CUSTOMER SEGMENTATION ANALYSIS IN THE INSURANCE INDUSTRY

PROBLEM STATEMENT

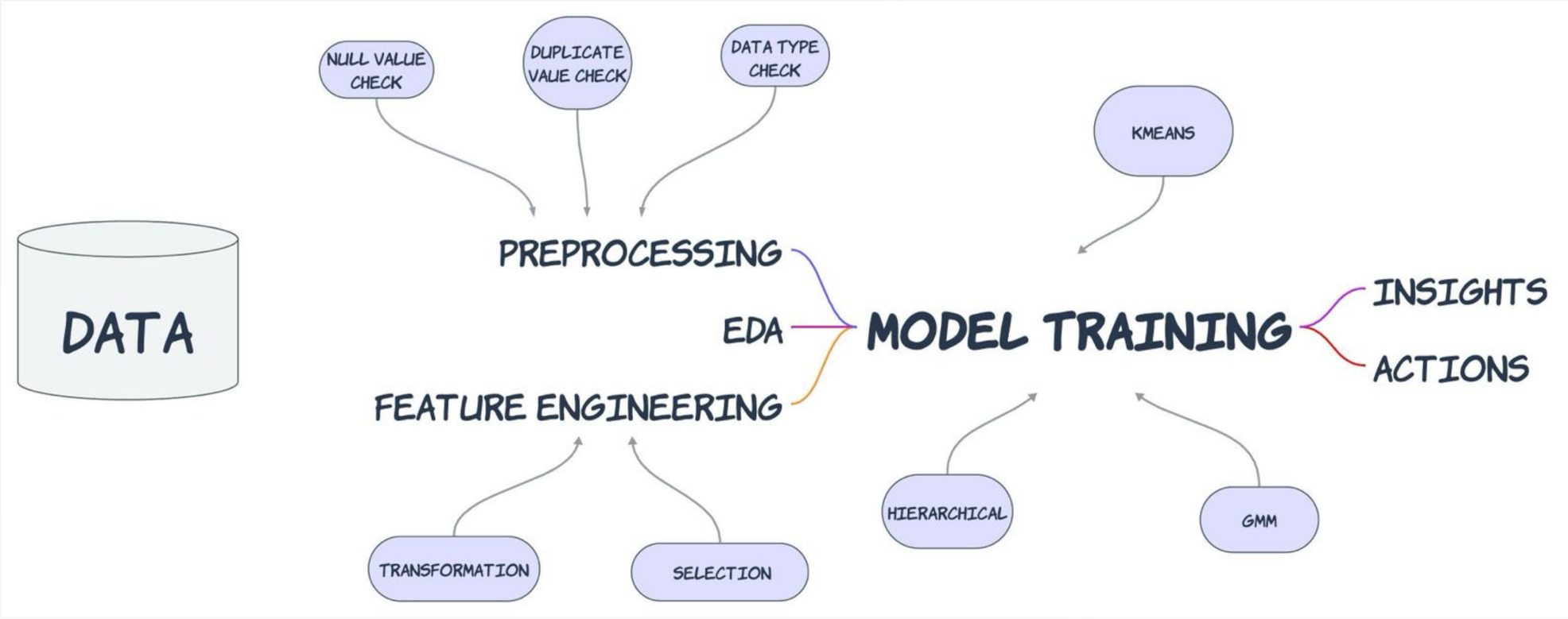
Investigate customer segmentation in the insurance dataset to identify distinct groups with varying product interests, market participation, and responsiveness to marketing. Utilize insights to optimize strategic decisions regarding opportunities promotions, pricing, and targeted marketing efforts.

