LOGISTIC REGRESSION ANALYSIS MARKETING CAMPAIGN

PRESENTED BY: THE DEBUGGING DRAGONS
DAVID LAYNE | JENNA BELLAGAMBA | S M SIDDIQUR RAHMAN LIKHON | RASHI RATAN

THE BUSINESS PROBLEM:

Background	A large retail company operates across a variety of products - food, grocery items, beverages, household items and much more.
Goal	The company seeks to optimize and enhance its marketing efforts by improving marketing effectiveness and customer targeting.
Action	To support this goal, the organization is providing a dataset that provides customer demographics like age, marital status, education, income, and purchase behavior.

What is the purpose of the provided dataset?

Helps to understand:	Customer behavior Segment customers for better targeting Increase retention and overall revenue and, Gauge the efficiency and effectiveness of marketing campaigns.
Demonstrate marketing interactions through:	Amount spent by category, frequency, and recency A variety of contacts Responses to marketing campaigns.

Our Solution

- The business objectives requires a predictive solution that can anticipate customer responses (Accepted/Declined) for future marketing campaigns.
- This classification will enable more targeted outreach, better allocation of marketing resources, and higher conversion rates.

REVIEW OF OUR DATA

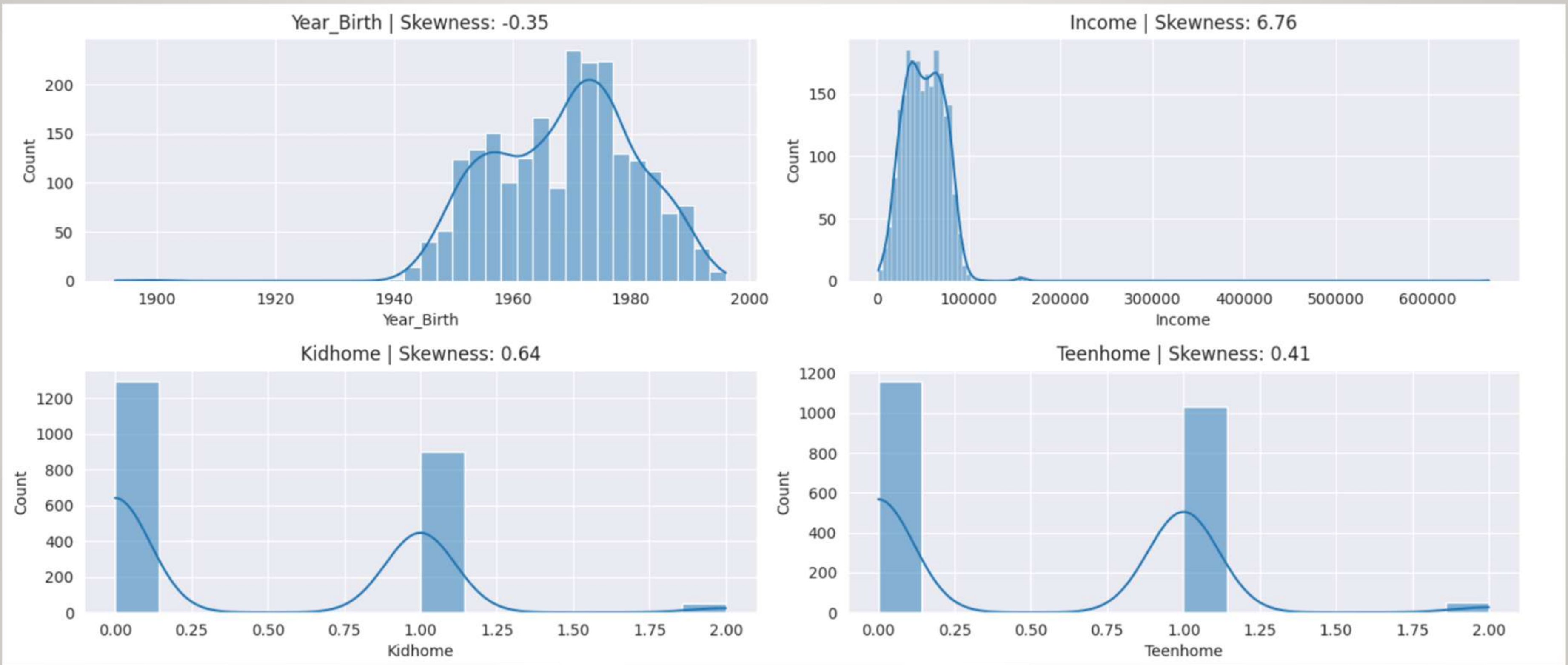
	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines
0	5524	1957	Graduation	Single	58138.0	0	0	2012-09-04	58	635
1	2174	1954	Graduation	Single	46344.0	1	1	2014-03-08	38	11
2	4141	1965	Graduation	Together	71613.0	0	0	2013-08-21	26	426
3	6182	1984	Graduation	Together	26646.0	1	0	2014-02-10	26	11
4	5324	1981	PhD	Married	58293.0	1	0	2014-01-19	94	173

NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	Z_CostContact	Z_Revenue	Response
7	0	0	0	0	0	0	3	11	1
5	0	0	0	0	0	0	3	11	0
4	0	0	0	0	0	0	3	11	0
6	0	0	0	0	0	0	3	11	0
5	0	0	0	0	0	0	3	11	0

OUR PROCESS

1. Eliminate null values
2. Declare features and target
3. Variance Visualization
4. Data Modeling & Scaling
5. Results and Evaluation
6. Accuracy, Precision, Recall, F-1 Score
7. Confusion Matrix
8. Important Metrics
9. Summary
10. Conclusion

VARIANCE VISUALIZATIONS

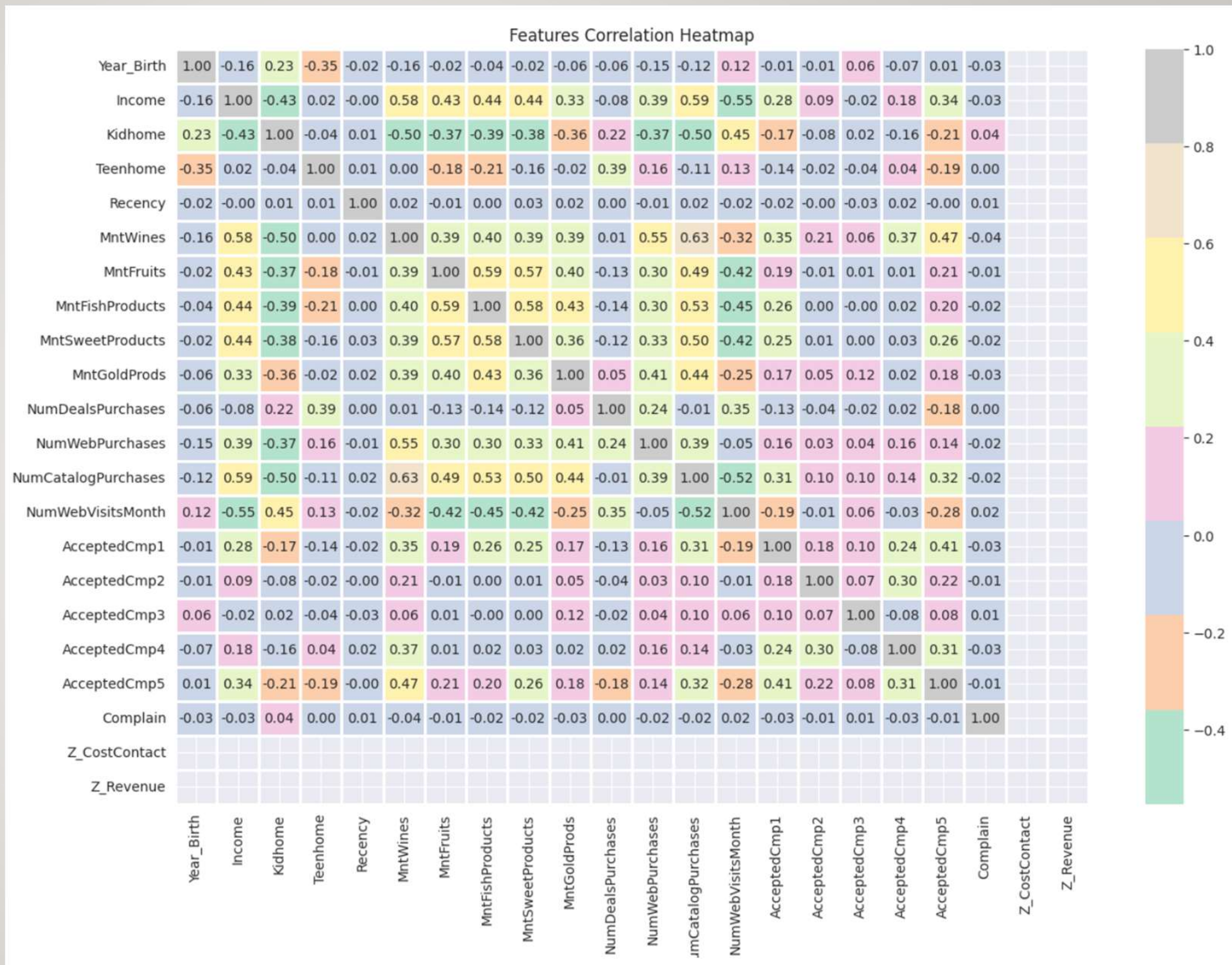


FEATURES CORRELATION HEATMAP

Values close to +1 indicates strong positive correlation

-1 indicates a strong negative correlation

0 indicates suggests no linear correlation.



ACCURACY

Accuracy: Test Set: 0.8761

Interpretation:
The model correctly classified 87.61% of the customer responses in the test set.

This indicates a very high overall performance in distinguishing an Acceptance over a Declination for the last marketing campaign.

PRECISION

Precision: Test Set: 0.7308

Interpretation:
When the model predicts an accepted response (Response=1), it is correct 73.08% of the time.

