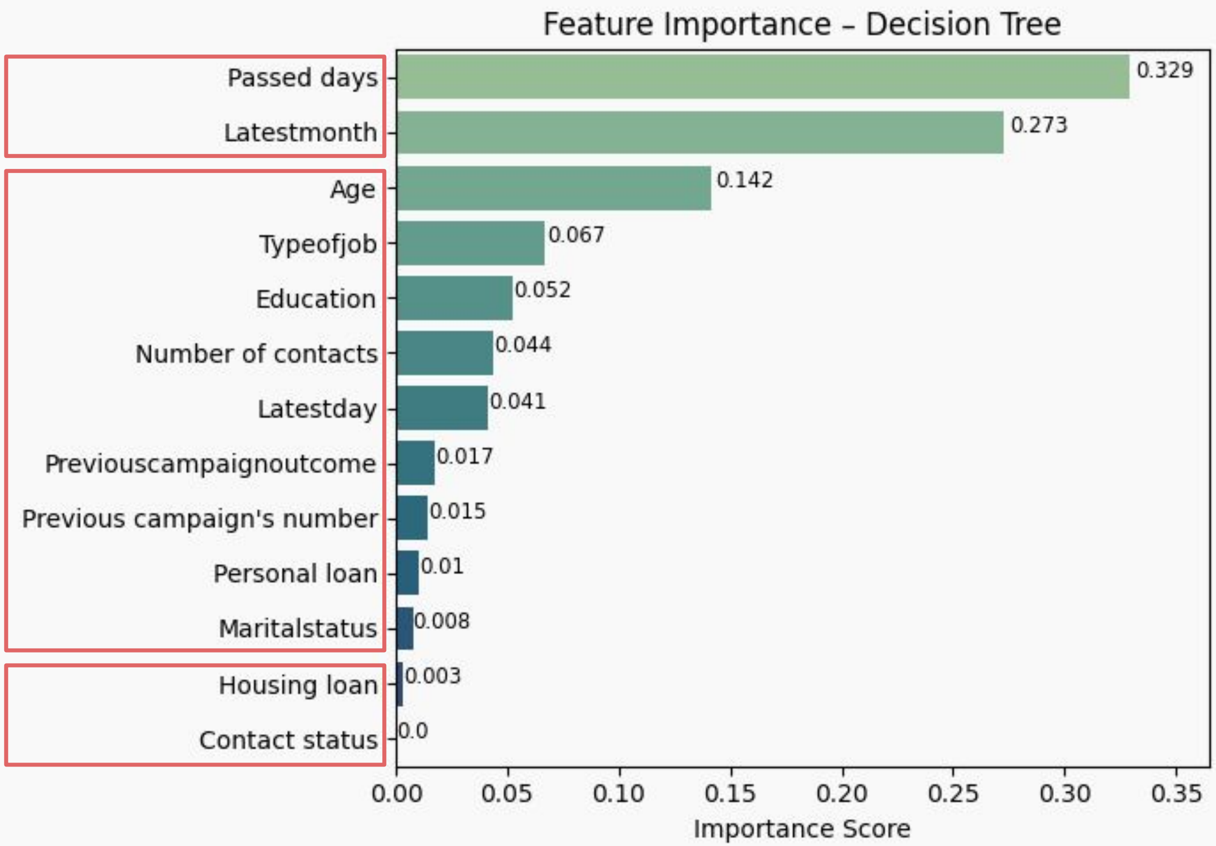
Feature Importance
(Logistic Regression)

- **Latest month** and **Contact status** are key features — useful for filtering high-potential customers.
- Variables such as **Job type** and interaction history enhance segmentation — ideal for designing personalized content.
- **Age** and **Education** have marginal impact — may be excluded to simplify the model.



