OPTIMIZING CAMPAIGN STRATEGY WITH PREDICTIVE ANALYTICS

Client Subscription Patterns in Retail Banking

Moro, S., Rita, P., & Cortez, P. (2014). Bank Marketing [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5K306.

YU-JEN WANG May/2025

SEP 2023 - SEP 2024

Master's degree Financial Technology | University of Bristol

Learn about big data analysis of financial data and AI applications

AUG 2021 - MAY 2023

Senior Engineer, Optical Design Dept., INNOLUX Co. Ltd

Set up optical automated inspection equipment and develop image inspection software using C#

SEP 2017 - JUN 2020

Master's degree Applied Mathematics Computational Science | National Chung Hsing University

Learn numerical methods and image processing principles

SEP 2013 - JUN 2017

Bachelor's degree Applied Mathematics | National Chung Hsing University

Develop logical training and algorithm knowledge

BACKGROUND

PROJECT

Installation of automated optical inspection production line in China

Purpose: Improve production yield through optical inspection.
Setup: Installed optical hardware, customized software to meet factory specs, and integrated result feedback to the production line.
Role: Acted as liaison between the Taiwan and China factories.
Challenge: Managed the entire setup independently, requiring cross-department collaboration and communication.
Outcome: Completed installation within deadline, resulting in improved yield.

Compety Hackathon: UX A/B Test for Conversion Lift

Purpose: Improve conversion rates via A/B testing of food image sizes.
Setup: Cleaned 300K+ records, applied stratified sampling, and ran T-tests/Chi-Square tests.
Role: Led end-to-end analysis and insights presentation.
Challenge: Dealt with missing data, platform bias, and external seasonal effects.
Outcome: Identified significant lift in conversions ($p < 0.001$); ranked in 6 of 42.

PROJECT

Cross-National Analysis of Mobile Money Adoption

Purpose: Analyze socioeconomic drivers of mobile money adoption using panel data.

Setup: Applied FE and GMM models to address heterogeneity and endogeneity across 27 countries.

Role: Led the GMM analysis; collaborated with teammates from the UK, Spain, and Indonesia to align research direction.

Challenge: Managed cross-cultural coordination and synthesized diverse analytical perspectives.

Outcome: Identified key dynamic drivers such as prior adoption and urbanization; the project received a high score of 72/100.

Analyst Sentiment and Corporate Investment Strategy

Purpose: Explore how analyst sentiment affects corporate investment efficiency under macroeconomic and policy pressures.

Setup: Analyzed 232 reports from 24 S&P 500 firms (2019–2024); applied VADER sentiment analysis and panel regressions with interaction effects.

Role: Led the full research process, including data collection, sentiment scoring, and econometric modeling.

Challenge: Overcame data inconsistencies across Bloomberg reports and tight project timelines.

Outcome: Identified significant links between analyst sentiment and investment behavior; received a final grade of 65 for the dissertation.

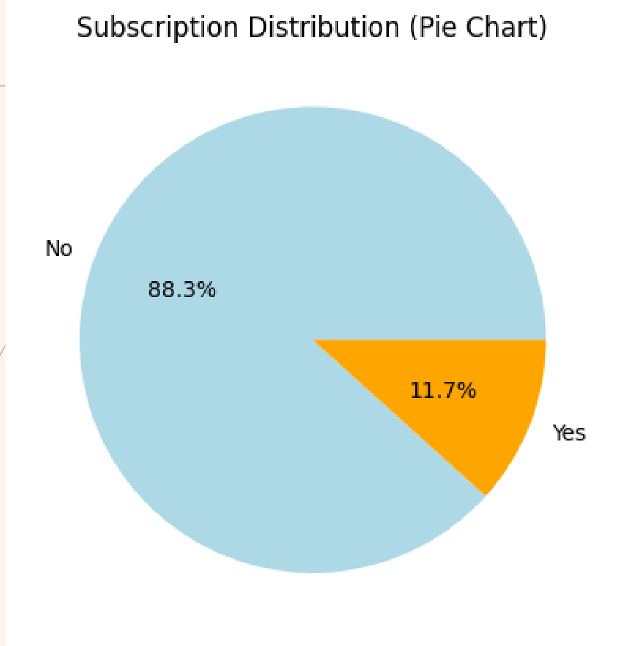
University of BRISTOL

OPTIMIZING CAMPAIGN STRATEGY WITH PREDICTIVE ANALYTICS

TABLE OF CONTENTS

EXPERIMENTAL PURPOSE AND METHODS

Purpose: Understand which client segments are most likely to subscribe to term deposits, using UCI Bank Marketing Dataset.



Exploratory data analysis

Statistical testing (t-tests and chi-square) to identify key drivers

Predictive modeling (Logistic Regression & Random Forest)

Feature importance comparison and business recommendations

Rows	45211
Features	16
Feature Type	Categorical, Integer

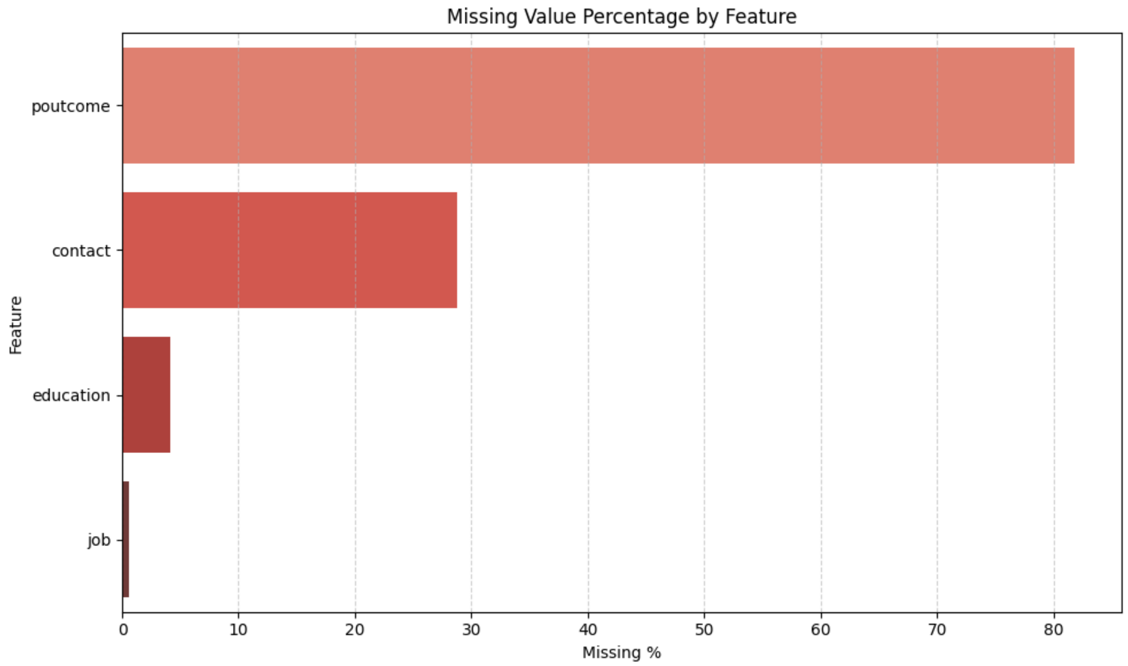
DESCRIPTIVE ANALYSIS – CLIENT PROFILES

Categorical Insights: Appendix 1

- Most clients are married, blue-collar, and have a secondary education.
- Campaigns were mostly conducted in May, July, and August, indicating strong seasonality.

