

Business Understanding

Cross-Selling Marketing Strategy:

Cross-selling refers to the strategy of encouraging existing customers to purchase additional products or services that complement what they have already bought. Essentially, it is a way for businesses to increase sales by offering related or compatible items to enhance the customer's experience or meet more of their needs. The specific approach to cross-selling can vary depending on factors such as the company's size, industry type, and overall business goals. Each organization may interpret and implement cross-selling differently based on its market conditions and financial objectives.

Objective:

The objectives of cross-selling can be twofold:

1. To boost the revenue generated from existing customers, or
2. To strengthen and maintain the relationship with current clients.

The method adopted for cross-selling can differ across businesses.

Unlike acquiring new customers, cross-selling carries a certain risk, as it may potentially affect the established relationship with the client.

Therefore, it is crucial to ensure that the additional product or service offered genuinely adds value to the customer's experience and benefits them through their association with the organization. In reality, large enterprises often integrate both cross-selling and up-selling strategies to maximize their income.

For the seller, the advantages are significant. The most evident is the increase in overall sales. Moreover, there are operational efficiencies gained by managing a single account instead of multiple ones.

Most importantly, businesses that successfully cross-sell more products or services to a client are less likely to lose them to competitors. The greater the number of offerings a client purchases from a vendor, the higher the cost and effort involved in switching to another provider.

In the case of United India Insurance Company, there are two types of customers: Existing Customers and New Customers.

Hence to maximise the benefits of cross-selling, existing customers are targeted.

For Existing Customers, three cases arise:

Case-1: When New Coverage is greater than the Current Coverage

- In this case, an existing customer who is already covered by one product insurance is offered an insurance coverage for another new product.
- The current coverage provided to a customer under a current product (1) may or (2) may not be equal to the coverage he/she receives under another new product
 1. Consider an example where a 30 year old customer's current product coverage is Rs. 10,00,000 and he is provided with Rs. 10,00,000 as coverage under another new product. The New Coverage for this customer is depicted as Rs.20,00,000 i.e. the algebraic sum of the current product coverage and the coverage offered under another new product
 2. Consider an example where the same 30 year old customer's current product coverage is Rs. 10,00,000 and this time he is provided with Rs. 30,00,000 as coverage under another new product. The New Coverage for the same customer in this case is Rs.40,00,000 i.e. the algebraic sum of the current product coverage and the coverage offered under another new product.

Case-2: When New Coverage is equal to the Current Coverage

- In this case, the validity or rather the lifetime period of a current product insurance under which a particular customer is covered is terminated or completed.
- Through the cross-selling strategy, the same customer, who is no longer covered under the current product insurance, is offered another new product insurance with a coverage equal to his/her expired or previous product coverage.
- Consider an example where the lifetime period of a particular customer's current product insurance ("ANS") with a current coverage of Rs. 10,00,000 has been completed or his current product insurance has expired.
- As part of cross-selling, the same customer is offered another new product insurance ("TLE") with a coverage equal to the previous expired product i.e. Rs 10,00,000. Hence, now the customer's New Coverage is Rs. 10,00,000

Case-3: When New Coverage is less than the Current Coverage:

- In this case, the validity or rather the lifetime period of a current product insurance under which a particular customer is covered is terminated or completed.
- Through the cross-selling strategy, the same customer, who is no longer covered under the current product insurance, is offered another new product insurance with a coverage not equal to his/her expired or previous product coverage.
- Consider an example where the lifetime period of a particular customer's current product insurance ("INV") with a current coverage of Rs. 40,00,000 has been completed or his current product insurance has expired.
- As part of cross-selling, the same customer is offered another new product insurance ("END") with a coverage equal to Rs 20,00,000. Hence, now the customer's New Coverage is Rs. 20,00,000.

Types of Cross-Selling:

Broadly speaking, cross-selling takes three forms:

- a) First, while servicing an account, the product or service provider may hear of an additional need, unrelated to the first, that the client has and offer to meet it.
- b) Selling add-on services - That happens when a supplier convinces a customer that it can enhance the value of its service by buying another from a different part of the supplier's company.
- c) The third kind of cross-selling can be called selling a solution. In this case, the customer buying air conditioners is sold a package of both the air conditioners and installation services.

Benefits of Cross-Selling:

- Cross selling builds customer loyalty
- Strengthens the current customer relationship with the firm involved
- Increased sales revenue
- Improves customer and client satisfaction even in B2B businesses where business is conducted between companies rather than between a company and individual consumers Increases Customer Lifetime Value (CLV)

Customer Lifetime Value:

- The lifetime value of a customer, or customer lifetime value (CLV), represents the total amount of money a customer is expected to spend in a business, or on its products, during his/her lifetime.
- To calculate customer lifetime value you need to calculate average purchase value of a customer, and then multiply that number by the average purchase frequency rate to determine customer lifetime value.

Services:

- Motor Insurance
- Health Insurance
- Travel Insurance
- Personal Accident Insurance
- Householder's Insurance
- Shopkeeper's Insurance
- Fire,Marine,Industry,Liability,Micro and Credit Insurance

Data Understanding/Description

The data provided by the insurance company is spread across one dataset with 100000 records and 16 variables; the contents of which are:

- **Age:** Listing Customers' age ranging from 18 years to 60 years
- **Gender:** Listing whether the customer is a male or a female
- **Marital Status:** Listing whether a customer is married, divorced or single
- **Family Members:** Listing the number of members in the family of a customer i.e. ranging from 1 to 10 members
- **Education:** Listing the education background of customers
- **Occupation and Job Title:** Listing the occupations and job titles of customers
- **Income:** Listing the income of customers income in Lakhs of Rupees per Annum
- **Current Product:** Listing whether a customer has currently chosen a product insurance or not i.e. it indirectly implies whether a customer is new or already existing.
- **Current Product Type:** Listing the product type (code) of a current product
- **Current Coverage:** Listing the amount (in rupees) current product insurance covers under the current policy or scheme

- **New Product Type:** Listing the product type (code) of a new product
- **New Coverage:** - If a customer is already provided coverage for a current product and opts for a new product: New Coverage exhibits the cumulative total coverage(in rupees) for both the products chosen by a customer
- **Rating:** Listing the rating given by customers as Cold, Warm or Hot for the services provided by the insurance company
- **Converted:** Listing whether the cross-selling strategy of the insurance company has been successful or not i.e. whether the new and existing customers have converted or not
- **Status:** Listing whether the cross-selling strategy of the insurance company has been successful or not i.e. whether the new and existing customers have converted or not. If customers have not converted then various codes have been allocated by the company based on the insurance policy of a customer

Data Audit

DataAudit involves the following steps:

1. Formatting changes – Examining variables that need to be converted to an appropriate data type before analyzing them
2. Validation – Finding out whether the variables contain any missing values or not
3. Checking for Errors – Searching for Variables that contain any outliers or negative values
4. Subsetting – Removal of unnecessary or insignificant variables from the data set

Age

- Missing Data: 44 records
- Range: 18 to 60
- Outliers: None
- Data Type Conversion: Double → Numeric
- Imputation Method: Minimum of Mean and Median
 - Mean > Median (Median = 39 years)
 - **Median** used for imputing missing values

