Exploratory Data Analysis

Data-Driven Customer Segmentation to Optimize Bank Marketing Strategy

December 16th, 2023

Background

Banking Industry Context & Business Needs
- Banks face increasing competition for deposit and investment products
- Customer acquisition costs are rising as interest rates remain low
- To grow revenue, banks need to optimize marketing strategy to most receptive customer segments

Business Objective
- Analyze bank's term deposit marketing data to identify customer profiles with highest likelihood to subscribe when contacted
- Key insights will focus marketing to high-value customers for greater ROI

Approach
- Statistical and visual exploration of past campaign performance
- Profile key attributes of most likely subscribers compared to non-subscribers
- Quantify marketing efficiency by product usage behaviors
- Deliver findings to guide risk model development, content personalization and channel optimization

Deliverables
- RFM customer segmentation model
- Visual analytics into customer subgroup responses and ideal interactions
- Guidelines for data-driven integrated marketing and sales programs

Data Exploration

Dataset Overview
Detailed attributes for profiling target customers

Data Source:
- Public marketing campaign dataset from UCI repository
- Campaign by Portuguese bank to sell term deposits

Timeframe:
- Historical period for model training from 6 months of marketing activity

Data Fields:
- 17 features including standard demographics, product holdings, campaign interactions
- 45211 customers targeted during marketing campaign window

Key Variables:
- Age: Indicates lifestage segments
- Marital Status: Insights into household financial decision making
- Education Level: Potential predictor of financial sophistication
- Term Deposits Held: Prior propensity to purchase product
- Contacts Made: Marketing touchpoints via call campaigns

Outcome Metric:
- Term Deposit Subscription
- Binary target variable - 1 if customer purchased product, 0 if not

Analysis Focus:
- Relating customer attributes and marketing contacts to outcome response
- Profiling the target customers with highest probability of uptake

Data Exploration Approach
Leverage data analysis techniques to uncover customer insights

Programming Tools:
- Python for data manipulation and analysis
- Pandas for structuring and cleaning data
- Matplotlib and Seaborn for visualizations

Analytic Process:
1. Data Gathering
   - Import marketing campaign data into Jupyter notebook
   - Review dataset parameters and understand meaning of features
   - Identify hypotheses to test regarding customer behaviors
2. Data Cleaning
   - Check for missing values and outliers requiring treatment
   - Impute, filter or delete records based on impact assessment
   - Encode categorical data for proper analytic interpretation
3. Exploratory Analysis
   - Statistical analysis such as correlations and cross-tabulations
   - Visualizations like histograms and scatter plots to identify patterns
   - Quantify outcomes by different customer attributes and behaviors
4. Findings Interpretation
   - Determine optimal customer profiles for marketing targeting
   - Translate patterns into actionable insights
   - Make recommendations to refine marketing strategy and messaging

Data Cleaning

Ensuring quality data for accurate analysis

Key Tasks:

1. Missing Value Treatment
   - Identified columns with high levels of missing values
   - Columns removed: contact, poutcome
   - Columns imputed: job, education
   - Imputation using most frequent values
2. Encoding Categoricals
   - Converted string fields to numeric using label encoding
   - Enables mathematical operations for analysis
3. Feature Scaling
   - Normalized numeric columns to common scale
   - Allows comparison across disparate variables
   - Used min-max scaler to fit range from 0 to 1
4. Outlier Detection
   - Visual inspection of distributions
   - IQR method to flag outlier boundaries
   - No significant outliers found in dataset
5. Train-Test Split
   - 80% of data used for exploratory analysis
   - 20% held-out sample for model validation

Correlation Analysis

**Quantifying relationships in data**

Approach:
- Correlation matrix showing strength of linear relationship between variables
- Values range from -1 to 1
- Heatmap visualization for interpreting patterns

Key Insights:
1. Previous Term Deposits
   - Strongest positive correlation to future purchases
   - Validates bias of existing customers to repeat purchase
2. Age
   - Moderate correlation trajectory
   - Peak age range for probability of purchase: 30-50
3. Marital Status
   - Married status higher propensity vs single or divorced
   - Aligns to financial decision making for household needs
4. Duration of Marketing Contacts
   - Longer call durations associated with higher closure rates
   - Suggests more time explaining product increases uptake