Numeric Insights: Appendix 2

- Clients are mainly aged between 30–40.
- Most have low campaign exposure and short contact duration, but a few outliers skew the distribution.
- Many clients have no previous or recent contact, suggesting untapped potential.



"Contact" and "Poutcome" features have over 25–80% missing values.

STATISTICAL TESTS – WHAT DRIVES SUBSCRIPTIONS?

$H_0 =$ The means of (y=1 and y=0) are equal

$H_1 =$ The means of (y=1 and y=0) are different

T-Test Results (Numeric Variables):

- Longer client conversations are positively linked with conversion, emphasizing quality engagement over quantity.
- Follow-ups and recent contact (pdays, previous) also increased success.
- Excessive contact attempts may trigger client fatigue, reducing subscriptions.

Feature	t_stat	df	p_value	mean_0	mean_1	std_0	std_1	n_0	n_1	Significant
age	4.3183	6109.2	1.60E-05	40.84	41.67	10.17	13.5	39922	5289	Yes
balance	9.9335	6339.8	0	1303.71	1804.27	2974.2	3501.1	39922	5289	Yes
duration	57.5141	5685.31	0	221.18	537.29	207.38	392.53	39922	5289	Yes
campaign	-22.8007	9751.56	0	2.85	2.14	3.21	1.92	39922	5289	Yes
pdays	18.9435	6251.5	0	36.42	68.7	96.76	118.82	39922	5289	Yes
previous	18.118	6430.14	0	0.5	1.17	2.26	2.55	39922	5289	Yes

T-Test Summary

$H_0 =$ The variable is independent of y

$H_1 =$ The variable is associated with y

Chi-Square Results (Categorical Variables):

- Month had the strongest association, supporting observed seasonality in successful campaigns.
- Job, education, and marital status also played a significant role, aligning with earlier demographic insights.

Feature	Chi2	df	p_value	Significant	n_categories
job	836.1448	10	0	Yes	11
marital	196.4959	2	0	Yes	3
education	233.7465	2	0	Yes	3
month	3061.8389	11	0	Yes	12
day_of_week	574.0506	30	0	Yes	31

Chi-Square Test Summary

PREDICTIVE MODELING AND FEATURE INTERPRETATION

Model Performance Comparison:

- XGBoost had higher overall accuracy (86.7%) and better F1-score for subscribers (57.5%), showing strong balance between precision and recall.
- Logistic Regression had lower accuracy (82.9%) but achieved the highest recall for subscribers (79.6%), useful for identifying more potential customers.
- XGBoost also led in macro and weighted averages, indicating better overall performance across both classes.

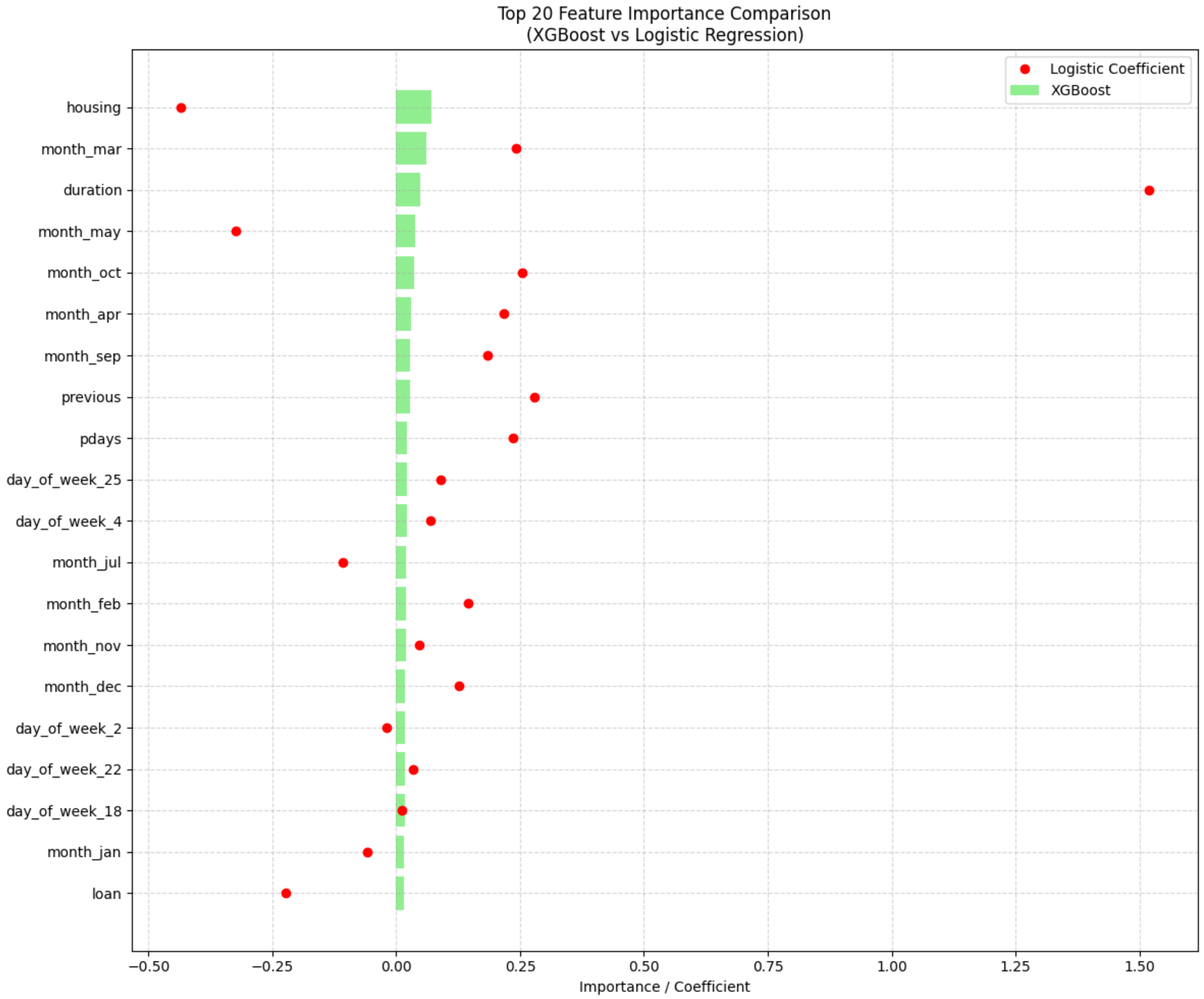
Choose XGBoost for balanced performance and higher precision, while Logistic Regression is preferred when maximizing subscriber recall is the primary goal.

