

Presentation by Team Data 200

Health Insurance Cross Sell Prediction

Rakamin Academy

Team Members



Theofilus Arifin
Project Leader



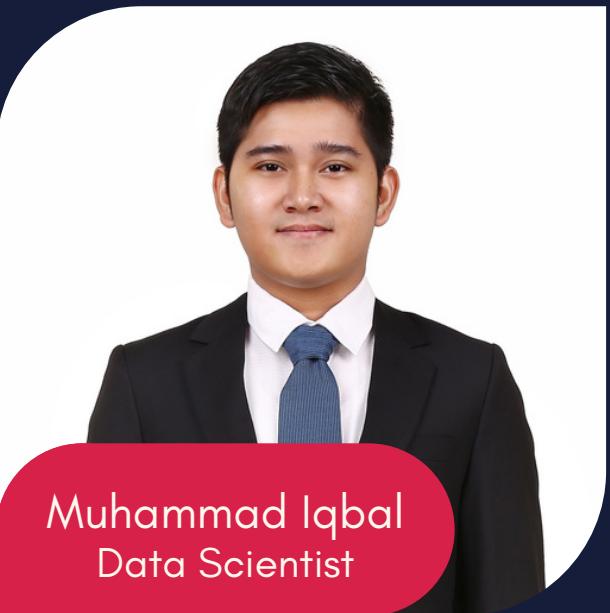
Hanifah Arrasyidah
Data Analyst



Christopher Stephen
Business Analyst



Ujang Pian
Data Engineer



Muhammad Iqbal
Data Scientist



Ramlan Apriyansyah
Business Analyst



Christofer Bryan N. K.
Business Intelligence

Table Of Content



Project Background



EDA & Insight



Pre-Processing



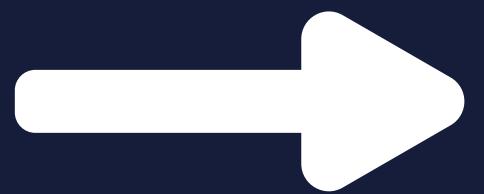
Modeling



Recommendations



Background

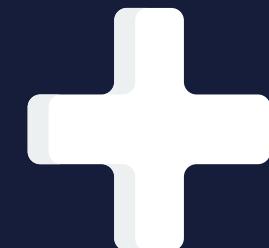


PT APPA

Diversifikasi



Asurasi Kesehatan



Cross Selling



Asurasi Kendaraan

⋮
⋮



12.3% Orang Tertarik

Problem Importance



Pengaruh langsung
pada Pendapatan



Efisiensi Pemasaran



Keberlanjutan dan
Pertumbuhan Bisnis

Goals

Meningkatkan Conversion Rate Pemasaran
Asuransi Kendaraan PT APPA

Objective

Membuat model klasifikasi untuk memprediksi pemegang polis asuransi kesehatan yang tertarik untuk berlangganan asuransi kendaraan.



Business Metric



Conversion Rate

Persentase jumlah pemegang polis yang beralih dari minat menjadi pembeli aktif produk asuransi kendaraan.

Table Of Content



Project Background



EDA & Insight



Pre-Processing



Modeling



Recommendations



Dataset

Health Insurance Cross Sell Prediction

Analytics Vidhya

- Dataset ini berisi informasi demografis dan perilaku nasabah asuransi kesehatan yang berkaitan dengan minat terhadap asuransi kendaraan
- Dataset ini terdiri dari 12 kolom dan 381.109 baris

Link to Dataset:

