TARGETED MARKETING CAMPAIGN

By Adrian, Kritika, Mohamed,
Sheida, Vincent

OUR TEAM

Name	Role	GitHub ID
Adrian Alarcon Delgado	Data Engineer	aladelca
Kritika Nayyar	Business Analyst	knayyar0416
Mohamed Elenany	Data Engineer	MohamedElenany
Sheida Majidi	Data Analyst	SheidaMajidi
Vincent El-Ghoubaira	Data Modeler	Vincent-elgh

GITHUB REPOSITORY

Predicting Subscription Outcomes in Direct Marketing Campaigns

ID: targeted_marketing

TABLE OF CONTENTS

BUSINESS VALUE

- **Predict Clients Likely to Subscribe to Term Deposits:** Utilize predictive models to identify clients who are more inclined to subscribe to term deposits.
- **Optimize Marketing Efforts:** Segment clients based on relevant criteria (e.g., demographics, behavior) to tailor marketing messages effectively.
- **Timing the Marketing Campaign Effectively:** Determine the optimal timing for marketing campaigns, considering fiscal year-end and tax-related factors.
- **Tailoring Strategies for Existing vs. New Clients:** Develop distinct marketing strategies for existing clients versus new prospects.

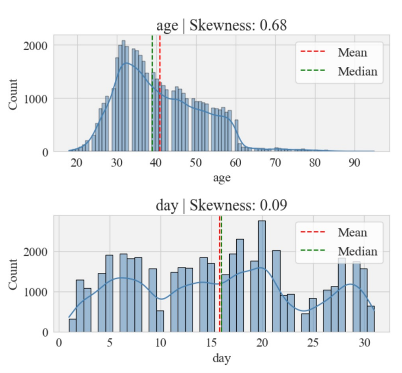
EDA

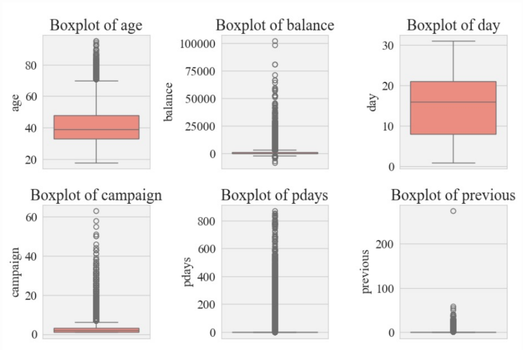
- The dataset, containing a total of 45k observations, is derived from direct marketing campaigns conducted by a Portuguese banking institution.
- With a total of 17 features, there are 10 object and 7 integer datatypes.
- There are no missing or duplicate values.

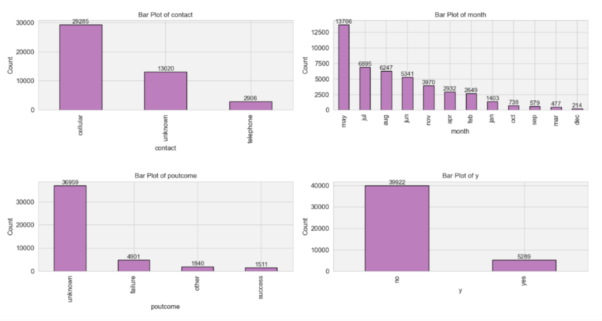
Job | Marital | Education
Default | Housing | Loan | Contact
Month | Poutcome | Conversion
Age | Balance | Day | Duration
Campaign | Pdays | Previous

UNIVARIATE ANALYSIS

Univariate analysis of numerical features helped us analyze skewed data in balance, campaign, pdays, and previous. Categorical feature poutcome and contact contain the most unknown values. Our target variable, conversion, is highly skewed with 89% clients who did not subscribe to bank term deposit.

