OPTIMIZING CAMPAIGN STRATEGY WITH PREDICTIVE ANALYTICS

Client Subscription Patterns in Retail Banking

YU-JEN WANG May/2025

BACKGROUND

SEP 2023 - SEP 2024
Master's degree Financial Technology | University of Bristol

AUG 2021 - MAY 2023
Senior Engineer, Optical Design Dept., INNOLUX Co. Ltd

SEP 2017 - JUN 2020
Master's degree Applied Mathematics Computational Science | National Chung Hsing University

SEP 2013 - JUN 2017
Bachelor's degree Applied Mathematics | National Chung Hsing University

PROJECT

CLOSE THE GAP

Our product makes consumer lives easier, and no other product on the market offers the same features

TARGET AUDIENCE

Our target audience is Gen Z (18-25 years old)

COST SAVINGS

Reduce expenses for replacement products

EASY TO USE

Simple design that gives customers the targeted information they need

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.

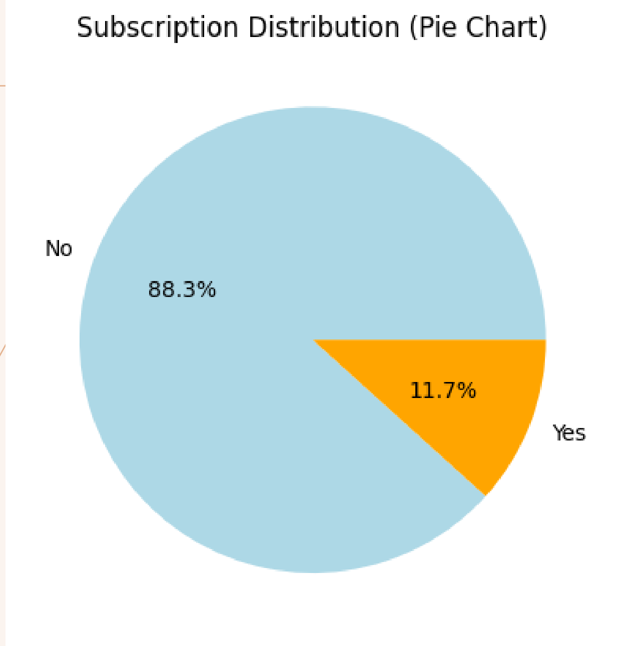
Exploratory data analysis

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



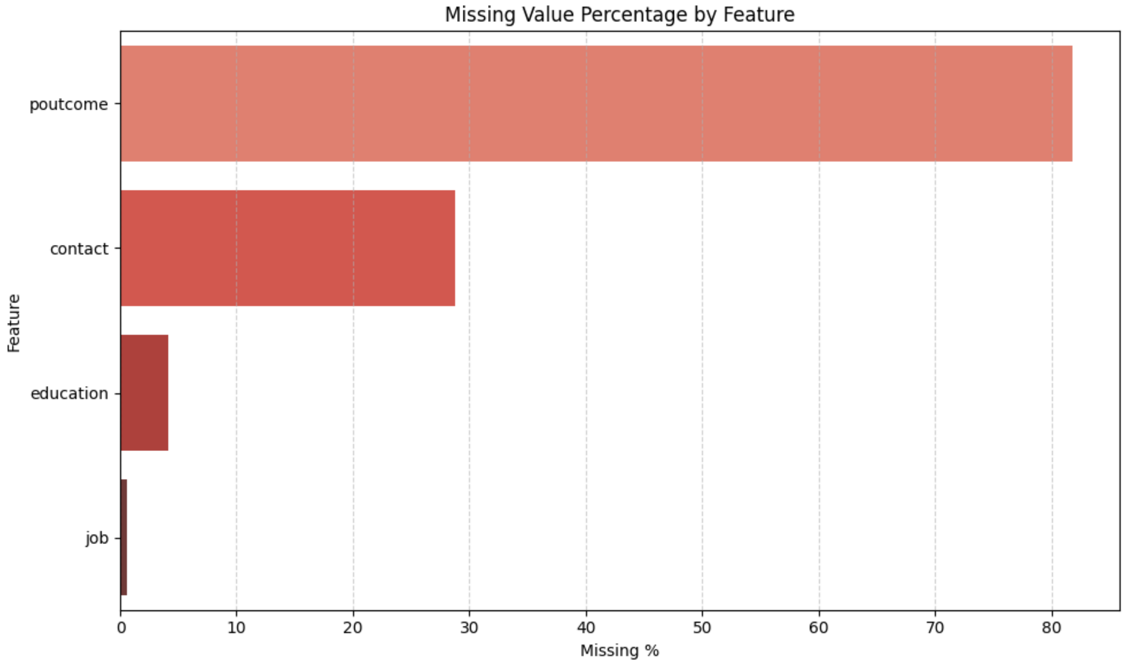
DESCRIPTIVE ANALYSIS – CLIENT PROFILES

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?

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

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:

- Random Forest achieved higher overall accuracy (89.9%) and was more reliable in identifying non-subscribers, reducing false positives.
- Logistic Regression had stronger recall for subscribers (79.6%), making it better for targeting potential customers.
- However, both models struggled with subscriber prediction due to class imbalance (F1-scores: 0.43 vs. 0.52).
- Macro metrics show a trade-off: Random Forest is more balanced, while Logistic Regression better captures the minority class.

Overall, Random Forest suits general performance needs; Logistic Regression is preferred for customer acquisition.