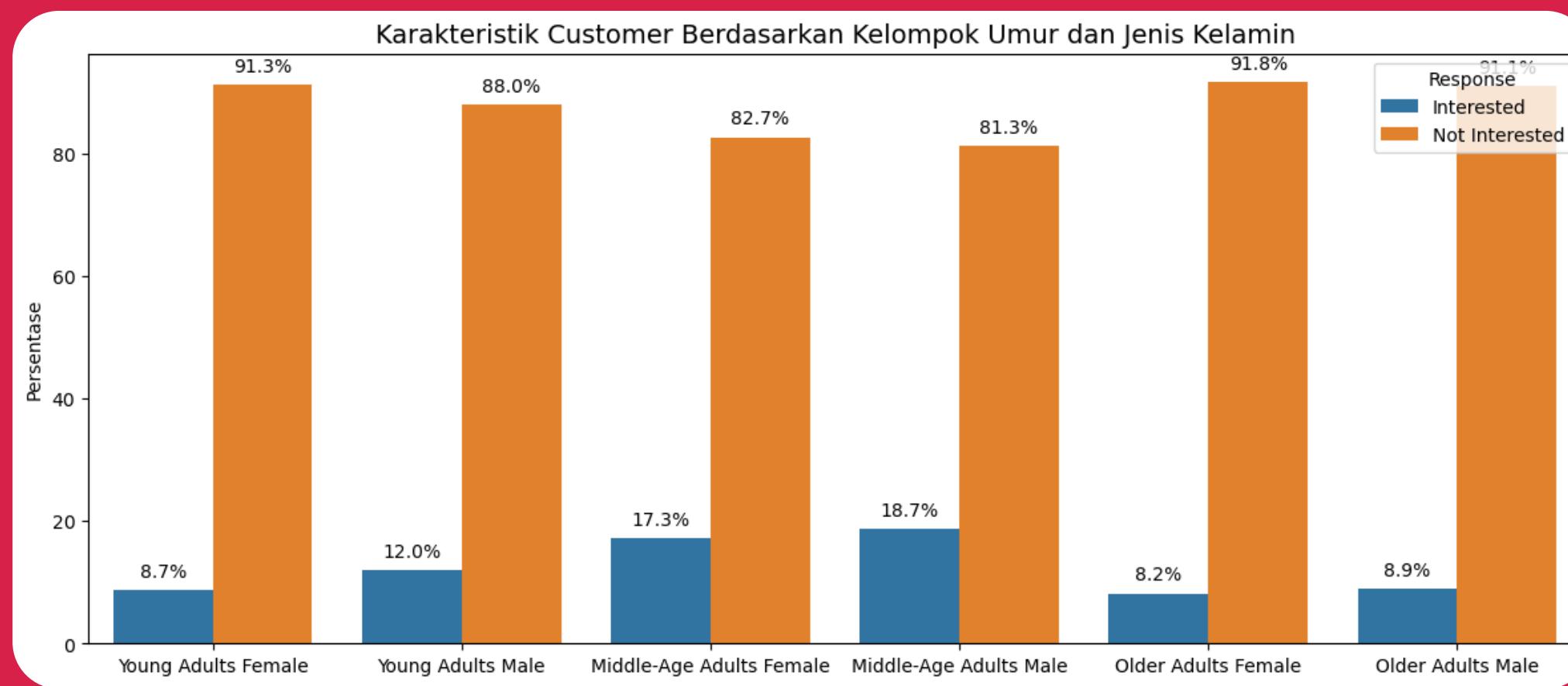
kaggle



Dataset

Kolom Feature	Deskripsi	Kolom Feature	Deskripsi
ID	ID unik tiap nasabah	Vehicle Age	Umur kendaraan nasabah
Gender	Jenis kelamin nasabah	Vehicle Damage	Apakah nasabah pernah mengalami kerusakan kendaraan
Age	Umur nasabah	Annual Premium	Biaya premi tiap tahun
Driving License	Kepemilikan SIM nasabah	Policy Sales Channel	Saluran pemasaran
Region Code	Kode wilayah	Vintage	Lama nasabah berlangganan
Previously Insured	Kepemilikan asuransi kendaraan sebelumnya	Response	Ketertarikan nasabah terhadap asuransi kendaraan

Age Group & Gender to Response



Hubungan antara Kelompok Umur dan Jenis Kelamin Terhadap Asuransi Kendaraan

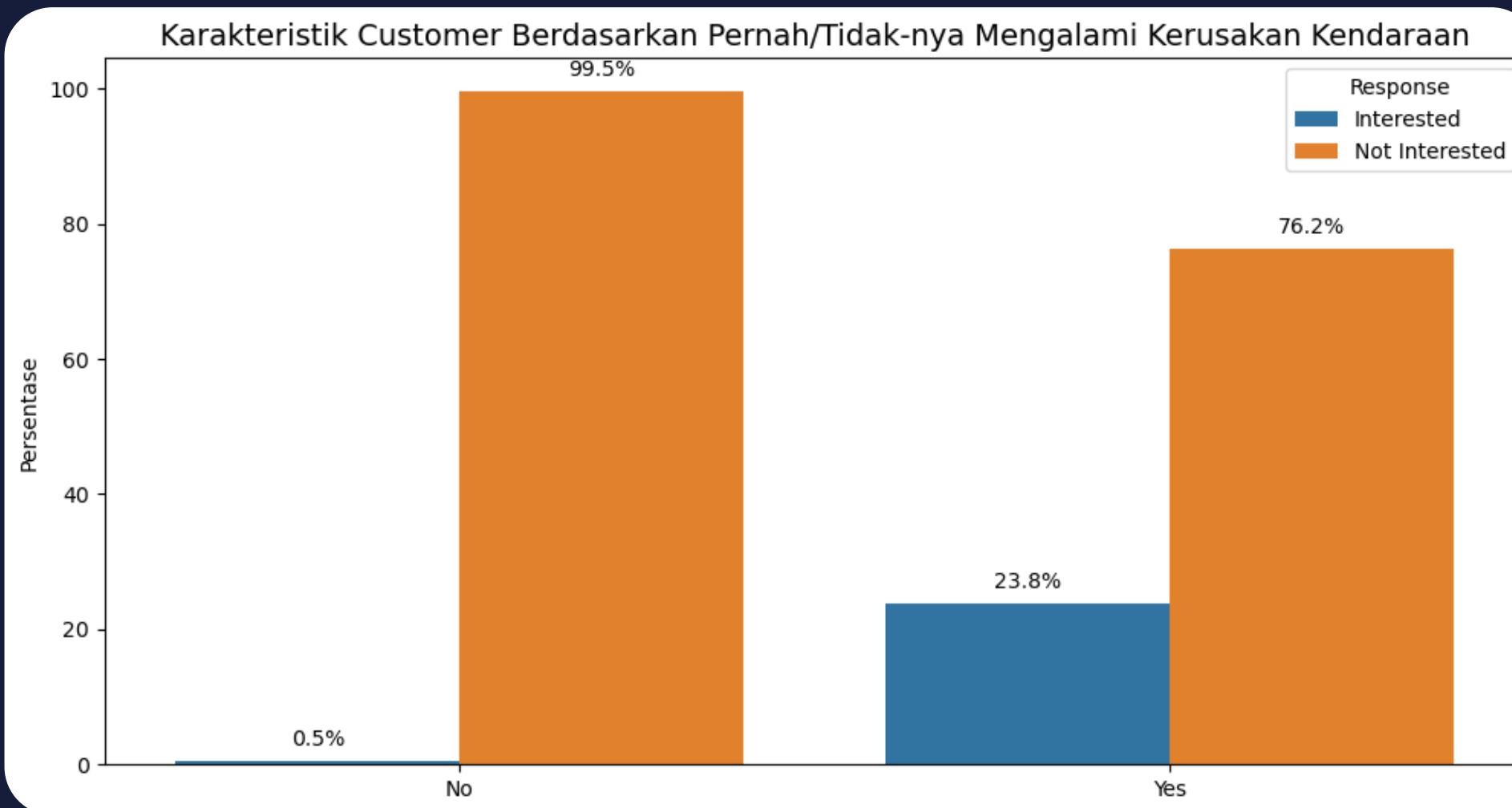
Kelompok umur Middle-Age Adults baik perempuan maupun laki-laki memiliki tingkat ketertarikan tertinggi terhadap asuransi kendaraan, yaitu masing-masing sebesar 17.3% dan 18.7%