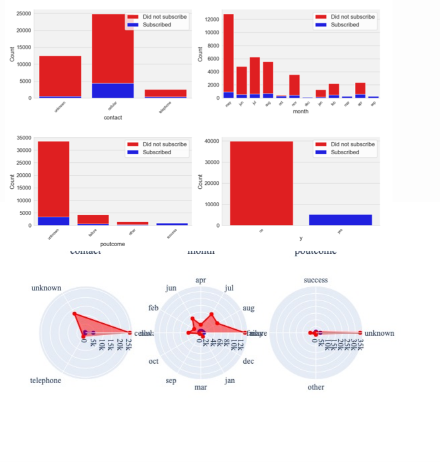


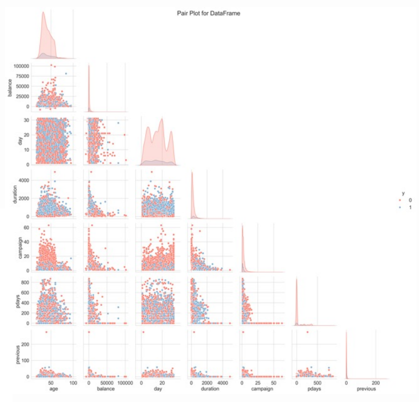


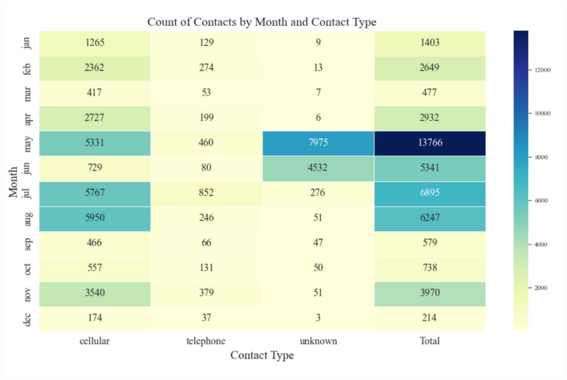


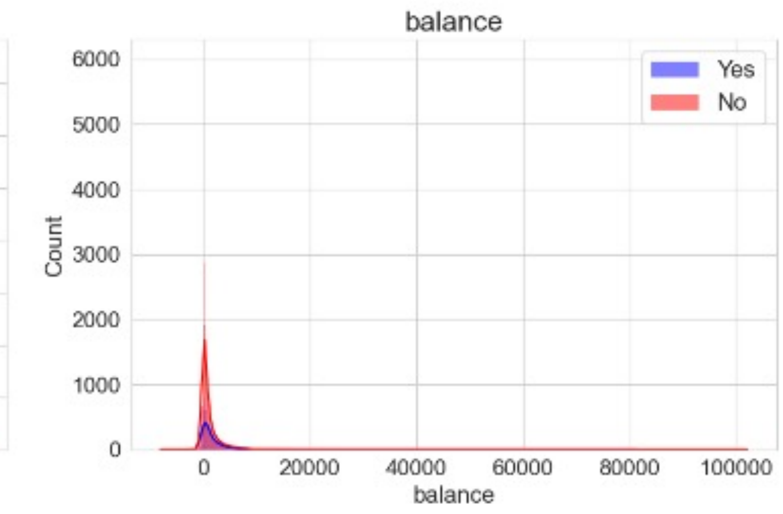
BIVARIATE ANALYSIS

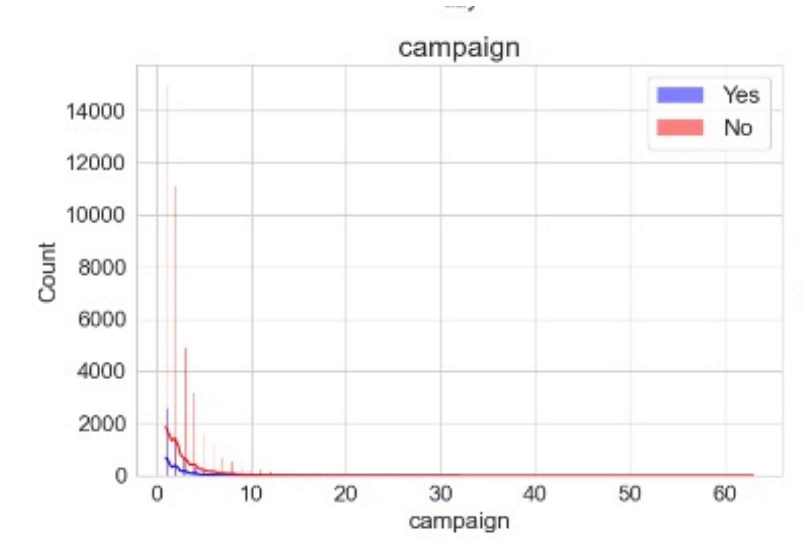
Bivariate analysis helps us find that majority of direct calls were made in May to August. It is interesting to see months March and September resulted in 100% conversion. Our hypothesis is that these are organic converts.

