

Banking on Data: Marketing Analysis

Benjamin Gerber

June 2024

OUR MISSION

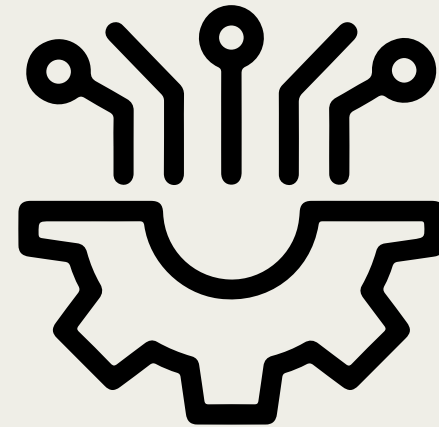
Analyze the marketing campaign of a bank to find actionable insights for improvement

HOW ARE WE GOING TO DO IT?



1

Data Analysis



2

**Machine
Learning**



3

Visualization

OUR DATASET

11,162 Rows

17 Features

TARGET VARIABLE

Deposit: Whether someone deposited money into their account during the marketing campaign period

WHY DOES THIS MATTER?

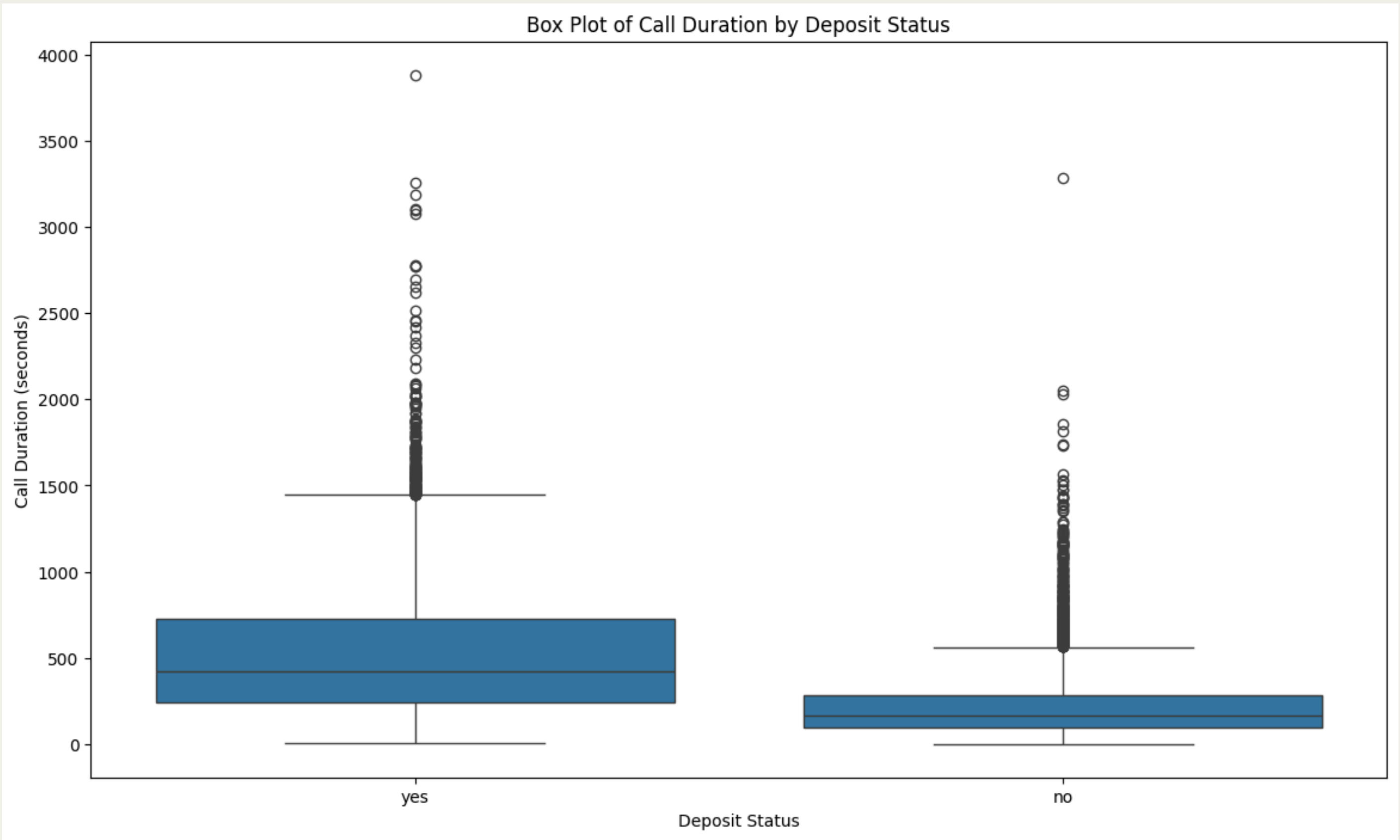
- Banks need money from deposits to invest in order to make money
- The effectiveness of this marketing campaign is to get people to deposit
 - The dataset did not measure amount deposited