*Notes:

Pembagian kelompok umur (WHO, 2013)

- 20-44 = Young Adults
- 45-64 = Middle-Age Adults
- > 65 = Older Adults or Seniors

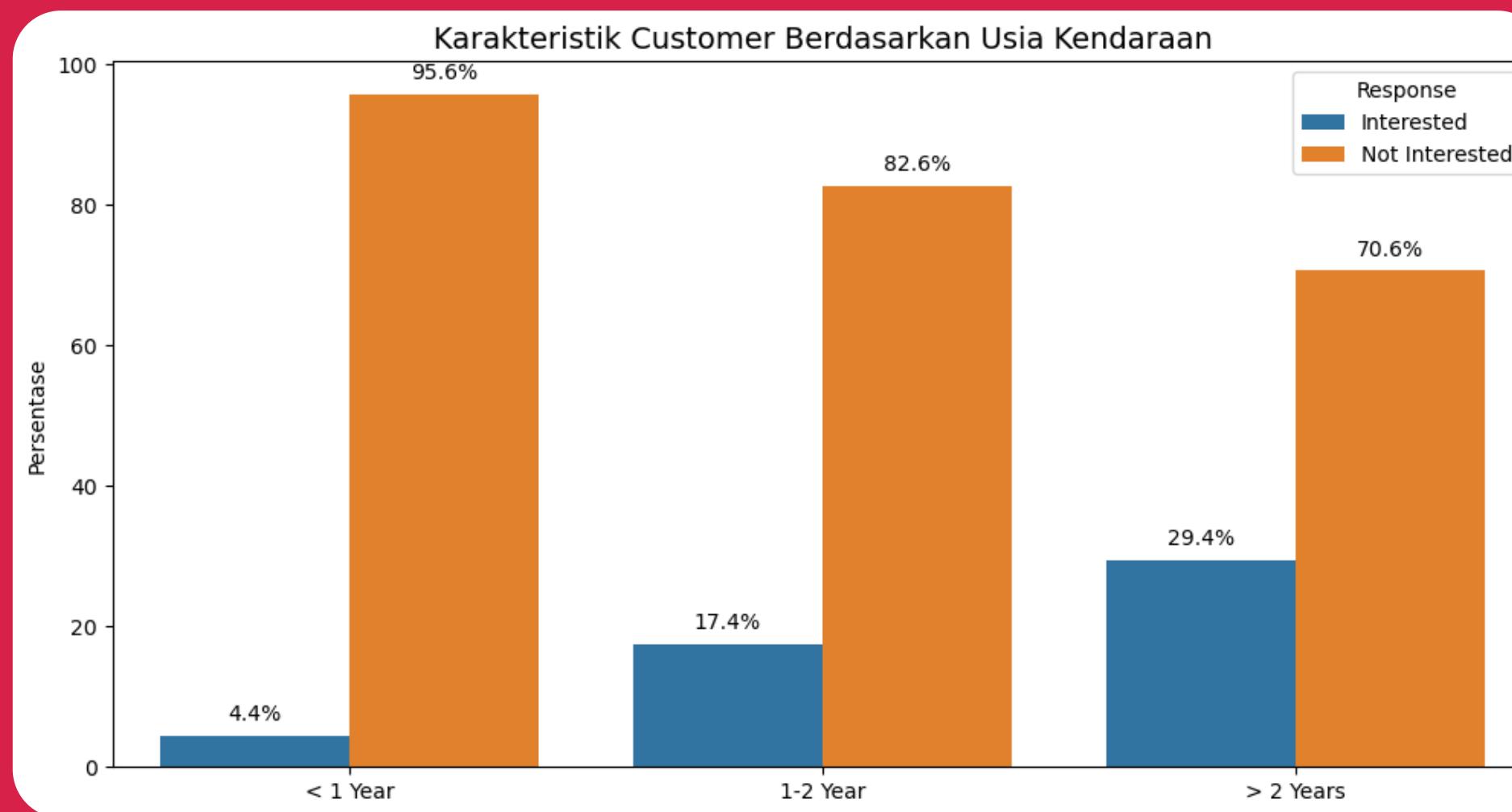
Vehicle Damage and Response



Hubungan antara kerusakan kendaraan nasabah dan Ketertarikan Terhadap Asuransi Kendaraan

Hampir keseluruhan nasabah yang tertarik terhadap asuransi kendaraan, pernah mengalami kerusakan terhadap kendaraannya di masa lalu.

