

A blurred background image featuring a financial chart. It includes a line graph with white circular markers and a bar chart with blue bars. A numerical value '289.33' is visible on the chart. The overall color scheme is dark blue and black.

# BBI MARKETING STRATEGY PRESENTATION

**Mary Shas**

Director of Marketing Analytics for Buckeye Bank Inc.

# PROBLEM STATEMENT

- We want to improve the marketing campaigns at BBI that are based on phone calls.
- How?
  - We'll predict if the client will subscribe a term deposit using classification models.
  - If yes, we'll contact.



WHICH GROUP OF CLIENTS ARE  
LIKELY TO SUBSCRIBE TO OUR  
SERVICES?

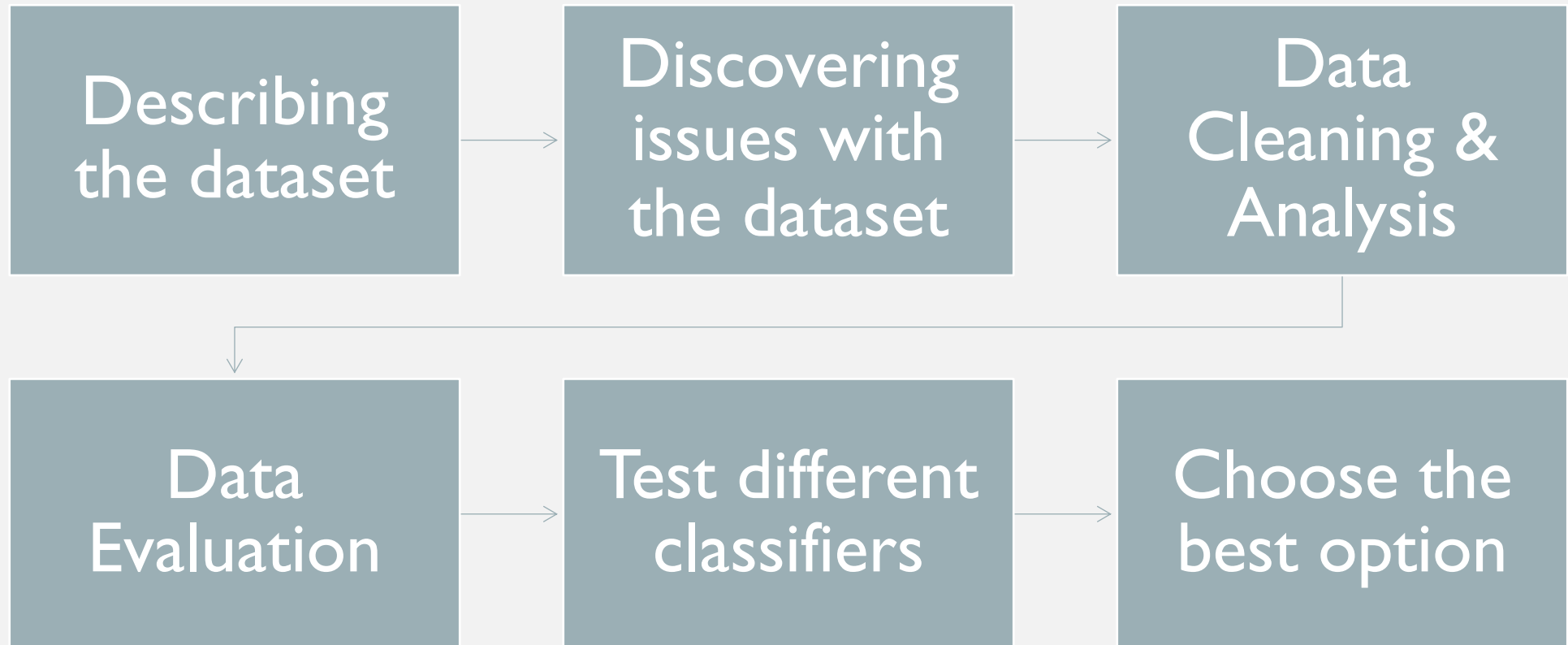


TOTAL SALES COST: \$50 (FIXED)  
ONLY CONTACT IF WE PREDICTED  
THE CLIENT WILL SUBSCRIBE.



SALES BENEFIT: \$200

# THE PROCESS



# DATASET DESCRIPTION

- Dataset taken from existing marketing campaigns.
- The dataset contains 21 attributes, in which more than half of the attributes are numerical, and the other half are categorical with mostly nominal traits. Three of the attributes are binary type.
- The attributes are a combination of the client's personal information and our marketing history with them. We also included rates and indexes related to our business.

Age
Job
Marital Status
Education
Has credit?
Has housing loan?
Has personal loan?
Contact Type
Day of the Week
Month
Last contact duration
Number of contact during campaign
Number of days passed after contact
Previous marketing outcome
Employment variation rate
Consumer price index
Consumer confidence index
Rate of deposits
Number of employees
Has the client subscribed to term deposit?

# ISSUES WITH THE DATASET

## UNBALANCED CLASS

- The dataset is dominated by 'no' class.