High precision is important if the cost of incorrectly flagging a customer response as accepted (FP) is high over a Declination for the last marketing campaign.

RECALL

Recall: Test Set: 0.2836

Interpretation:
The model identified 28.36% of all the truly accepted responses present in the test set.

High recall is crucial if the cost of missing a declination (FN) is very high. This model has a low recall score, leading to declinations being wrongfully shown as accepted(FN)

F-1 SCORE

F1-Score: Test Set: 0.4086
Interpretation: The F1-Score of 0.4086 is low.

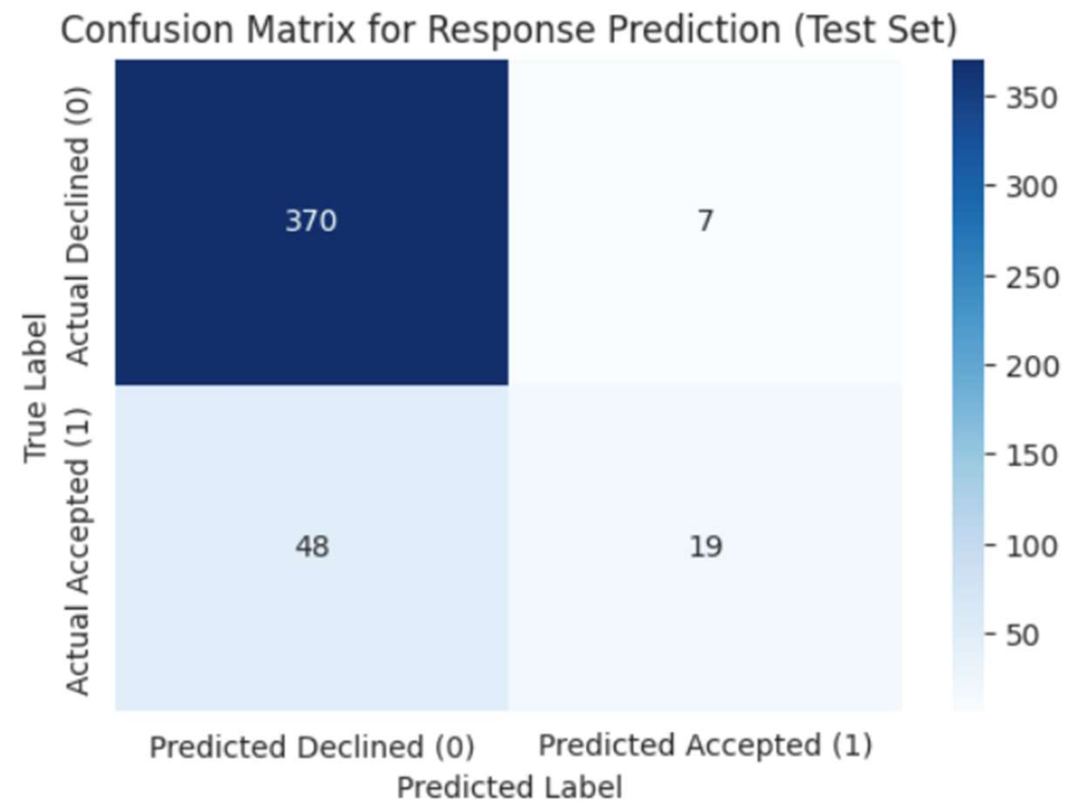
The model is flawed in finding actual customer acceptances (TP)

```
--- Classification Report (Comprehensive Summary - Marketing Campaign Response Model) ---
                       precision    recall  f1-score   support

Declined Response (0)       0.89      0.98      0.93       377
Accepted Response (1)       0.73      0.28      0.41        67

             accuracy                           0.88       444
            macro avg       0.81      0.63      0.67       444
         weighted avg       0.86      0.88      0.85       444
```