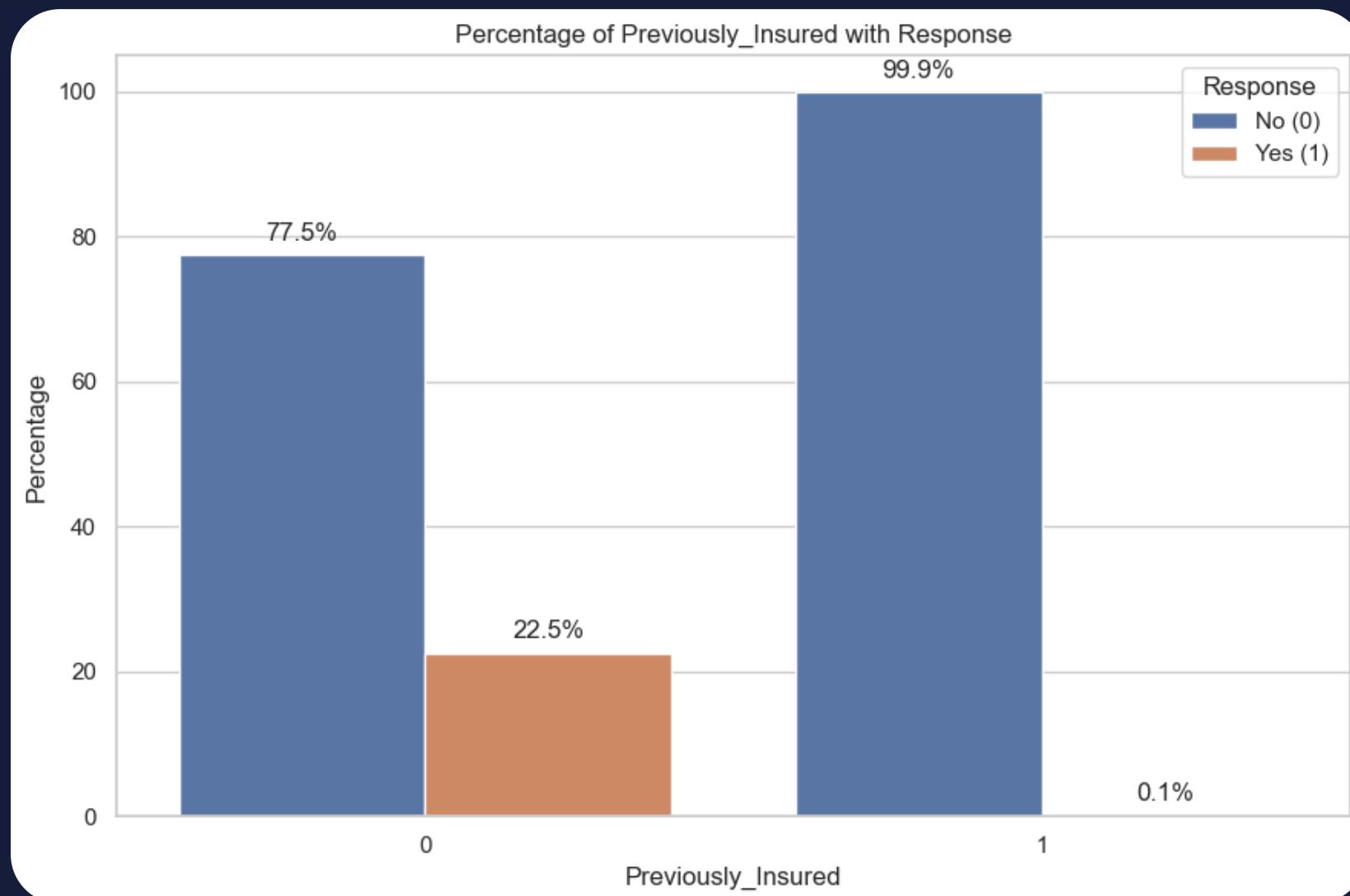
Vehicle Age and Response



Hubungan antara Umur Kendaraan dan Ketertarikan Terhadap Asuransi Kendaraan

Pemilik kendaraan dengan usia lebih dari 2 tahun memiliki tingkat ketertarikan asuransi tertinggi, mencapai 29.4%. Sementara itu, kendaraan berusia 1-2 tahun menduduki posisi kedua dengan tingkat ketertarikan sebesar 17.4%, sementara kendaraan di bawah 1 tahun menunjukkan ketertarikan sebesar 4.4%.