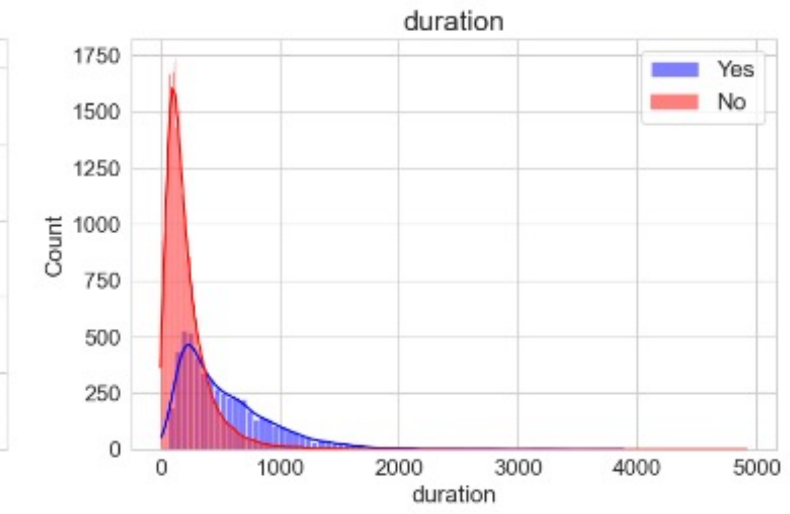


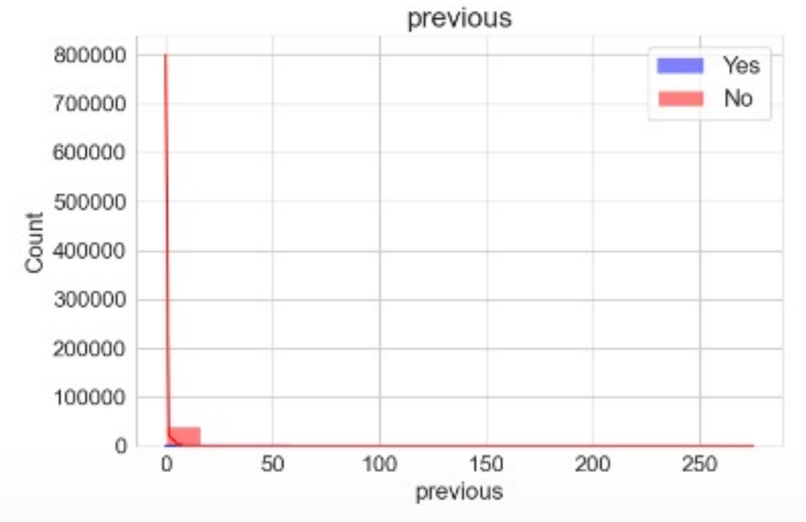


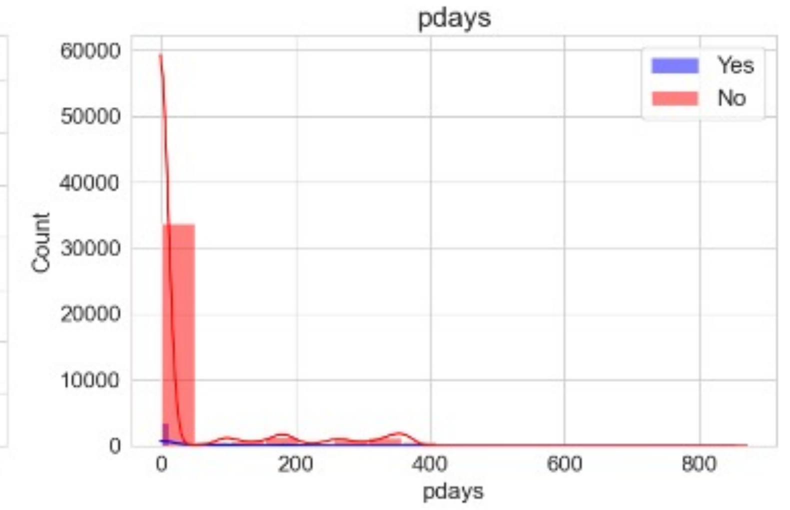




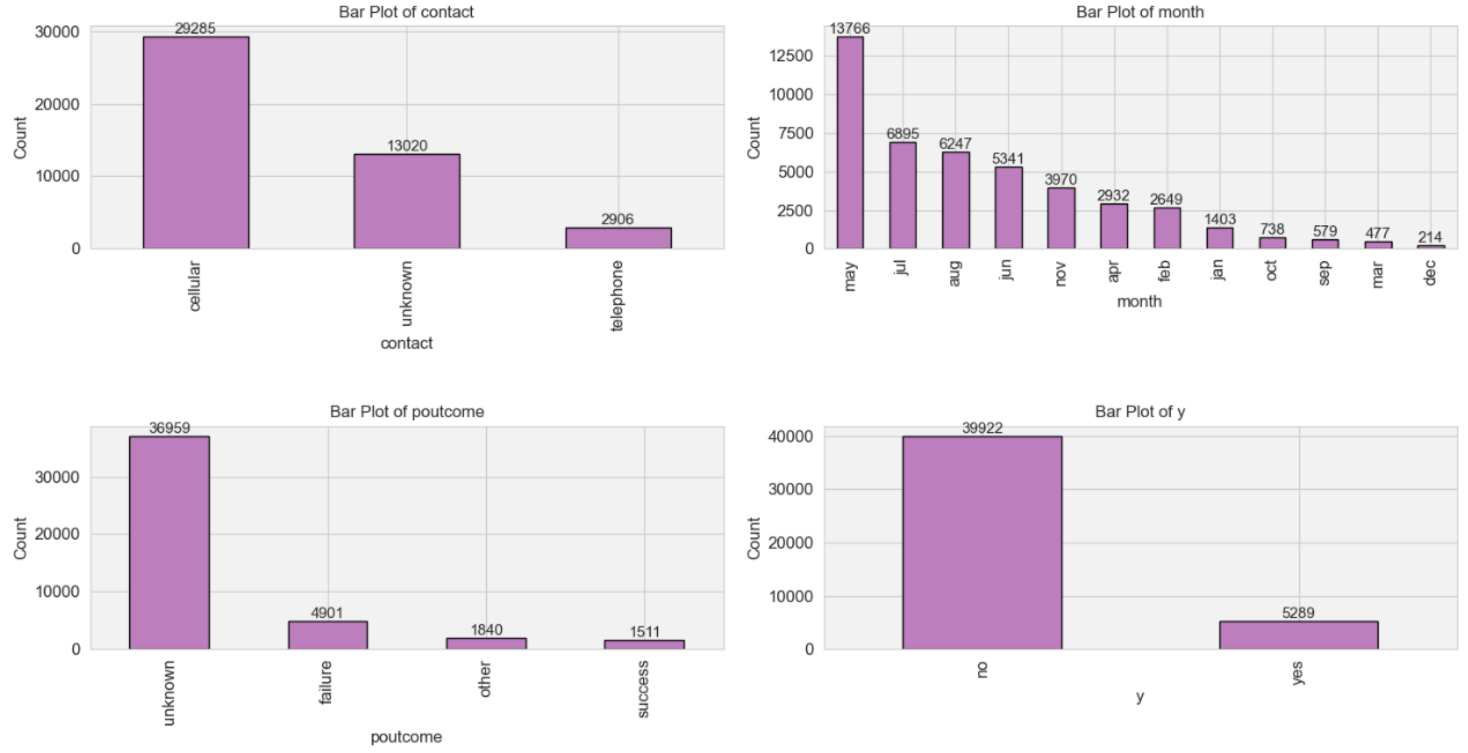




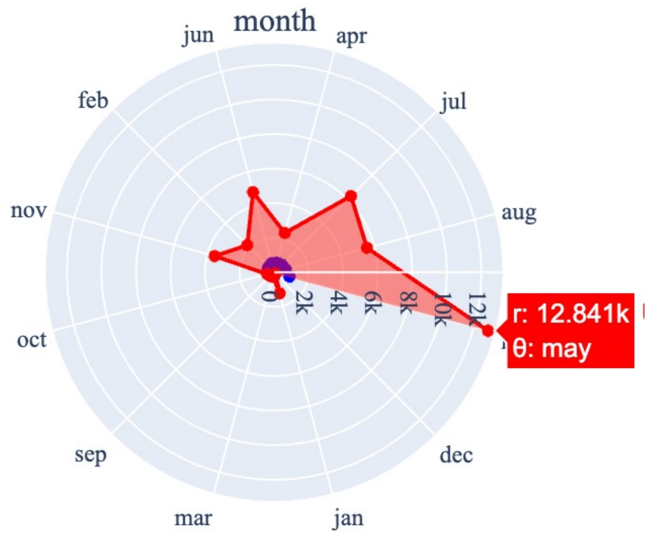


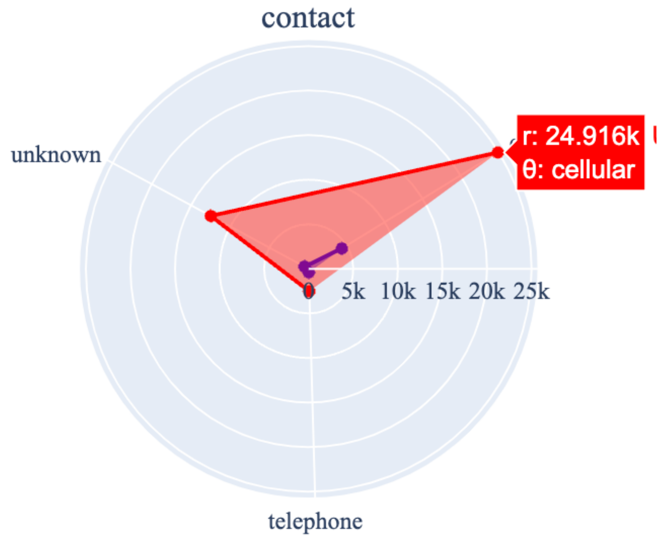


Features balance, duration and campaign are highly skewed. Further, ~82% of values in features pdays and previous are unknown.



Features job, education, contact and poutcome contain unknown values. While the former three are still acceptable for analysis purposes, we removed poutcome.





We can see most of the contacts were made in the months May, June, July and August. Further, majority of the contacts for the months May and June are unknown.

