PHASE-I: EDA Presentation and proposed modeling technique

Team Member's Details:

• Group Name: "I am the TEAM"

• Name: Krishnakant Sonji

• Email: cronjobs alo.edu, krishnasonji@gmail.com

• Country: United States

• College/Company: University of New York at Buffalo

• Specialization: Data Science

Problem Description: XYZ Bank wants to send personalized Christmas offers to different customer groups instead of sending the same o er to all. The goal is to create no more than 5 customer segments using machine learning to improve the efficiency of the campaign.

GitHub Repo link: https://github.com/Thir13een/DG-week-11.git

EDA presentation for business users:

Slide 1:
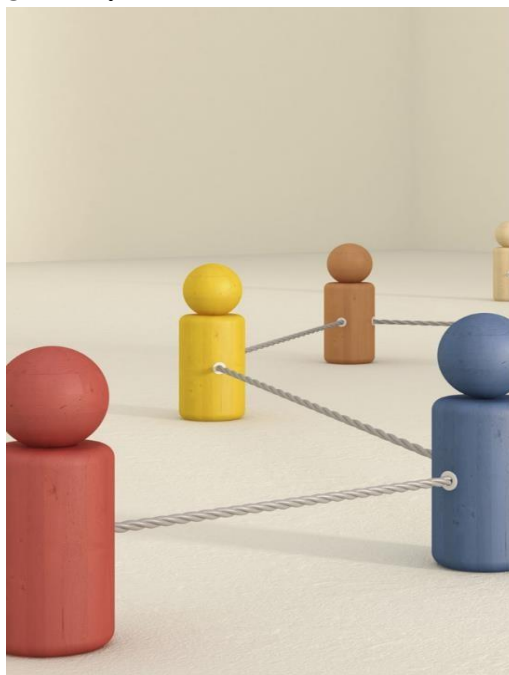
Customer Segmentation

BY KRISHNAKANT SONJI

Slide 2:

Understanding Our Customer Base

|	
🖥	This presentation aims to:
🔗	uncover patterns within our customer base.
💡	providing insights to enhance marketing strategies.
▦	product development.
⚙	customer experience.
🔄	The analysis covers customer demographics, product usage, and segmentation to uncover hidden opportunities.

Slide 3:

	data_partition_date	customer_id	employee_status	residence_country	gender	customer_age	contract_start_date	new_customer_index	customer_s
0	2015-01-28 00:00:00	1375586	N	ES	H	35	2015-01-12	0.0	
1	2015-01-28 00:00:00	1050611	N	ES	V	23	2012-08-10	0.0	
2	2015-01-28 00:00:00	1050612	N	ES	V	23	2012-08-10	0.0	
3	2015-01-28 00:00:00	1050613	N	ES	H	22	2012-08-10	0.0	
4	2015-01-28 00:00:00	1050614	N	ES	V	23	2012-08-10	0.0	
...	...	...	...	...	...	...	...	...	
976315	2015-02-28 00:00:00	1183296	N	ES	H	27	2013-09-25	0.0	
976316	2015-02-28 00:00:00	1183295	N	ES	H	56	2013-09-25	0.0	
976317	2015-02-28 00:00:00	1183294	N	ES	V	39	2013-09-25	0.0	
976318	2015-02-28 00:00:00	1183293	N	ES	V	36	2013-09-25	0.0	
976319	2015-02-28 00:00:00	1183289	N	ES	H	38	2013-09-25	0.0	

976320 rows × 44 columns

Data Overview

1. The dataset contains 976,320 customer records with 44 columns

2. Features: age, income, financial product usage, and customer status.

Slide 4:

