



Understanding and Predicting Customer Deposits: **A Data-Driven Analysis Proposal**



Capstone Experience 2025

Prepared By:
Rajashekar Reddy Muskula

Professor:
Kristin Marie Greenwalt

Project Overview



Objective

To understand which variables drive deposit growth by analyzing customer demographics, financial behaviors, interaction history, and external economic factors.



Technical Implementation

Leveraging Python, Excel, Tableau, and RapidMiner for robust data analysis and predictive modeling capabilities.

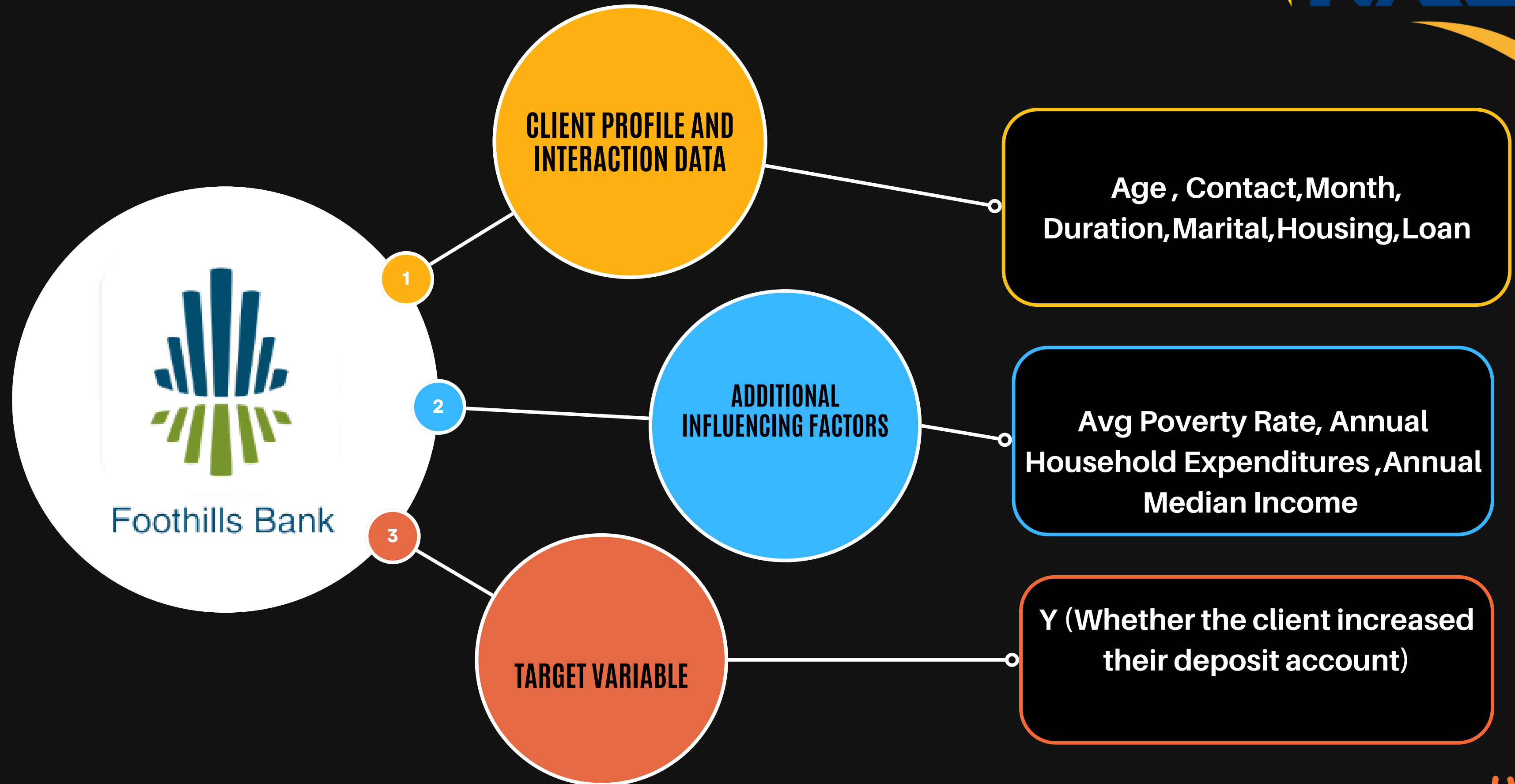


Outcome

The project will provide insights into the key factors driving deposit changes, enabling better-targeted strategies for financial planning, customer engagement, and service optimization.



About The Data



Key Variables



01

Age

Client age in years



02

Marital

What is the relationship status of the client



03

Housing

Does the client have a home loan?