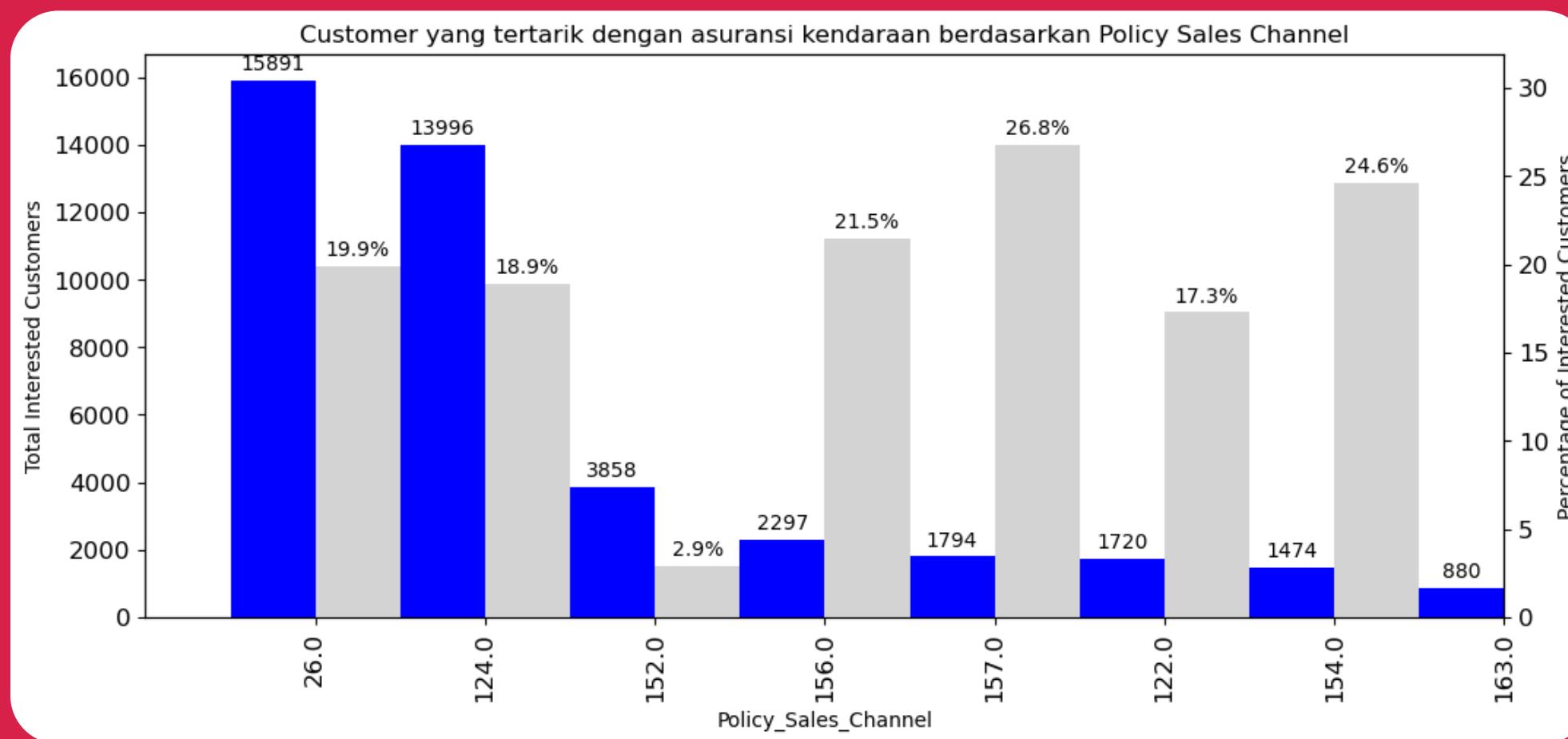
Previously Insured and Response



Hubungan antara Pernah Berlangganan Asuransi Sebelumnya dan Ketertarikan Terhadap Asuransi Kendaraan

Nasabah yang belum pernah berlangganan asuransi kendaraan sebelumnya memiliki ketertarikan terhadap asuransi kendaraan lebih tinggi yaitu mencapai 22.5%. Sedangkan hampir keseluruhan nasabah yang sudah pernah berlangganan asuransi kendaraan, tidak tertarik lagi dengan pembelian asuransi kendaraan.