Distribution of Customer Age
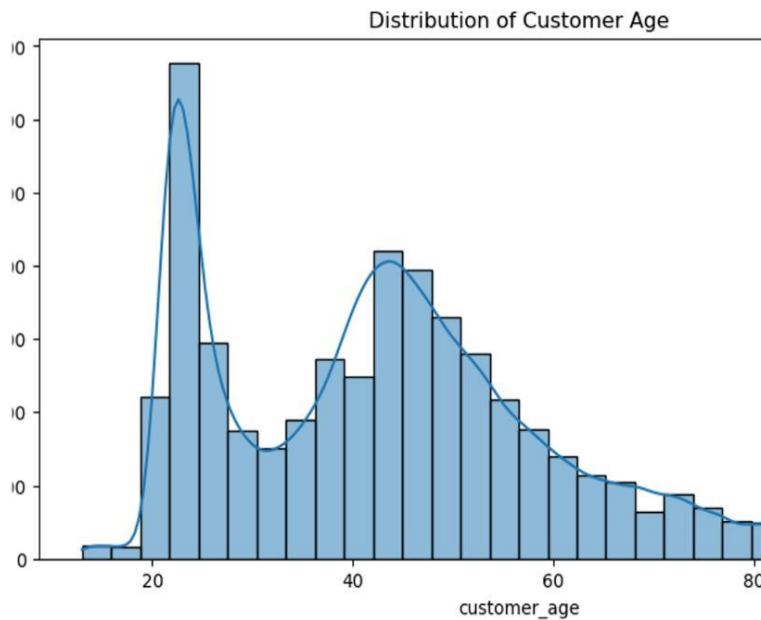
Age Demographics of Customers

1. The customer age distribution is bimodal, with prominent age groups around 20 and 40-50 years.

2. The first peak indicates a large presence of younger customers, while the second peak points to middle-aged customers.

3. Opportunity: Develop tailored offers and experiences to cater to the distinct age demographics.

Slide 5:

New Customers Over Time
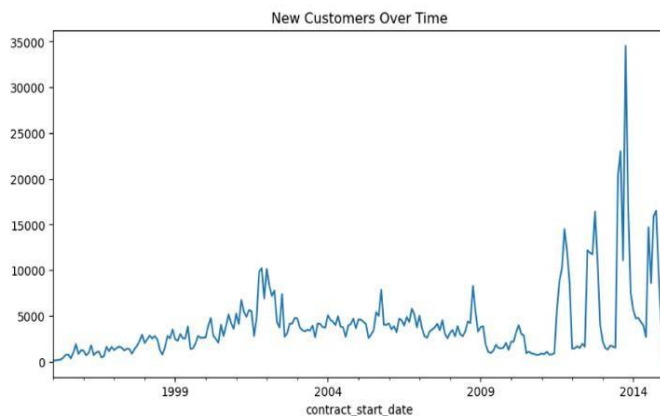
Customer Trends Over Time

1. New Customer Growth: Analyze how the number of new customers has evolved monthly or annually.

2. Age Trends: Explore changes in the age distribution of new vs. existing customers over time.

3. Income Changes: Monitor any shifts in the average household income of customers over various time periods.

4. Opportunity: Identifying growth periods or shifts in customer demographics can inform when to launch new products, campaigns, or retention efforts.

Slide 6:

Income Profile of Customers

$ The household income distribution is highly right-skewed, with the majority of customers in the low to moderate income range.
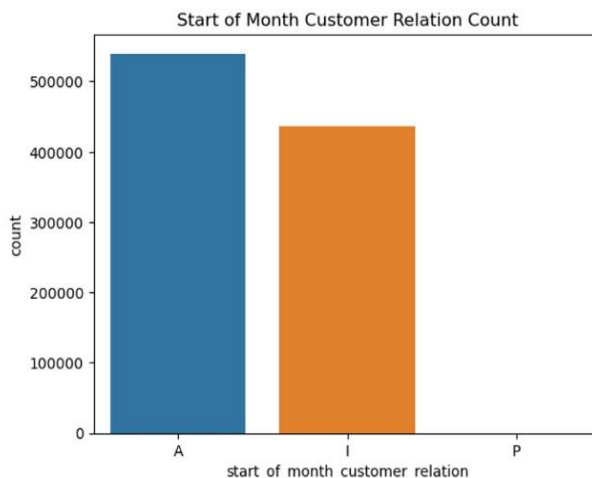
A small number of outliers represent very high-income households.

Opportunity: Segment customers by income to design differentiated product offerings.