CLASSIFICATION: PREPROCESSING

- **General Preprocessing Steps:**
  - One hot encoding categorical variables
  - Scaling data using MinMax Scaler
  - Detecting and eliminating outliers using an Isolation Forest
- **Feature Elimination:**
  - "Poutcome": outcome of the previous marketing campaign
    - Eliminated for having 90% missing values
  - "Contact": contact method of customer
    - Eliminated for being dominated by only one unique value
  - "Day": day of the month the contact was made
    - Eliminated for contributing to model multicollinearity
  - "Duration": duration of the contact with the customer
    - Eliminated for causing data leakage
- **Class imbalance in target variable:**
  - Only 9% of the instances represent customers who registered successfully for a term deposit
  - Diagnosis: up-sampling of minority class
- **Feature Engineering:**
  - Lasso Regression was used to detect and eliminate features with nearly no predictive power

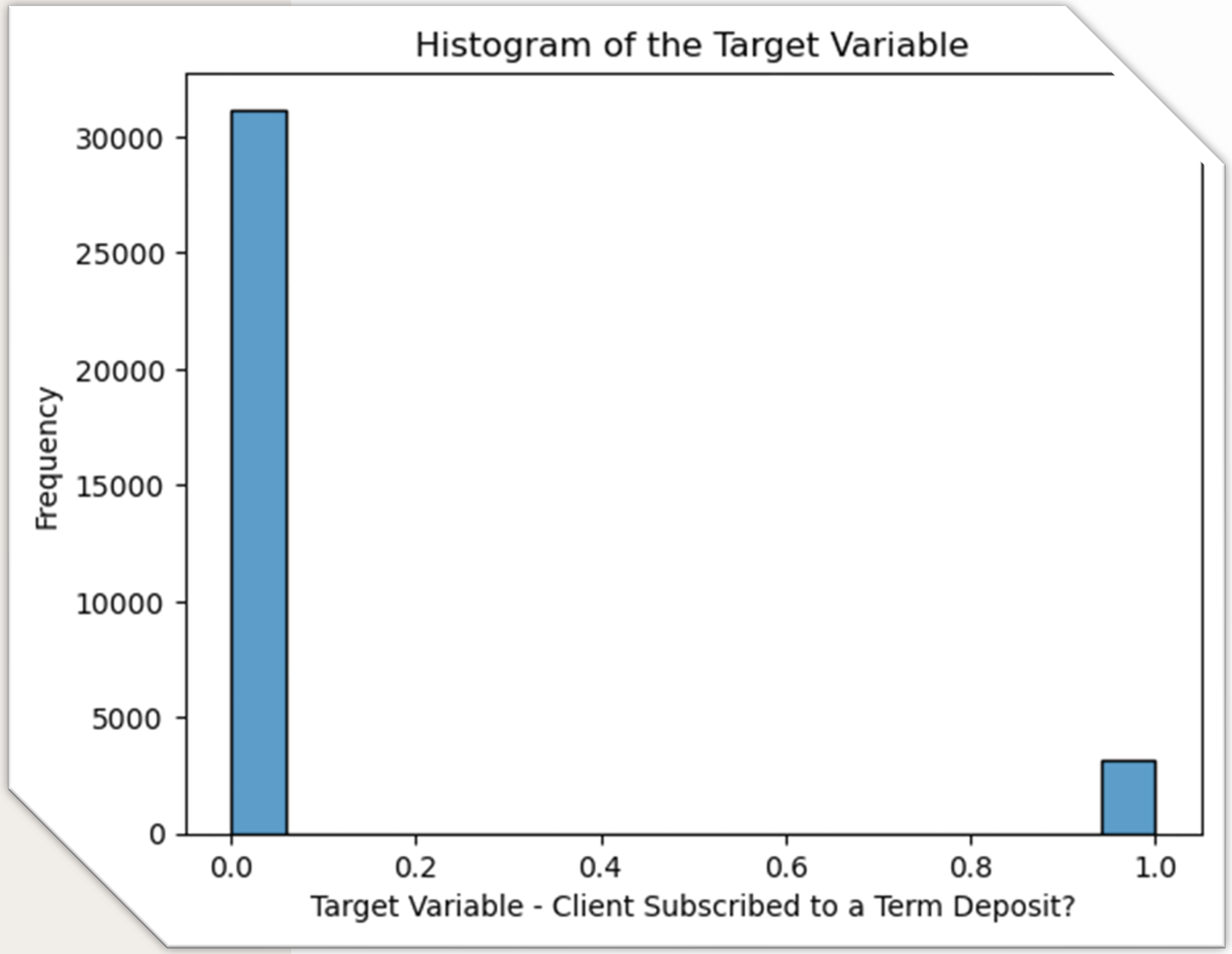


Figure – Class Imbalance in Target Variable

MODELLING

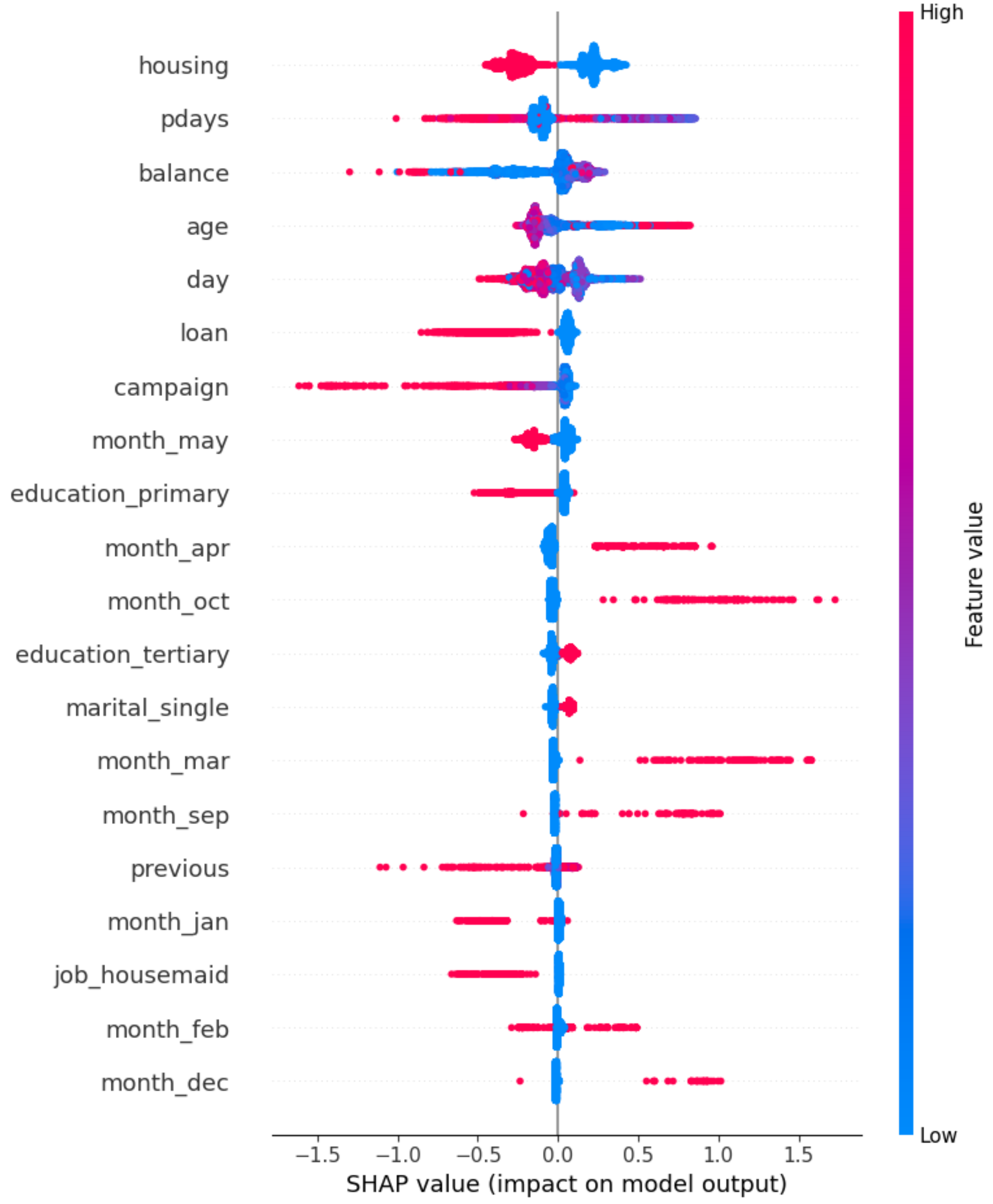
Models

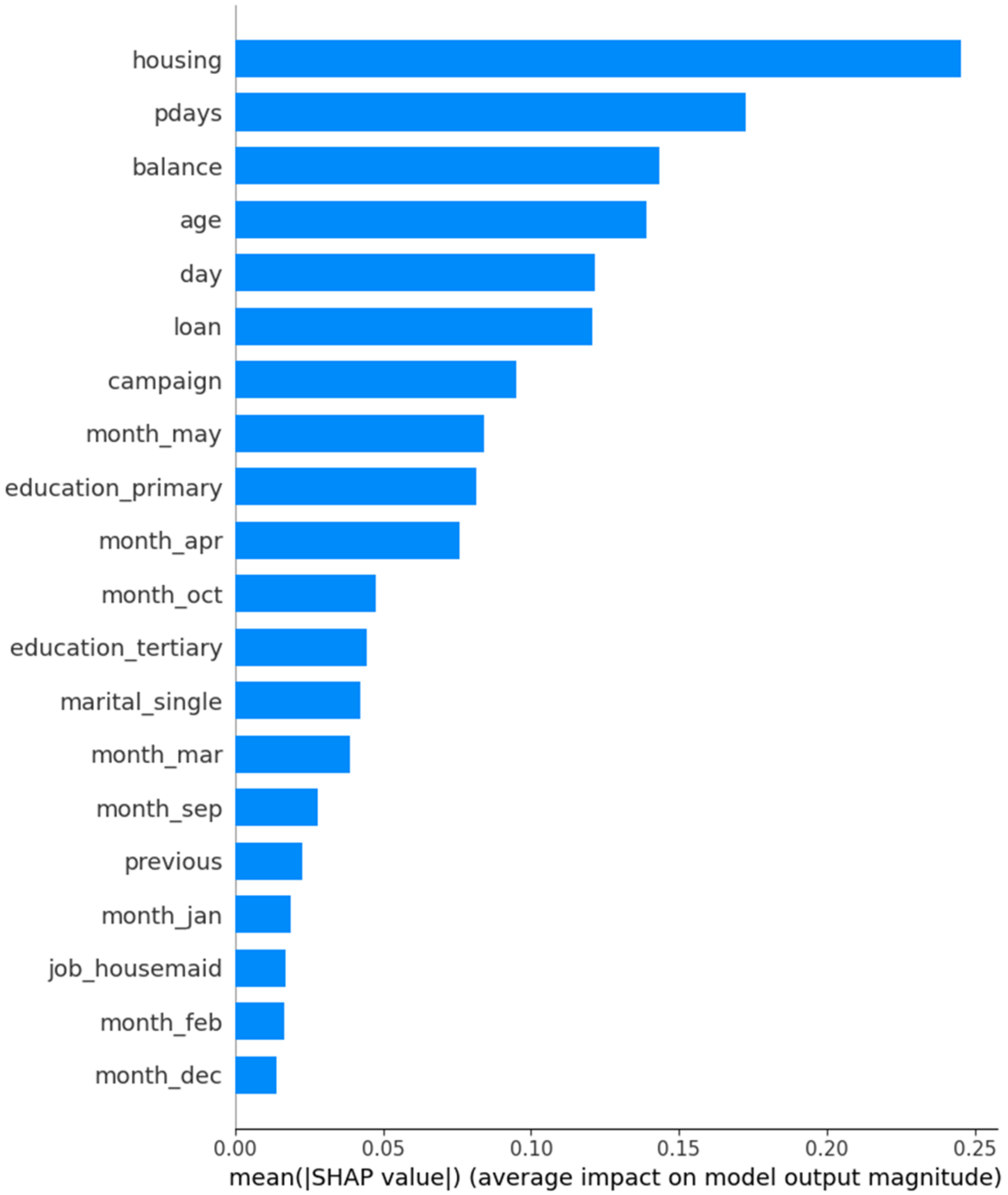
- Used min max scalar
- Performed predictive modeling on the preprocessed data using these five methods:
  - Logistic Regression
  - Random Forest
  - Gradient Boosting
  - LightGBM
  - XGBoost
- For our case we need to capture the most possible conversions ➔ Higher Recall Rate
- Aimed at having the highest F2 Score

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	F2 Score
Logistic Regression	0.68	0.22	0.64	0.33	0.66	0.46
Random Forest	0.88	0.49	0.20	0.28	0.59	0.23
Gradient Boosting	0.76	0.28	0.62	0.39	0.70	0.50
LightGBM	0.80	0.32	0.57	0.41	0.70	0.49
XGBoost	0.81	0.31	0.48	0.38	0.67	0.43

