

INTERNSHIP PROJECT

## TERM DEPOSIT SUBSCRIPTION PREDICTION

# BANK MARKETING ANALYTICS

DATA ANALYTICS & MACHINE LEARNING PROJECT

PREPARED BY:  
LAXMAN HADAGALI

## BUSINESS PROBLEM

# PROBLEM STATEMENT

BANKS CONDUCT LARGE-SCALE PHONE MARKETING CAMPAIGNS TO PROMOTE TERM DEPOSIT PRODUCTS. HOWEVER, MOST CUSTOMERS DO NOT SUBSCRIBE, RESULTING IN HIGH OPERATIONAL COSTS AND LOW CONVERSION RATES.

THE KEY CHALLENGE IS TO IDENTIFY POTENTIAL CUSTOMERS IN ADVANCE SO THAT MARKETING EFFORTS CAN BE FOCUSED ON CUSTOMERS WHO ARE MORE LIKELY TO SUBSCRIBE.



## DATA DESCRIPTION

# DATASET OVERVIEW

THE DATASET USED IN THIS PROJECT IS RELATED TO BANK MARKETING CAMPAIGNS CONDUCTED THROUGH DIRECT PHONE CALLS.

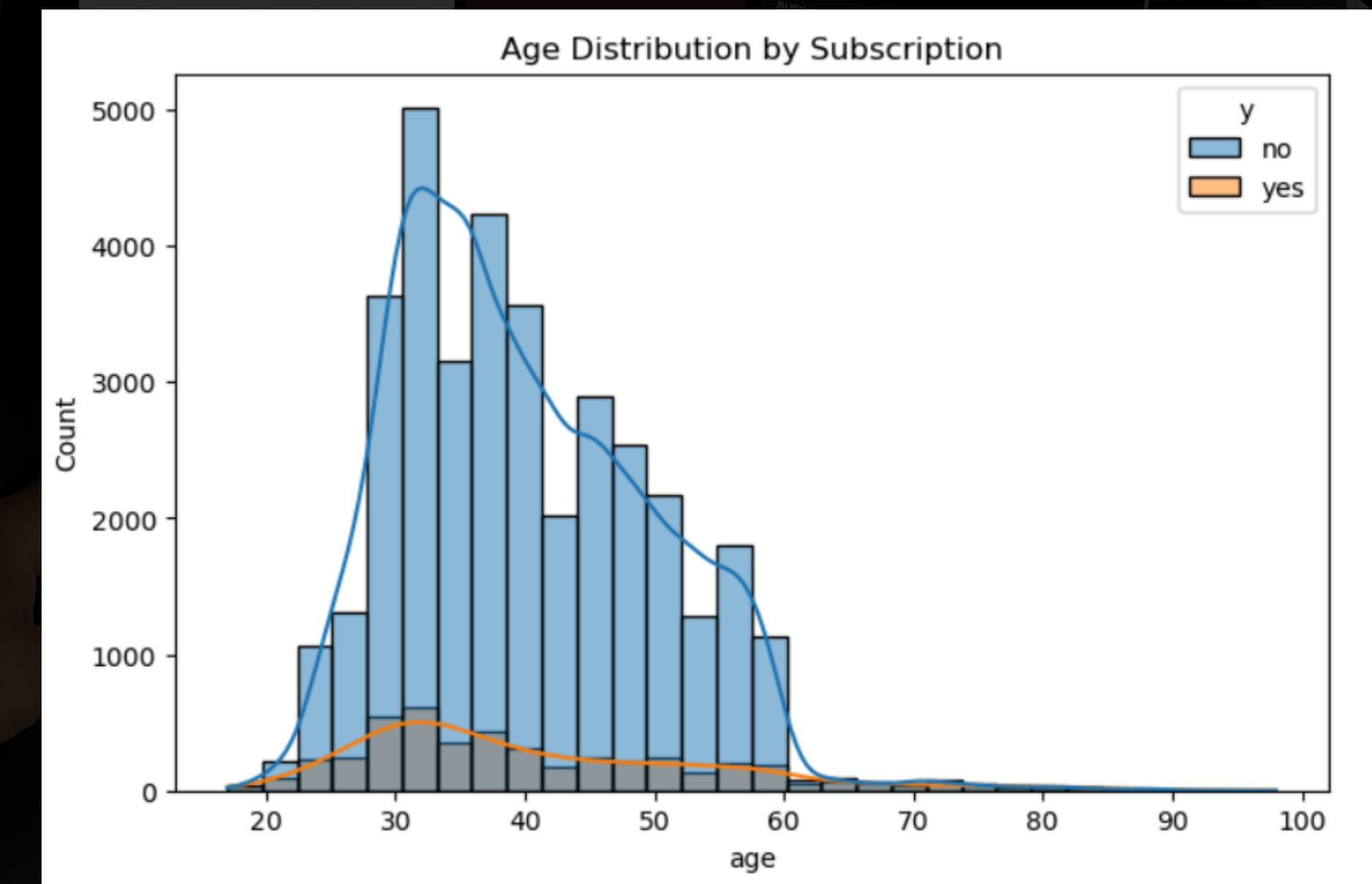
IT CONTAINS CUSTOMER DEMOGRAPHIC DETAILS, FINANCIAL INFORMATION, AND PREVIOUS CAMPAIGN OUTCOMES. THE TARGET VARIABLE INDICATES WHETHER A CUSTOMER SUBSCRIBED TO A TERM DEPOSIT (YES / NO).