	Logistic Regression	XGBoost
Accuracy	0.8294	**0.8669**
Precision (Class 0)	0.9686	0.9665
Precision (Class 1)	0.3883	0.4589
Recall (Class 0)	0.8338	0.8797
Recall (Class 1)	0.7958	0.7700
F1-score (Class 0)	0.8962	0.9211
F1-score (Class 1)	0.5219	**0.5751**
Macro Avg Precision	0.6784	**0.7127**
Macro Avg Recall	0.8148	**0.8248**
Macro Avg F1-score	0.7090	**0.7481**
Weighted Precision	0.9007	**0.9071**
Weighted Recall	0.8294	**0.8669**
Weighted F1-score	0.8524	**0.8806**

PREDICTIVE MODELING AND FEATURE INTERPRETATION

Key Feature Interpretation:

- Call duration was the strongest predictor, longer calls led to more subscriptions.
- Follow-up history (pdays, previous) played a key role in XGBoost, showing past contact boosts success.
- Loan and housing status had moderate influence, reflecting financial readiness.
- Month-related features (e.g., month_mar, month_may) were more emphasized in Logistic Regression, highlighting seasonal effects.



BUSINESS INSIGHT – WHEN & WHO TO TARGET

Top 5 Customer Groups by Subscription Rate

Loan Status	Education	Job	Subscription Rate
yes	primary	unknown	1
no	primary	student	0.363636
no	secondary	student	0.300797
no	tertiary	retired	0.29321
no	unknown	retired	0.272727

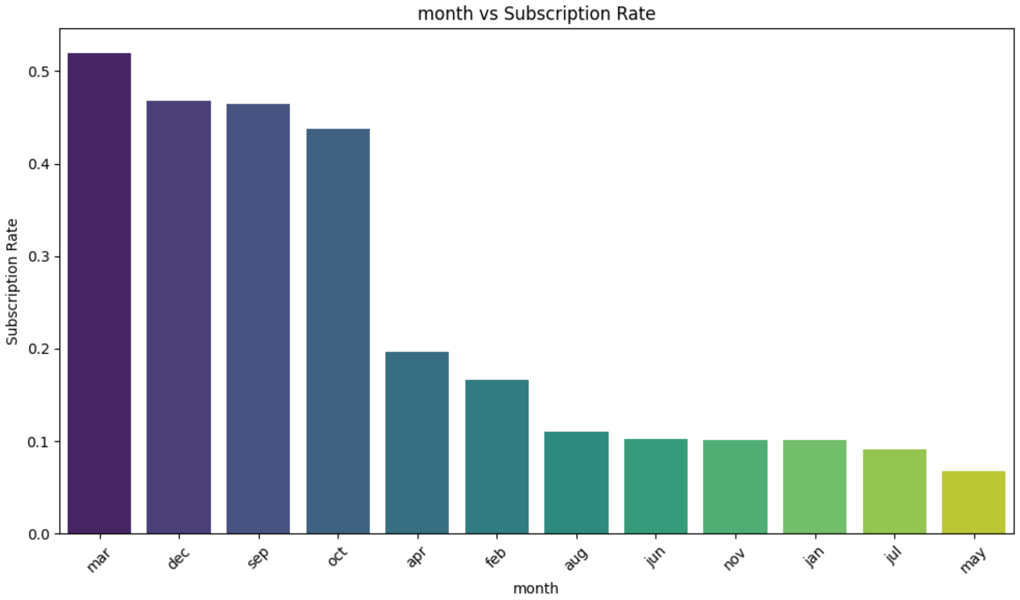
Top 5 Customer Segments:

- High engagement: Students, retirees, and managers.
- Education: Primary and tertiary levels often linked to higher education.
- No loan: Common among top segments, suggesting less debt encourages deposit interest.

When to Target

•**Best months:** March, December, and September — highest subscription rates.

•**Avoid:** May to July — low conversion despite frequent contact, indicating potential fatigue.



Business Implications

Target wisely: Focus on students and retirees without loans for better conversion.

Time it right: Run campaigns in March, December, and September.

Tailor messages: Highlight savings for retirees and easy entry for students.

Avoid waste: Skip broad outreach in May–July due to low response.

CONCLUSIONS & RECOMMENDATIONS

Conclusions:

Target the right profiles (age 30–40, no loan, students/retirees), focus campaigns in March, September, and December, and prioritize longer, high-quality calls over frequent contact.

Recommendations:

1. **Target high-potential groups**
   Focus on students and retirees with low debt and primary/tertiary education.
2. **Time campaigns wisely**
   Run promotions in March, December, and September. Avoid May–July.
3. **Improve call strategy**
   Longer calls work better. Avoid frequent short contacts.
4. **Use XGBoost for targeting**
   It balances precision and recall well, helping identify likely subscribers.

THANK YOU

YU-JEN WANG

+886 905514547/+44 7774652480

Ztex060033153@gmail.com

https://github.com/Naenaedadader

https://www.linkedin.com/in/yu-jen-wang-7692b2268/

APPENDIX 1

Categorical Variables:

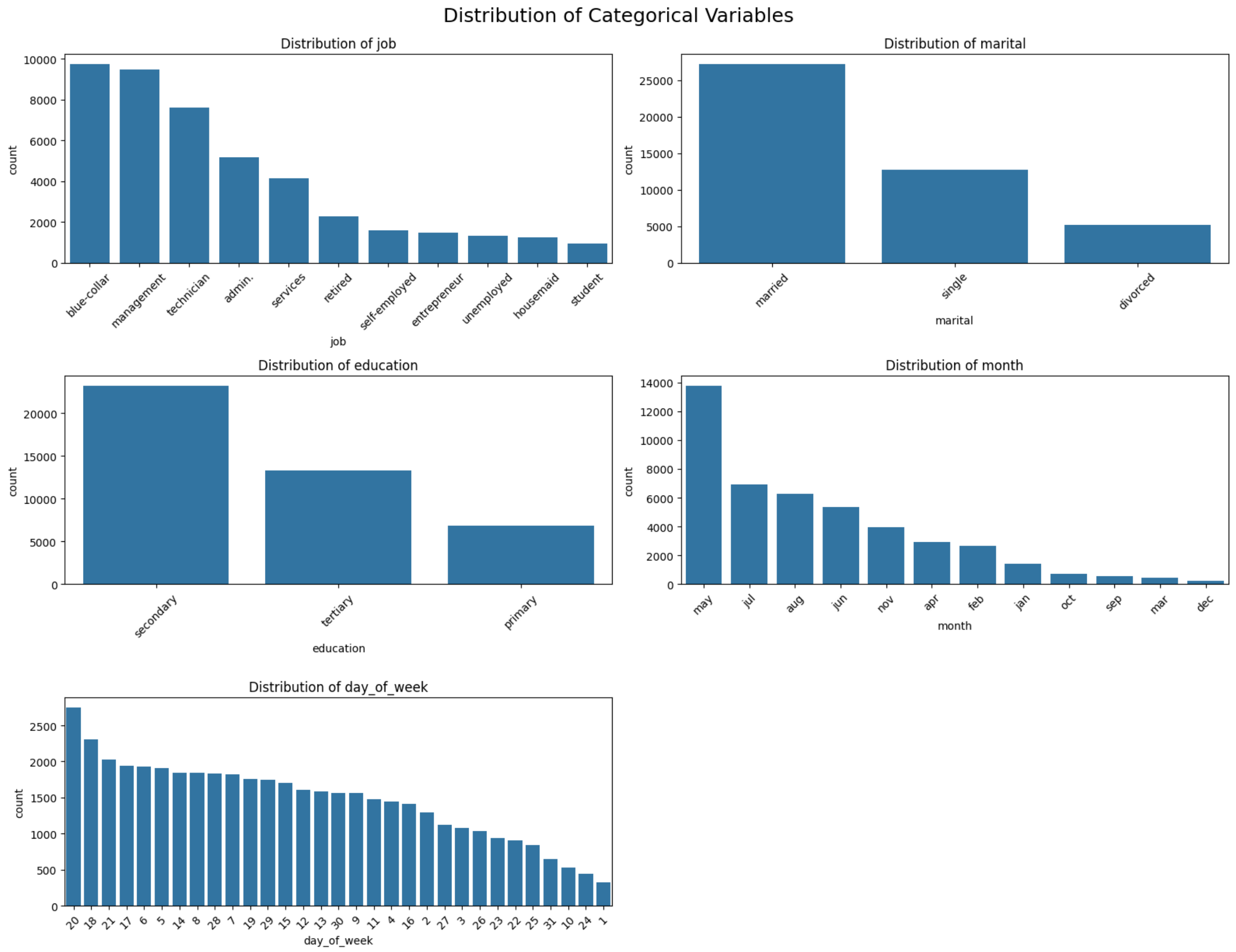
The most common job types are blue-collar, management, and technician, while students and housemaids are least represented.