Gender

- Missing Data: 182 records
- Data Type Conversion: To Character
- Distribution:
 - Female: 39,893
 - Male: 59,925
- Imputation Method: Mode (Male)
- Post-Imputation Conversion: Character → Factor

Marital Status

- Missing Data: 42 records
- Data Type: Character
- Distribution:
 - Divorced: 23,160
 - Married: 53,950
 - Single: 22,848
- Imputation Method: Mode (Married)
- Post-Imputation Conversion: Character → Factor

Family Members

- Missing Data: 22 records
- Range: 1 to 10
- Outliers: None
- Data Type Conversion: Double → Numeric
- Imputation Method: Minimum of Mean and Median
 - Mean = 4.659, Median = 4.0
 - **Median** used for imputing missing values

Education

- Missing Data: 46 records
- Data Type: Character
- Distribution:
 - BD: 29,586
 - LHS: 13,399

- MD: 22,543
- NE: 7,232
- PD: 14,746
- UHS: 12,448
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Occupation

- Missing Data: 46 records
- Data Type: Character
- Distribution:
 - SE: 39,443
 - SFT: 8,522
 - SPT: 51,992
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Job Title

- Missing Data: 123 records
- Data Type: Character
- Distribution:
 - BA: 7,980
 - CB: 5,396
 - CF: 2,142
 - DD: 9,795
 - FH: 23,780
 - OC: 1,660
 - OM: 5,935
 - OT: 6,530
 - PA: 1,609
 - PG: 19,623
 - PR: 7,666
 - RR: 7,761
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Income

- Missing Data: 22 records
- Range: 1 to 10
- Outliers: None
- Data Type Conversion: Double → Numeric
- Imputation Method: Minimum of Mean and Median
 - Mean = 4.96, Median = 3.0
 - **Median** used for imputing missing values

Current Product

- Missing Data: 18 records
- Data Type: Character
- Distribution:
 - No: 42,862
 - Yes: 57,120
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Current Product Type

- Missing Data: 33 records
- Data Type: Character
- Distribution:
 - ANS: 18,461
 - END: 5,487
 - INV: 10,346
 - NA: 42,855
 - PMT: 8,477
 - TLE: 14,341
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Current Coverage

- Missing Data: 49 records

- Range: 0 to 15,000,000
- Outliers: Positive Outliers Present
- Data Type Conversion: Character → Numeric
- Imputation Method: Minimum of Mean and Median
 - Mean = 3,633,770
 - Median = 50,000
 - **Median** used for imputing missing values

New Product Type

- Missing Data: 47 records
- Data Type: Character
- Distribution:
 - ANS: 29,151
 - END: 12,718
 - INV: 20,429
 - PMT: 20,429
 - TLE: 25,067
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

New Coverage

- Missing Data: 127 records
- Range: 1,000,000 to 15,000,000
- Outliers: None
- Data Type Conversion: Character → Numeric
- Imputation Method: Minimum of Mean and Median
 - Mean = 6,106,250
 - Median = 3,000,000
 - **Median** used for imputing missing values

Rating

- Missing Data: 44 records
- Data Type: Character
- Distribution:

- Cold: 47,551
- Hot: 22,988
- Warm: 29,417
- Imputation Method: Mode
- Post-Imputation Conversion: Character → Factor

Converted

- Missing Data: 49 records
- Data Type: Character
- Distribution:
 - Converted: 38,296
 - Not Converted: 61,655

Status

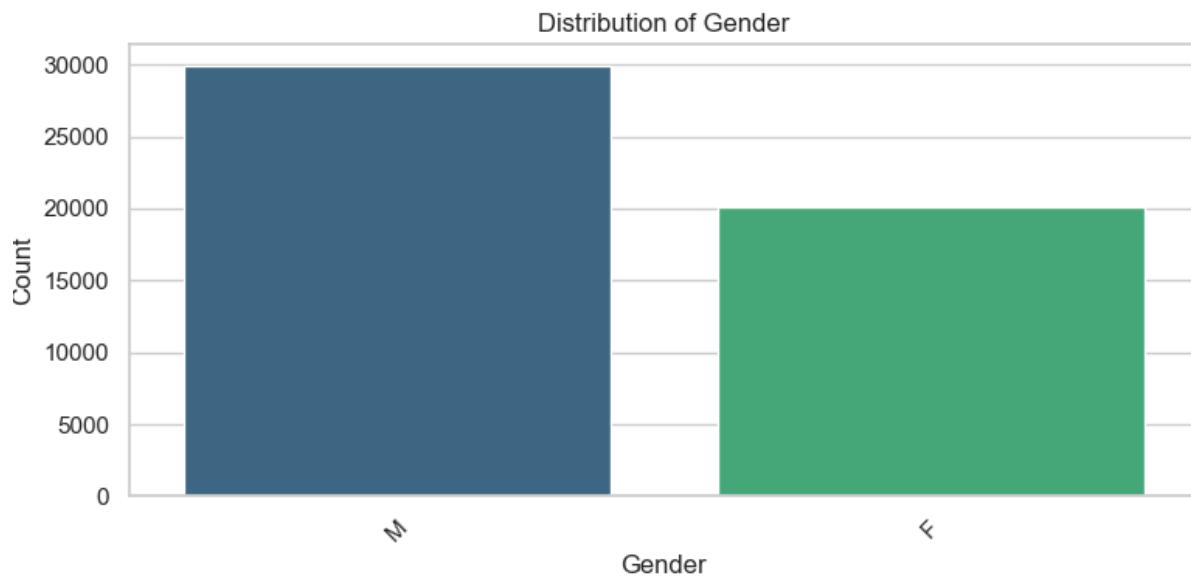
- Missing Data: None
- Transformation:
 - All coded entries replaced with “**Not Converted**”
 - Created a new variable **Status_new** for predictive modeling
 - Data Type Conversion: Character → Factor

Data Exploration and Model Building Report

Gender

- Out of **1,00,000 customers**: **39,893 Females** and **60,107 Males**
- Among Females, **51 years (1004)** age group is maximum
- Among Males, **50 years (1512)** age group is maximum