DATA

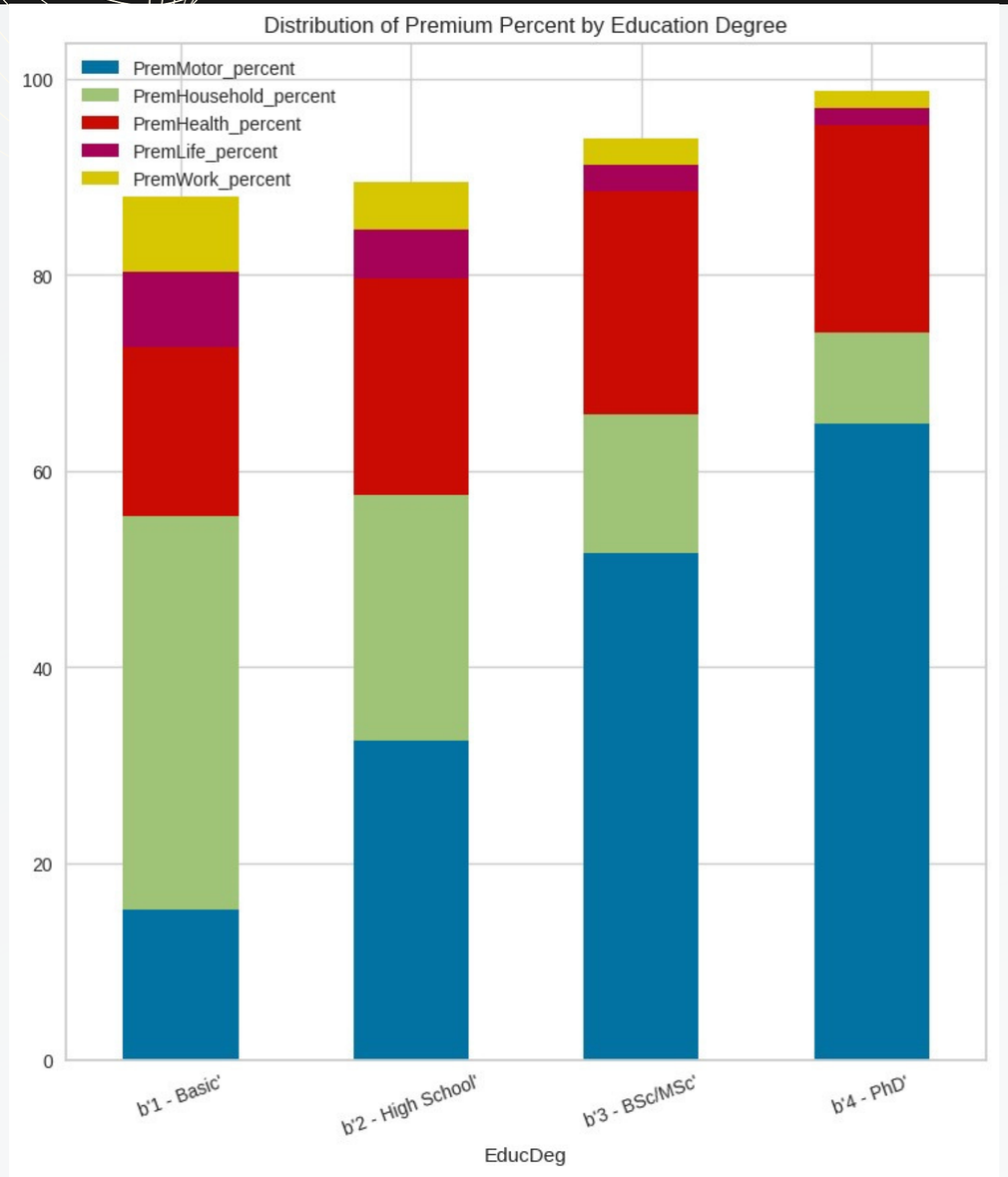
Variable	Description
ID	ID
First Policy	Year of the customer's first policy
Birthday	Customer's Birthday Year
Education	Academic Degree
Salary	Gross monthly salary (€)
Area	Living area
Children	Binary variable (Y=1)
CMV	Customer Monetary Value
Claims	Claims Rate
Motor	Premiums (€) in LOB: Motor
Household	Premiums (€) in LOB: Household
Health	Premiums (€) in LOB: Health
Life	Premiums (€) in LOB: Life
Work Compensation	Premiums (€) in LOB: Work Compensations