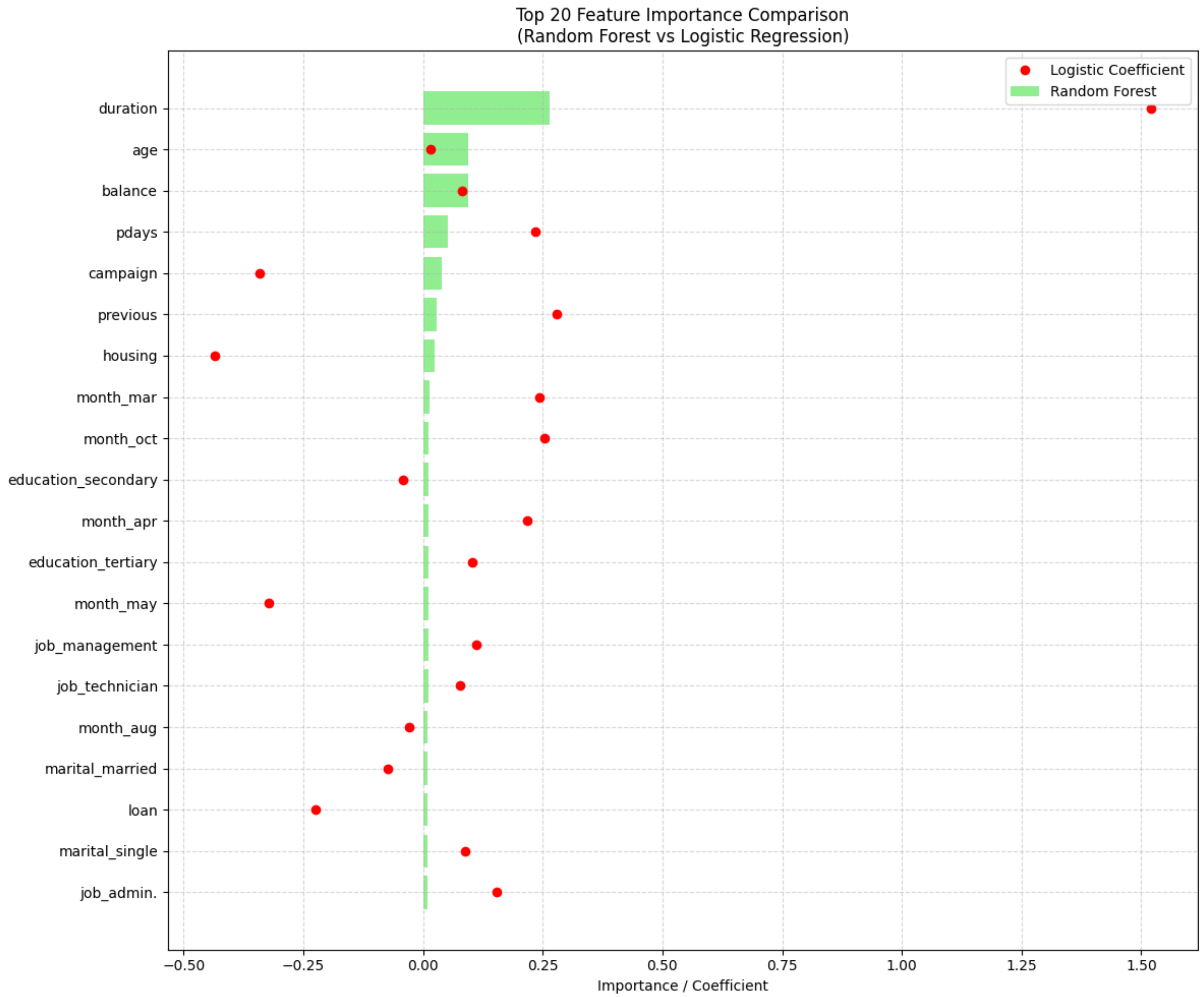
	Random Forest	Logistic Regression
Accuracy	0.8992	0.8294
Precision (Class 0)	0.9164	0.9686
Precision (Class 1)	0.6335	0.3883
Recall (Class 0)	0.9748	0.8338
Recall (Class 1)	0.3289	0.7958
F1-score (Class 0)	0.9447	0.8962
F1-score (Class 1)	0.433	0.5219
Macro Avg Precision	0.775	0.6784
Macro Avg Recall	0.6519	0.8148
Macro Avg F1-score	0.6889	0.709
Weighted Avg Precision	0.8833	0.9007
Weighted Avg Recall	0.8992	0.8294
Weighted Avg F1-score	0.8848	0.8524

PREDICTIVE MODELING AND FEATURE INTERPRETATION

Key Feature Interpretation:

- Call duration was the strongest predictor—longer calls led to higher subscription rates.
- Follow-up history (pdays, previous) also improved success chances.
- Balance and age had moderate influence, linked to financial readiness.
- Seasonality and demographics (month, job, marital status) were more pronounced in Logistic Regression, useful for customer segmentation.

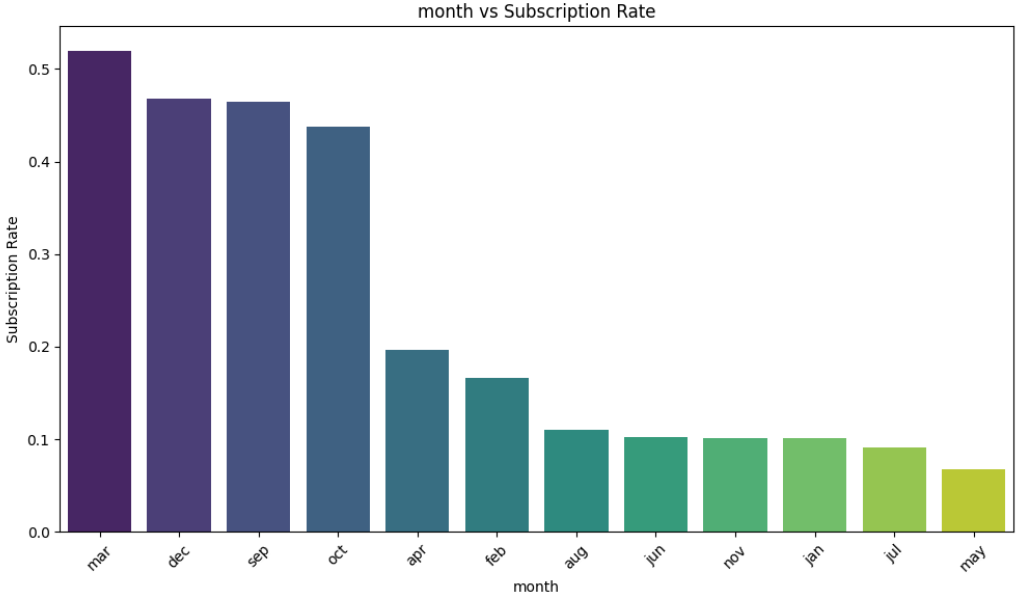
Models highlight key traits to focus on for effective targeting, beyond just prediction.