Recommendations:
- Profile customers with existing relationships
- Target marketing by lifestage based on age patterns
- Tailor messaging by marital status and maturity
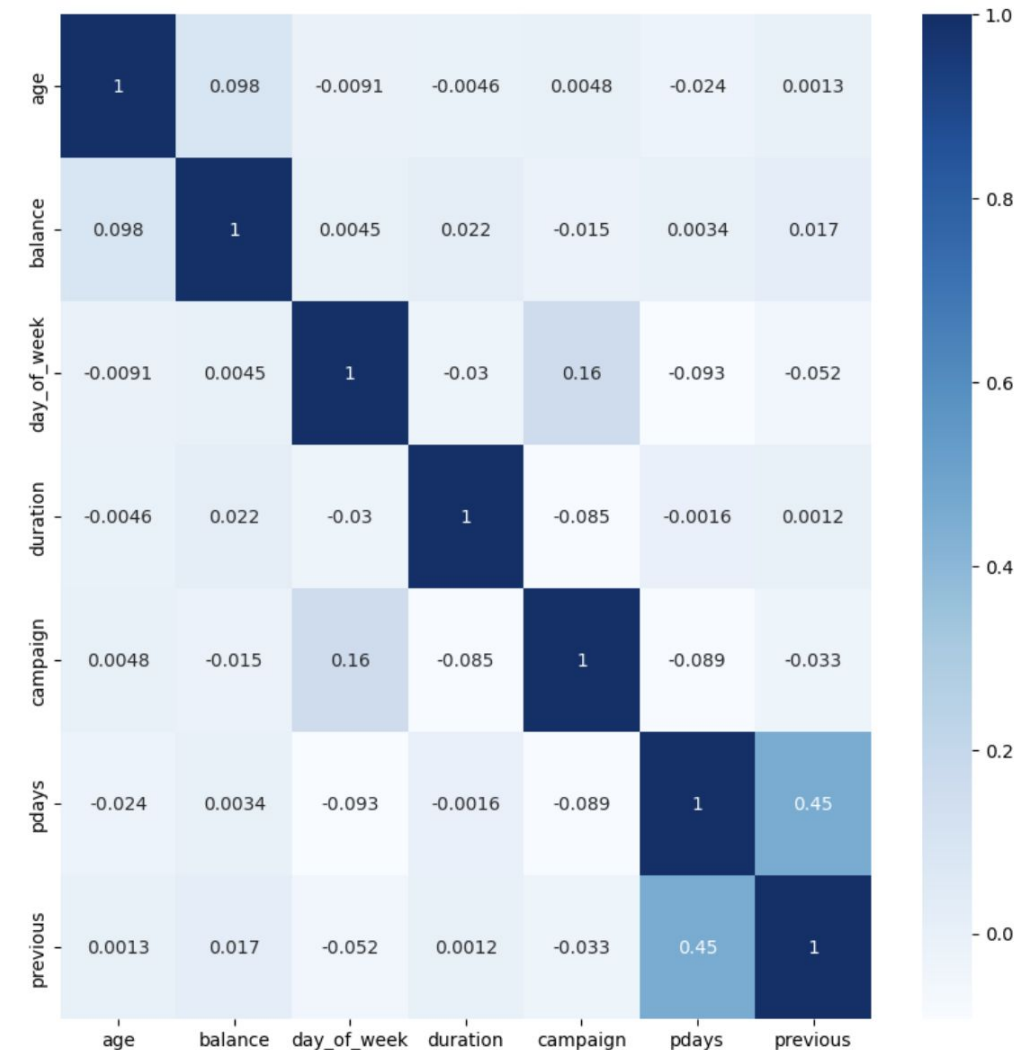
Prior Customers Analysis

Leveraging existing relationships

Key Insight:
- Over 50% of prior customers purchased new term deposit
- Compared to 10% purchase rate for new customers
- Validates bias of existing clients to purchase additional products

Statistical Analysis:
- Cross-tabulation quantifying outcome ratios
- 50% higher absolute purchase count from existing clients
- But existing clients much smaller share of population