EDA

EXPLORATORY DATA ANALYSIS

CALL DURATION (SECONDS) VS DEPOSIT



CALL DURATION (SECONDS) VS DEPOSIT

Deposit?	Min	25%	Median	Mean	75%	Max
No	2.0	94.0	163.0	223.1	282.0	3284.0
Yes	8.0	244.0	426.0	537.2	725.0	3881.0

Longer Calls are More Likely to Deposit

BALANCE (\$) VS DEPOSIT

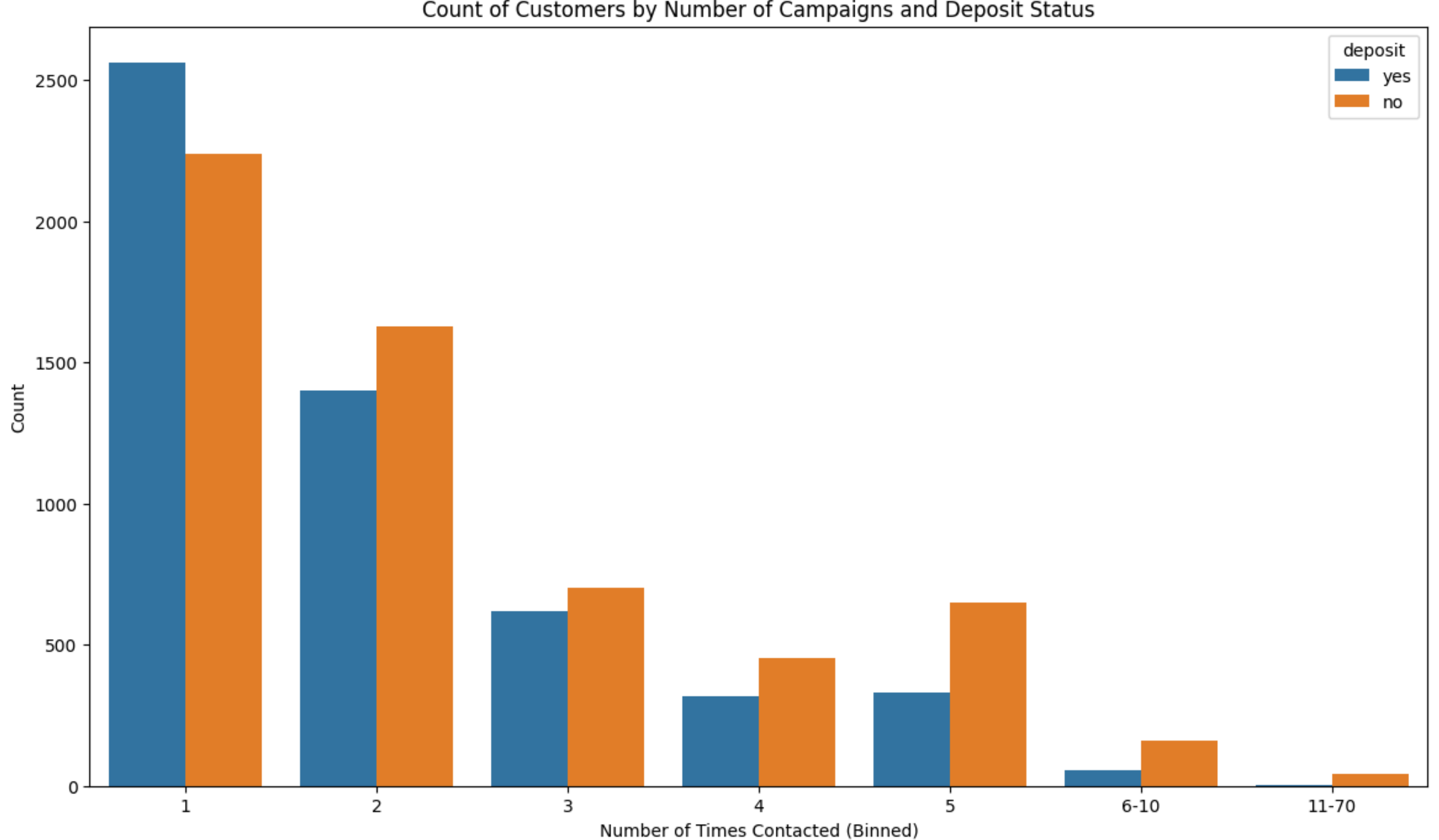
Deposit?	Min	25%	Median	Mean	75%	Max
No	-\$6847.00	\$64.00	\$414.00	\$1280.23	\$1324.00	\$66653.00
Yes	-\$3058.00	\$210.00	\$733.00	\$1804.27	\$2159.00	\$81204.00