Policy Sales and Response

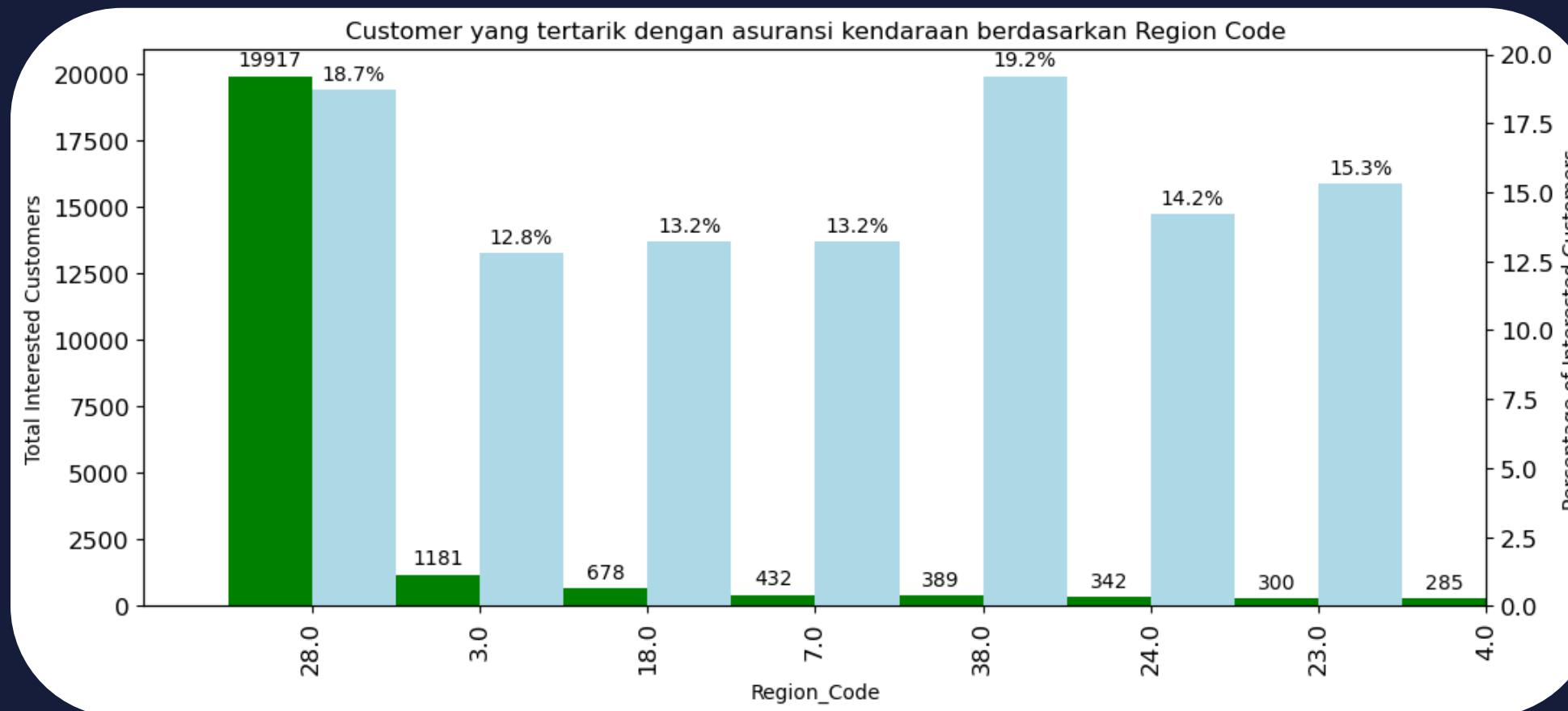


Hubungan antara Saluran pemasaran Terhadap Asuransi Kendaraan

Channel yang paling banyak memperoleh customer adalah channel 26 dan 124.

Adapun dari segi persentase, channel 157 memiliki persentase tertinggi yaitu 26.8%

Region Code and Response



Hubungan antara Kode Wilayah Terhadap Asuransi Kendaraan

Sebagian besar customer yang tertarik dengan asuransi kendaraan berasal dari Region Code 28 yaitu sebanyak 19.917 customer

Dari Segi persentase Region Code 38 memiliki tingkat persentase tertinggi yaitu mencapai 19.2%

Table Of Content

 Project Background

 EDA & Insight

 Pre-Processing

 Modeling

 Recommendations



Data Cleansing

Handle Missing And Duplicate

Tidak ada missing value dan duplikat data pada dataset

Handle Outlier

Handling outlier hanya dilakukan pada fitur 'Annual_Premium' menggunakan IQR karena fitur lain tidak memiliki outlier

Feature Transformation

Pada fitur numerik dilakukan standard scaling untuk membuat distribusi lebih simetrik

Feature Encoding

Feature encoding dilakukan terhadap fitur kategorikal supaya menjadi fitur numerik dengan label encoding

Handle Class Imbalance

Menggunakan undersampling dengan mengurangi jumlah sampel dari kelas mayoritas

Feature Extraction

Age Category

Fitur ini hasil dari pengelompokan usia menjadi beberapa kategori berdasarkan aturan berikut ini:

- Pengemudi Muda (16-24 tahun)
- Pengemudi Berpengalaman (25-39 tahun)
- Pengemudi Usia Tengah (40-64 tahun)
- Pengemudi Senior (65+ tahun)



Feature Extraction

Vehicle Damage and Age Interaction

Fitur ini hasil perkalian dari usia pelanggan dan kerusakan kendaraan.

Premium to Age Ratio

Fitur ini hasil dari pembagian premi tahunan (Annual Premium) dan usia (Age) pelanggan.



Feature Selection

01 Pearson Correlation

- Menghilangkan fitur redundan yang saling terkait ($\text{korelasi} > 0.85$)
- Menghapus fitur yang tidak memiliki pengaruh yang signifikan terhadap variabel target ($\text{korelasi mendekati nol}$).

02 Mutual Information

- $\text{Mutual Information} < 0.02$: Tidak dapat digunakan untuk prediksi.
- $0.02 \leq \text{Mutual Information} < 0.1$: Prediktor lemah.
- $0.1 \leq \text{Mutual Information} < 0.3$: Prediktor sedang.
- $0.3 \leq \text{Mutual Information} < 0.5$: Prediktor kuat.

Feature Selection

Berikut fitur-fitur terpilih dari seleksi pearson correlation dan mutual information. Fitur-fitur ini akan dipakai untuk modeling