Slide 7:

Start of Month Customer Relation Count

Length of Customer Relationship

1. Most customers have low seniority (<50 months), indicating a high proportion of recently acquired clients.

2. Seniority gradually decreases as tenure increases, with a smaller proportion of long-term customers.

3. Opportunity: Implement retention strategies to nurture these newer customers into long-term loyal clients.
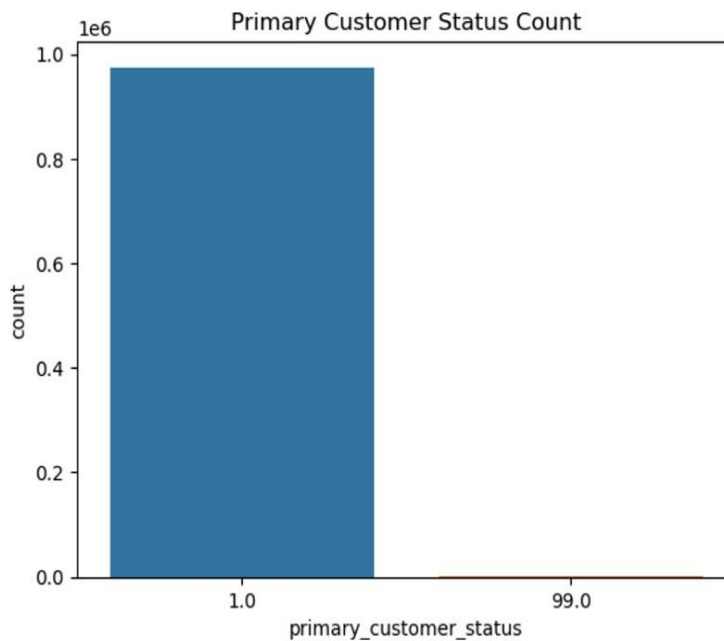
Slide 8:

Gender Breakdown

1. The gender distribution is balanced between male and female customers.

2. This balance enables gender-neutral product development and marketing strategies, ensuring inclusive targeting across all genders.

Slide 9:

Primary Customer Status Overview

1. Nearly all customers are classified as 1.0, suggesting a uniform classification of active customers.

2. The minor category (99.0) is insignificant, which indicates either a special condition or potential data anomaly..

Slide 10:

Current vs. Savings and Credit Accounts

1. Current account has the highest usage among customers, while saving account and credit cards are significantly underutilized.

2. Opportunity: Promote savings and credit accounts to diversify financial product engagement and increase revenue streams.
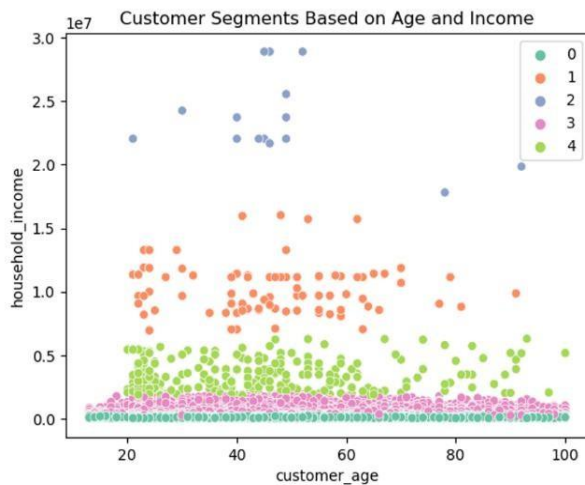
Slide 11:

Segmenting Customers Based on Age, Income, and Seniority

1. Clustering analysis groups customers into distinct segments based on their age, income, and seniority.

2. This segmentation helps identify specific customer needs and preferences, enabling more focused marketing strategies.

Slide 12:

Visual Clusters of Customer Segments

Customer Segments Based on Age and Income

1. Five customer segments are identified with unique combinations of age, income, and seniority.

2. The scatter plot reveals the visual distinction between these groups, indicating potential market segments to target.

Slide 13:

Recap of EDA Insights

1. The customer base skews younger with mostly low to moderate income.

2. Heavy usage of current accounts; savings and credit products are underutilized.

3. Clear customer segments offer opportunities for targeted product offerings and marketing.

4. Takeaway: Strategies should focus on product diversification, tailored marketing for segments and retaining newer customers.

Slide 14:

Actionable Next Steps

Slide 15:

For Business Users

Boost Product Adoption: Create campaigns to promote savings and credit products.

Enhance Customer Retention: Develop loyalty programs to retain newly acquired customers.

Personalize Offers: Leverage customer segments to design personalized products and communication.

Slide 16:

For Technical Users

Modeling Suggestions: Use KMeans for refined clustering; apply Logistic Regression or Decision Trees for product adoption prediction; use Random Forest/XGBoost for complex classification.

Handle Data Skewness: Transform skewed variables (e.g., income) for improved model performance.

Next Steps: Explore dimensionality reduction (PCA) and model tuning to enhance accuracy.

Slide 17:

Thank You!