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 conversion.
- 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 campaigns to specific segments (e.g., students or retired clients with no loans) in high-conversion months to maximize ROI.
- Tailor messaging by group – e.g., emphasize long-term savings for retirees, or easy entry terms for students.
- Avoid broad, unfocused outreach during low-conversion windows.

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.Prioritize high-potential segments
Focus on student and retired groups with primary or tertiary education and no loan obligations.

2.Optimize campaign timing
Allocate more marketing resources in **March, September, and December**, when customers are most receptive.

3.Refine call strategies
Encourage longer, personalized conversations instead of frequent short calls. Avoid over-contacting to reduce fatigue.

4.Leverage predictive modeling
Use models like Random Forest to score clients and prioritize outreach to those with high predicted conversion probability.

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:

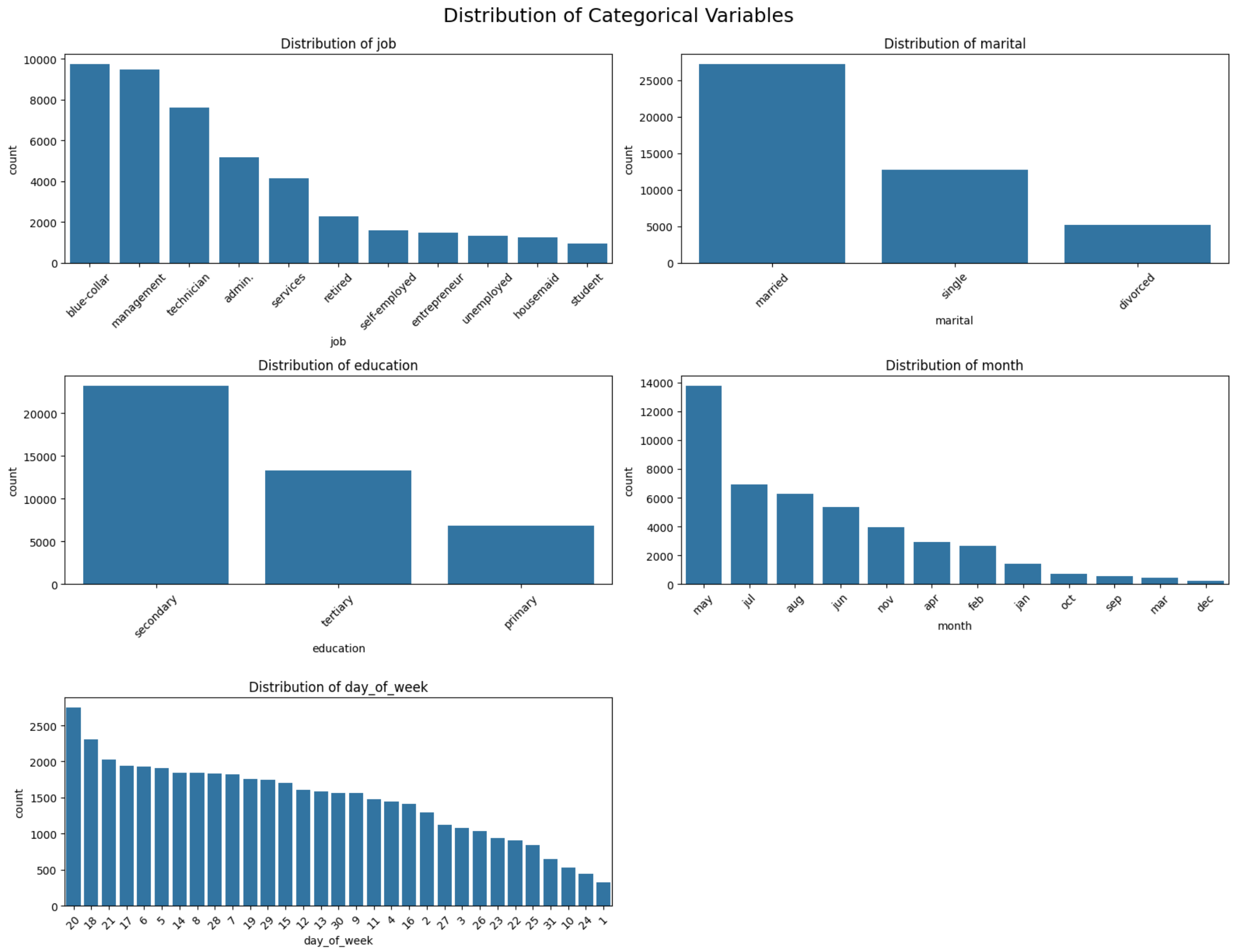
The most common job types are blue-collar, management, and technician, while students and housemaids are least represented.