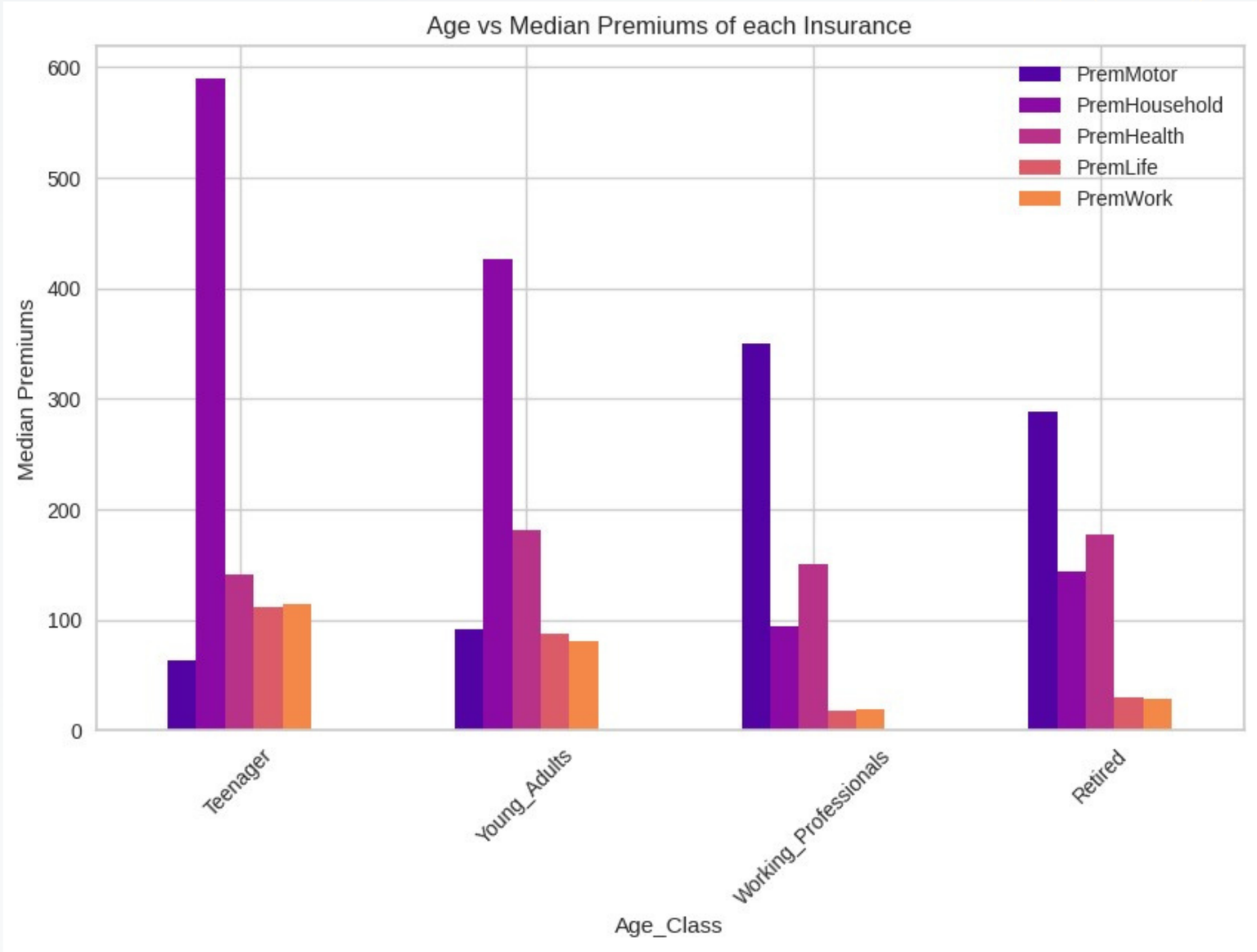
- DATASET CONTAINS 10,290 ROWS AND 13 COLUMNS
- DATASET EXHIBITS APPROXIMATELY 2% MISSING VALUES
- LIFETIME VALUE = (ANNUAL PROFIT FROM THE CUSTOMER) X (NUMBER OF YEARS THAT THEY ARE A CUSTOMER) - (ACQUISITION COST)
- ANNUAL PREMIUMS (2016). NEGATIVE PREMIUMS MAY MANIFEST REVERSALS OCCURRED IN THE CURRENT YEAR, PAID IN PREVIOUS ONE(S)

PROJECT FLOW



ANALYSIS


Distribution of Premium Percent by Education Degree
PremMotor_percent
PremHousehold_percent
PremHealth_percent
PremLife_percent
PremWork_percent
100
80
60
40
20
0
b'1 - Basic'
b'2 - High School'
b'3 - BSc/MSc'
b'4 - PhD'
EducDeg