04

Loan

Does the client have a personal loan?



05

Contact

Method of last communication with the Client



06

Month

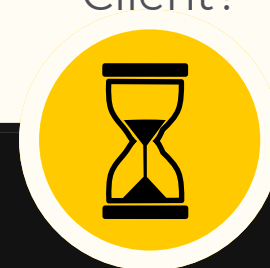
When was the known contact with the client?
Which month?



07

Duration

What was the duration (seconds) of the last known contact with the Client?



08

Avg Poverty Rate

represents the middle Poverty value of all clients



09

Annual Household expenses

represents the middle House Hold Expense value of all clients

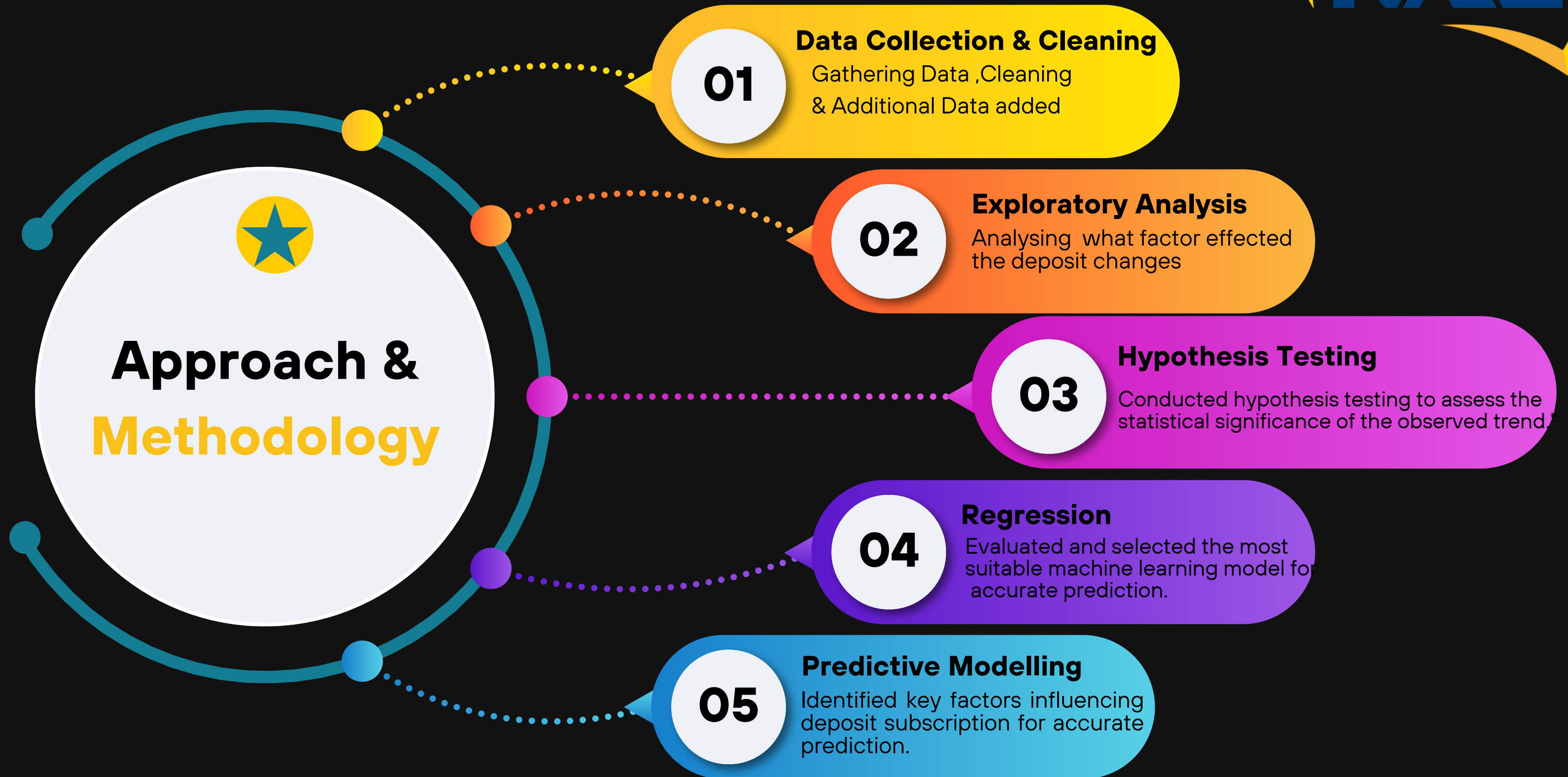


10

Annual Median Income

represents the middle income value of all clients





Exploratory Analysis



Descriptive Statistics

Descriptive Statistics for Avg Poverty Rate	
Mean	11.7
Median	10.6
Mode	10.6
Range	8.5
Minimum	5.1
Maximum	13.6
Sum	479364.5
Count	41017.0