## OUTLIERS IS AGE

- The min and max value in age is too low and too high, respectively.

## MISSING VALUES

- Marital, Default, Month, Poutcome, Y contains missing values.

## IRRELEVANT ATTRIBUTE

- Discard cons.conf.idx attribute because has zero relation to class variable.

## EVALUATION METHODS FOR CLASSIFIER

- Benefit-Cost Model: a table that lists the benefit and cost amounts for a decision on a single prospect
- Accuracy: comparison of actual class values to predicted values

Actual Class	Predicted Class	Sale Benefit	Sale Cost	Net Benefit (Benefit – Cost)
FALSE	FALSE	0	0	0
FALSE	TRUE	0	-50	-50
TRUE	FALSE	0	0	0
TRUE	TRUE	200	50	150

# CLASSIFIER EVALUATION

Attributes chosen to use in the classifier: AGE, JOB, MARITAL, DURATION

## **KNN (Neighbors = 5)**

Accuracy: 0.88  
True Positive: 377  
False Positive: 459

Net Benefit per Prospect:  
\$2.72

Net Benefit per Year:  
\$108,993.59

## **DecisionTree (Max Depth = 7)**

Accuracy: 0.90  
True Positive: 353  
False Positive: 287

Net Benefit per Prospect:  
\$3.13

Net Benefit per Year:  
**\$125,212.88**

## **Gaussian Naive Bayes (Neighbors = 5)**

Accuracy: 0.84  
True Positive: 516  
False Positive: 1024

Net Benefit per Prospect:  
\$2.12

Net Benefit per Year:  
\$84,989.05



## ANALYSIS RESULT

### Best Result: **Decision Tree**

- The number of true positive is higher compared to false positive, which indicates Decision Tree best predict which client will subscribe to the term deposit.
- It also has the highest number of total net benefit per year and the highest accuracy.
- It works well with the dataset because it handles both categorical and numerical attributes well.



# RECOMMENDATION

## **Marketing Approach**

Using classifier to help us predict if the client will or will not subscribe to the term deposit.

- ✓ Can proceed with the marketing approach
- ✓ The model's accuracy is high
- ✓ Revenue
  - ✓ The company can make good money from this marketing approach

## **POSSIBLE IMPROVEMENTS AND NEXT STEP**



**Use a more balanced dataset for training.**

Dataset with more 'yes' result would help better in training the model.



**Don't 100% rely on calling as the marketing strategy.**

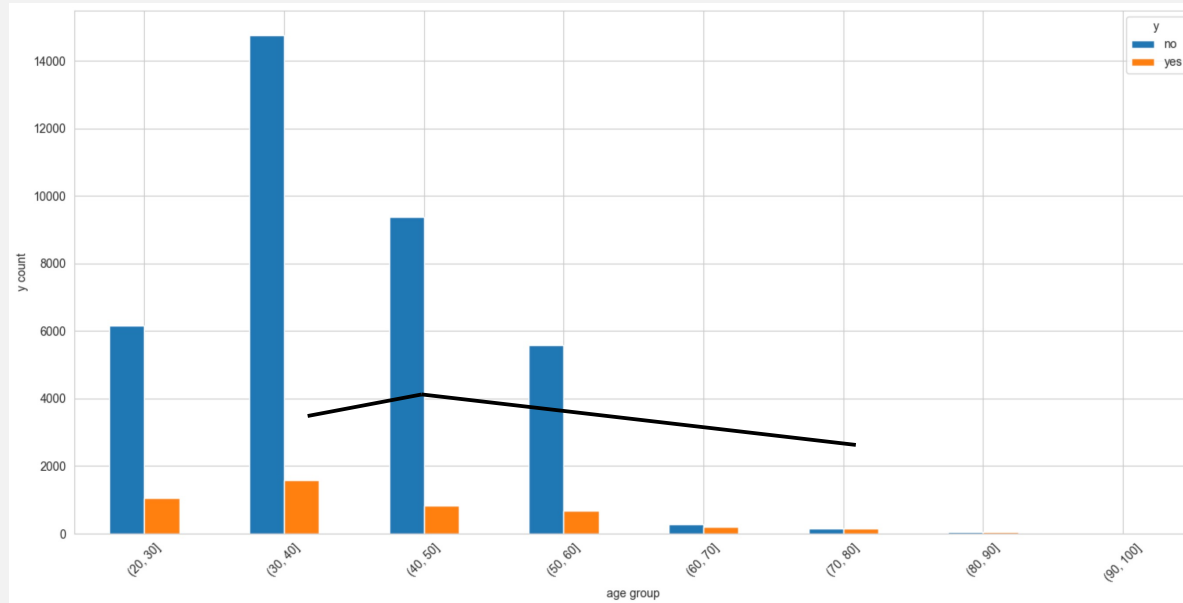
Use online ads or pass out brochures to clients in the bank.

# CONCLUSION

Using classifier to predict if our client will subscribe to our term deposit is a good marketing strategy because it helps us gain guaranteed profit. However, we should train our models with more balanced dataset in the future to have more accurate result. We should also be more open to other marketing strategy so we can further improve our market share.