Age vs Median Premiums of each Insurance
PremMotor
PremHousehold
PremHealth
PremLife
PremWork
600
500
400
300
200
100
0
Median Premiums
Teenager
Young_Adults
Working_Professionals
Retired
Age_Class

DATA MUNGING

HANDLING MISSING VALUES

- About 400 missing values were present
- Imputed continuos variables using mean
- Imputed categorical variables using mode

COHERENCE CHECK

- Removed Invalid Birthyears
- Replaced records which had Firstpolicy year before birth_year
- Removed records which had sum of all premiums greater than Annual Salary

OUTLIER REMOVAL

- We incorporated IQR range , Manual thresholding techniques
- Treated columns include all_premiums,monthly_salary,cmv and claim rate

CATEGORIZING FEATURES

- The education degree has been transformed into numeric values (1, 2, 3, 4) to enhance interpretation and facilitate management
- The salary column has been categorized into six bins

FEATURE ENGINEERING

SCALING : AFTER EXPERIMENTING WITH VARIOUS SCALING TECHNIQUES, WE DETERMINED THAT MIN-MAX SCALER PRODUCED THE MOST FAVORABLE OUTCOMES

ADDITION : WE EXPANDED THE DATASET BY INCORPORATING NEW FEATURES CALCULATED USING DEFINED FORMULAS AND ARE REPRESENTED BELOW. THIS APPROACH AIMS TO REVEAL INSIGHTS AND OFFER A BETTER INTERPRETATION OF THE DATA

$$\text{total\_premium} = \sum_{i}^{n} (\text{premium})_i$$

$$\text{Profit\_percent} = \left(\frac{\text{total\_premiums} - \text{claim\_amount}}{\text{claim\_amount}} \right) \times 100$$

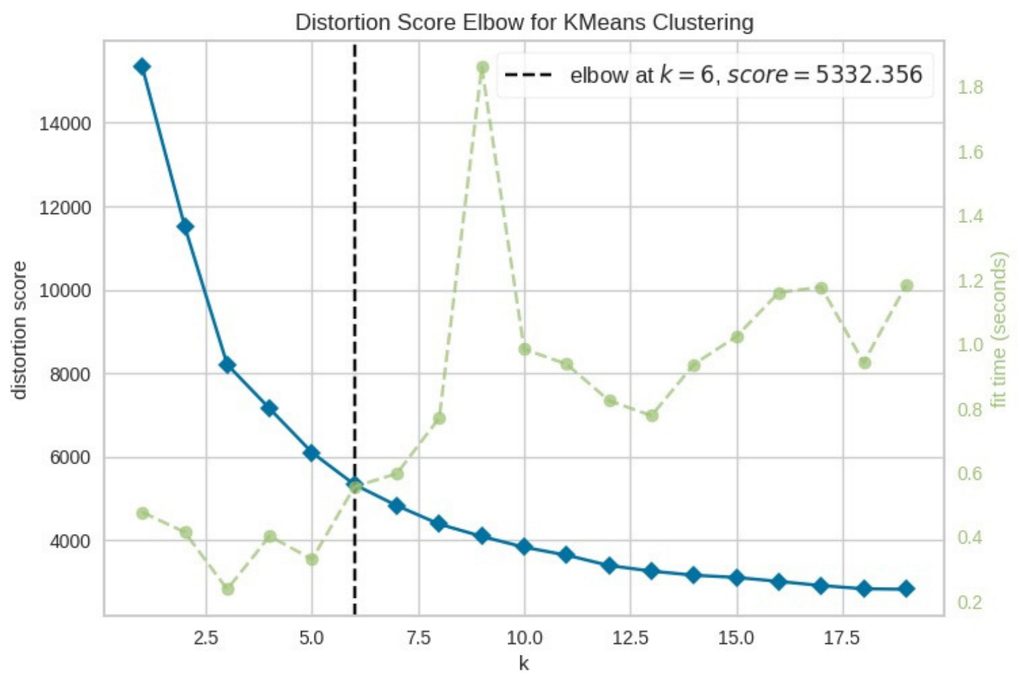
$$\text{retention\_cost} = (\text{profit} \times \text{no of years}) - \text{cmv}$$