## EDA OVERVIEW

# EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS WAS PERFORMED TO UNDERSTAND CUSTOMER BEHAVIOR AND IDENTIFY KEY FACTORS INFLUENCING TERM DEPOSIT SUBSCRIPTION DECISIONS.

THE ANALYSIS FOCUSED ON CUSTOMER DEMOGRAPHICS, FINANCIAL STATUS, PREVIOUS CAMPAIGN OUTCOMES, AND CAMPAIGN-RELATED ATTRIBUTES.



INSIGHT: CUSTOMERS BETWEEN THE AGE GROUP OF 30–50 SHOW HIGHER SUBSCRIPTION LIKELIHOOD COMPARED TO OTHER AGE GROUPS.

# INSIGHTS SUMMARY

# KEY INSIGHTS FROM EDA

- CUSTOMERS BETWEEN THE AGE GROUP OF 30–50 SHOW HIGHER SUBSCRIPTION LIKELIHOOD.
- CUSTOMERS WITH SUCCESSFUL OUTCOMES IN PREVIOUS MARKETING CAMPAIGNS ARE MORE LIKELY TO SUBSCRIBE.
- CUSTOMERS WITHOUT PERSONAL OR HOUSING LOANS SHOW BETTER RESPONSE TO TERM DEPOSIT OFFERS.
- EXCESSIVE CAMPAIGN CALLS REDUCE CUSTOMER INTEREST.



## PROJECT CHALLENGES

# CHALLENGES FACED

- THE TARGET VARIABLE WAS HIGHLY IMBALANCED, WITH FEWER CUSTOMERS SUBSCRIBING TO TERM DEPOSITS.
- MANY FEATURES WERE CATEGORICAL AND REQUIRED PROPER ENCODING BEFORE MODEL BUILDING.
- THE CALL DURATION FEATURE CAUSED DATA LEAKAGE AND HAD TO BE REMOVED FOR REALISTIC PREDICTIONS.
- SELECTING THE RIGHT EVALUATION METRIC WAS CHALLENGING, AS ACCURACY ALONE WAS MISLEADING.



# PREDICTIVE MODELING MODEL BUILDING

- MULTIPLE MACHINE LEARNING MODELS WERE BUILT TO PREDICT WHETHER A CUSTOMER WOULD SUBSCRIBE TO A TERM DEPOSIT.

THE MODELS IMPLEMENTED INCLUDE:

- LOGISTIC REGRESSION – USED AS A BASELINE MODEL
- DECISION TREE – TO CAPTURE NON-LINEAR PATTERNS
- RANDOM FOREST – TO IMPROVE PREDICTION PERFORMANCE AND HANDLE CLASS IMBALANCE



# LOGISTIC REGRESSION MODEL EVALUATION

LOGISTIC REGRESSION WAS USED AS A BASELINE CLASSIFICATION MODEL.

- ACHIEVED HIGH OVERALL ACCURACY
- PERFORMED WELL IN IDENTIFYING NON-SUBSCRIBERS
- RECALL FOR SUBSCRIBERS WAS LOW DUE TO CLASS IMBALANCE

KEY OBSERVATION: HIGH ACCURACY BUT LOW RECALL FOR SUBSCRIBERS,  
INDICATING CLASS IMBALANCE.