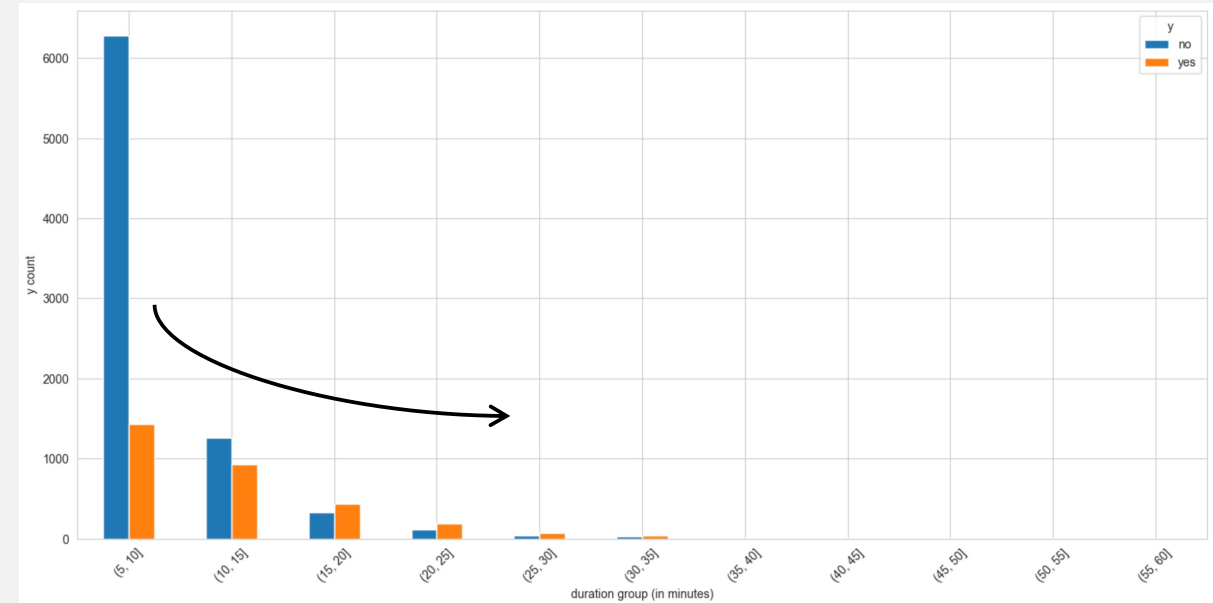
# **APPENDICES**

# Analyzing the Relationship



subscribed count vs age

- Client within the age 30-40 has the highest count of subscribed
- Clients that are  $\geq 60$  years old, unlikely to subscribe



subscribed count vs duration

- Calls with shorter duration leads to higher number of subscribed
- Good predictor; useful for classification

# Analyzing the Relationship

y_numeric	0.0	1.0	All
marital			
divorced	4135	476	4611
married	22398	2532	24925
single	9946	1620	11566
All	36474	4628	41102

subscribed count vs marital status

- Clients that are married are likely to subscribed to the term deposit.

y_numeric	0.0	1.0	All
job			
admin.	9057	1350	10407
blue-collar	8602	635	9237
entrepreneur	1330	123	1453
housemaid	951	106	1057
management	2593	328	2921
retired	1282	433	1715
self-employed	1267	149	1416
services	3639	323	3962
student	599	275	874
technician	6002	728	6730
unemployed	865	144	1009
unknown	287	34	321
All	36474	4628	41102

subscribed count vs job

- Clients who work an administrative job accounts for the highest number of subscribed count.

# KNN Classifier

Confusion Matrix:

	0	1
0	10483	459
1	1012	377

Classification Report:

	precision	recall	f1-score	support
0.0	0.91	0.96	0.93	10942
1.0	0.45	0.27	0.34	1389
accuracy			0.88	12331
macro avg	0.68	0.61	0.64	12331
weighted avg	0.86	0.88	0.87	12331

Net Benefit per Record: \$ 2.72

Net Benefit per Year: \$ 108993.59

# Gaussian Naive Bayes

Confusion Matrix:

	0	1
0	9894	1024
1	897	516

Classification Report:

	precision	recall	f1-score	support
0.0	0.92	0.91	0.91	10918
1.0	0.34	0.37	0.35	1413
accuracy			0.84	12331
macro avg	0.63	0.64	0.63	12331
weighted avg	0.85	0.84	0.85	12331

Net Benefit per Record: \$ 2.12

Net Benefit per Year: \$ 84989.05



# **APPENDIX**

# References

1. PresentingDataAnalysisResults\_VI.pdf, Thomas Bihari's Additional Guidance for Homework 3
2. PresentingDataAnalysisResults\_Examples\_VI.pdf, Thomas Bihari's Additional Guidance for Homework 3
3. Homework 1
4. Homework 2

# Adjustments

1. Adjusted the Benefit-Cost model in Homework 2.

# Reflections

I underestimate how hard it is to make a short but informational slides, especially slides for presenting my findings from my own analysis. This homework have taught me the importance of really understanding my dataset and what does my findings mean and challenge myself to effectively communicate my analysis and results to those who needs it.