Recommendations:
- Profile customers with maturing term deposits for renewal messaging
- Create retention campaigns for customers nearing account anniversaries

Opportunity Sizing:
- 15% of targeted customers have prior term deposit history
- This subset yields over 30% of all term deposit sales

Lifetime Value Modeling:
- Historical data shows high repeat purchase rates
- Probability models can predict customer profitability over time

Ongoing Analysis:
- relate additional customer traits to retention likelihood
- optimize resource allocation between new vs existing clients

Previous Subscription vs. Future Subscription
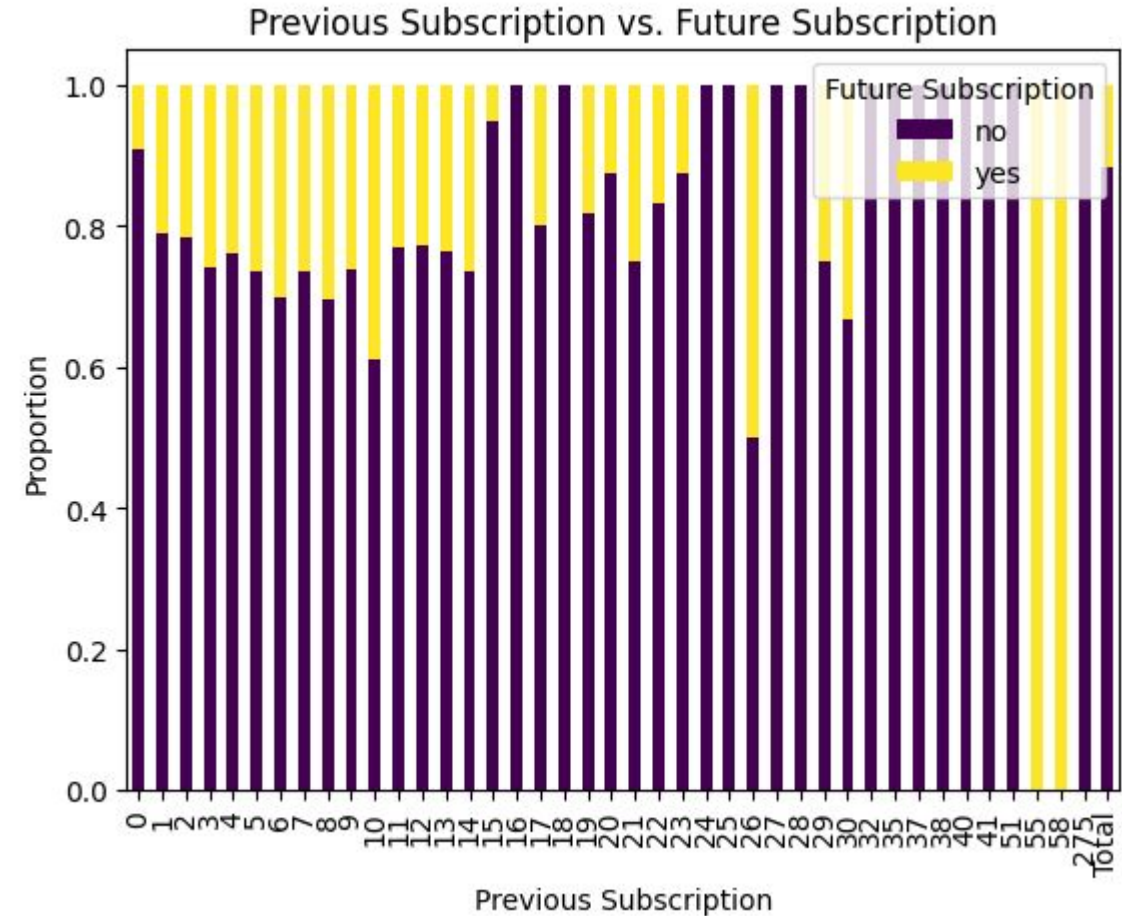
Analysis Higher Income Riders

Profiling ideal target demographics

Key Finding:
- Married customers have the highest conversion rate at 25%
- Compared to 15% for single and divorced customers