THIS MODEL HELPED ESTABLISH A BASELINE FOR COMPARISON WITH  
MORE ADVANCED MODELS.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.91      | 0.99   | 0.95     | 7310    |
| 1            | 0.70      | 0.22   | 0.33     | 928     |
| accuracy     |           |        | 0.90     | 8238    |
| macro avg    | 0.80      | 0.60   | 0.64     | 8238    |
| weighted avg | 0.88      | 0.90   | 0.88     | 8238    |

# DECISION TREE

# MODEL EVALUATION

DECISION TREE WAS USED TO CAPTURE NON-LINEAR RELATIONSHIPS  
IN CUSTOMER BEHAVIOR.

- IMPROVED RECALL FOR SUBSCRIBERS COMPARED TO LOGISTIC REGRESSION
  - EASY TO INTERPRET MODEL STRUCTURE
  - RISK OF OVERFITTING WITHOUT TUNING

THE MODEL PROVIDED BETTER INSIGHTS INTO DECISION PATTERNS  
BUT REQUIRED CAREFUL PARAMETER CONTROL.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.91      | 0.98   | 0.95     | 7310    |
| 1            | 0.67      | 0.26   | 0.37     | 928     |
| accuracy     |           |        | 0.90     | 8238    |
| macro avg    | 0.79      | 0.62   | 0.66     | 8238    |
| weighted avg | 0.89      | 0.90   | 0.88     | 8238    |

# RANDOM FOREST

# MODEL EVALUATION

RANDOM FOREST WAS USED TO IMPROVE PREDICTIVE PERFORMANCE AND HANDLE CLASS IMBALANCE EFFECTIVELY.

- ACHIEVED BETTER RECALL FOR SUBSCRIBERS
- REDUCED OVERRFITTING COMPARED TO DECISION TREE
  - MORE ROBUST AND STABLE PREDICTIONS

THIS MODEL WAS SELECTED AS THE FINAL MODEL DUE TO ITS BALANCED PERFORMANCE ACROSS EVALUATION METRICS.

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.91      | 0.98   | 0.95     | 7310    |
| 1            | 0.67      | 0.26   | 0.37     | 928     |
| accuracy     |           |        | 0.90     | 8238    |
| macro avg    | 0.79      | 0.62   | 0.66     | 8238    |
| weighted avg | 0.89      | 0.90   | 0.88     | 8238    |

PERFORMANCE COMPARISON

# MODEL COMPARISON

| Model | Accuracy | Precision (Yes) | Recall (Yes) | F1-score (Yes) |
|-------|----------|-----------------|--------------|----------------|
|-------|----------|-----------------|--------------|----------------|

|   |                     |          |      |      |      |
|---|---------------------|----------|------|------|------|
| 0 | Logistic Regression | 0.901068 | 0.70 | 0.22 | 0.33 |
|---|---------------------|----------|------|------|------|

|   |               |          |      |      |      |
|---|---------------|----------|------|------|------|
| 1 | Decision Tree | 0.902039 | 0.67 | 0.26 | 0.37 |
|---|---------------|----------|------|------|------|

|   |               |          |      |      |      |
|---|---------------|----------|------|------|------|
| 2 | Random Forest | 0.847779 | 0.39 | 0.64 | 0.49 |
|---|---------------|----------|------|------|------|

RANDOM FOREST ACHIEVED THE BEST RECALL AND F1-SCORE FOR  
SUBSCRIBERS, MAKING IT THE MOST SUITABLE MODEL FOR THIS  
BUSINESS PROBLEM.

RANDOM FOREST WAS SELECTED AS THE FINAL PREDICTIVE MODEL.

# BEST MODEL SELECTION

BASED ON THE EVALUATION RESULTS, RANDOM FOREST WAS SELECTED AS THE FINAL PREDICTIVE MODEL.

- ACHIEVED THE HIGHEST RECALL FOR IDENTIFYING SUBSCRIBERS
- PROVIDED BETTER BALANCE BETWEEN PRECISION AND RECALL
- REDUCED OVERTFITTING COMPARED TO DECISION TREE

THIS MODEL ALIGNS WELL WITH THE BUSINESS OBJECTIVE OF IDENTIFYING POTENTIAL CUSTOMERS FOR TERM DEPOSIT OFFERS.