Descriptive Statistics for Annual Household Expenditures	
Mean	61590.86
Median	66500.00
Mode	54000.00
Range	34500.00
Minimum	34500.00
Maximum	69000.00
Sum	2526272500.00
Count	41017.00

Descriptive Statistics for Annual Median Income	
Mean	79019.32
Median	85000.00
Mode	70000.00
Range	35070.00
Minimum	54930.00
Maximum	90000.00
Sum	3241135310.00
Count	41017.00

"Descriptive statistics of the supplementary data used for analysis."



Deposit Subscriptions by Month and Contact Method



Contact	may	jun	aug	dec	Yes Total	Grand Total
cellular	612	349	614	78	1653	1653
telepho ne	272	210	36	11	529	529
Grand Total	884	559	650	89	2182	2182



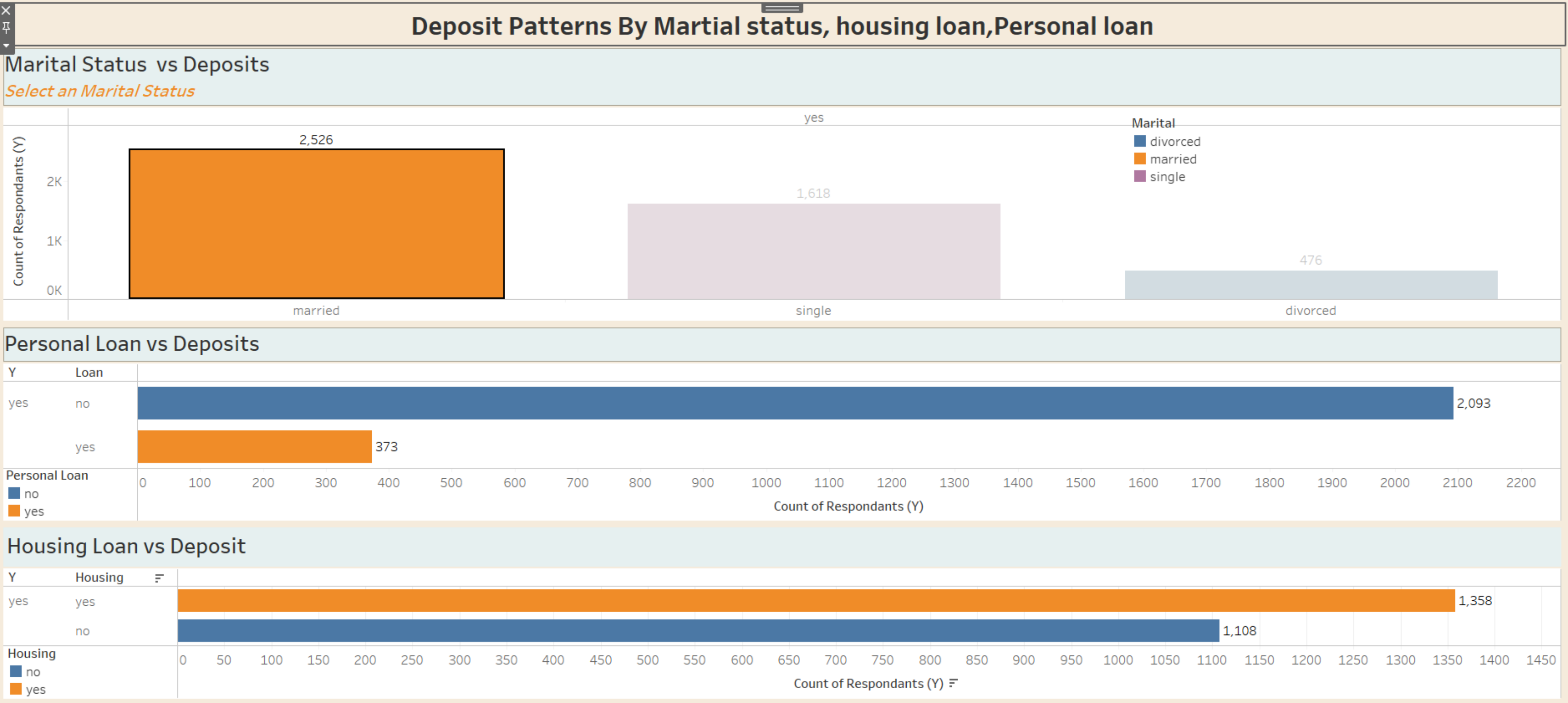
Summary

The table displays the distribution of deposit subscriptions (Y) by month and contact method, with the highest deposits in May (884) and the lowest in December (89) across both cellular and telephone contact methods.