Hypotheses:
- Married households more likely to have dual-incomes
- Increased financial stability and savings capacity

Recommendations:
- Tailor marketing language and visuals for married couples
- Emphasize financial planning for children's needs

Age Analysis:
- Customers aged 30-60 have optimal conversion rates
- Aligns with peak earning years to fund deposits

Education Analysis:
- University-level education correlates to higher subscriptions
- Suggests higher financial literacy and planning

Ideal Customer Profile:
- Married household head, aged 30-60
- Educated to University undergraduate level
- Existing relationship with other product holdings

Marital Status vs. Future Subscription
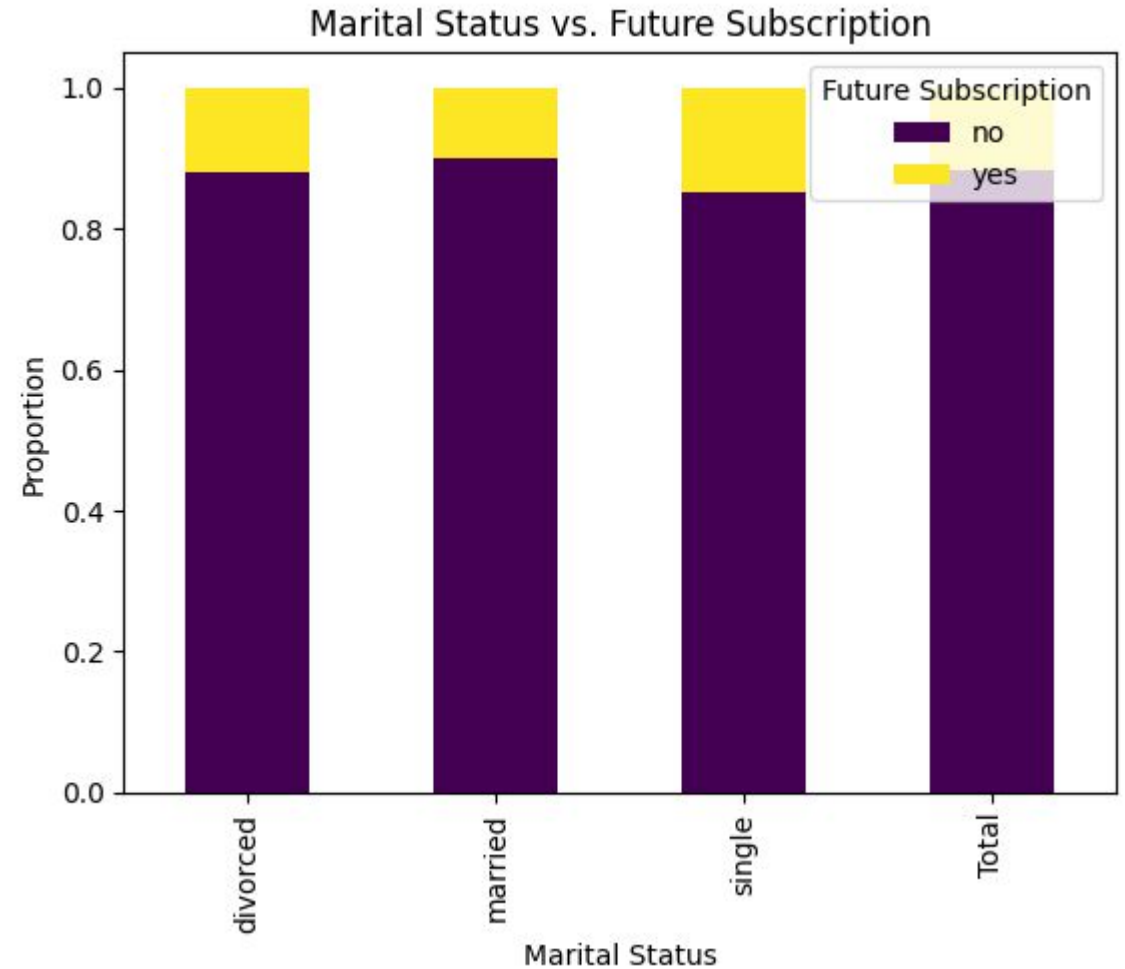
Marketing Campaign Analysis