The majority of clients are married, with fewer being single or divorced.

In terms of education, secondary education dominates, followed by tertiary and primary.

Contact month shows strong concentration in May, July, and August, suggesting campaign seasonality.

Day-of-week distribution is roughly uniform but slightly skewed toward day 20 and day 12 (likely due to encoding practices).



APPENDIX 2

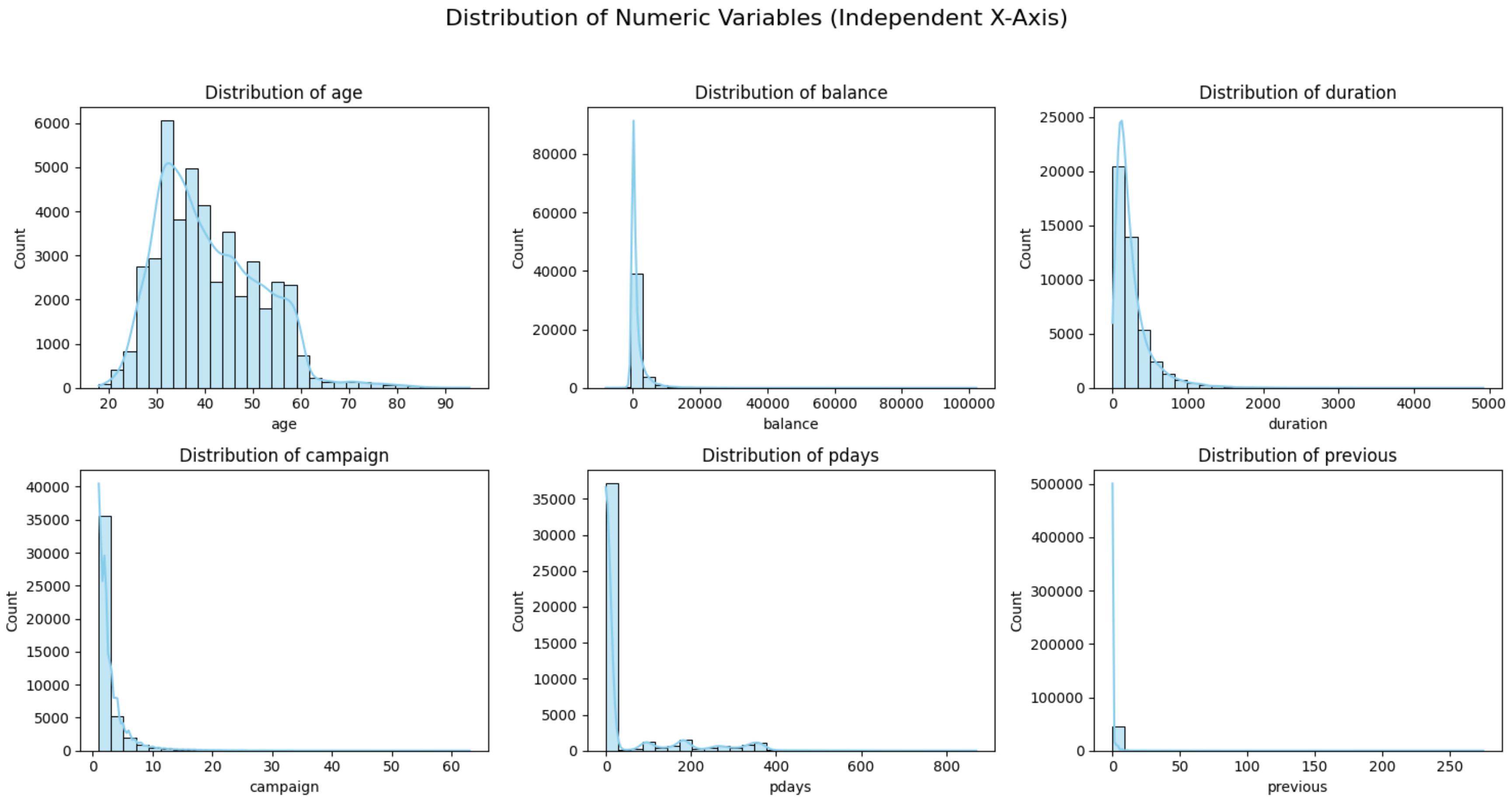
Numeric Variables:

Age follows a right-skewed distribution with a peak around 30–40 years old.

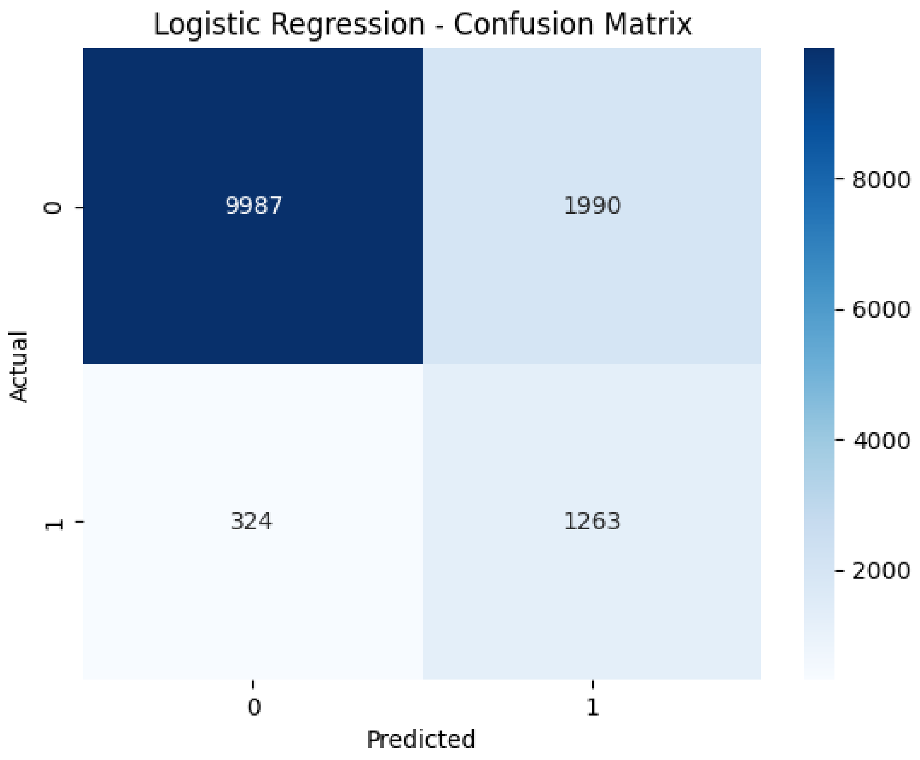
Balance and duration exhibit extreme right skewness with many low values and a long tail of high values.