Deposit Patterns Analysis by Marital ,Housing, Personal Loan Status



Yes {Y} (Since last contact, did the client increase their deposit account?)

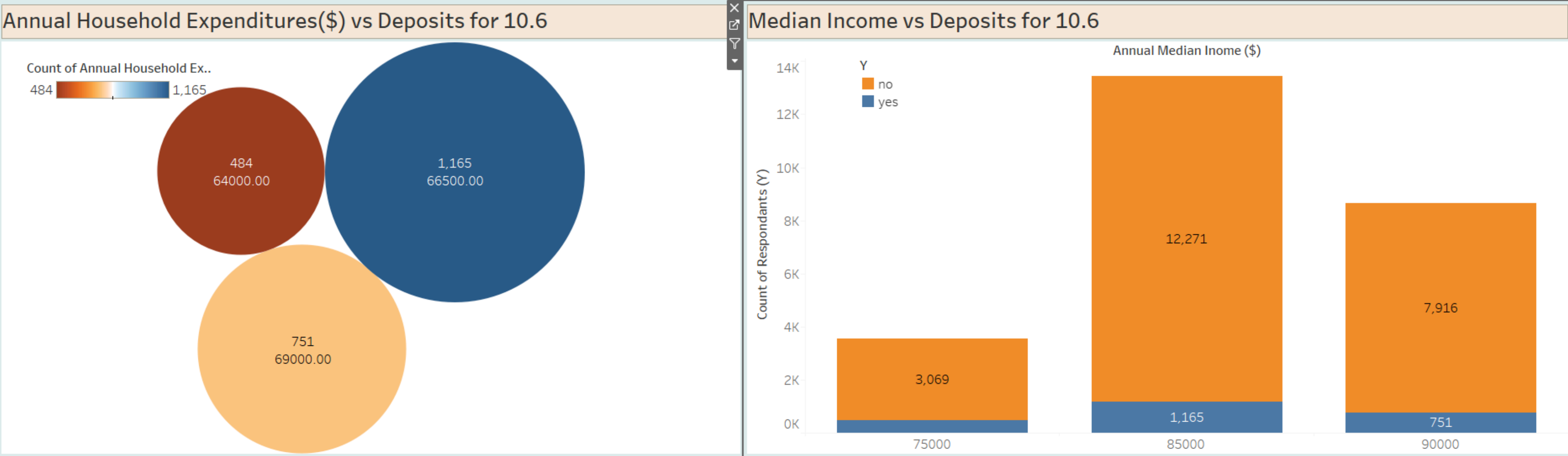
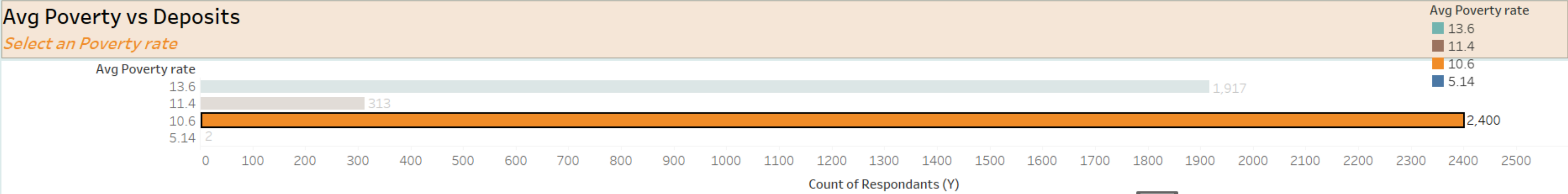


Deposit Patterns Analysis by Avg Poverty,Expenditures, Median Income Status

Yes {Y} (Since last contact, did the client increase their deposit account?)