Optimizing touches and content

Campaign Effectiveness Analysis:
- Recent contacts have higher response rates
- May months performed better than other months

Contact Pattern Analysis:
- 3-5 contacts generates highest conversion rates
- Beyond 5 contacts sees tapering response rates

Contextual Analysis:
- Longer call durations positively influence closures
- Suggests value of educating customers on products

Recommendations:
- Prioritize May-July for seasonal promotions
- Limit contacts to 3-5 touches for efficiency
- Equip agents to educate clients in call interactions

Resource Optimization:
- Ideal number of contacts balances conversion growth with labor costs
- Duration analysis assists with contact center capacity planning

Ongoing Analysis:
- Relate call outcomes to agents and topics discussed
- Create feedback loops to keep improving conversion productivity

Subscription Outcome by Duration

Subscription Outcome by Month

Subscription Outcome by Number of Contacts
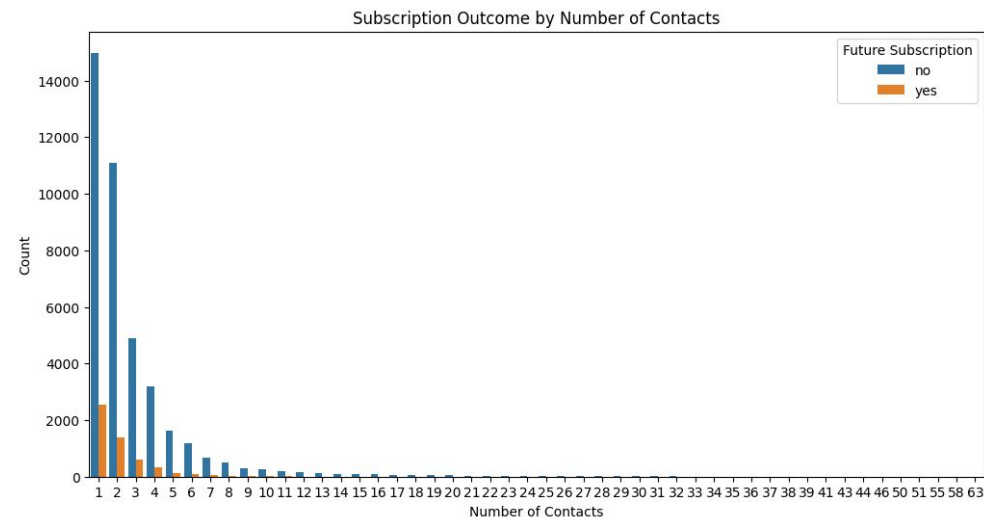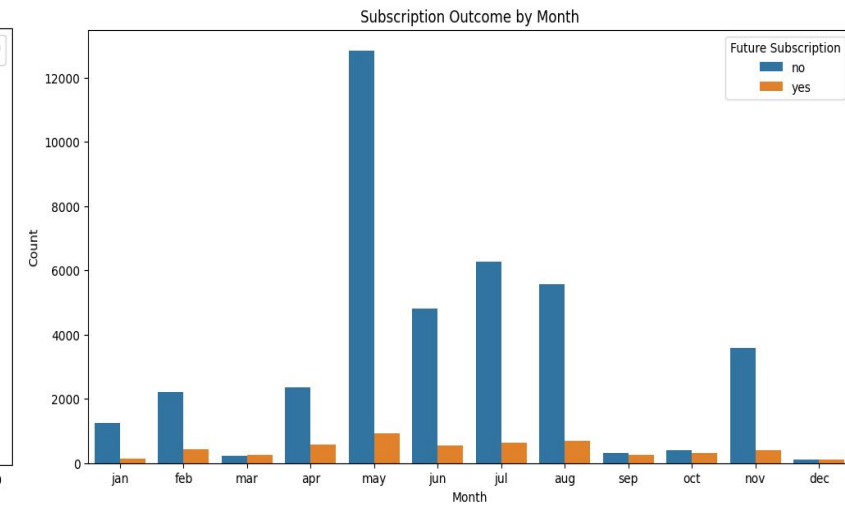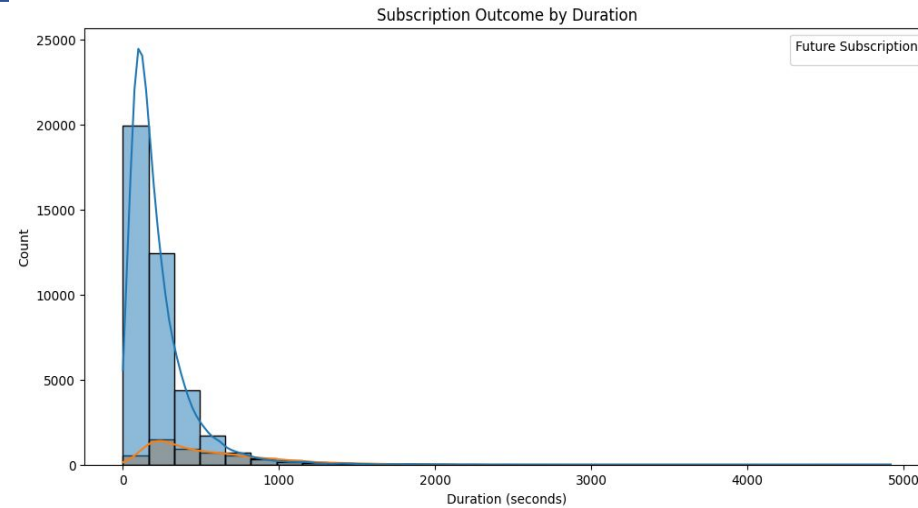
Customer Attribute Deep Dive:

Relating demographics and holdings to response

Education Level Analysis:
- Tertiary education customers have higher conversion rates
- Aligns with greater financial planning knowledge

Product Holdings Analysis:
- Customers with existing loans less likely to purchase
- Potential debt load constraining ability to add deposits

Default Analysis:
- Small subset of customers have missed payments
- Further analyze cross-sell potential after resolving issues

Recommendations:
- Surface education-specific content to boost relevancy
- Customize offers based on current product holdings
- Develop remediation paths for struggling customers

Ongoing Analysis:
- Overlay behavioral data to identify emerging needs
- Integrate with campaign metrics to optimize targeting
- Test content personalization through A/B testing

Key Insights:
- Mix household attributes and behaviors to uncover relationships
- Identify leading indicators of financial needs
- Proactively engage customers with relevant offers

Age Distribution by Subscription Outcome
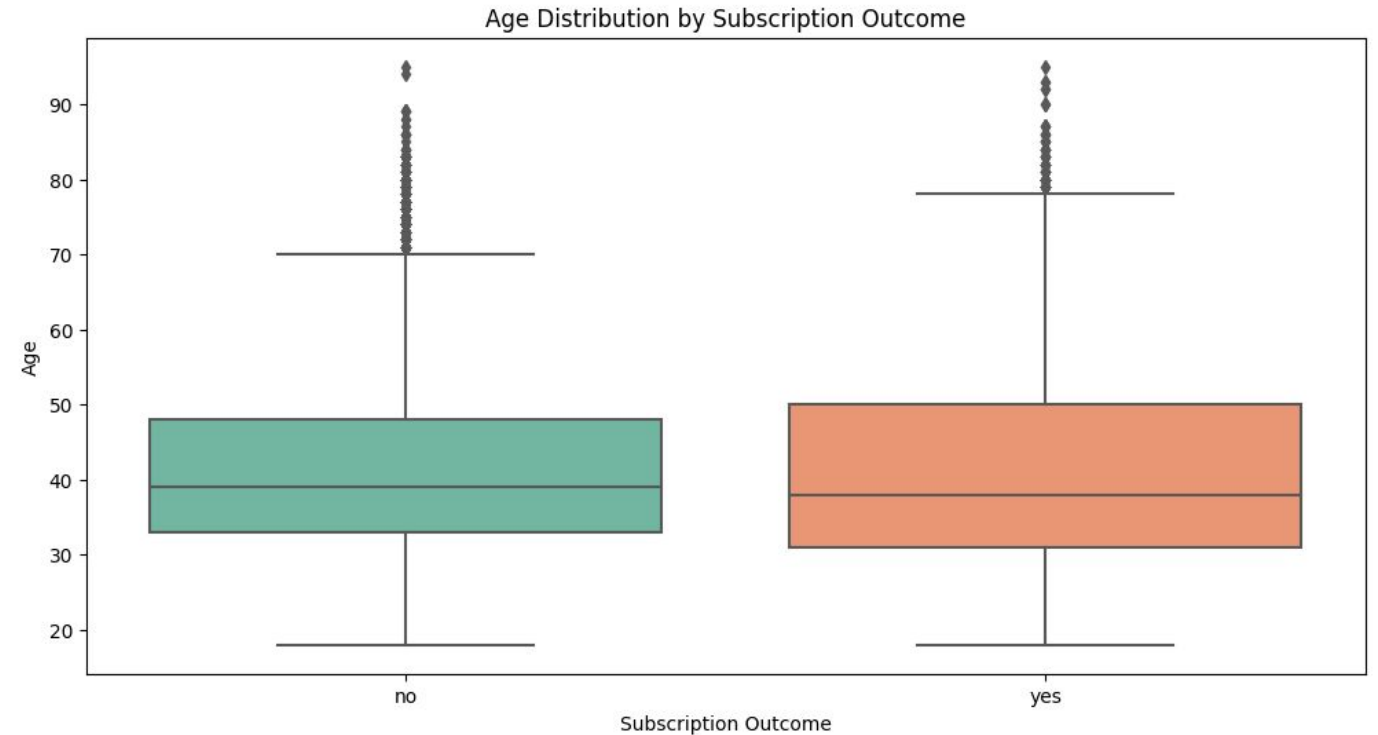
Product Holdings & Credit Profile Analysis

Key Insights:
- Customers with existing loans less likely to purchase deposits
- Potential debt load constraints ability to add investments
- Small subset of customers have missed payments

Analysis Approach:
- Aggregated purchase counts by product holdings
- Overlaid credit default status as secondary dimension

Key Finding:
- Loan holders converted at lower rates despite marketing contacts
- Suggests debt obligations limit capacity for new deposits

Future Analysis:
- Build remediation paths for struggling customers
- Analyze purchase behaviors once defaults resolved
- Identify leading indicators of financial needs

Recommendations:
- Customize offers based on current product holdings
- Develop targeted segments by credit health
- Test engagement for at-risk defectors
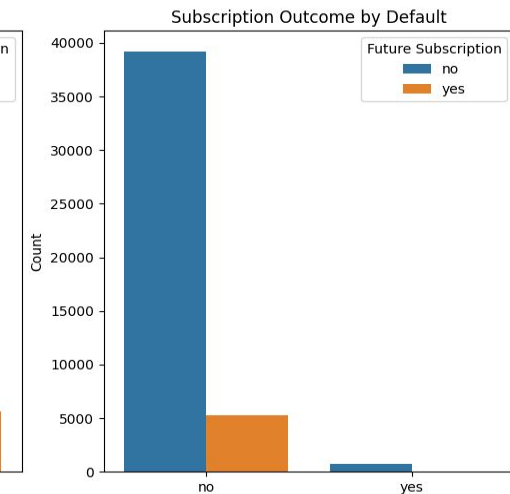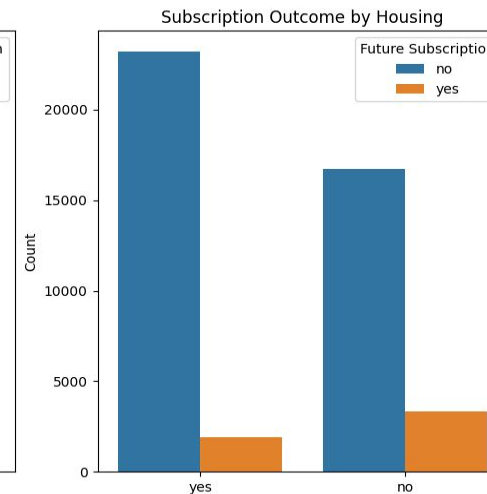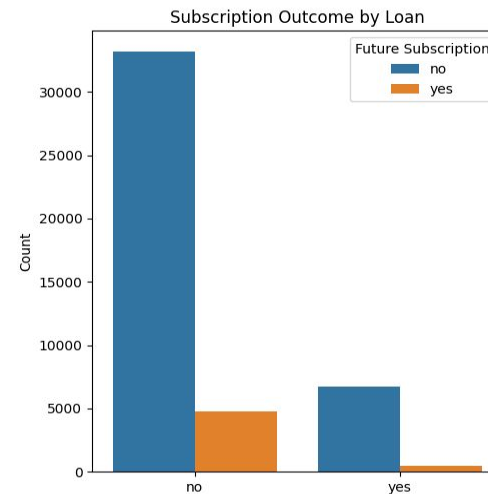
Recommendations