Feature Importance
(Decision Tree)

- **Passed days** and **Latest month** are high-impact time-based features — helpful for designing campaigns based on customer engagement recency.
- Demographic variables (**Age, Type of job, Education**) assist in distinguishing potential customer segments.
- Some features like **Contact status** and **Housing loan** show negligible impact — may be excluded to simplify the model.



Model Recommendation

★ **Logistic Regression:**

- Stronger in Recall & AUC → Better at covering buyers
- Key features focus on customer interaction status (contact history, previous campaign outcomes)

★ **Decision Tree:**

- Stronger in Accuracy & Precision → Better at identifying high-potential customers
- Key features combine customer interaction status and customer demographics (age, education, job type) → Offers more intuitive prediction logic

→ **Recommended Model: Decision Tree — preferred for targeting potential buyers effectively**

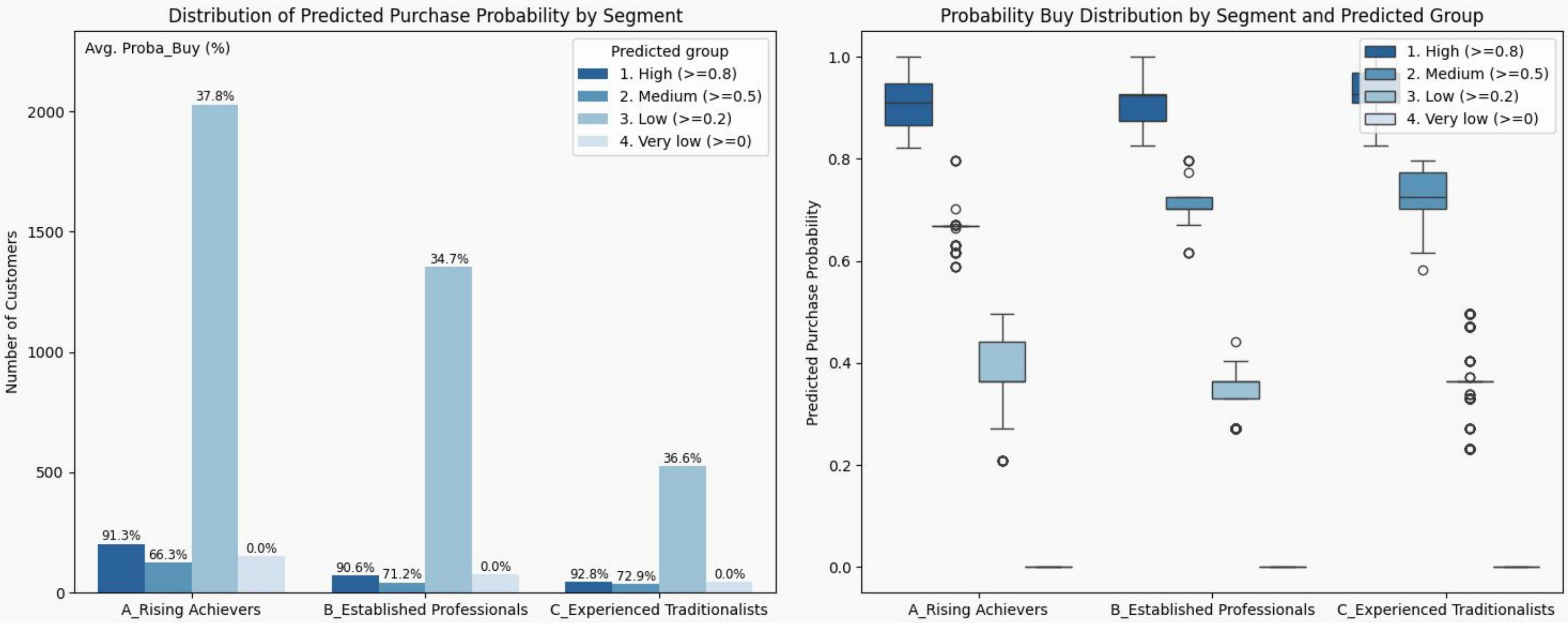
4. KEY FINDINGS & RECOMMENDATIONS

Applying the Predictive Model to Test Data

★ **Assumption:** This analysis is based on **a bank marketing** campaign targeting **term deposit** subscriptions. The campaign is designed for **long-term objectives** and aims to **optimize resource allocation**.