Deposit Trends by Avg Poverty, Annual Household Expenditures and Median Income



Hypothesis



Variable's : Duration, Marital Status

	Divorced	married
Mean	254.31	257.835
Variance	62140.15	67064.19
Observations	4595	24813
P(T<=t) two-tail	0.38	

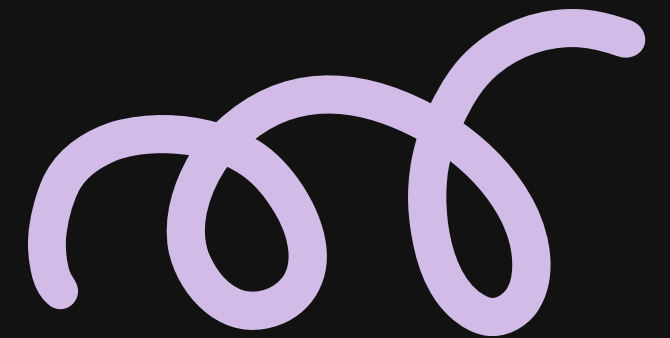
	Single	Divorced
Mean	261.92	254.31
Variance	69793.55	62140.15
Observations	11529	4595
P(T<=t) two-tail	0.085	

	married	Single
Mean	257.84	261.97
Variance	67064.19	69793.55
Observations	24813	11529
P(T<=t) two-tail	0.17	

The Hypothesis is done between Duration & Marital status

Summary:

All three scenarios show P-values greater than 0.05, meaning none of them provide statistically significant evidence to reject the null hypothesis. However, the first scenario (Single vs. Divorced) with a P-value of 0.085



Regression

Single Vs Divorced



Model	Accuracy
Linear Regression	0.88%
Logistic Regression	0.99%
Decision Tree	0.91%

Since the **Logistic Regression** model provided the highest accuracy, it will be utilized for prediction.

Prediction of New Customer

Deposit Increase



Single Vs Divorced

Duration(Seconds)	Duration(Min)	Single	Divorced	Deposit
824<=	17.33<=	Yes	Yes	No
824>=	17.33>=	Yes	Yes	Yes

Since the **Logistic Regression** model provided the highest accuracy, We Used for prediction.

Regression



Variables used in Regression are **Age ,Annual Household Expenditures ,Annual Median Income,Avg Poverty Rate,Contact,Month,Loan,Marital,Housing**

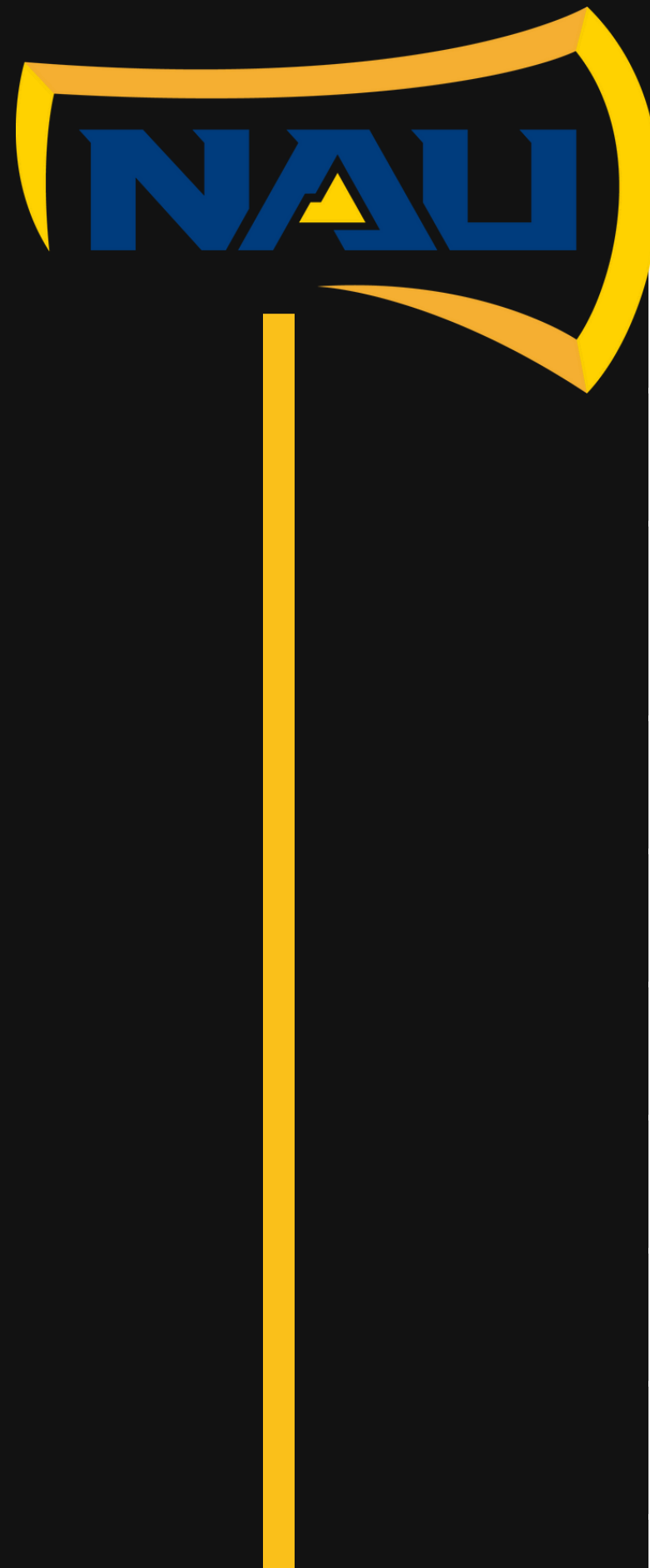
Model	Accuracy
Linear Regression	89.06%
Logistic Regression	89%
Decision Tree	88.05%
Random Forest	87.96%