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

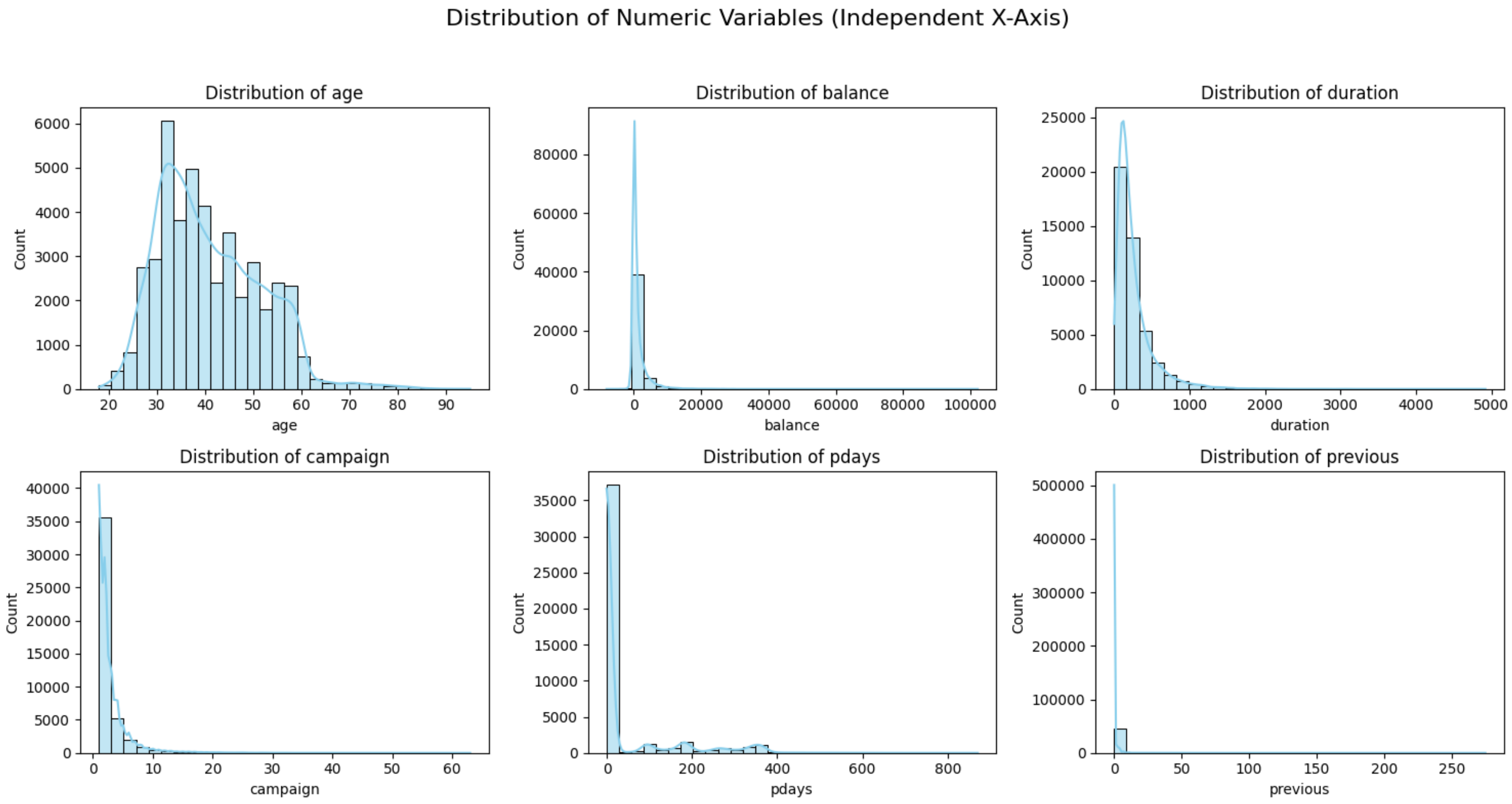
Numeric Variables:

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

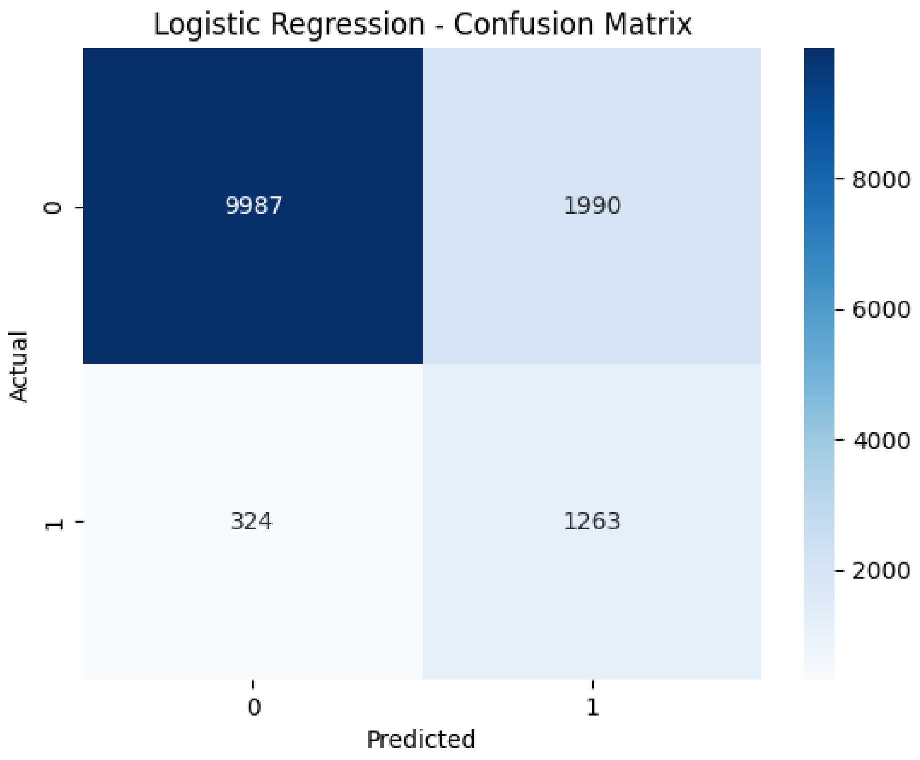
Balance and duration exhibit extreme right skewness with many low values and a long tail of high values.