★ **Description:**

- The test dataset was categorized into the 3 customer segments and apply the predictive model to estimate purchase probabilities:
  + Segment A_Rising Achievers: 2509 customers
  + Segment B_Established Professionals: 1543 customers
  + Segment C_Experienced Traditionalists: 652 customers
  + Unassigned: 6767 customers



- **High group (90–93%):** Very high predicted rate. The model distinguishes clearly, strong concentration in prediction accuracy. → Should be prioritized as the primary target for campaign efforts
- **Medium group (66–72%):** Variable purchase rate by segment. The model's classification is moderately dispersed. → Should have tailored and personalized messaging, promising group if resources allow
- **Low group (34–38%):** Low predicted rate. The model doesn’t clearly distinguish, predictions are still a bit noisy. → Should be nurtured slowly or excluded
- **Very low (0%):** No purchase likelihood. The model distinguishes clearly → Should not pursuing

Key Findings & Actionable Recommendations
(Tailored Messaging & Channel Guide)

	Segment A *Rising Achievers*	**Segment B** *Established Professionals*	**Segment C** *Experienced Traditionalists*
Customer Demographics	Age: 26–35 Job: admin, technician, blue-collar, services Status: single, married Education: university, high school, basic.9y	Age: 36–50 Job: admin, technician, blue-collar, services Status: married Education: university, high school, basic.9y	Age: 51–60 Job: admin, blue-collar, retired, technician, management Status: married Education: university, high school, basic.4y
Key Messaging	- Financial freedom - Competitive interest rates - Flexible short-term options	- Highlight long-term goals - Financial security - Risk-free growth	- Safety - Stable interest rates - Direct customer support
Suggested Channels	- Digital campaigns - In-app notifications - Social media ads	- Email marketing - Direct consultation - Loyalty/member programs	- Phone consultations - In-branch brochures - Exclusive customer seminars

Key Findings & Actionable Recommendations

(Interaction Strategy by Customer Probability Segment)

High (0.8 - 1) Direct offers to trigger purchase	**Medium (0.5 - 0.8)** Personalized content, nurturing engagement	**Low (0.2 - 0.5)** Light nurturing, exclude if necessary	**Very low (0 - 0.2)** Maintain brand awareness only
Phone call ***(can be based on the optimal customer interaction analysis discussed above)***			
Strongly recommended — effective for quick conversion	Recommended selectively — use for personalized guidance	Not prioritized — phone calls are not cost-effective here	Not recommended — little to no return from calling efforts
Email / Messaging			
Direct offers, strong call-to-action messages	Personalized messaging based on profile and behavior	Occasional brand-focused emails — maintain light connection	Inspirational, community-oriented content — avoid sales pitch

Key Findings & Actionable Recommendations

(Interaction Strategy by Customer Probability Segment)

High (0.8 - 1) Direct offers to trigger purchase	**Medium (0.5 - 0.8)** Personalized content, nurturing engagement	**Low (0.2 - 0.5)** Light nurturing, exclude if necessary	**Very low (0 - 0.2)** Maintain brand awareness only
Advertising / Retargeting			
Targeted ads with clear CTA to drive conversions	Behavioral retargeting with tailored messaging	Optional light retargeting — nurture gently	Not recommended — low ROI for ad spend
Chatbot / Automated Assistance			
Conversion-focused — order placement, offer activation	Quick answers, informative guidance, support logic	Light touchpoint — minimal resources needed	Not necessary — avoid deploying chatbot resources here

THANK YOU FOR YOUR LISTENING!

(Q&A)