Since the **Linear Regression** model provided the highest accuracy, it will be utilized for prediction.

Prediction of New Customer Deposit Increase

Prediction of New Customer : **Yes**

Metrics	Value
Age	>=36
Annual Household Expenditures	<= 15000\$
Annual Median Income	>90000\$
Avg Poverty Rate	<=5.14
Contact	Cellular
Month	May
Loan	No
Marital	Yes
Housing	No

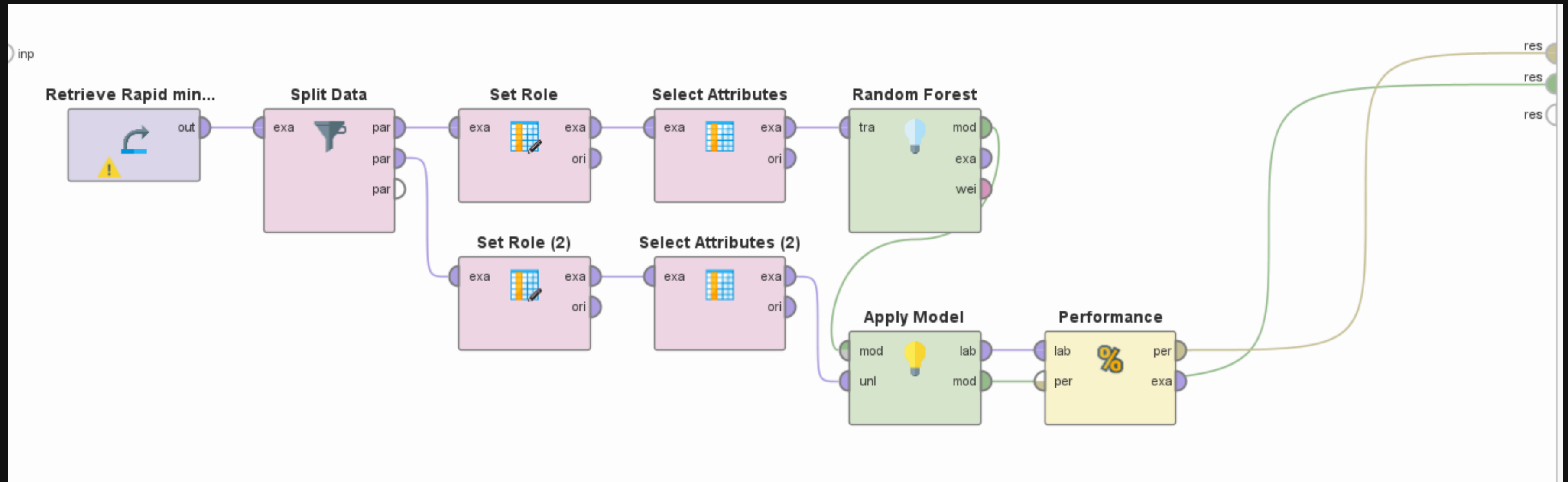


Prediction of New Customer : **No**

Metrics	Value
Age	<=35
Annual Household Expenditures	> 15000\$
Annual Median Income	<90000\$
Avg Poverty Rate	>5.14
Contact	Telephone
Month	May
Loan	No
Marital	Yes
Housing	Yes

Rapid Minner

Random Forest



Set Role



Selected Attributes

Selected Attributes

+

×

Annual Household Expenditures(\$)

Annual Median Income (\$)