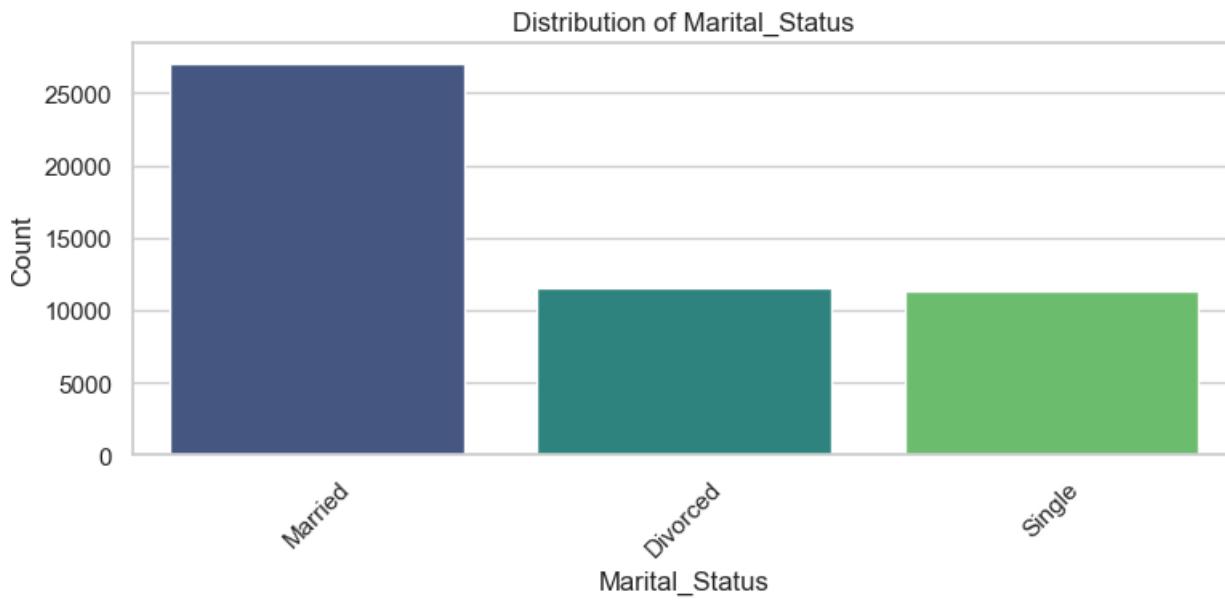
 Bar Plot showing Number of Customers (Y-Axis) vs Gender of Customers (X-Axis)



Marital Status

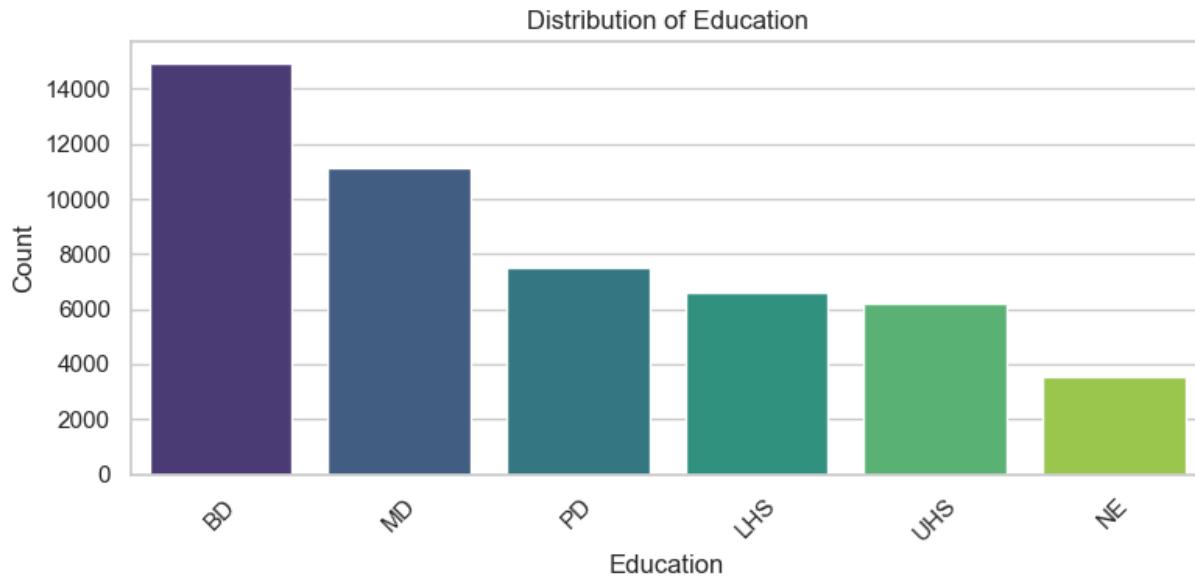
- Divorced: **23,160**, Married: **53,992**, Single: **22,848**
- **21,519 Females** and **32,473 Males** are Married
- **9,075 Females** and **13,773 Males** are Single
- **5,602 customers aged 19–21** are Single
- No customers aged 18–21 are Divorced

 Bar Plot showing Number of Customers (Y-Axis) vs Marital Status (X-Axis)



Education

- Bachelor's Degree: **29,586**, LHS: **13,399**, Master's Degree: **22,543**, NE: **7,232**, Professional Degree: **14,746**, UHS: **12,448**
 - **1207 customers aged 19** not enrolled in any program
 - **18–20-year-olds** have no Master's or Professional Degrees
 - **Males:** 17,863 Bachelor's, 13,592 Master's, 4309 NE
 - **Females:** 11,769 Bachelor's, 8951 Master's
 - **Married:** 16,868 Bachelor's, 12,653 Master's
 - **905 Divorced** customers have no educational qualification
- Bar Plot showing Number of Customers (Y-Axis) vs Education Background (X-Axis)



Occupation

- SE: **39,443**, SFT: **8,522**, SPT: **52,035**
- **Females**: SE-15,731, SFT-3458, SPT-20,704
- **Males**: SE-23,712, SFT-5064, SPT-31,331
- **Divorced**: SE-9483, SFT-1665, SPT-12012
- **Married**: SE-21887, SFT-4228, SPT-27877
- **Single**: SE-8073, SFT-2629, SPT-12146
- Customers with Bachelor's: SE-12310, SPT-17317
- LHS group: only **4 customers** have SPT occupation

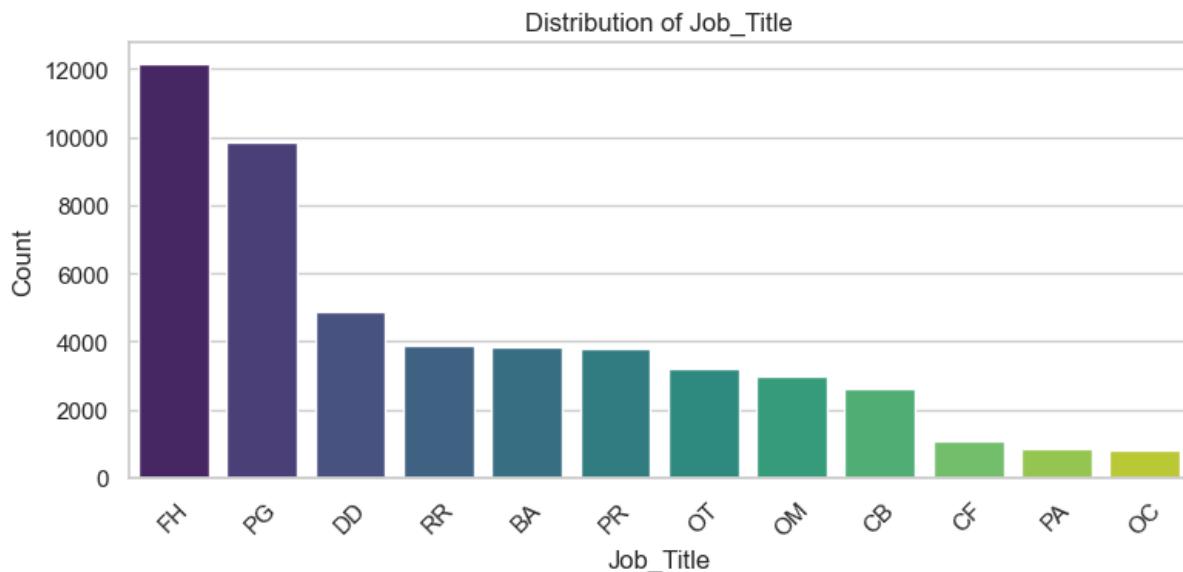
 Bar Plot showing Number of Customers (Y-Axis) vs Occupation (X-Axis)