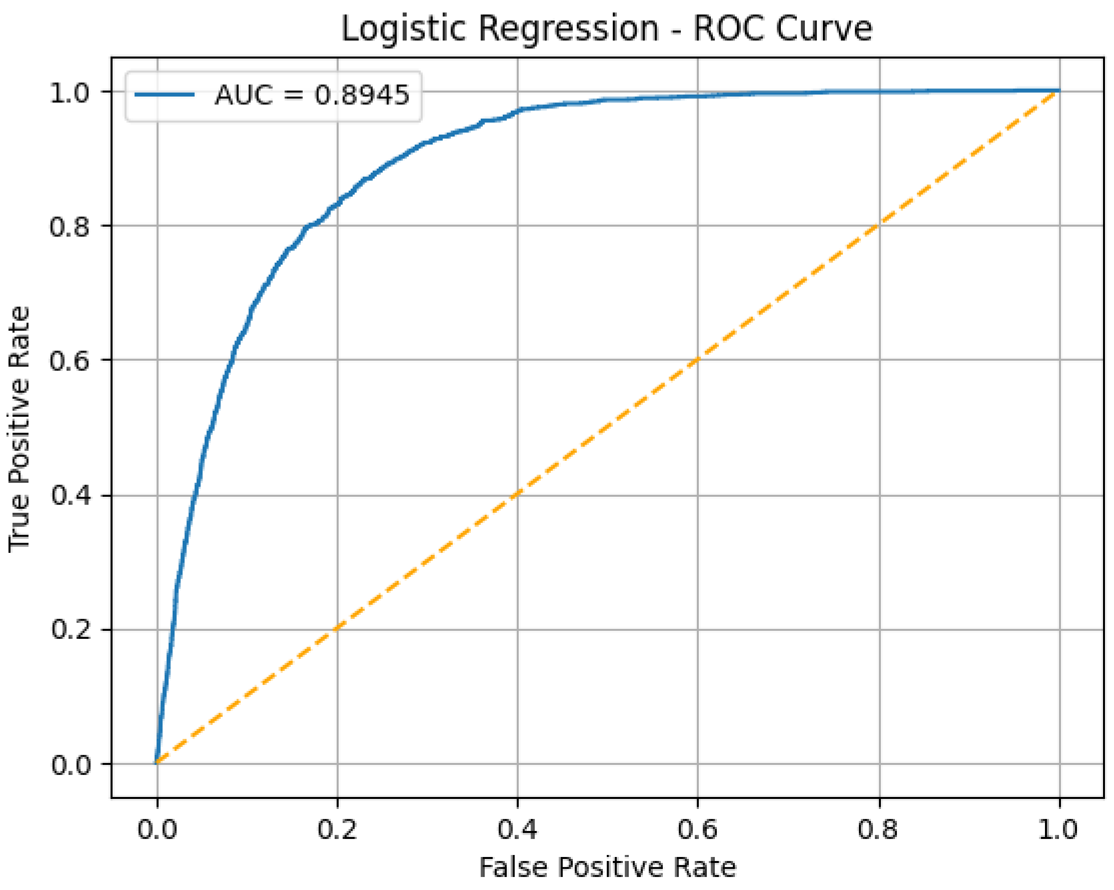
Campaign, pdays, and previous also have heavy tails, indicating that most clients were contacted only a few times.

Many clients had zero previous contact or no recent contact (pdays = 999), highlighting a large untouched base.