Vehicle Age

Vehicle Damage

Region Code

Age Category

Previously Insured

Policy Sales Channel

Premi to Age Rasio

Vehicle Damage Age Interaction



Table Of Content



Project Background



EDA & Insight



Pre-Processing



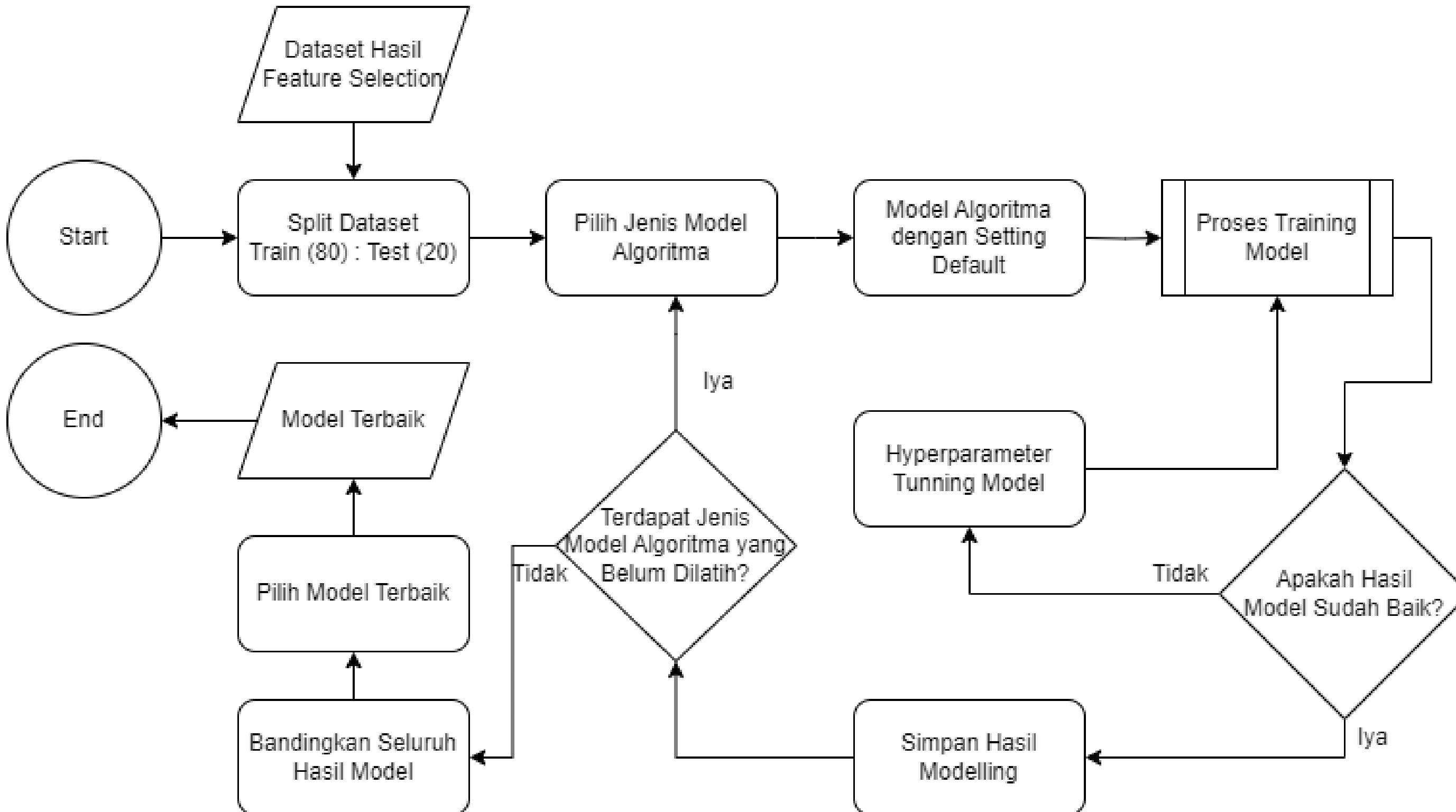
Modeling



Recommendations



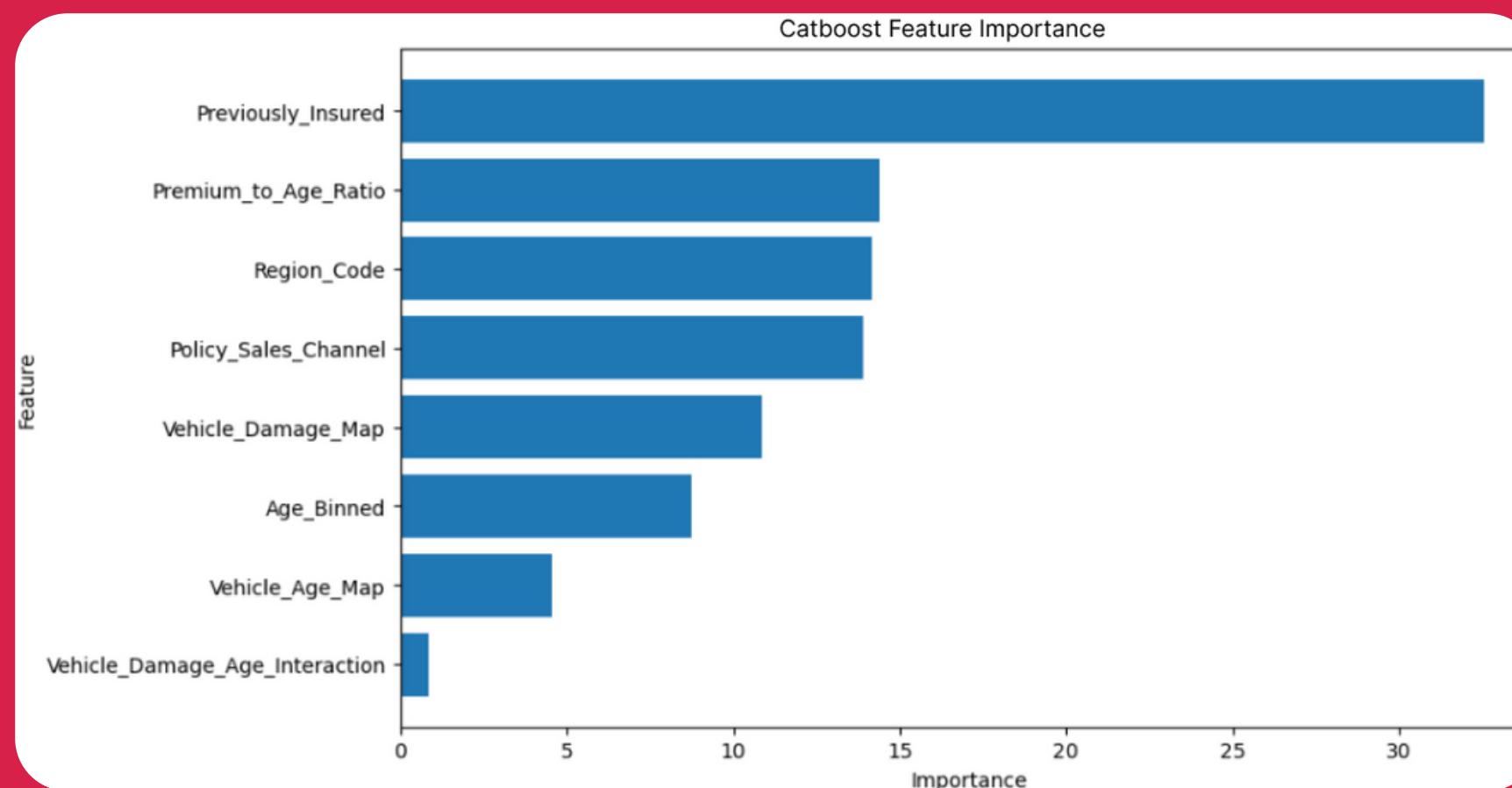
Flowchart Modeling



Model Comparison

Metric	XGboost	Random Forst	Adaboost	LightGBM	CatBoost	KNN
Accuracy	0.8	0.8	0.8	0.8	0.8	0.8
Precision	0.73	0.73	0.73	0.73	0.74	0.73
Recall	0.94	0.93	0.95	0.94	0.93	0.94
F1 Score	0.82	0.82	0.82	0.82	0.82	0.82
Mean ROC-AUC (Train)	0.84	0.84	0.84	0.84	0.85	0.84
Std ROC-AUC (Train)	0.01	0	0.01	0.01	0	0
Mean ROC-AUC (Test)	0.85	0.85	0.85	0.85	0.84	0.65
Std ROC-AUC (Test)	0.01	0	0.01	0.01	0.01	0

Feature Importance



Pengaruh tiap fitur pada ketertarikan nasabah terhadap asuransi kendaraan

Fitur Previously Insured merupakan fitur yang paling berpengaruh pada ketertarikan nasabah terhadap asuransi kendaraan. Hal ini menandakan dominan nasabah yang belum pernah berlangganan memiliki tingkat ketertarikan yang tinggi terhadap asuransi kendaraan.