Job Title

- BA-7980, CB-5396, CF-2142, DD-9795, FH-23903, OC-1660, OM-5935, OT-6530, PA-1609, PG-19623, PR-7666, RR-7761
- **Females:** FH-9517, PG-7762
- **Males:** FH-14386, PG-11861
- Bachelor's Degree holders: 10,629 FH
- Master's Degree holders: 8,052 FH
- Professional Degree holders: 5196 FH
- SE occupation: 23,831 FH

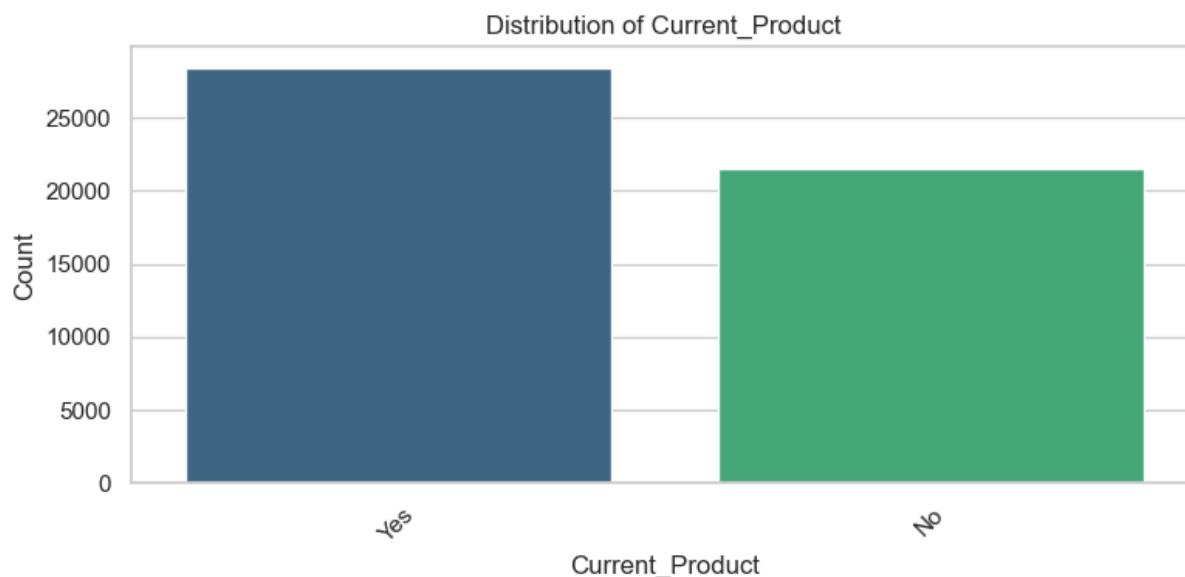
 Bar Plot showing Number of Customers (Y-Axis) vs Job Title (X-Axis)



Current Product

- **Existing Customers:** 57,129
- **New Customers:** 42,871
- Existing: Females–17174, Males–25697
- New: Females–22719, Males–34410
- Existing Married–32437, Single–9253
- New Married–21555, Single–13595
- Existing customers have higher cumulative and average income (**₹5.2 Lakh**)

Bar Plot: Number of Customers vs Current Product (Existing/New)



1.Univariate Analysis

1. Age and Income follow roughly normal-like distribution but slightly right-skewed → middle-class workforce dominates.
2. Current_Coverage shows extreme skew — indicates strong inequality in insurance coverage.
3. New_Coverage seems clustered around ₹10–30 lakh, showing the company's preferred selling range.

	Age	Income	Family_Members	Current_Coverage	New_Coverage
count	49976.000000	49976.000000	49976.000000	4.997600e+04	4.997600e+04
mean	38.979161	4.965280	4.658656	3.592711e+06	6.088942e+06
std	12.124556	3.292877	2.603157	5.635161e+06	5.949610e+06
min	18.000000	0.893503	1.000000	-2.165380e+05	7.555480e+05
25%	28.000000	2.038755	2.000000	6.463000e+03	1.091648e+06
50%	39.000000	3.032735	4.000000	1.499985e+05	2.986876e+06
75%	49.000000	9.106073	7.000000	3.943459e+06	1.494404e+07
max	60.000000	10.122132	10.000000	1.527844e+07	1.522913e+07

Business Recommendations

1. Segmented Targeting

----->Young (25–35 yrs): Pitch affordable “starter” health plans.

----->Mid-age (36–50 yrs): Offer retirement savings or child education linked plans.

----->Older (50+): Promote senior care and medical coverage.

2. Income-based Personalization

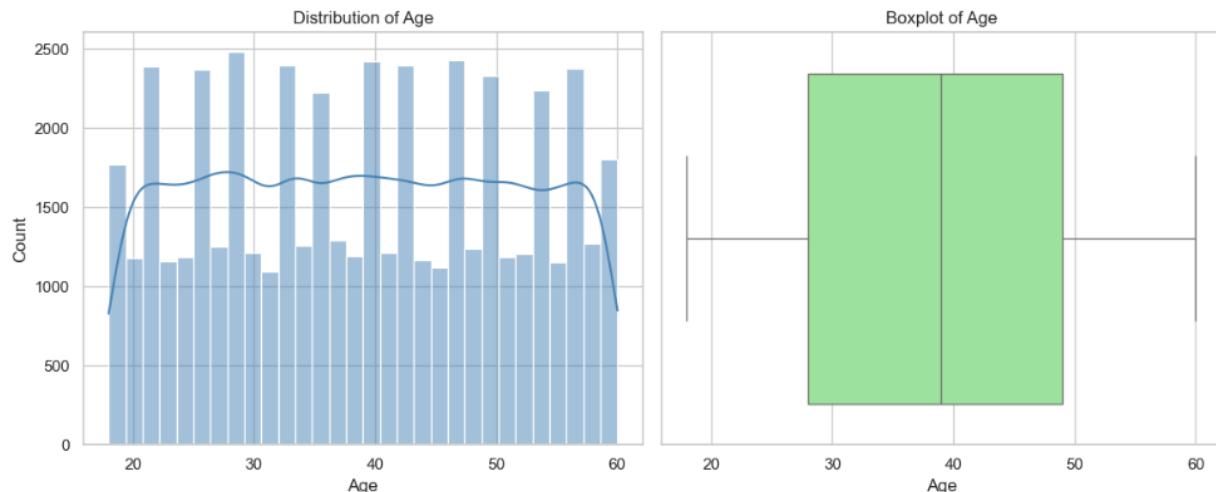
----->Low–mid income → focus on term and family coverage.