MODELLING

Feature Importance Gradient Boosting



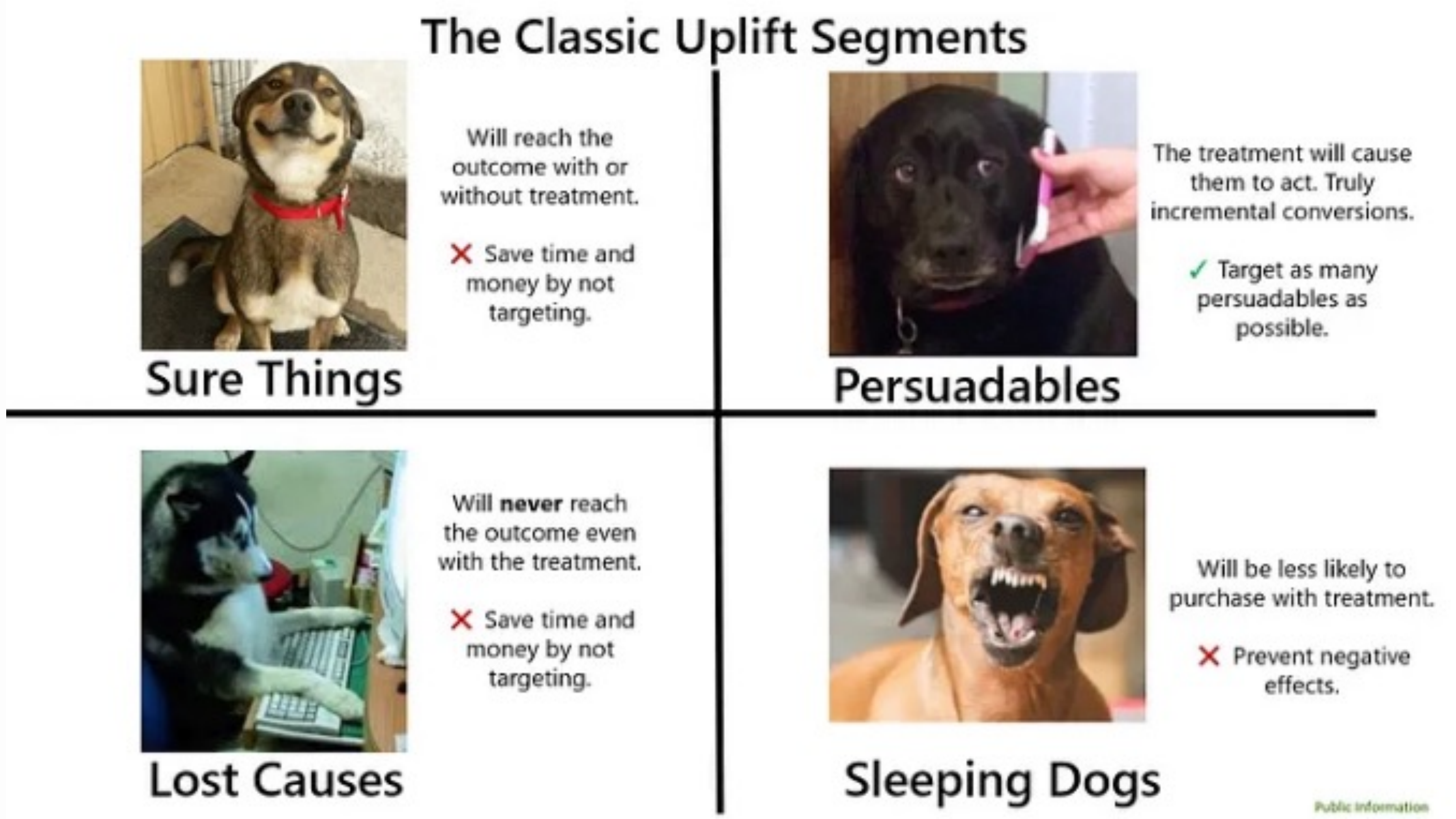


UPLIFTING MODELING

How to target correctly our marketing campaigns?

Avoid

Focus

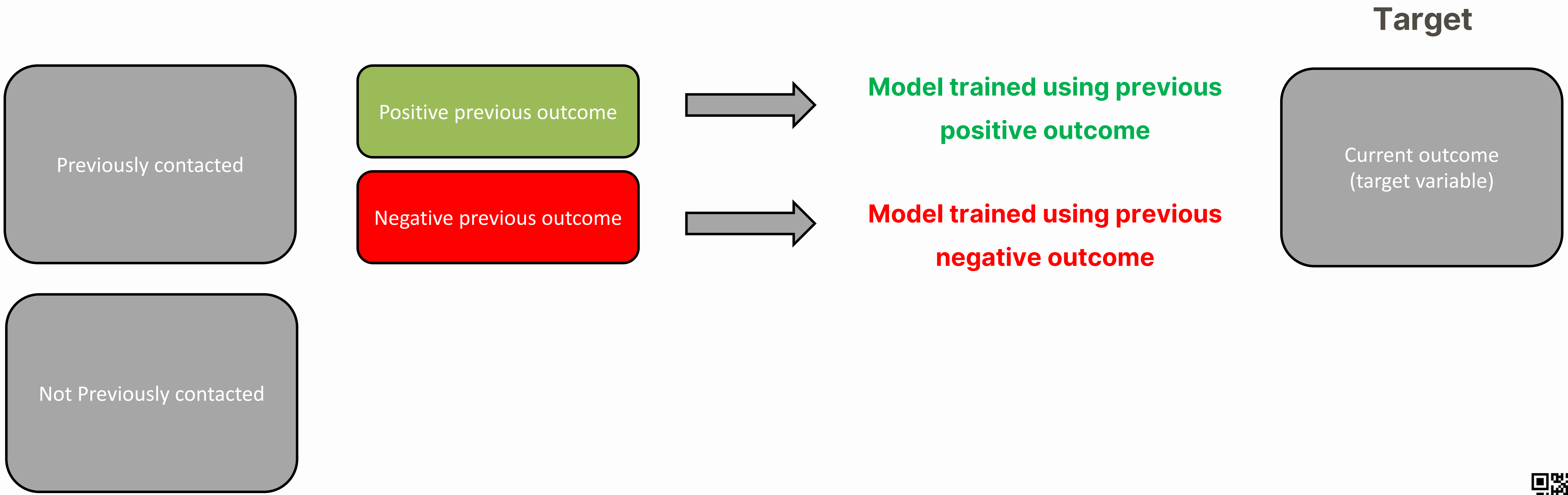


Avoid

Avoid

UPLIFTING MODELING: METHODOLOGY

How to target correctly our marketing campaigns?



UPLIFTING MODELING: METHODOLOGY

How to target correctly our marketing campaigns?