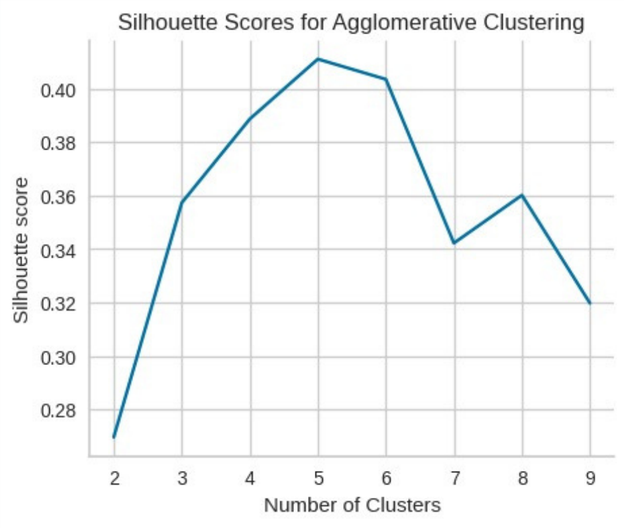
$$\text{commitment} = \frac{\text{annual salary}}{\text{total\_premium}}$$

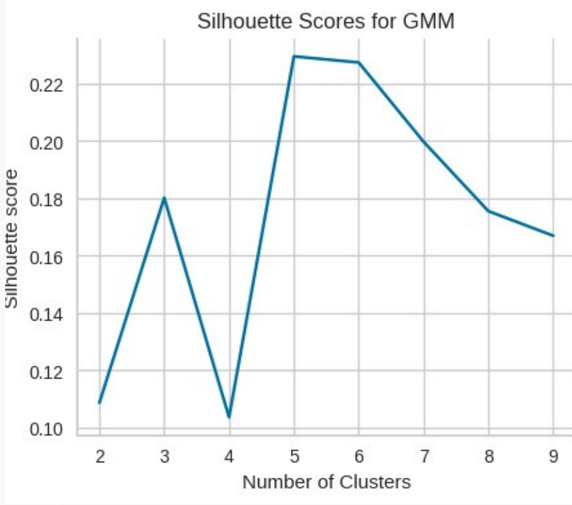
$$\text{loss} = \begin{cases} 0 & \text{if no loss occurred} \\ 1 & \text{if loss occurred} \end{cases}$$

CLUSTERING

- *THE PROCESS BEGAN BY APPLYING VARIOUS CLUSTERING ALGORITHMS, INCLUDING K-MEANS, HIERARCHICAL CLUSTERING, AND GAUSSIAN MIXTURE MODELS, TO THE DATASET.*
- *TO ASSESS THE CLUSTERING QUALITY OF EACH ALGORITHM, WE UTILIZED THE SILHOUETTE SCORE AS A METRIC.*
- *AFTER CALCULATING THE SILHOUETTE SCORE FOR EACH ALGORITHM, WE COMPARED THEM TO DETERMINE THE ONE WITH THE HIGHEST SCORE, INDICATING THE MOST SUITABLE CLUSTERING SOLUTION.*
- *SUBSEQUENTLY, WE CONDUCTED FURTHER OPTIMIZATION BY FINE-TUNING THE PARAMETERS OF THE SELECTED CLUSTERING ALGORITHM USING GRIDSEARCHCV.*