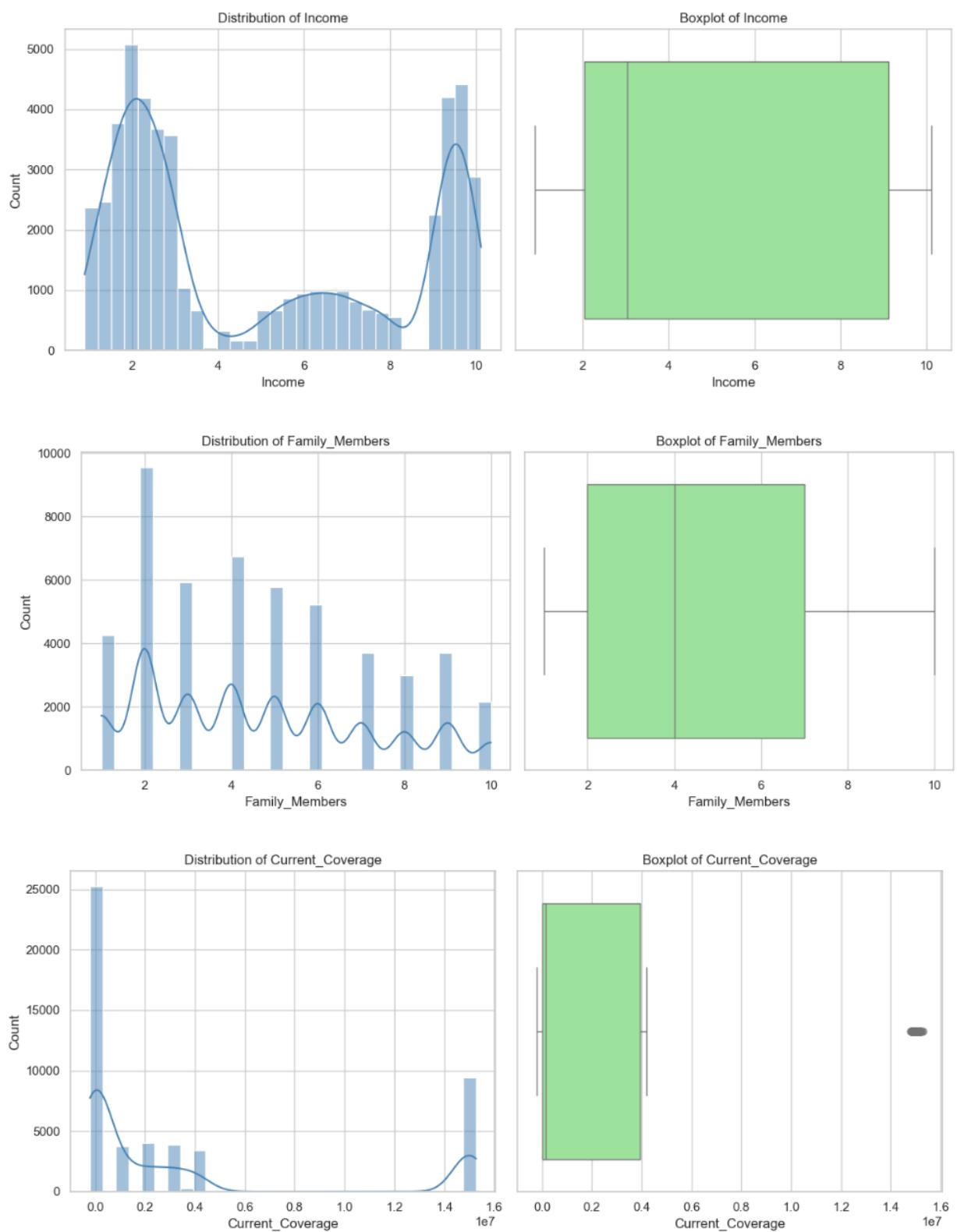
----->High income → focus on investment + protection hybrids (ULIPs, annuities).

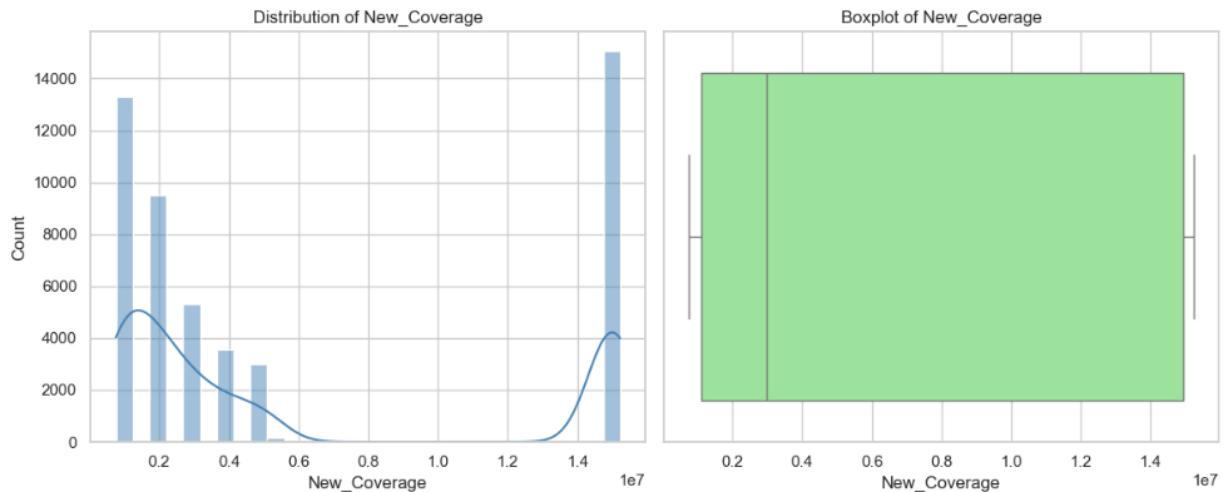
3. Coverage Gap Opportunity

----->Customers with low or zero Current_Coverage are prime cross-sell targets.

----->Marketing can filter customers where Current_Coverage < ₹2 lakh and pitch “Upgrade your protection” offers.