Higher Balances are More Likely to Deposit

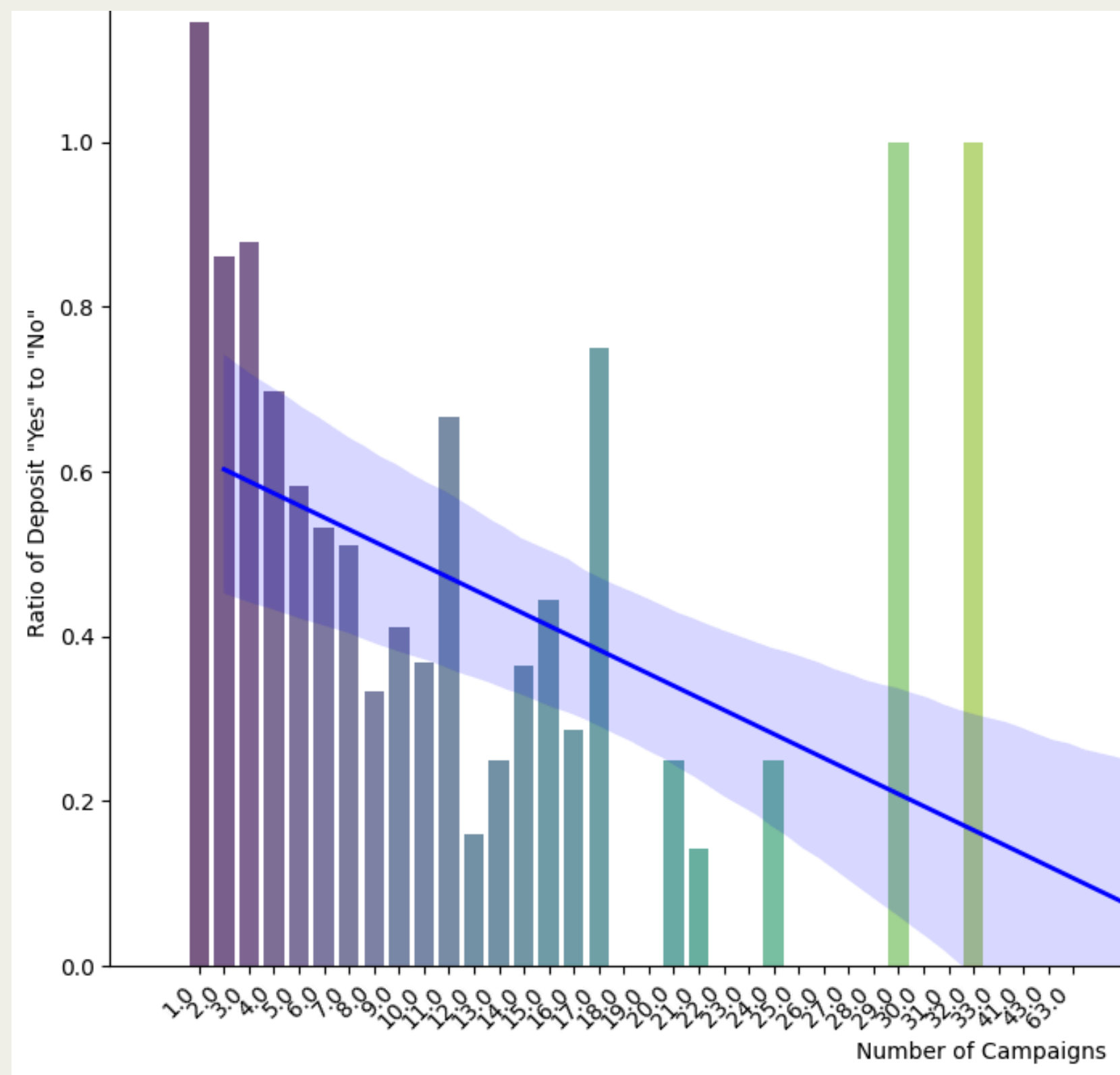
CAMPAIGN VS DEPOSIT

Campaign: The number of marketing calls made to a customer during the marketing campaign

CAMPAIGN VS DEPOSIT



CAMPAIGN VS DEPOSIT



When people are contacted more, they tend to not deposit.

THREE ACTIONABLE INSIGHTS

1

EXTEND THE
CALL

2

TARGET
ACCOUNTS
WITH HIGH
BALANCES

3

LESS IS
MORE

Machine Learning

OUR MODELS

Decision Tree Classifier

- Advantages
 - Step By Step Model
 - Interpretable
- Disadvantages
 - Possible Overfitting
 - Tree Pruning / Simplifying Model

Logistic Regression

- Advantages
 - Provides Probability for Output
 - Low Training Time
- Disadvantages
 - Struggles With Multicollinearity (when different features are correlated)

OUR MODELS

Both Models Have One Goal:

Accurately predict whether or not
someone will deposit during the
marketing campaign

DECISION TREE ANALYSIS

Deposit?	Precision	Recall	F-1 Score	Accuracy
No	0.86	0.75	0.80	0.80
Yes	0.75	0.87	0.81	

DECISION TREE ANALYSIS

Is this Accuracy Good?

Random Guessing

- 1 Guess = 50% Chance

Decision Tree

- 1 Guess = 80% Chance
- 2 Guesses = 64% Chance
- 3 Guesses* = 51.2% Chance

*Chance of getting all 3 predictions correct

It take our model four predictions (and getting all correct) before random guessing has a higher chance of just getting one guess correct.

DECISION TREE ANALYSIS

Does the Model Back the EDA Insight?

- The model includes these variables to make decisions:
 - Duration
 - Balance
 - Campaign
- Although those variables are in the model, we don't know how impactful they are on the effect to deposit

So, we create a second model to try to gain more insight.

LOGISTIC REGRESSION ANALYSIS

1

**CREATE
MODEL WITH
ALL
FEATURES**

2

**SIMPLIFY
MODEL TO
GAIN DEEPER
INSIGHT**

LOGISTIC REGRESSION ANALYSIS

Initial Model

- 42 Features
- Accuracy: 83.0%

Simplified Model

- 10 Features
- Accuracy: 79.2%

Our simplified model is much more interpretable while being nearly just as accurate!