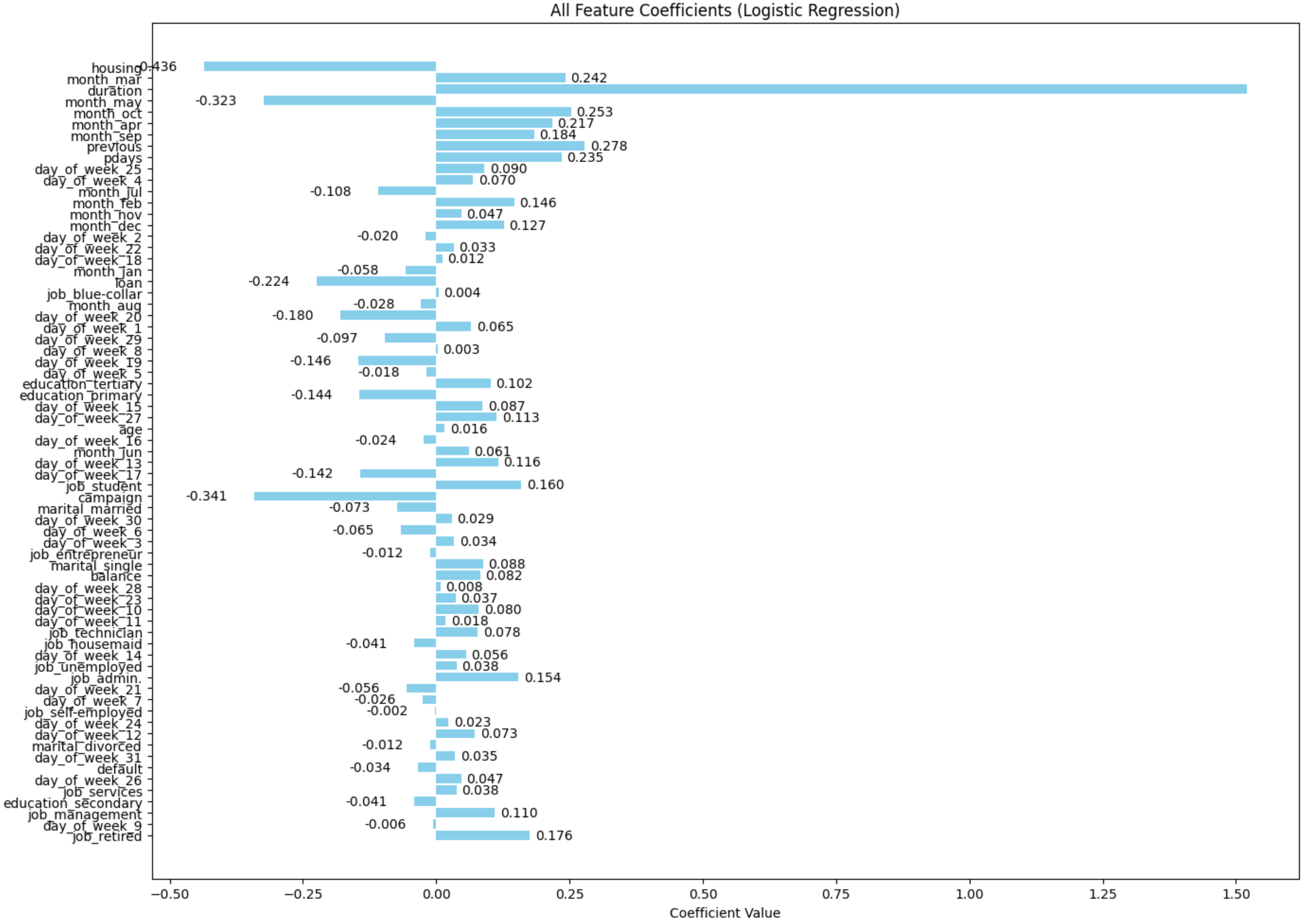


APPENDIX 3

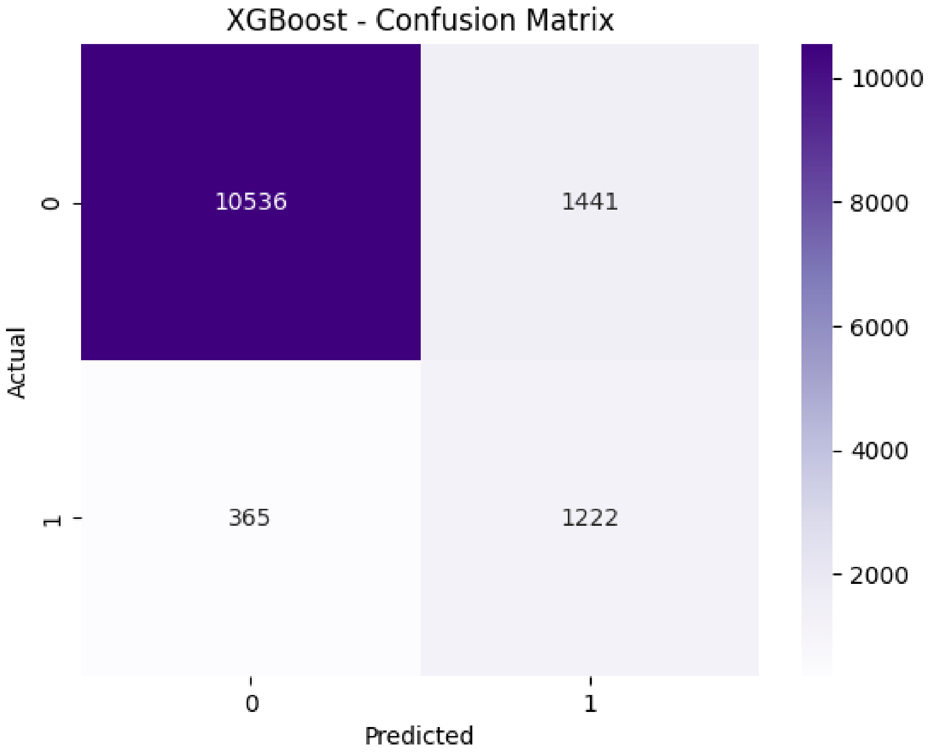

Logistic Regression - Confusion Matrix
9987
1990
324
1263
8000
6000
4000
2000
0
1
Actual
0
1
Predicted

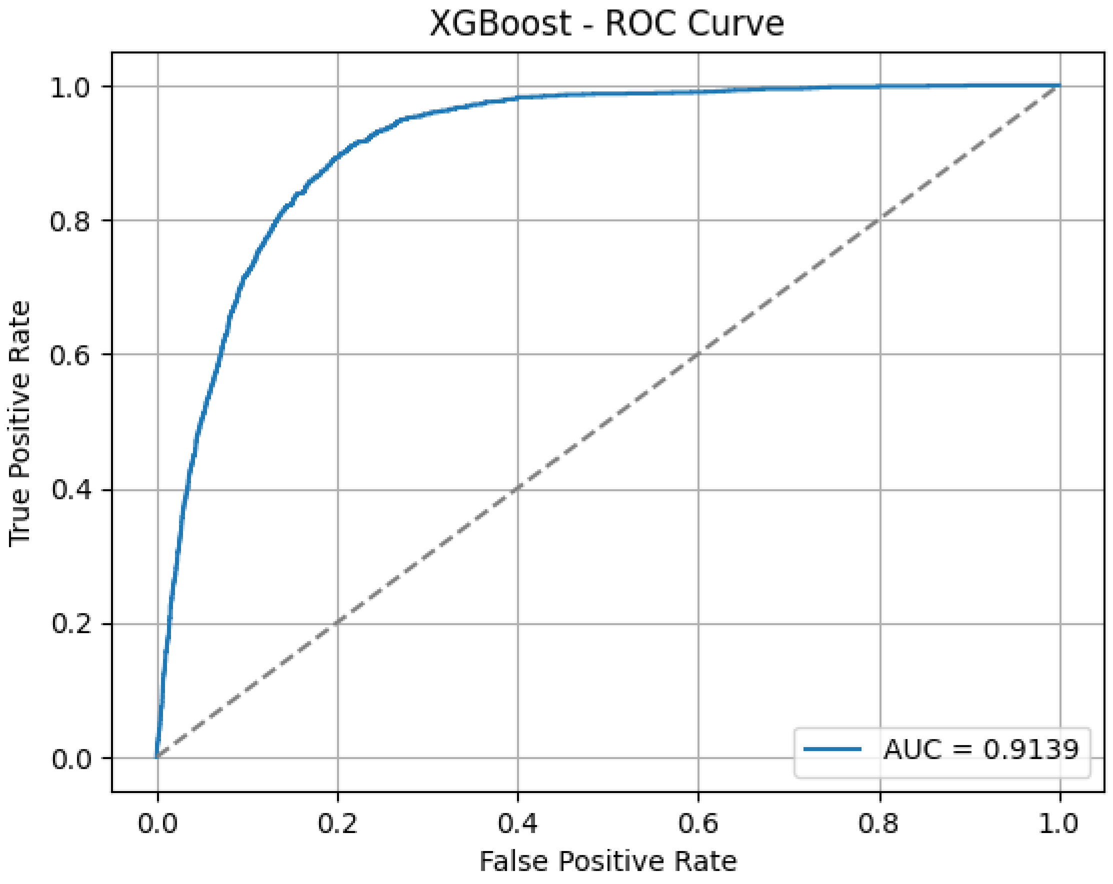

Logistic Regression - ROC Curve
AUC = 0.8945
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate

APPENDIX 4

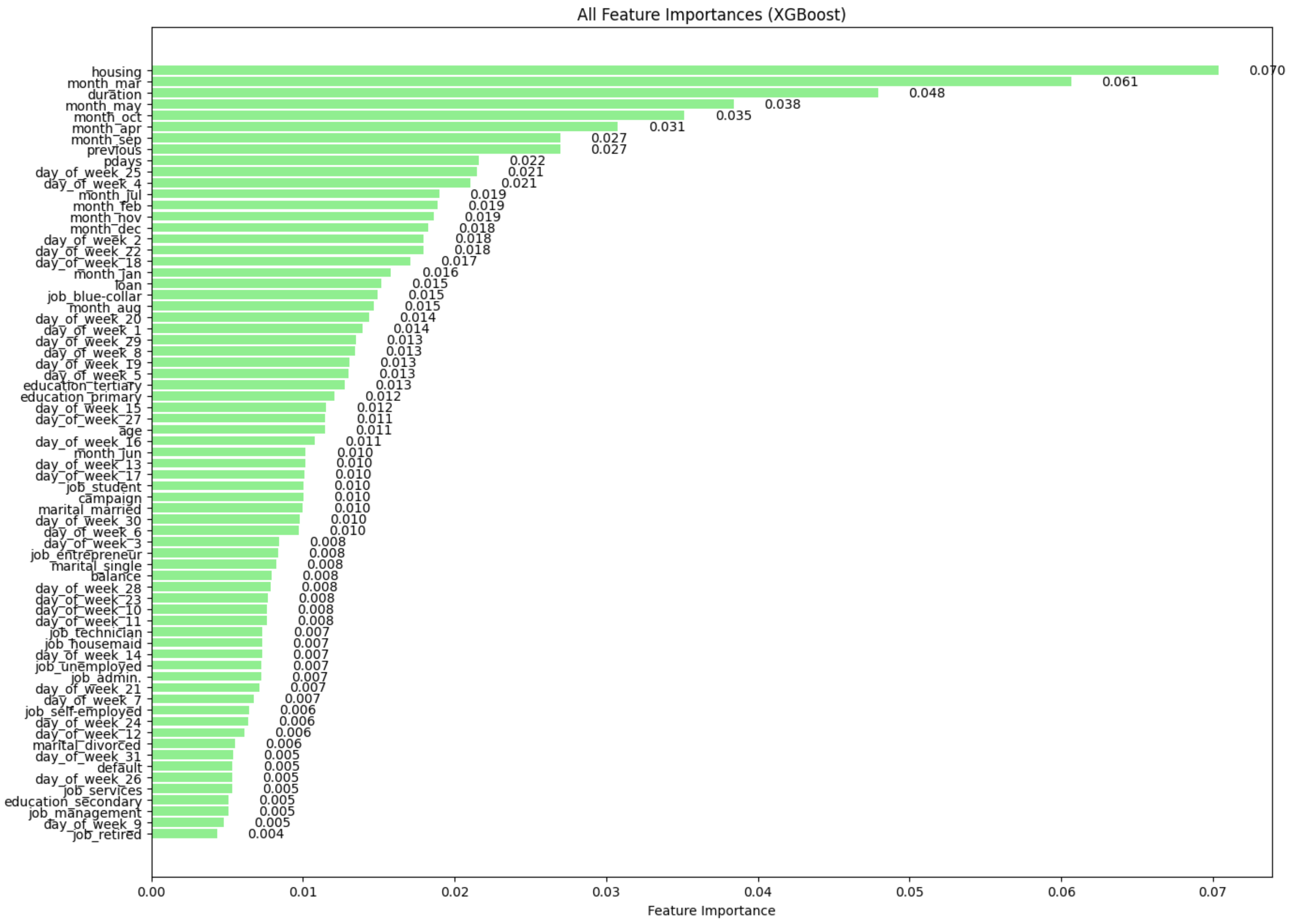

All Feature Coefficients (Logistic Regression)
housing -0.436
month_mar 0.242
duration
month_may -0.323
month_oct 0.253
month_apr 0.217
month_sep 0.184
previous 0.278
pdays 0.235
day_of_week_25 0.090
day_of_week_4 0.070
month_jul -0.108
month_feb 0.146
month_nov 0.047
month_dec 0.127
day_of_week_2 -0.020
day_of_week_22 0.033
day_of_week_18 0.012
month_jan -0.058
loan -0.224
job_blue-collar 0.004
month_aug -0.028
day_of_week_20 -0.180
day_of_week_1 0.065
day_of_week_29 -0.097
day_of_week_8 0.003
day_of_week_19 -0.146
day_of_week_5 -0.018
education_tertiary 0.102
education_primary -0.144
day_of_week_15 0.087
day_of_week_27 0.113
age 0.016
day_of_week_16 -0.024
month_jun 0.061
day_of_week_13 0.116
day_of_week_17 -0.142
job_student 0.160
campaign -0.341
marital_married -0.073
day_of_week_30 0.029
day_of_week_6 -0.065
day_of_week_3 0.034
job_entrepreneur -0.012
marital_single 0.088
balance 0.082
day_of_week_28 0.008
day_of_week_23 0.037
day_of_week_10 0.080
day_of_week_11 0.018
job_technician 0.078
job_housemaid -0.041
day_of_week_14 0.056
job_unemployed 0.038
job_admin. 0.154
day_of_week_21 -0.056
day_of_week_7 -0.026
job_self-employed -0.002
day_of_week_24 0.023
day_of_week_12 0.073
marital_divorced -0.012
day_of_week_31 0.035
default -0.034
day_of_week_26 0.047
job_services 0.038
education_secondary -0.041
job_management 0.110
day_of_week_9 -0.006
job_retired 0.176
−0.50 −0.25 0.00 0.25 0.50 0.75 1.00 1.25 1.50
Coefficient Value