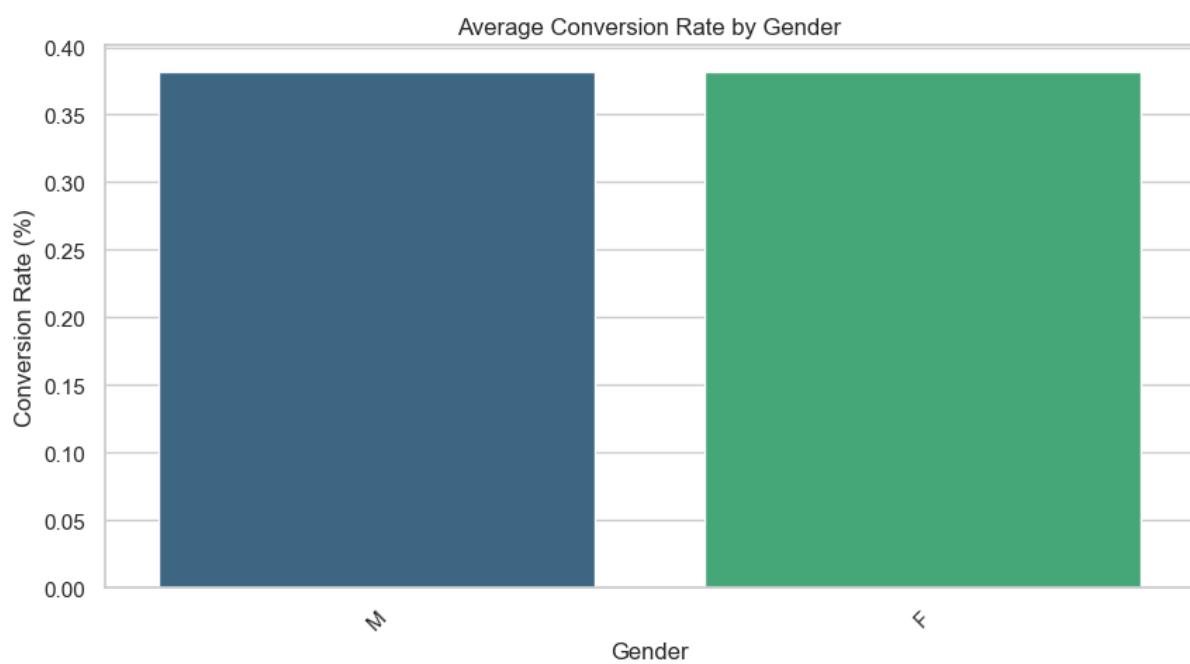
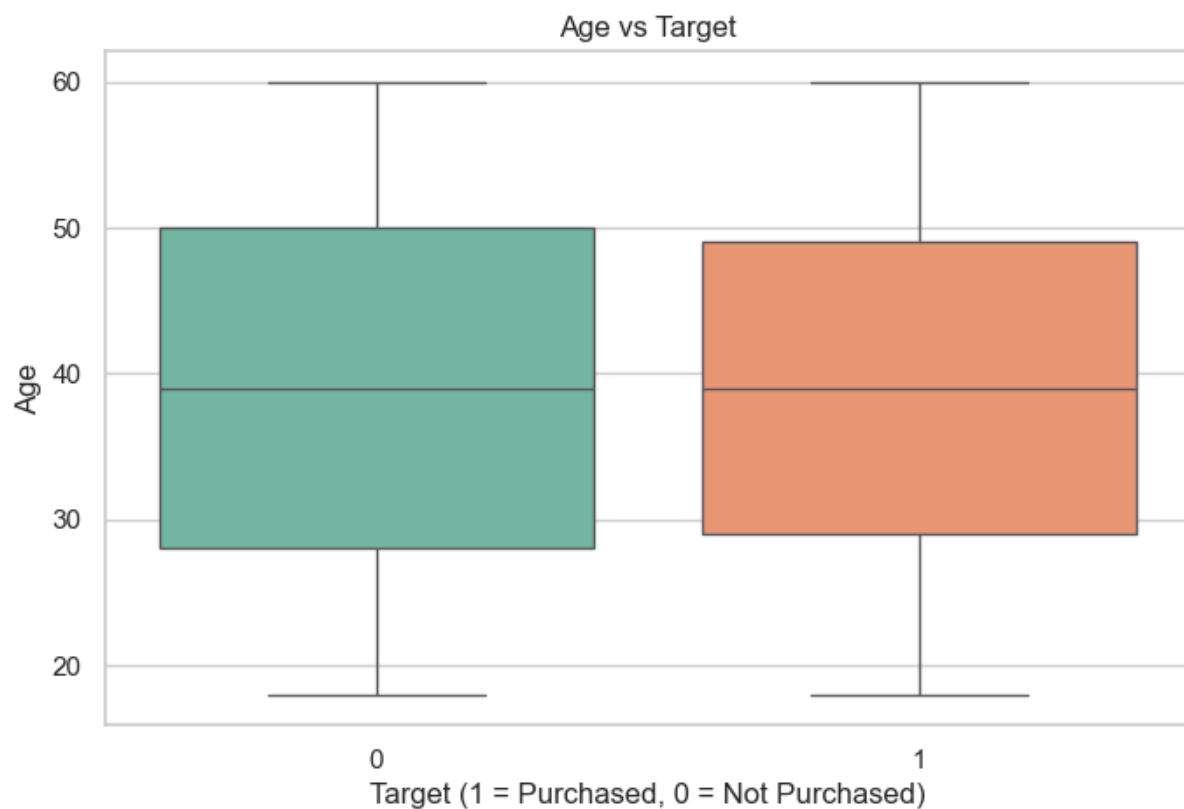


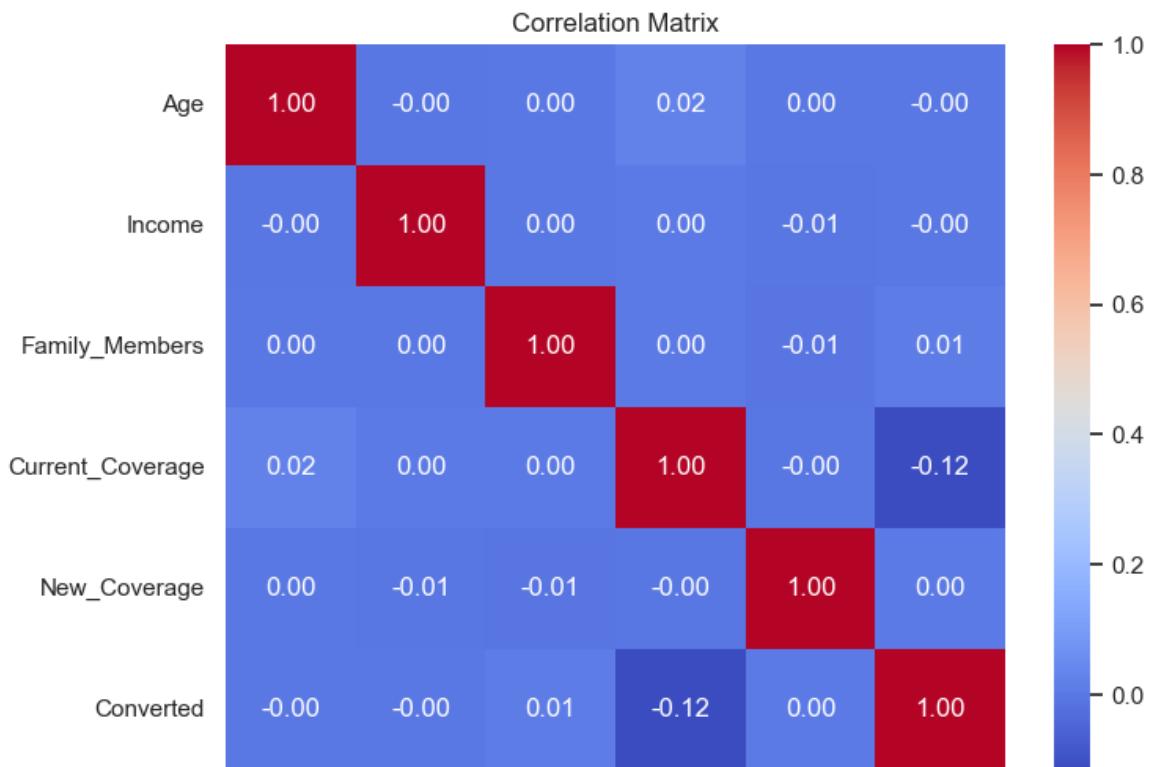


Bivariate Analysis

We'll start by exploring:

1. Numerical vs Target relationships (using boxplots & mean comparisons)
2. Categorical vs Target relationships (using grouped bar charts)
3. Correlations among numerical columns





Overall Business Recommendations

1. Target young, mid-income customers with moderate current coverage — they have the highest cross-sell conversion probability.
2. Use family-based segmentation to promote add-on or bundle products.
3. Leverage high-rated, satisfied customers for premium cross-sell campaigns — they already trust the brand.
4. Design differentiated messaging by occupation and education level.
5. Enhance digital engagement for younger customers — app notifications, email nudges, and quick purchase options.

Data-Driven Insights & Business Interpretations

1. Pairwise Relationships

Observation:

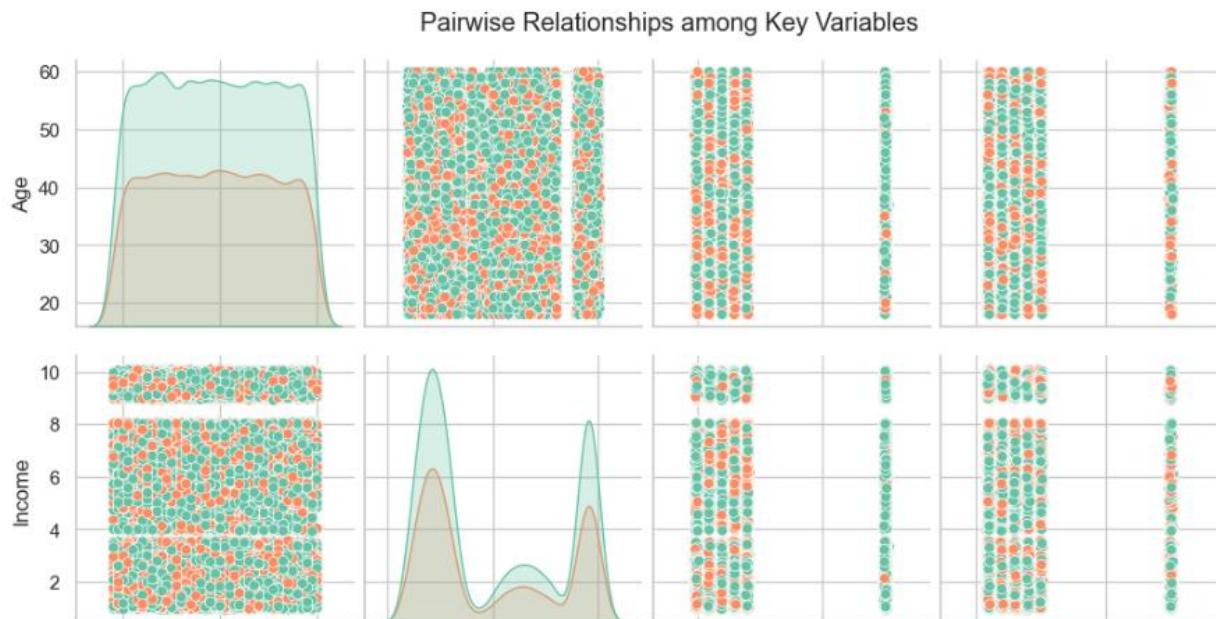
1. A clear cluster of younger, moderate-income customers shows higher cross-sell conversion (target=1).

2. Higher income + older customers are more likely to not purchase additional policies.
3. Customers with low current coverage and medium income tend to opt for new coverage more often.

Business Insight:

---> Focus cross-sell campaigns on young-to-mid-age, mid-income customers with smaller existing coverage.

---> Launch smart add-on packages with low-cost premium top-ups for this segment.



Income vs Age (Target Overlay)

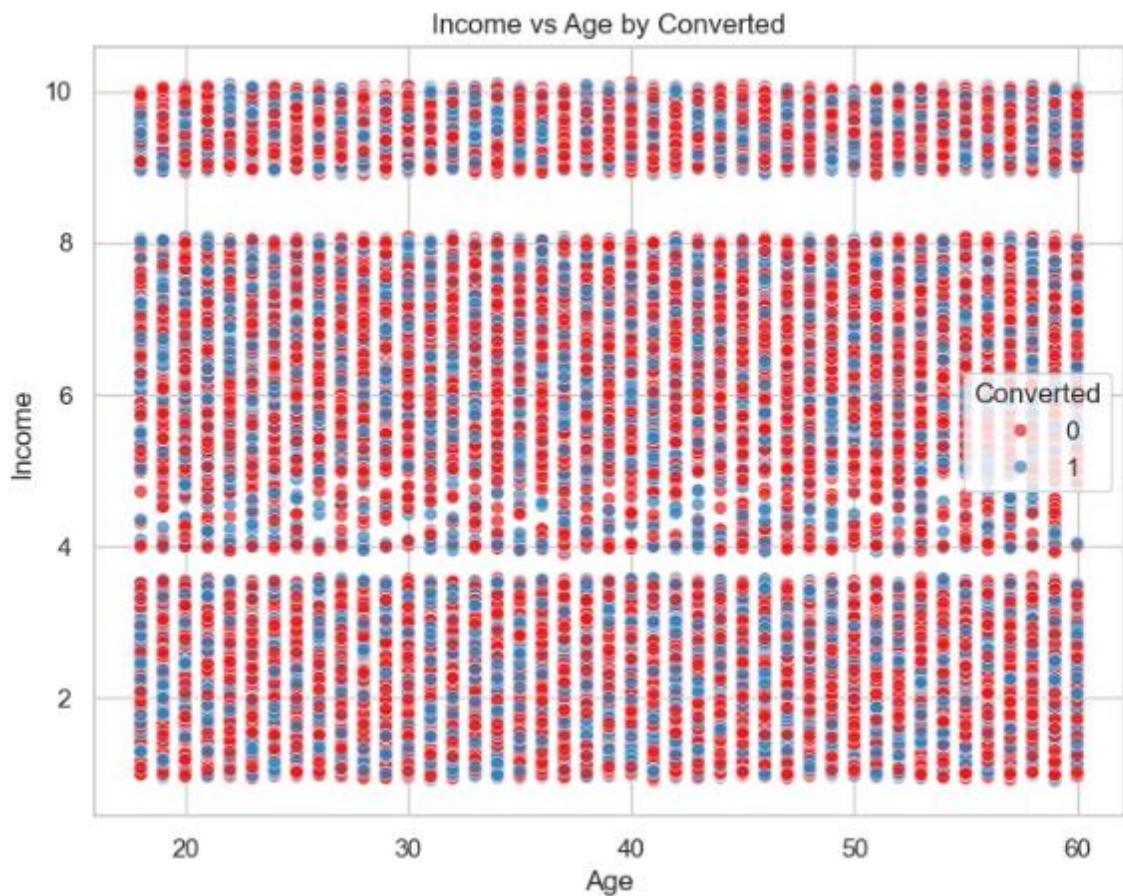
Observation:

1. The plot shows a non-linear trend:
 - a. Young customers (under 35) with income between ₹4L–₹10L respond best.
 - b. Older, higher-income customers are less influenced by marketing offers.