YOUR PARAGRAPH



MARKETING STRATEGY INSIGHTS

# BUSINESS RECOMMENDATIONS

BASED ON THE ANALYSIS AND PREDICTIVE MODELING RESULTS,  
THE FOLLOWING RECOMMENDATIONS ARE SUGGESTED FOR THE  
BANK MARKETING TEAM:

- FOCUS MARKETING EFFORTS ON CUSTOMERS AGED BETWEEN 30–50 WHO SHOW HIGHER SUBSCRIPTION LIKELIHOOD.
- PRIORITIZE CUSTOMERS WITH SUCCESSFUL OUTCOMES IN PREVIOUS MARKETING CAMPAIGNS.
- REDUCE EXCESSIVE CALL FREQUENCY TO AVOID CUSTOMER FATIGUE AND NEGATIVE RESPONSE.
- USE THE PREDICTIVE MODEL TO TARGET HIGH-PROBABILITY CUSTOMERS AND OPTIMIZE MARKETING COSTS.



## PROJECT SUMMARY

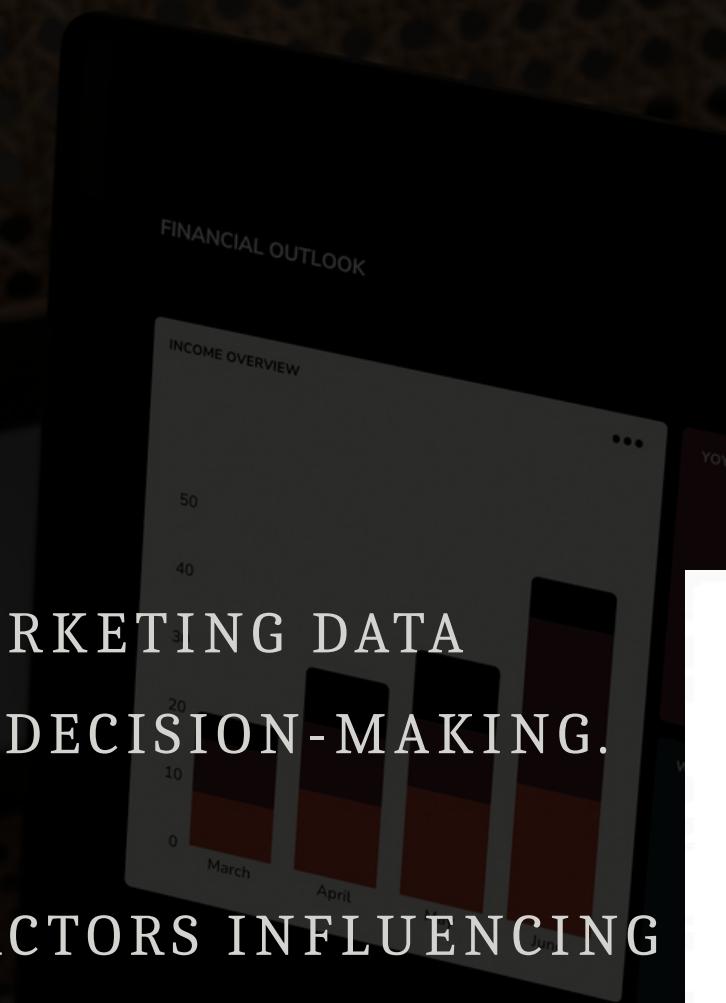
# CONCLUSION

THIS PROJECT SUCCESSFULLY ANALYZED BANK MARKETING DATA TO UNDERSTAND CUSTOMER BEHAVIOR AND IMPROVE DECISION-MAKING.

EXPLORATORY DATA ANALYSIS HELPED IDENTIFY KEY FACTORS INFLUENCING TERM DEPOSIT SUBSCRIPTION.

MULTIPLE MACHINE LEARNING MODELS WERE EVALUATED, AND RANDOM FOREST WAS SELECTED AS THE BEST MODEL FOR PREDICTION.

THE INSIGHTS AND RECOMMENDATIONS FROM THIS PROJECT CAN HELP THE BANK OPTIMIZE MARKETING STRATEGIES AND INCREASE CONVERSION RATES.



GROUND

THE INDUSTRY'S HISTORY

# THANK YOU