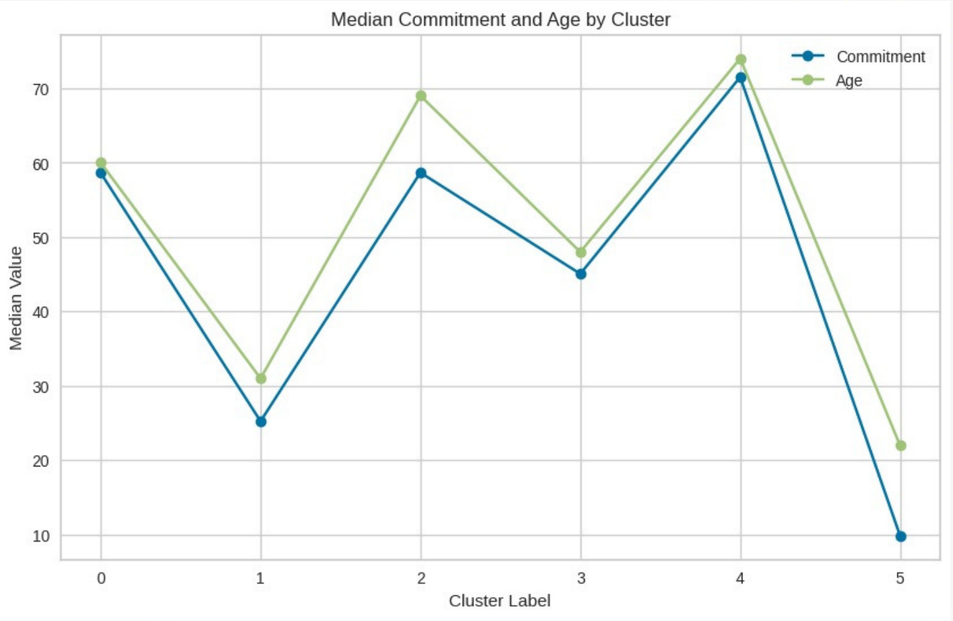


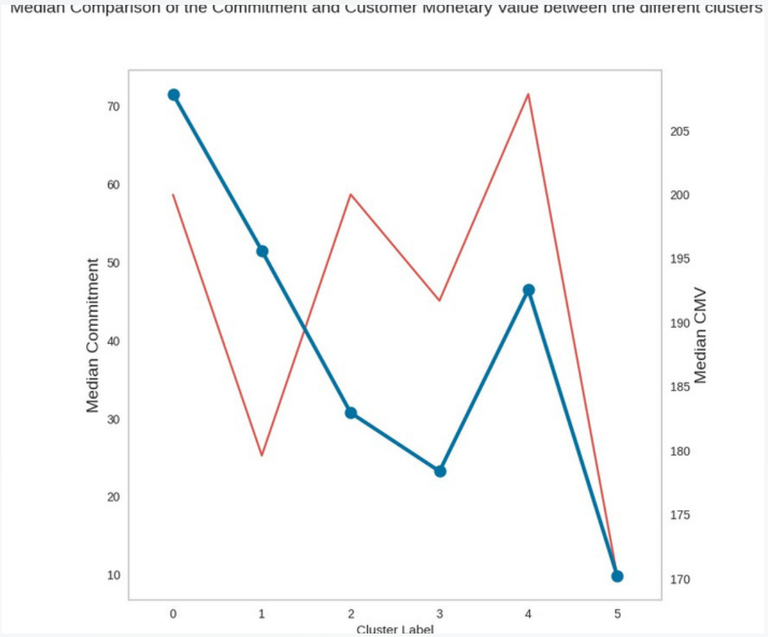


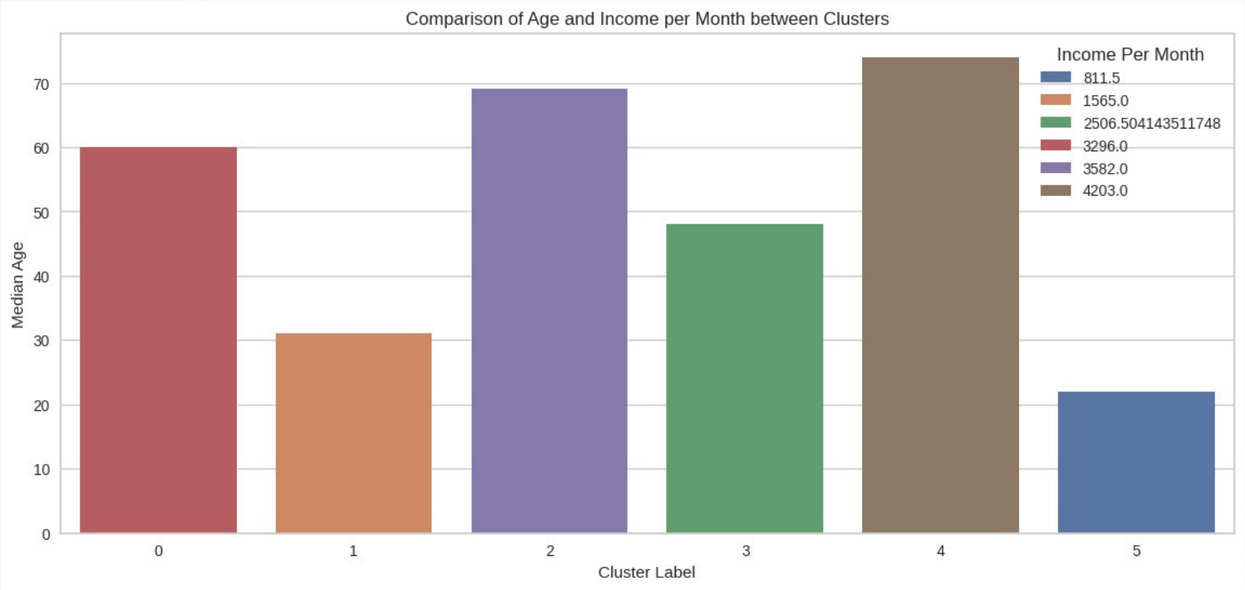


POST EDA


Median Comparison of the Profit_percent between the different clusters
Median Profit_percent
Cluster Label


Median Commitment and Age by Cluster
Commitment
Age
Median Value
Cluster Label


Median Comparison of the Commitment and Customer Monetary Value between the different clusters
Median Commitment
Median CMV
Cluster Label


Comparison of Age and Income per Month between Clusters
Income Per Month
811.5
1565.0
2506.504143511748
3296.0
3582.0
4203.0
Median Age
Cluster Label

INSIGHTS

Profitable Customers, in their middle age with lowest claim rate

---

** Middle-class salary between 3000 and 4000 euros*
** Most of have children and are educated*
** Largest Retention Cost*
** Second highest Motor premiums although presenting the lowest claims rate*
** Second Lowest Health premiums*

Educated young adults with low purchasing power

** Young adults, born around 1985*
** Educated people*
** Monthly salary between 1000 and 2000 euros*
** Contribute high premiums along with low retention cost*
** Overall a secure client*

Senior people with no children and high purchasing power

** All customers have a monthly income between 3000 and 4000 euros*
** Educated people (Bachelor/ Masters degree)*
** No children*
** Senior people, on average 69 years*
** Low percentage of claims rate*
** Focuses more on health premium*

Oldest customers with no children and high purchasing power