LOGISTIC REGRESSION ANALYSIS

DURATION OF CALLS
IS BY FAR THE MOST
IMPACTFUL THING ON
DEPOSIT LIKELIHOOD

PEOPLE THAT
DEPOSITED IN EARLIER
CAMPAIGNS ARE MUCH
MORE LIKELY TO
DEPOSIT AGAIN

PEOPLE WITH HOUSING
LOANS ARE MUCH
LESS LIKELY TO
DEPOSIT

THE NUMBER OF CALLS
(CAMPAIGN) HAS A
STRONG NEGATIVE
IMPACT ON DEPOSIT

LOGISTIC REGRESSION ANALYSIS

**GOAL 1: KEEP CALLS
LONG!**

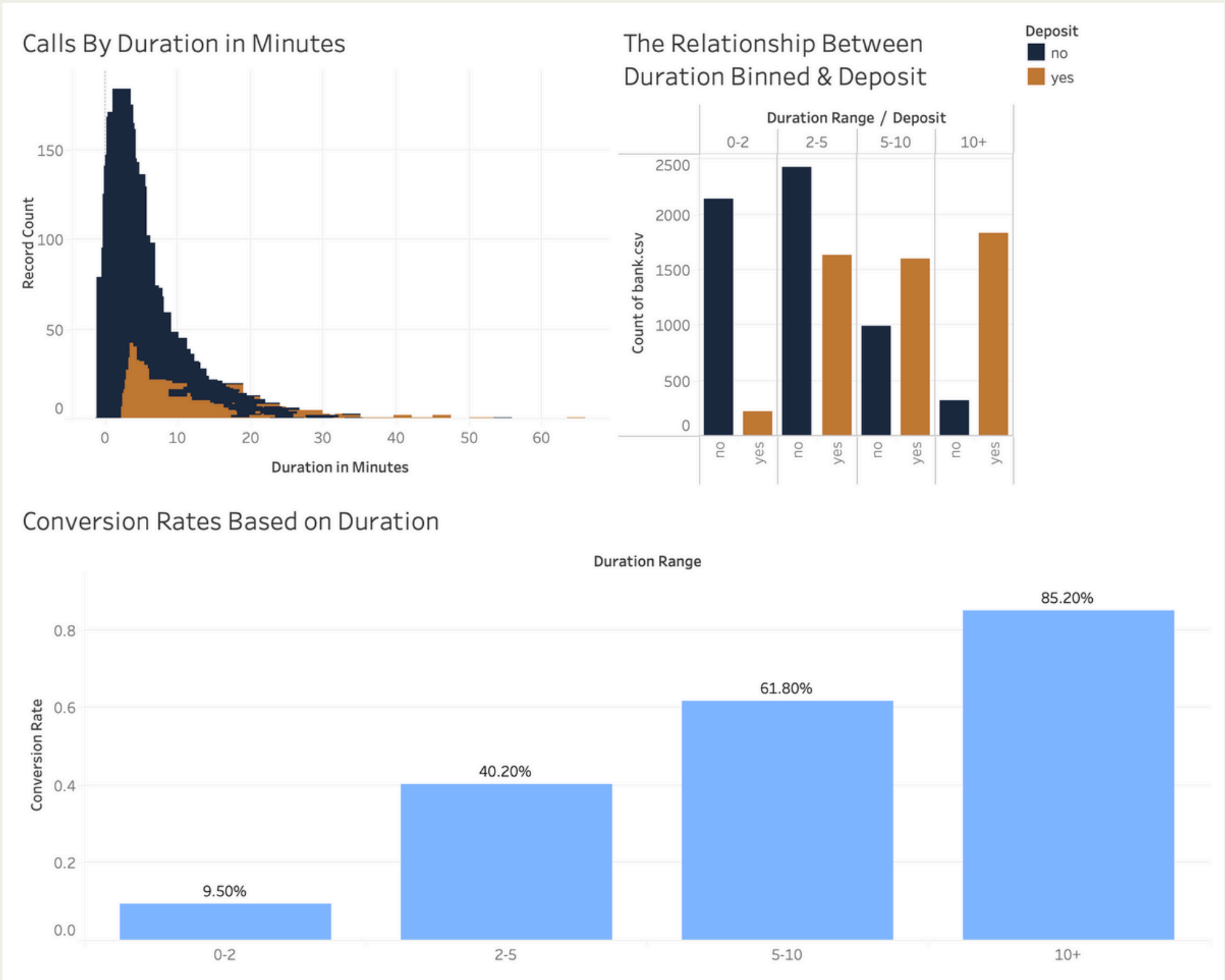
**GOAL 2: CONTACT
PEOPLE WHO
DEPOSITED IN
PREVIOUS CAMPAIGNS**

**GOAL 3: AVOID PEOPLE
WITH HOUSING LOANS**

**GOAL 4: LESS IS
MORE, DON'T WASTE
TIME CONTINUOUSLY
CALLING SOMEONE**

How long should calls be?

DASHBOARD ANALYSIS



> 5 Minutes

WHAT NEXT?

Time-Series Analysis

- Which months are we more successful during?
- What days of the week result in longer calls?

A/B Testing

- Do certain scripts lead to longer calls?

APPENDIX

GitHub Repository_ for Full Code (Link)