Predict

Model trained using previous positive outcome

Not Previously contacted

Model trained using previous negative outcome

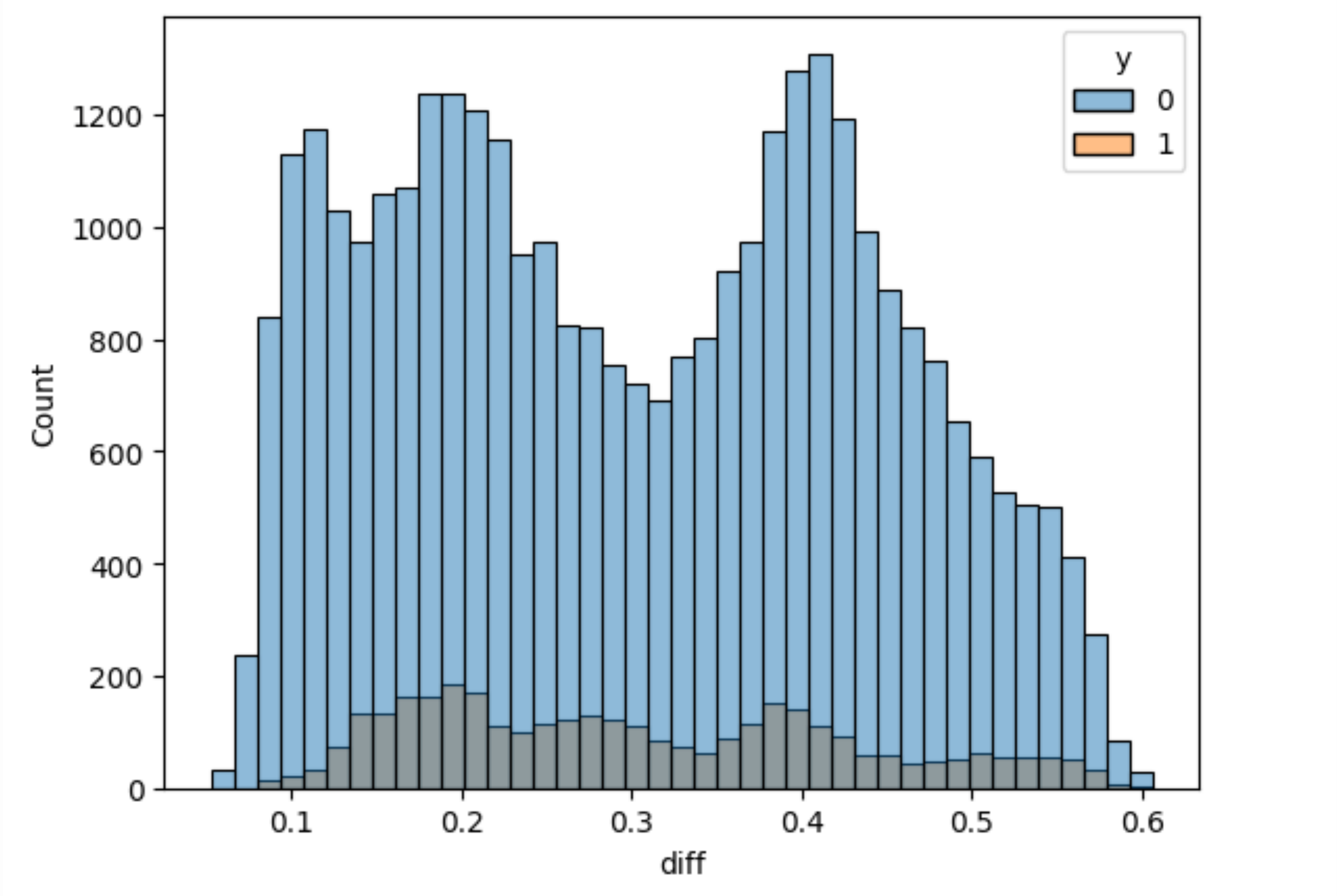
Not Previously contacted

Calculate differences in probabilities

UPLIFTING MODELING: METHODOLOGY

How to target correctly our marketing campaigns?