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

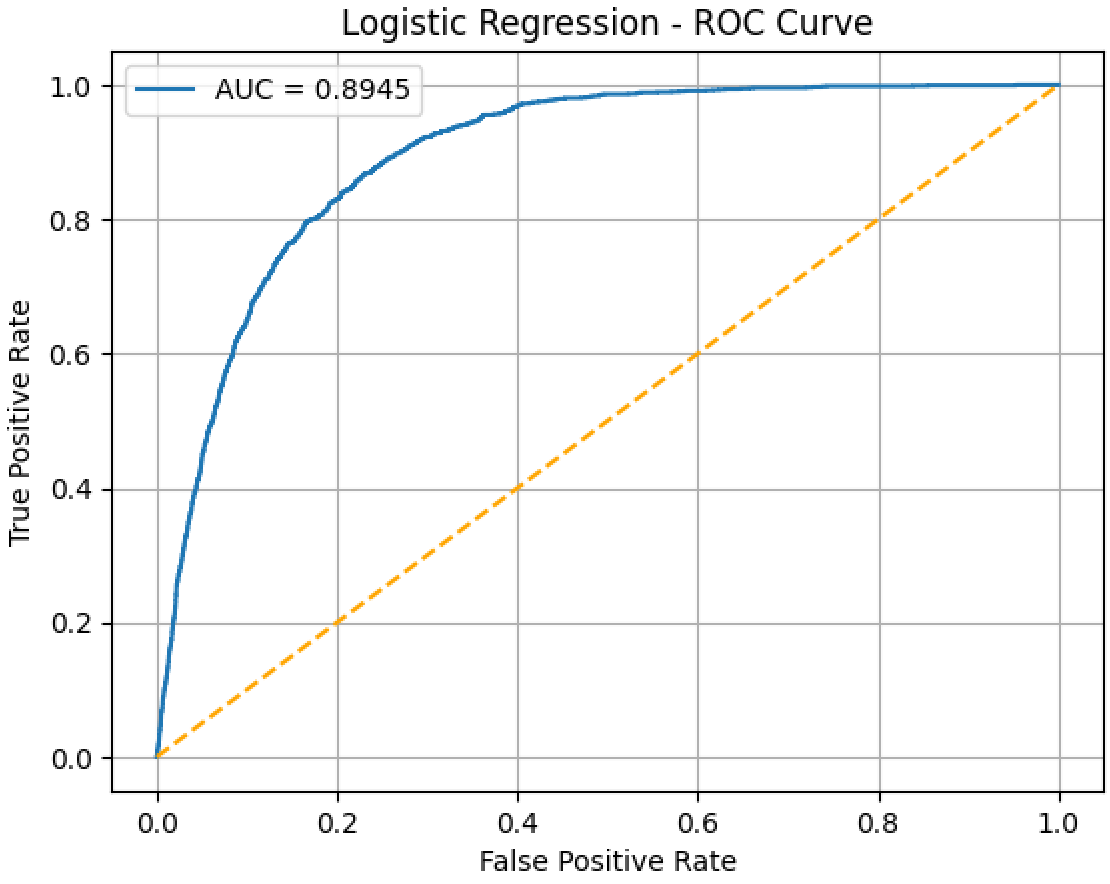
Many clients had zero previous contact or no recent contact (pdays = 999), highlighting a large untouched base.