CONFUSION MATRIX

Model Notes:

- True Negatives (TN): 370 (Correctly predicted 'Declined')
- False Positives (FP): 7 (Incorrectly predicted 'Accepted')
- False Negatives (FN): 48 (Incorrectly predicted 'Declined')
- True Positives (TP): 19 (Correctly predicted 'Accepted')




```
--- Confusion Matrix Components (Response Model) ---
True Negatives (TN): 370 (Correctly predicted 'Declined')
False Positives (FP): 7 (Incorrectly predicted 'Accepted')
False Negatives (FN): 48 (Incorrectly predicted 'Declined')
True Positives (TP): 19 (Correctly predicted 'Accepted')
```

COEFFICIENTS & INTERPRETATIONS:

TOP 5 SIGNIFICANT FEATURES IMPACTING CAMPAIGN RESPONSE

Feature	Impact (Odds-Ratio Interpretation)
AcceptedCmp3	+63% higher odds with each past acceptance
NumWebVisitsMonth	+55% higher odds with each extra website visit
AcceptedCmp5	+47% higher odds with each past acceptance
Teenhome	-45% odds with each additional teenager at home
Recency	-52% odds with each additional day since last purchase

```
Intercept (β₀ - Log-Odds when X=0): -2.3652

----Model Coefficients (Log-Odds):----
AcceptedCmp3: 0.49
NumWebVisitsMonth: 0.44
AcceptedCmp5: 0.39
Teenhome: -0.61
Recency: -0.74

----Odds-Ratios:----
AcceptedCmp3: 1.63
NumWebVisitsMonth: 1.55
AcceptedCmp5: 1.47
Teenhome: 0.55
Recency: 0.48
```

Comments/Analysis:

- Each coefficient holds all other features constant
- Positive coefficients = greater acceptance likelihood
- Negative coefficients = lower acceptance likelihood
- Recency and Teenhome decrease acceptance odds sharply

KEY MODEL PERFORMANCE METRICS

Accuracy Score

87.61% predictions correct

- Measures overall prediction success
- Indicates the model's reliability for business decisions

Precision Score

73.08% precision on positive predictions

- Of customers predicted to accept, 73% actually accepted
- Important for correctly identifying *true* responders

Precision vs Recall Trade-Off

- Precision prioritized over Recall
  ➔ Precision reduces *false positives*
  ➔ Minimizes overestimation of campaign success
- Critical for **budget-conscious** marketing strategies

MODEL SUMMARY

Declined Response Class:

Precision: 0.89 – Predictions of decliners are highly accurate.

Recall: 0.98 – Nearly all actual decliners are correctly identified.

Insight: Very effective at filtering out non-responders and minimizing wasted outreach.

Accepted Response Class:

Precision: 0.73 – Moderately accurate when predicting acceptances.

Recall: 0.28 – Fails to identify most people who actually would accept.

Insight: This is the model's critical weakness and the key barrier to improving campaign engagement.

CONCLUSIONS

Insights:

The model accurately and effectively identifies likely decliners with high precision (0.89) and recall (0.98). Despite the high accuracy of the model, lower precision (0.73) and significantly low recall (0.28) for accepted responses is a major concern. The model's ability to correctly identify campaign acceptance is limited and hindering its overall effectiveness. This conclusion is also supported by the F1-Score (0.41) reflecting a precision-recall imbalance.

Implications:

The current state of the model is not optimal for maximizing campaign reach and conversions due to missed opportunities in engaging potential acceptors.

Recommendations:

Prioritize improving the model's recall for the Accepted Response class through further investigation and analysis (e.g., exploring algorithms, feature engineering, threshold adjustments).

THANK YOU!

Review our findings!

Google Colab Link:
https://colab.research.google.com/drive/1a4oWpGlKEl0eHUD84xBdaIMiXYdfx0Mh?usp=sharing

Data Source:
https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign/data