APPENDIX 5

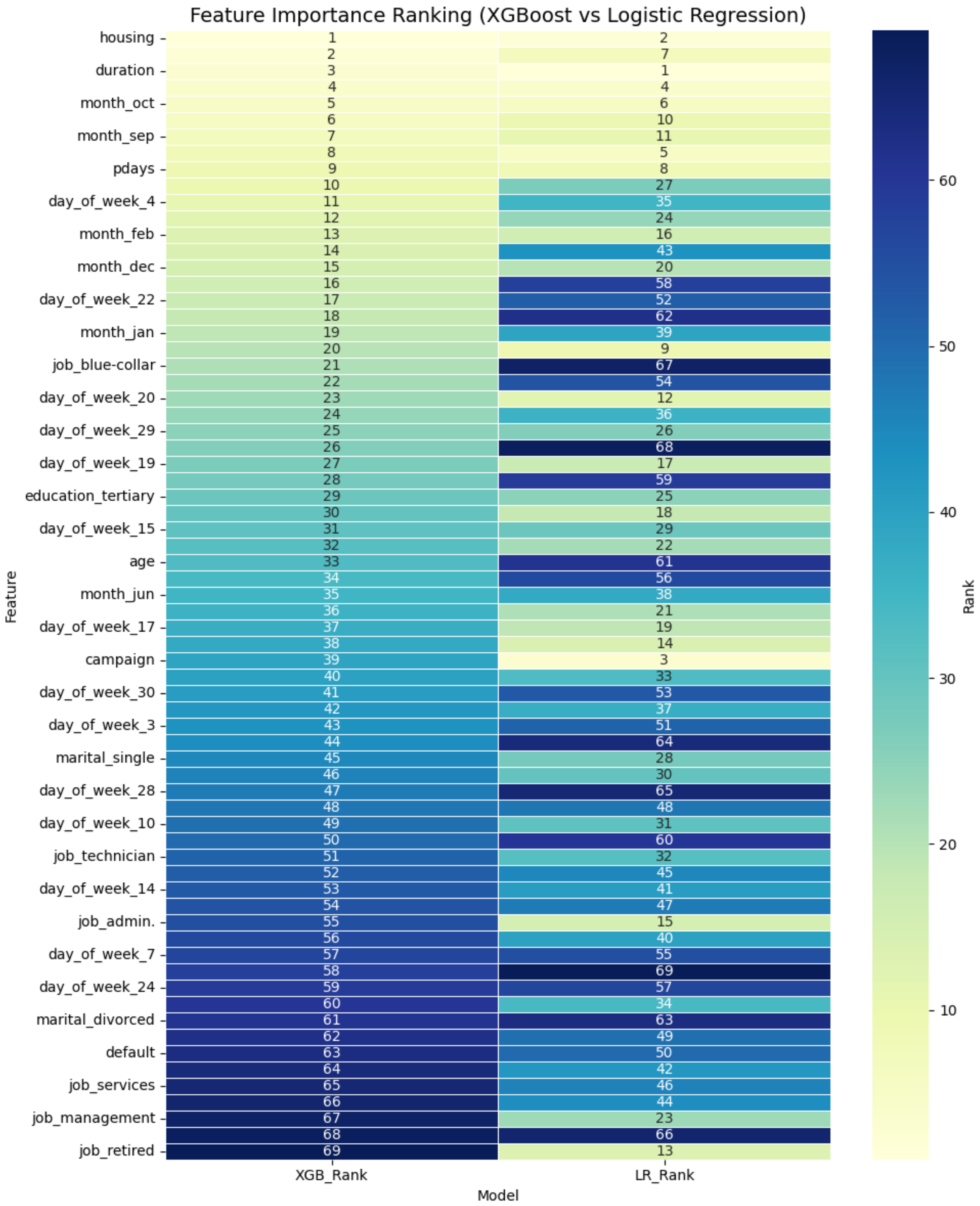

XGBoost - Confusion Matrix
10536
1441
365
1222
Actual
0
1
Predicted
0
1
10000
8000
6000
4000
2000


XGBoost - ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.9139

APPENDIX 6


All Feature Importances (XGBoost)
housing
month_mar
duration
month_may
month_oct
month_apr
month_sep
previous
pdays
day_of_week_25
day_of_week_4
month_jul
month_feb
month_nov
month_dec
day_of_week_2
day_of_week_22
day_of_week_18
month_jan
loan
job_blue-collar
month_aug
day_of_week_20
day_of_week_1
day_of_week_29
day_of_week_8
day_of_week_19
day_of_week_5
education_tertiary
education_primary
day_of_week_15
day_of_week_27
age
day_of_week_16
month_jun
day_of_week_13
day_of_week_17
job_student
campaign
marital_married
day_of_week_30
day_of_week_6
day_of_week_3
job_entrepreneur
marital_single
balance
day_of_week_28
day_of_week_23
day_of_week_10
day_of_week_11
job_technician
job_housemaid
day_of_week_14
job_unemployed
job_admin.
day_of_week_21
day_of_week_7
job_self-employed
day_of_week_24
day_of_week_12
marital_divorced
day_of_week_31
default
day_of_week_26
job_services
education_secondary
job_management
day_of_week_9
job_retired
0.070
0.061
0.048
0.038
0.035
0.031
0.027
0.027
0.022
0.021
0.021
0.019
0.019
0.019
0.018
0.018
0.018
0.017
0.016
0.015
0.015
0.015
0.014
0.014
0.013
0.013
0.013
0.013
0.013
0.012
0.012
0.011
0.011
0.011
0.010
0.010
0.010
0.010
0.010
0.010
0.010
0.010
0.008
0.008
0.008
0.008
0.008
0.008
0.008
0.008
0.007
0.007
0.007
0.007
0.007
0.007
0.007
0.006
0.006
0.006
0.006
0.005
0.005
0.005
0.005
0.005
0.005
0.005
0.004
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Feature Importance

APPENDIX 7


Feature Importance Ranking (XGBoost vs Logistic Regression)
housing
duration
month_oct
month_sep
pdays
day_of_week_4
month_feb
month_dec
day_of_week_22
month_jan
job_blue-collar
day_of_week_20
day_of_week_29
day_of_week_19
education_tertiary
day_of_week_15
age
month_jun
day_of_week_17
campaign
day_of_week_30
day_of_week_3
marital_single
day_of_week_28
day_of_week_10
job_technician
day_of_week_14
job_admin.
day_of_week_7
day_of_week_24
marital_divorced
default
job_services
job_management
job_retired
Feature
XGB_Rank
LR_Rank
Model
Rank

APPENDIX 8

Cross-National Analysis of Mobile Money Adoption

H0:The socioeconomic variables do not have a statistically significant effect on mobile money adoption.

H1:At least one socioeconomic variable has a statistically significant effect on mobile money adoption.

Conclusion:

The GMM analysis shows that factors such as lagged mobile money usage and urban population have statistically significant effects on mobile money adoption. This indicates that adoption is influenced by prior usage patterns and urbanization levels.

Therefore, we reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1):

At least one socioeconomic factor significantly affects mobile money adoption.

APPENDIX 9

Why GMM?

- The model includes a lagged dependent variable, raising endogeneity concerns.
- Hausman test indicated that Random Effects is inconsistent, supporting the presence of endogeneity.
- GMM handles this by using internal instruments and provides robust estimates under heteroskedasticity and autocorrelation.
- It is ideal for dynamic panel data with lag structures.

APPENDIX 10

Panel Data

- Combines data across multiple entities (e.g., countries, firms) over time. Useful for analyzing both individual differences and trends over time.

Fixed Effects (FE)

- Assumes entity-specific traits **correlate** with explanatory variables. Focuses on **within-entity changes** over time.

Random Effects (RE)

- Assumes entity-specific effects are **uncorrelated** with the regressors. Uses both **within and between** variation but may be biased if the assumption fails.

Hausman Test

Determines model choice:

- Significant → **Reject RE**, prefer **FE**
- Not significant → **RE is consistent**, may be used

APPENDIX 11

Analyst Sentiment and Corporate Investment Strategy

H1a: There is a significant positive relationship between analyst sentiment and company investment decisions.

H2a: There is a significant positive relationship between analyst sentiment and company investment decisions.

H2b: A company's over-investment or under-investment negatively affects its investment decisions.

H2c: Whether a company is over- or under-invested influences the relationship between analyst sentiment and investment decisions.

H3a: There is a significant relationship between positive sentiment scores and company investment levels.

H3b: There is a significant relationship between negative sentiment scores and company investment levels.

H4a: Antitrust policy investigations have a significant impact on corporate investment strategies.

The hypotheses H2a, H3a, and H4a are accepted, indicating that analyst sentiment positively influences investment when accounting for firm heterogeneity, that positive sentiment alone drives investment behavior, and that U.S. antitrust investigations significantly reduce investment.

APPENDIX 12

Compety Hackathon: UX A/B Test for Conversion Lift

Hypothesis

H_0: Enlarging food images on the app does not significantly affect user conversion rates.

H_1: Enlarging food images on the app does significantly increase user conversion rates.

Conclusion
The results show that enlarging food images significantly increases user conversion rates. Both T-tests and Chi-square tests confirmed this effect in the full dataset and across 30%, 50%, and 70% stratified samples, validating the robustness of the findings and supporting rejection of the null hypothesis.

APPENDIX 13

Compety Hackathon: UX A/B Test for Conversion Lift

Experimental Procedure

1. Data Cleaning: Removed records with missing shop_id or final_order_status
2. Sampling: Conducted stratified sampling (30%, 50%, 70%) for robustness checks
3. Statistical Tests:
   - T-Test: To compare means between control and test groups
   - Chi-Square Test: To assess relationship between group and conversion
4. Supplementary Analysis: Platform (iOS vs Android) and time-of-day effects were also examined