APPENDIX 3

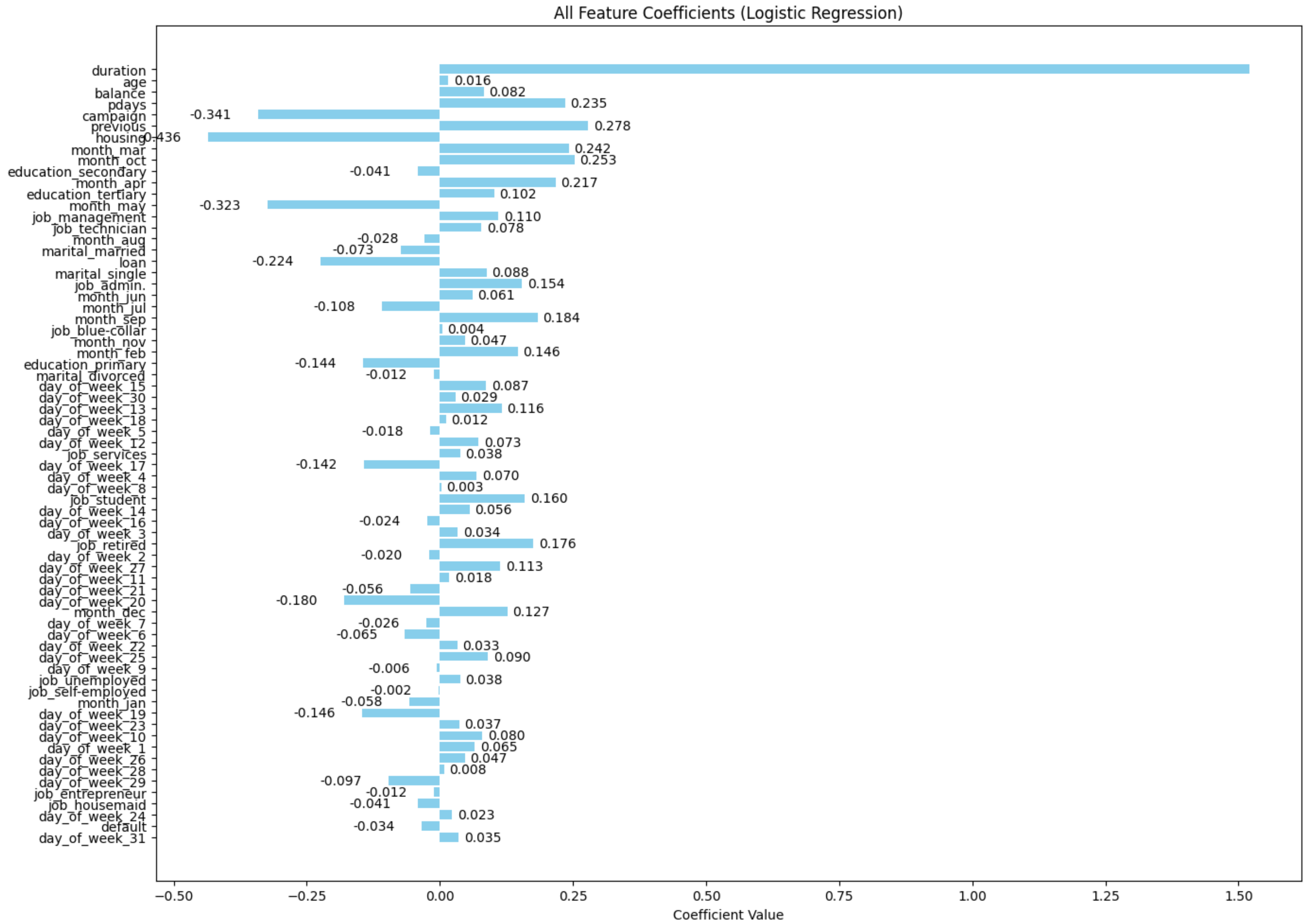


Logistic Regression - Confusion Matrix
9987
1990
324
1263
Actual
Predicted
0
1
0
1
8000
6000
4000
2000

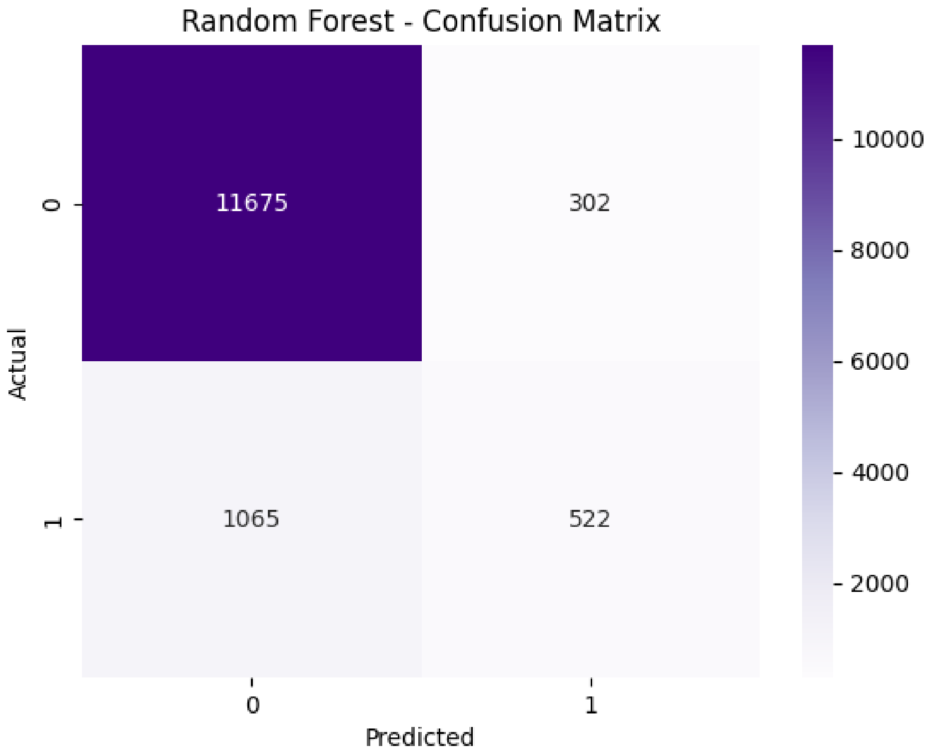


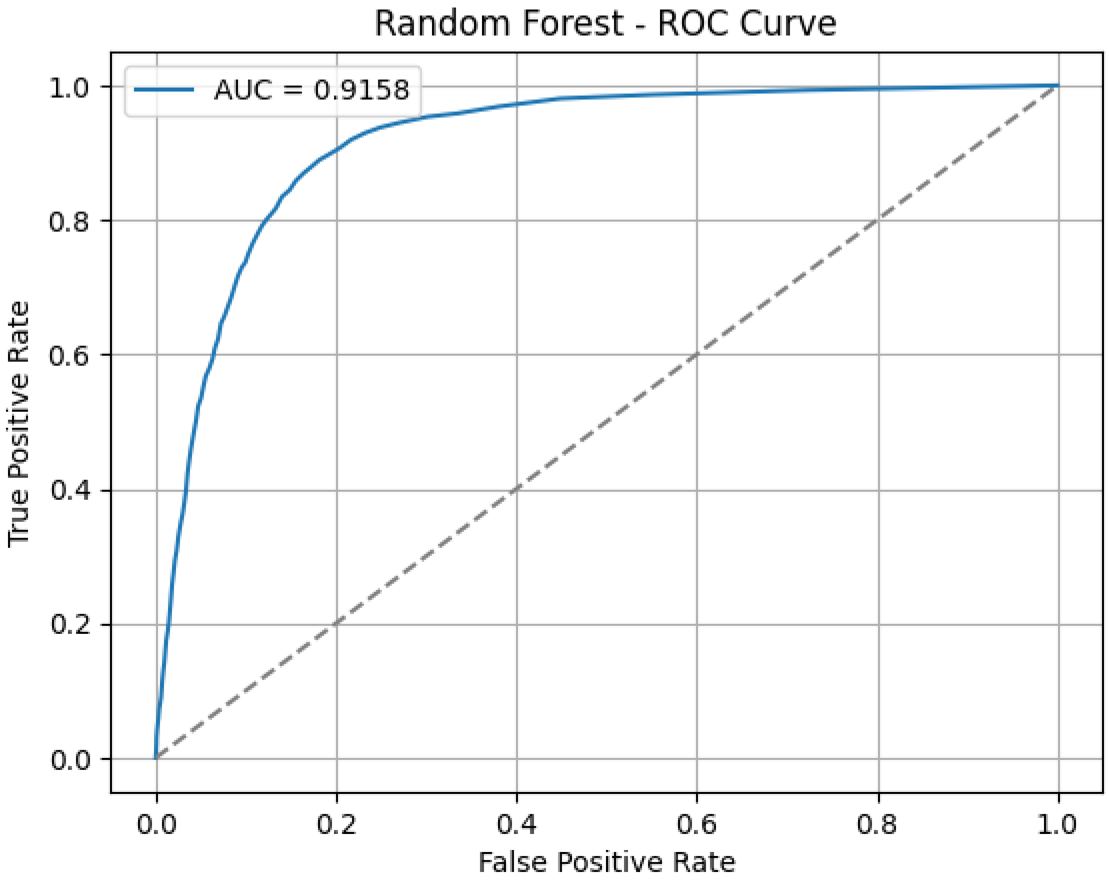
Logistic Regression - ROC Curve
AUC = 0.8945
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