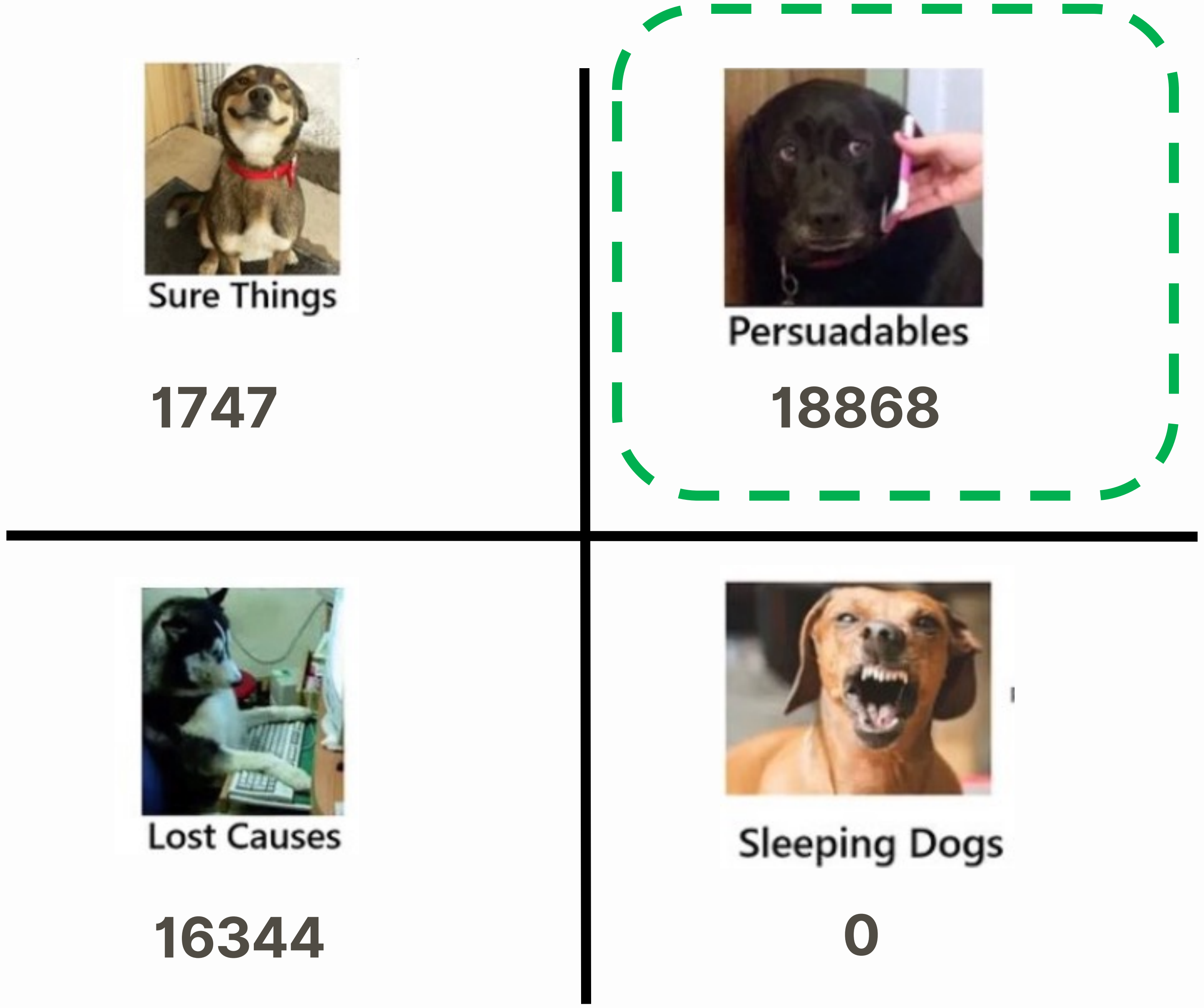
Differences in probabilities



All the differences are positive, which could be an indicator that the previous successful target campaign could lead to a future conversion. This would be analyzed further

UPLIFTING MODELING: METHODOLOGY

How to target correctly our marketing campaigns?



CAUSAL INFERENCE APPROACH 2

Aim: to analyze the causal impact of successful marketing campaigns on client decisions to subscribe to a term deposit among banking customers.
The project seeks to isolate the effect of the treatment (campaign success) from other confounding factors to make informed decisions.

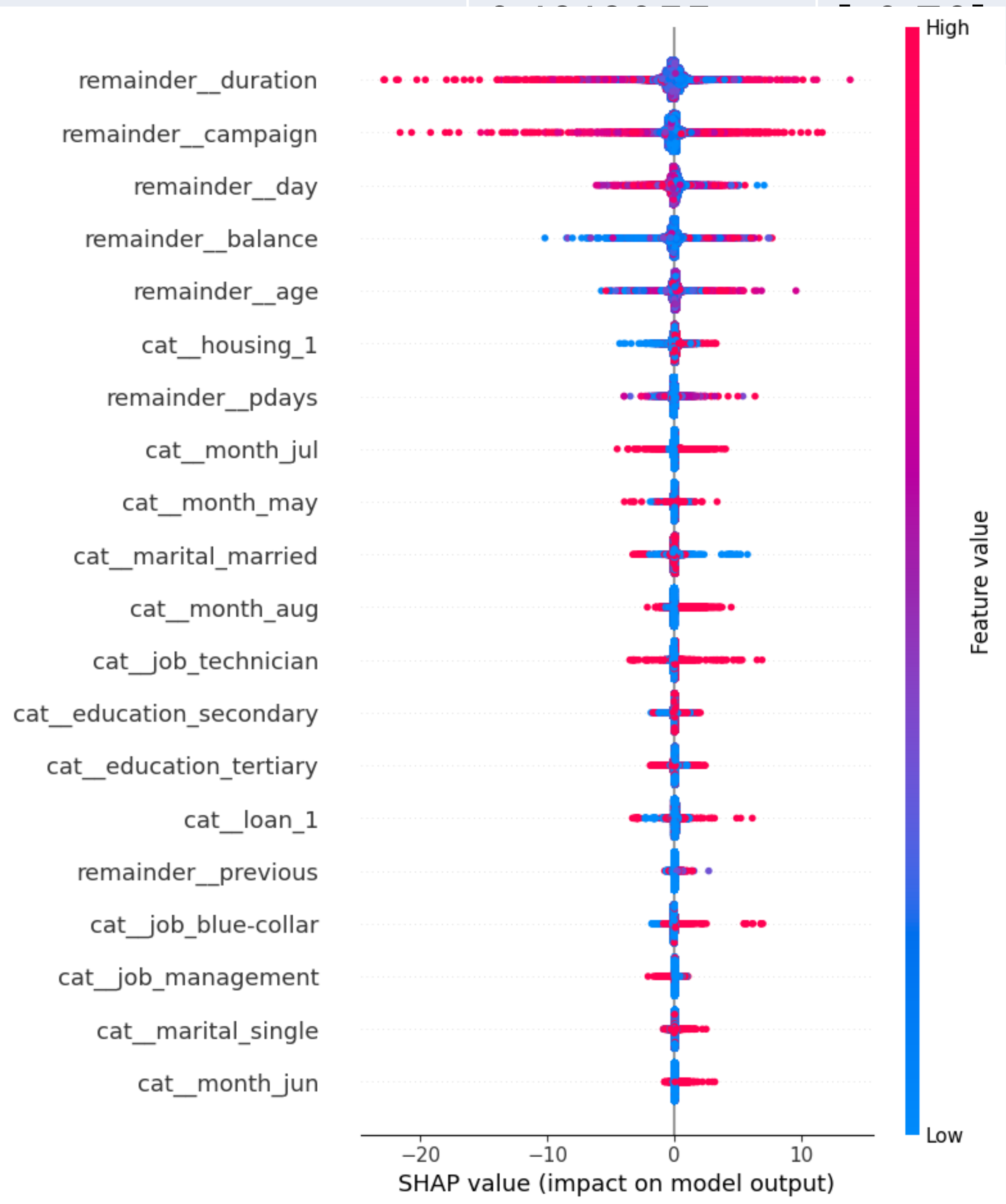
Approach	ATE	95% CI
LinearDMLCateEstimator	0.4995538	[0.47], [0.52]
Causal	[illegible]	[illegible] [1.14]

01 LinearDMLCateEstimator

- Estimate the average treatment effect (ATE) using a double machine learning approach with LassoCV models to control for confounders.

02 CausalForestDML

- Estimate the ATE using a causal forest approach to capture potential heterogeneity in treatment effects across the population.
- Utilize SHAP values to interpret the models and understand the contribution of each feature to the estimated treatment effects.



RECOMMENDATIONS

- Focus on people with good economic standings
- Time the campaign to taking in account the tax season
- Give balanced attention to the clients
- With a more targeted campaign we can have a better conversion rate
- Looking into the future

Thank You

Phone +1-514-514-0514

Address 1066 Sherbrook Blvd., Montreal, HA1 R27