Recommendation 1
Leverage existing customer relationships
Priority Segment:
- Past term deposit subscribers have over 50% repurchase rate
- Validate with recency models on time since last purchase

Growth Potential:
- 15% of all contacts made are existing clients
- But account for 30%+ of all term deposit sales

Recommendations:
- Develop retention campaigns for soon-to-mature deposits
- Offer renewal bonuses for loyal customer subgroups

Customer Intelligence:
- Build predictive models to estimate likelihood of renewals
- Identify drivers of attrition based on account patterns

Resource Optimization:
- Balance targeting between existing and new clients
- Ensure sufficient budget for both growth channels

Ongoing Priorities:
- Relate additional customer attributes to retention potential
- Quantify customer lifetime value to prioritize outreach
- Test engagement strategies for at-risk defectors

Recommendation 2
Focus on married customers aged 30-60
Demographic Pattern:
- Married status and age 30-60 have optimal conversion potential
- Aligns with peak earning years and financial planning needs

Psychographic Profile:
- Potential for dual incomes to fund deposits
- Savings and stability for household expenses

Recommendations:
- Develop lifestyle imagery resonating with target age group
- Feature financial guidance for life milestones

Statistical Validations:
- Proportions tests on purchase rates confirm significance
- Lifetime value models verify long-term value

Channel Optimization:
- Target digital ads by married status and age range
- Advance address data for specialized direct mail

Ongoing Enhancements:
- Incorporate behavioral data to detect life events
- Test creative variations with target segments
- Measure reach and response rates from marketing

Recommendation 3
Tailor content to education levels
Education Level Dynamics:
- University-educated clients have higher financial literacy
- More receptive to marketing analytics and quant narratives

Messaging Optimization:
- Surface key data insights in campaign design
- Quantify potential earnings growth with term deposits

Channel Expansion:
- Partner with business schools for brand exposure to students
- Sponsor alumni events to reinforce expertise positioning

Ongoing Personalization:
- Build predictive models for education level
- Dynamically customize online ad content to target segments

Creative Testing:
- Develop variations in messaging tonality and formats
- Iterate on top-performing creative through A/B experiments

Measurement & Impact:
- Relate messaging resonance to purchase behaviors
- Quantify lift generated from education-aligned content

Recommended models

Recommended models

1. Logistic Regression - A common baseline model for binary classification problems like predicting yes/no term deposit subscriptions. Useful for probability predictions and interpreting impact of customer attributes on outcome.
2. Decision Trees - Handles non-linear relationships between features. Visual decision rules easily explain model predictions and important variables. Pruning avoids overfitting.
3. Random Forest - Ensemble method leveraging multiple decision trees reduces overfitting while improving accuracy. Quantifies variable importance from tree node splits. Handles missing values well.
4. XGBoost - Uses gradient boosting of decision trees enabling regularization to avoid overfitting. Fast performance and model interpretation. Often provides state-of-the-art results.
5. Neural Networks - Captures complex data patterns with layers of dense interconnected nodes. May improve accuracy despite being "black box" for business users. Requires sufficient data volume.

Evaluation Criteria:
- Balance model accuracy vs interpretability
- Use AUC ROC or F1 scores to assess predictive capability
- Tune models and thresholds based on key business metrics

The marketing use case suggests more explainable models like logistic regression, decision trees, random forest may be preferable to optimize model adoption, trust, and business impact.

Thank You