APPENDIX 4

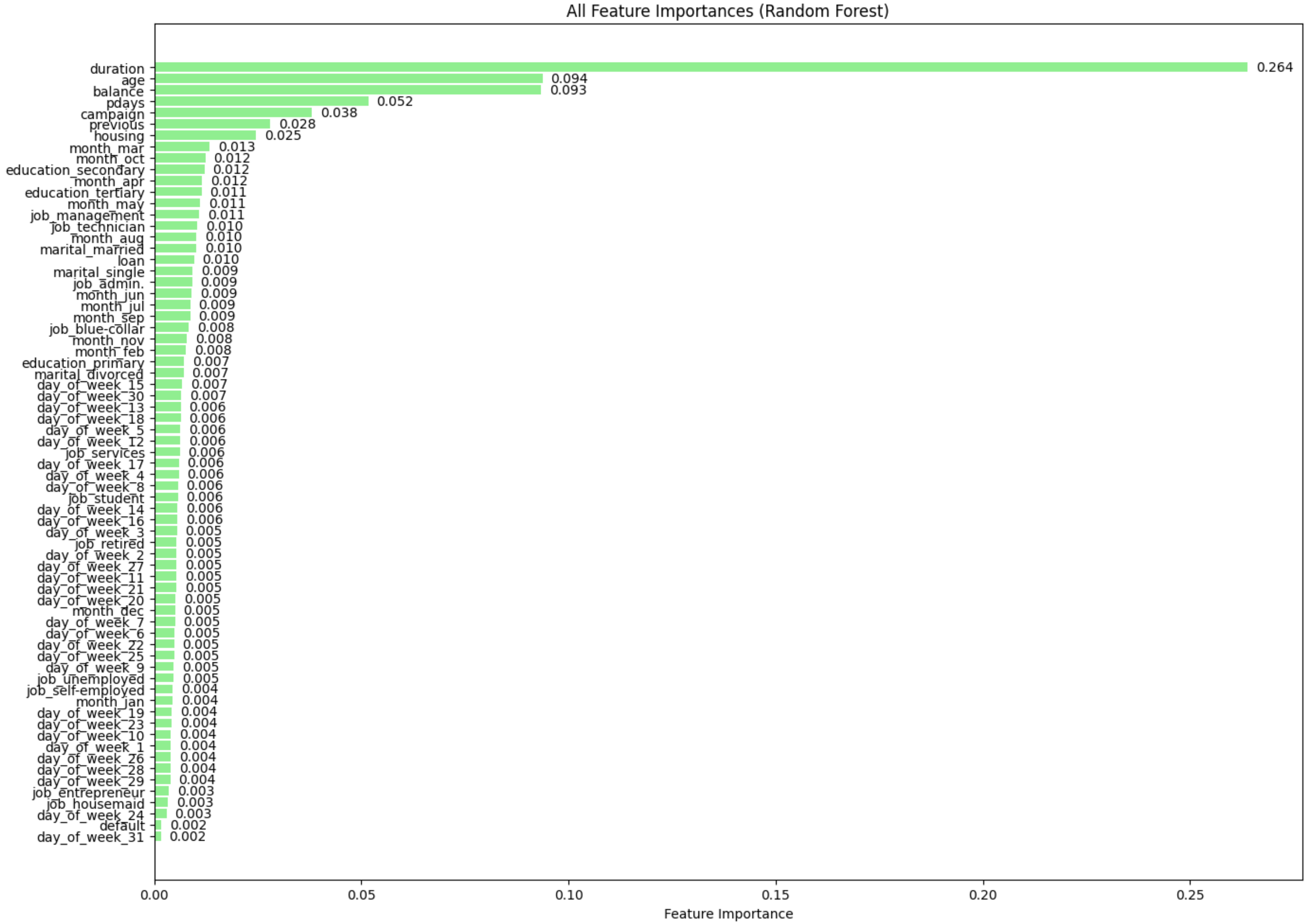

All Feature Coefficients (Logistic Regression)
duration
age
balance
pdays
campaign
previous
housing
month_mar
month_oct
education_secondary
month_apr
education_tertiary
month_may
job_management
job_technician
month_aug
marital_married
loan
marital_single
job_admin.
month_jun
month_jul
month_sep
job_blue-collar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