Business Suggestion:

1. Target digital-first campaigns (mobile app, WhatsApp reminders, email drip) for the 25–40 age group.

- For older, high-income customers, offer personalized advisory or one-to-one relationship management instead of generic cross-sell.



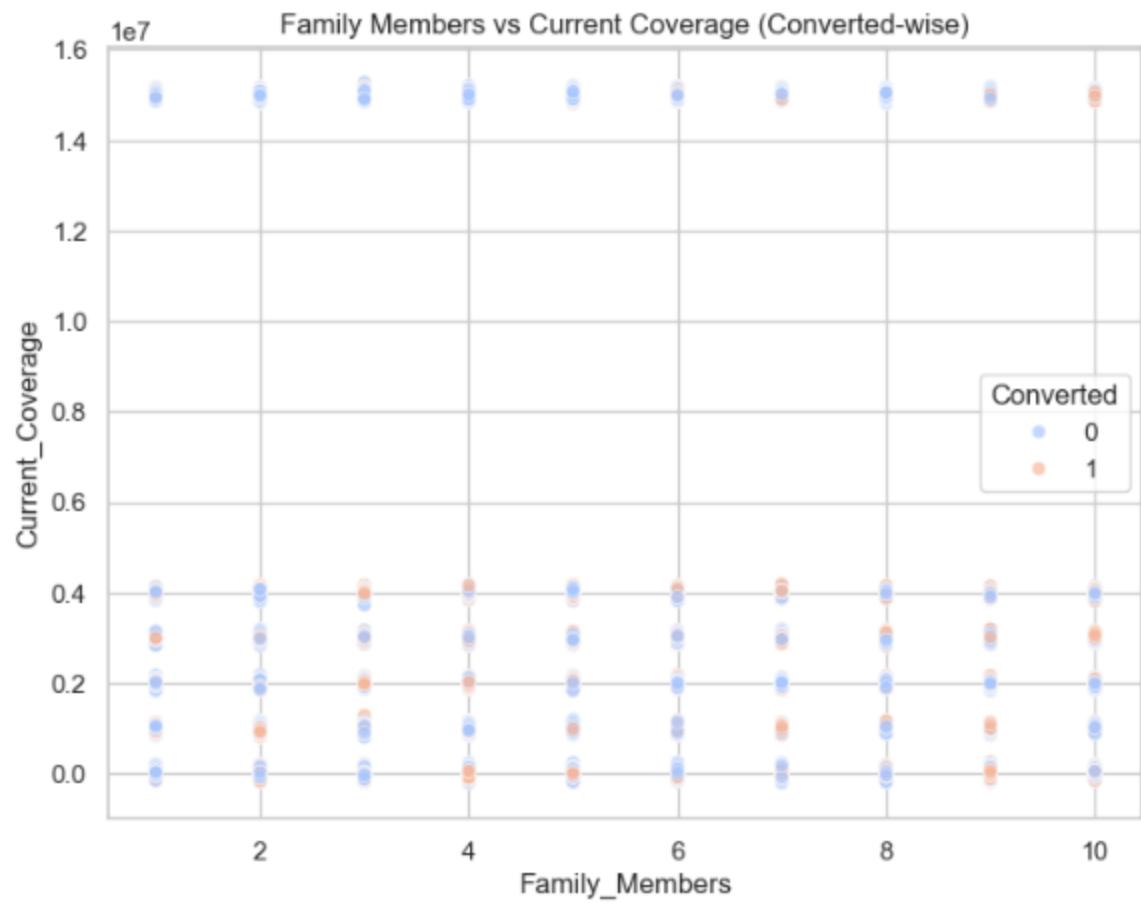
Family Members vs Current Coverage

Observation:

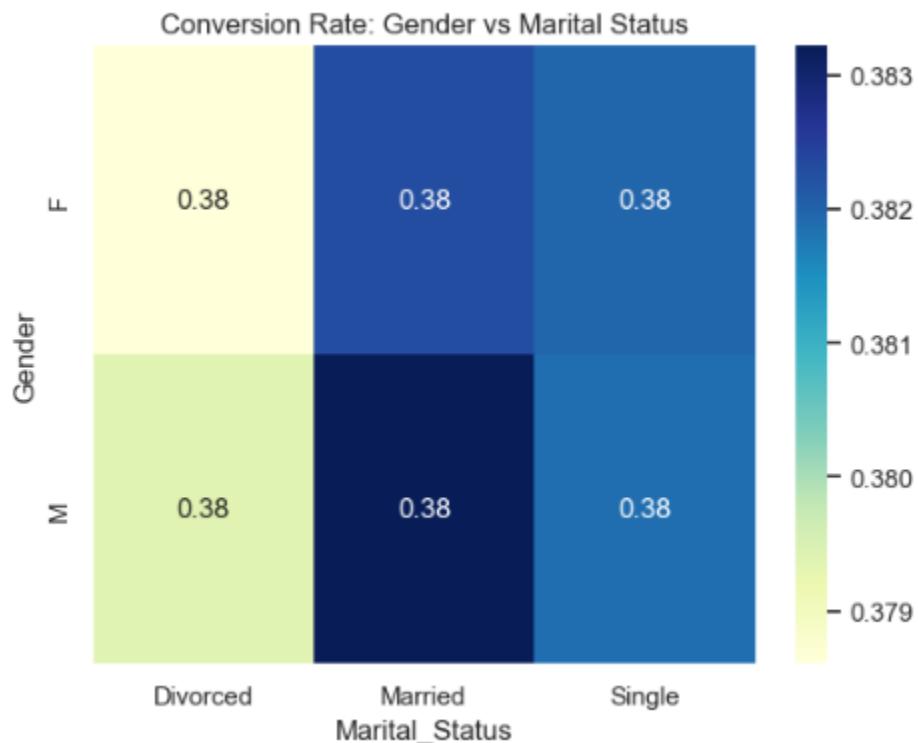
- Customers with more dependents (3–5 members) but low coverage show high interest in buying new products.
- Those with smaller families and already high coverage have lower conversion.

Business Suggestion:

- Introduce "Family Security Upgrade" plans — top-up benefits that increase protection for additional dependents.
- Bundle life + health insurance plans for families under the same premium umbrella.



Gender vs Marital Status (Pivot Heatmap)



Observation:

1. Married males have the highest conversion rate, followed by married females.
2. Single customers (both male & female) show significantly lower engagement.

Business Suggestion:

1. Develop “Family First” offers focusing on protection and dependents’ well-being.
2. For single customers, use investment-linked or travel-related insurance (appeal to independence and lifestyle).