Avg Poverty rate

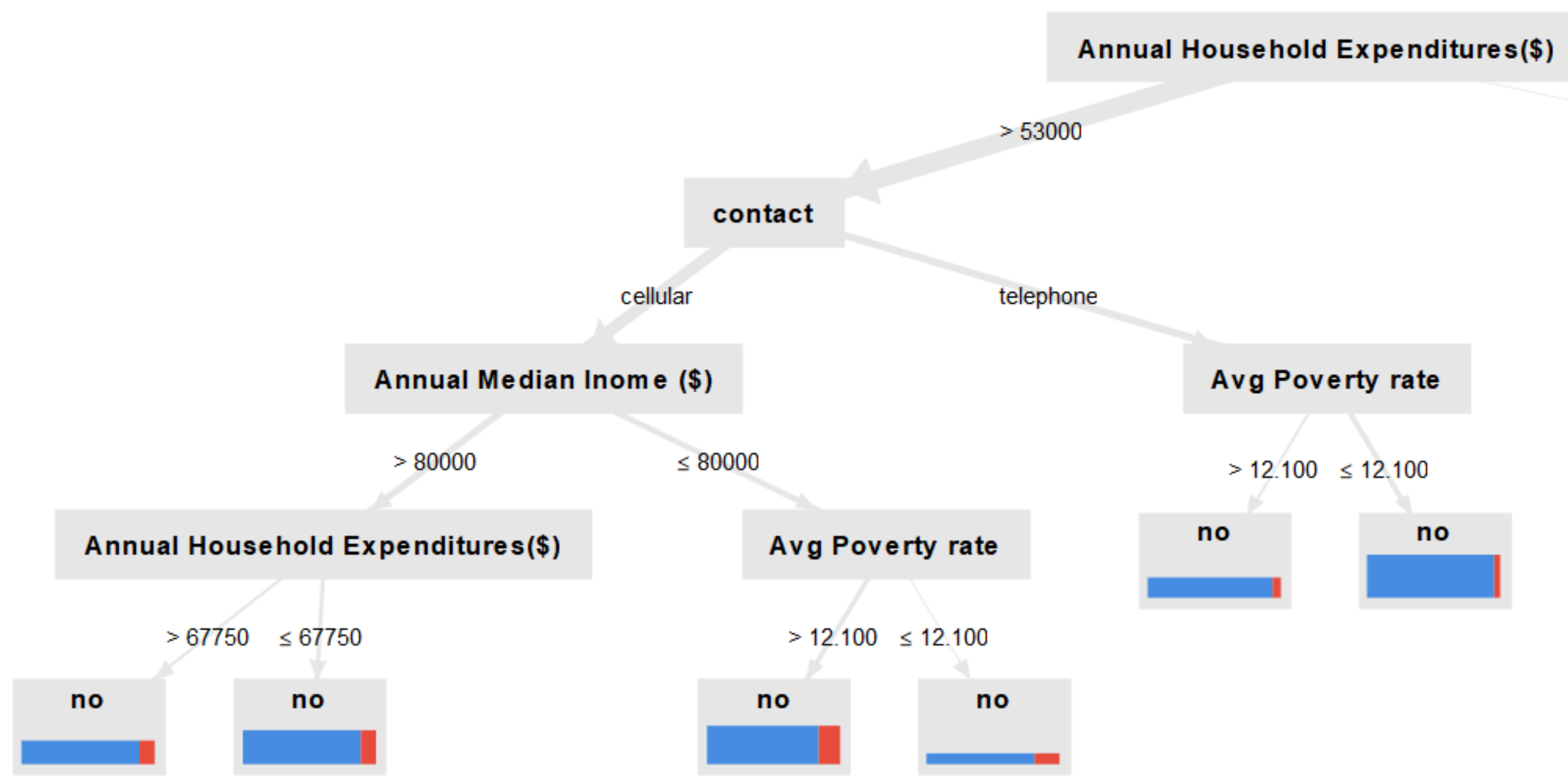
contact

y

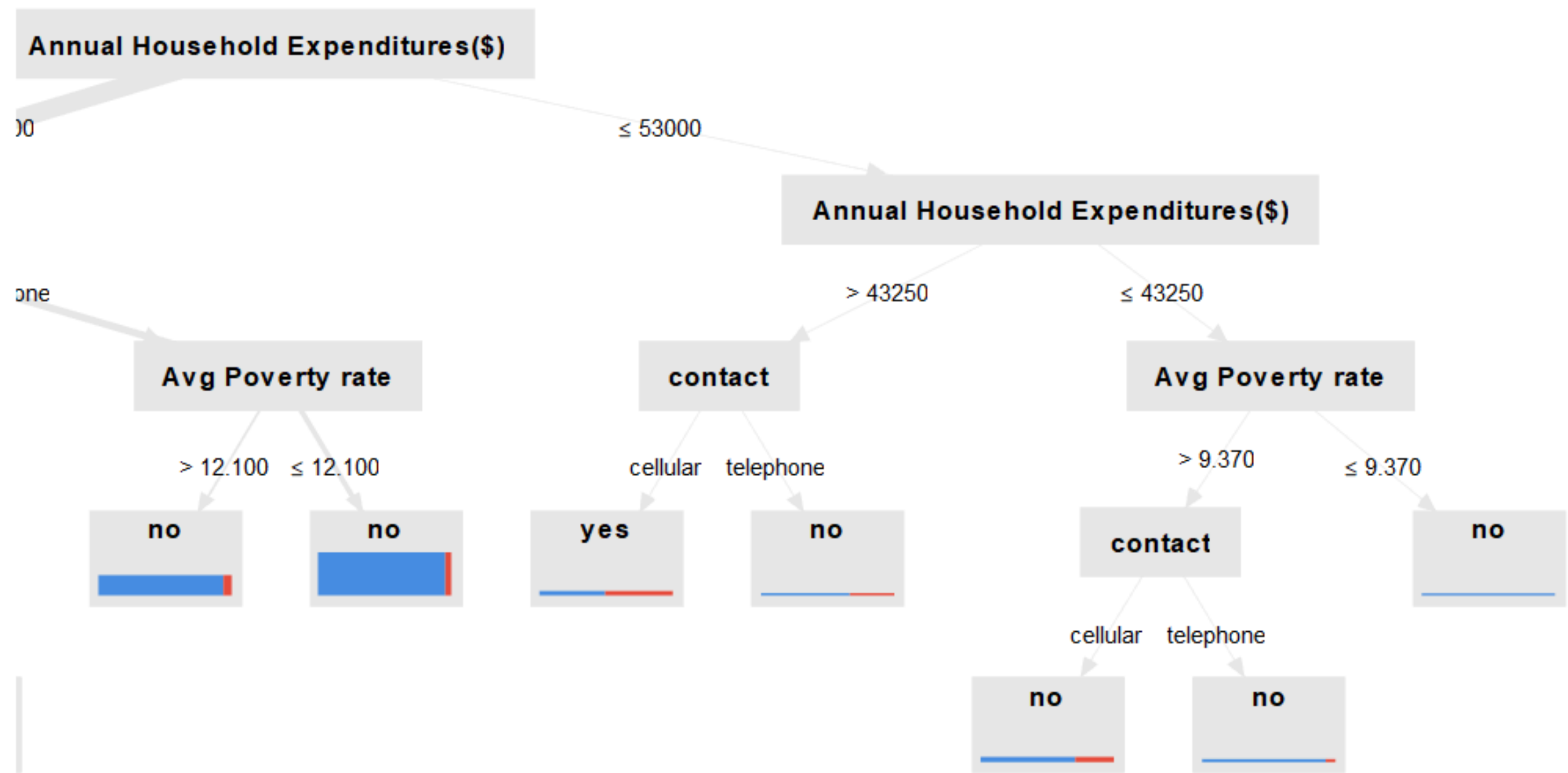
Target Attribute

attribute name	target role
y	label

Result - 1



Result - 2



The analysis reveals that factors such as poverty rate, household expenditures, and contact method influence deposit outcomes, with the highest "No" distribution (24.85%) occurring through telephone contact. The data was split into 70% for training and 30% for testing.

6 Recommendations to Increase Deposit Subscriptions :



- 1.Target High-Income Customers: Focus on customers with higher annual median incomes (above \$90,000) as they are more likely to subscribe to deposits.**
- 2.Leverage Mobile Communication: Use cellular contact to reach a broader audience, as it proves more effective in driving deposit subscriptions compared to telephone outreach.**
- 3.Focus on Peak Months: Capitalize on periods like May, which showed the highest deposit activity, by launching targeted campaigns during this month.**
- 4.Target Married Customers: Tailor campaigns towards married customers who are more likely to have stable financial conditions and goals, making them more inclined to increase deposits.**
- 5.Offer Incentives for Low Expenditures & High Poverty: Provide rewards or incentives for customers with lower annual household expenditures & High Poverty, as they may be more receptive to increasing their deposits & Also provide financial education on the importance of saving and depositing**
- 6.Incentivize Loan-Free Customers: Focus on customers who do not have any loans (as the data suggests they might have better financial stability), and provide them with deposit plans that offer better benefits or interest rates.**

Conclusion:



- 1.**The project identified key factors, such as income, Poverty rate, Expenditure, month of subscription, contact methods, and marital status, that significantly impact deposit growth.
- 2.** Predictive models and targeted strategies were developed using Python, Excel, and Tableau to optimize customer engagement and increase deposit subscriptions.
- 3.**Finally, recommendations were provided to enhance customer deposit subscriptions and attract new customers for deposit accounts.



Presentation By Rajashekar Reddy Muskula



Thank You So Much