APPENDIX 5

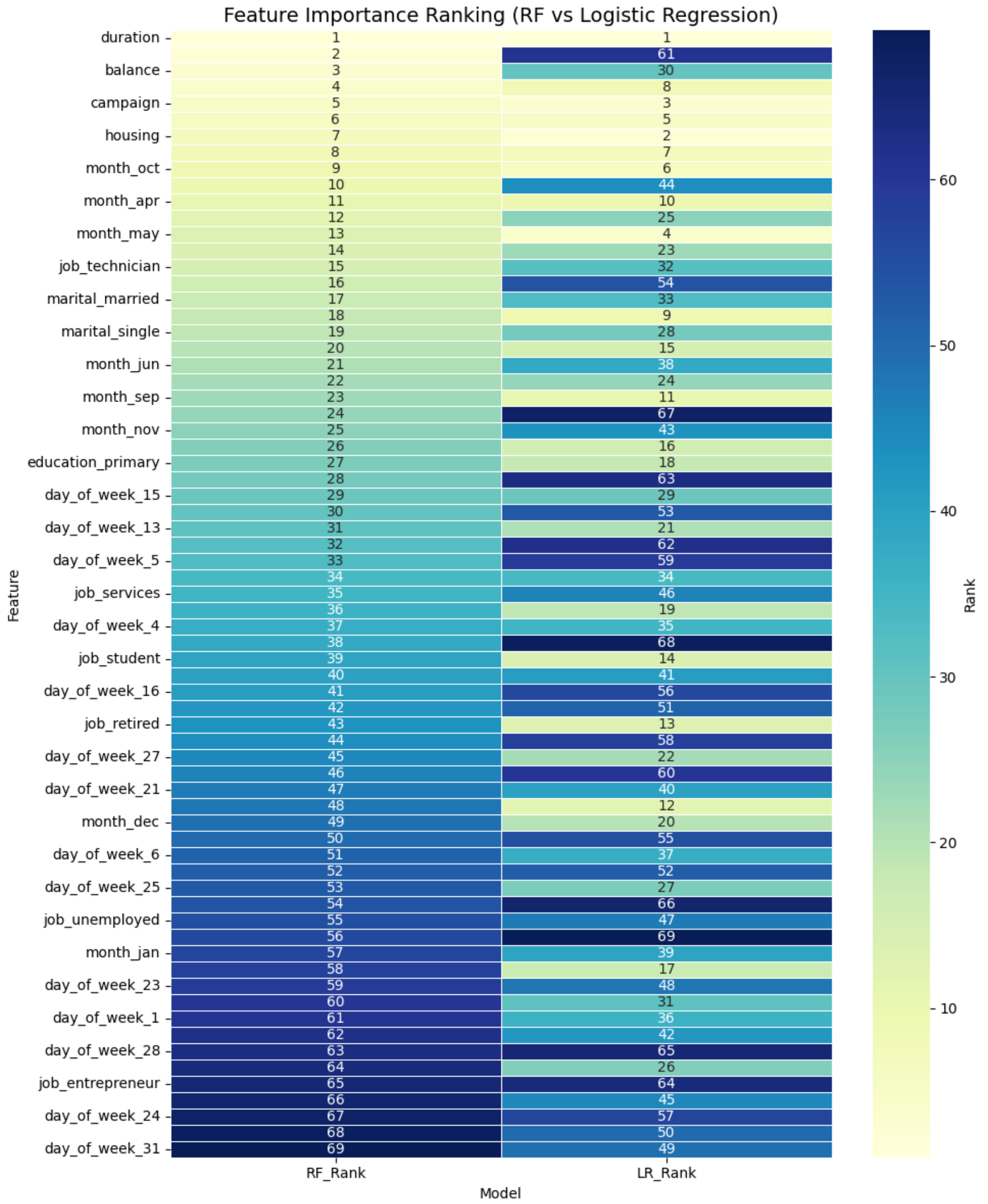

Random Forest - Confusion Matrix
11675
302
1065
522
Actual
0
1
Predicted
0
1
10000
8000
6000
4000
2000


Random Forest - ROC Curve
AUC = 0.9158
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

APPENDIX 6


All Feature Importances (Random Forest)
duration 0.264
age 0.094
balance 0.093
pdays 0.052
campaign 0.038
previous 0.028
housing 0.025
month_mar 0.013
month_oct 0.012
education_secondary 0.012
month_apr 0.012
education_tertiary 0.011
month_may 0.011
job_management 0.011
job_technician 0.010
month_aug 0.010
marital_married 0.010
loan 0.010
marital_single 0.009
job_admin. 0.009
month_jun 0.009
month_jul 0.009
month_sep 0.009
job_blue-collar 0.008
month_nov 0.008
month_feb 0.008
education_primary 0.007
marital_divorced 0.007
day_of_week_15 0.007
day_of_week_30 0.007
day_of_week_13 0.006
day_of_week_18 0.006
day_of_week_5 0.006
day_of_week_12 0.006
job_services 0.006
day_of_week_17 0.006
day_of_week_4 0.006
day_of_week_8 0.006
job_student 0.006
day_of_week_14 0.006
day_of_week_16 0.006
day_of_week_3 0.005
job_retired 0.005
day_of_week_2 0.005
day_of_week_27 0.005
day_of_week_11 0.005
day_of_week_21 0.005
day_of_week_20 0.005
month_dec 0.005
day_of_week_7 0.005
day_of_week_6 0.005
day_of_week_22 0.005
day_of_week_25 0.005
day_of_week_9 0.005
job_unemployed 0.005
job_self-employed 0.004
month_jan 0.004
day_of_week_19 0.004
day_of_week_23 0.004
day_of_week_10 0.004
day_of_week_1 0.004
day_of_week_26 0.004
day_of_week_28 0.004
day_of_week_29 0.004
job_entrepreneur 0.003
job_housemaid 0.003
day_of_week_24 0.003
default 0.002
day_of_week_31 0.002
0.00
0.05
0.10
0.15
0.20
0.25
Feature Importance

APPENDIX 7


Feature Importance Ranking (RF vs Logistic Regression)
Feature
duration
balance
campaign
housing
month_oct
month_apr
month_may
job_technician
marital_married
marital_single
month_jun
month_sep
month_nov
education_primary
day_of_week_15
day_of_week_13
day_of_week_5
job_services
day_of_week_4
job_student
day_of_week_16
job_retired
day_of_week_27
day_of_week_21
month_dec
day_of_week_6
day_of_week_25
job_unemployed
month_jan
day_of_week_23
day_of_week_1
day_of_week_28
job_entrepreneur
day_of_week_24
day_of_week_31
RF_Rank
LR_Rank
Model
Rank
60
50
40
30
20
10