** Highest purchasing power with a monthly income above 4000 euros*
** Educated people*
** Most have no children*
** Highest health premium*
** High Commitment required*
** Lowest loss percent*

Middle-aged customers with medium purchasing power

** Monthly Salaries between 2000 and 3000 euros*
** Educated people*
** Most have, at least, 1 child*
** On Average 48 years old.*
** Pays the highest Premiums for Motor.*
** Pays less for Life, Work and Health premiums.*

Young adults with low purchasing power

** Most have a max income of 1000 euros*
** They are less educated (most with only basic education)*
** Most have, at least, 1 child*
** Young adults, most being around 22 years old*
** High Claim Rate with Low Retention Cost*
** Contribute very Low to Motor Premiums but very High to other premiums.*
** Highest house premiums but with lowest the profit percent and highest loss percent.*

STRATEGIES

1. FOCUS ON MIDDLE-AGED CUSTOMERS FOR STANDARD INSURANCE PLANS

2. TARGET YOUNG ADULTS WITH BASIC INSURANCE PLANS

3. TAILOR PREMIUM INSURANCE OFFERINGS FOR RETIRED ADULTS

4. ALLOCATE MARKETING BUDGET PRUDENTLY, WITH LESS EMPHASIS ON CLUSTER 5 DUE TO LOWER PROFITABILITY AND HIGHER LOSS.

5. CAPITALIZE ON THE HIGH SPENDING POWER OF CUSTOMERS IN CLUSTER 4 BY PROMOTING CROSS-SELLING PRODUCTS.

6. TARGET MIDDLE-AGED INDIVIDUALS FOR MOTOR INSURA
NCE, AND YOUNG ADULTS EARNING AROUND 2000 EUROS PER MONTH FOR HOUSEHOLD AND WORK INSURANCE.

THANK YOU