Table Of Content

 Project Background

 EDA & Insight

 Pre-Processing

 Modeling

 Recommendations



Business Flow

Data Customer

Data customer yang diambil dari database perusahaan

Nasabah Potensial

Calon nasabah yang terprediksi tertarik dengan asuransi kendaraan

Peserta Polis

Peserta polis didapatkan dengan conversion rate yang lebih tinggi

Model Prediksi

Model melakukan prediksi siapa saja nasabah yang tertarik.

Penawaran

Hanya melakukan penawaran kepada nasabah potensial

Priority Customer Criteria

Previously Insured

Prioritaskan customer yang belum pernah berlangganan asuransi kendaraan sebelumnya.

Premium To Age Ratio

Prioritaskan customer dengan nilai premium to age ratio bertipe medium yaitu dari range 2.5 hingga 3.0 karena memiliki ketertarikan dengan suransi kendaraan yang tinggi.

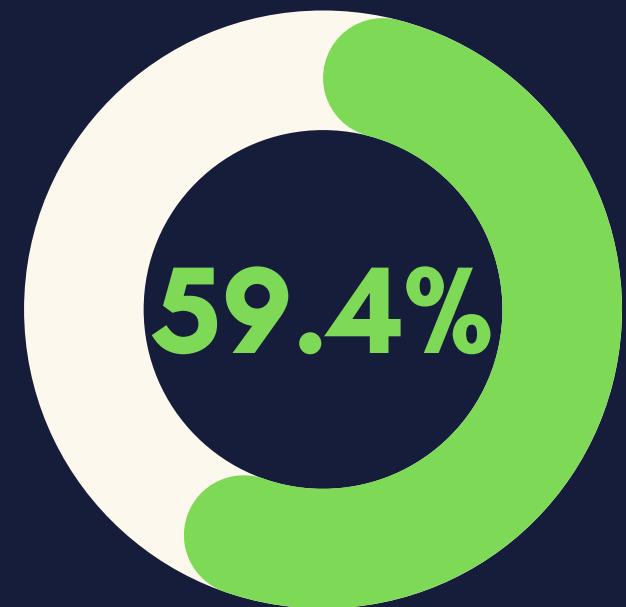
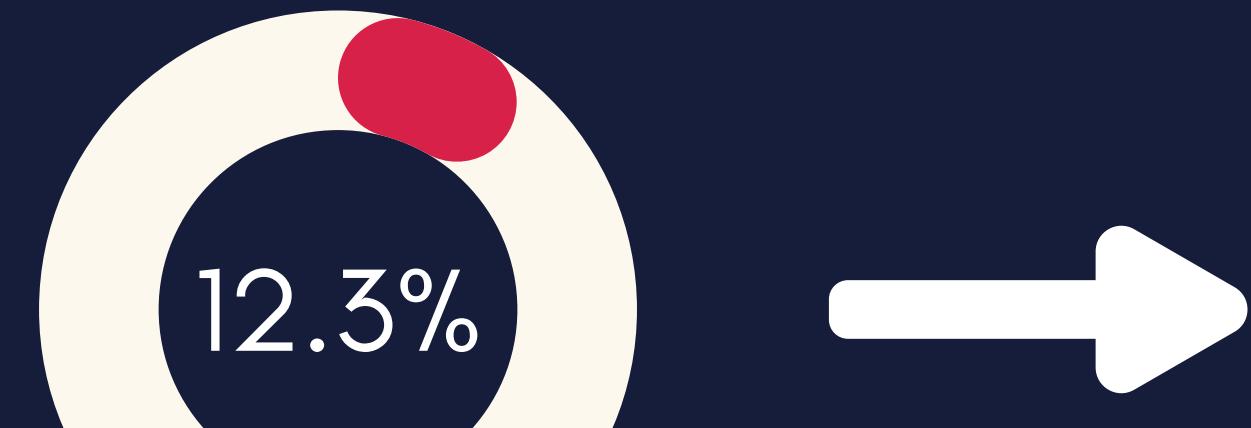
Region Code

Prioritaskan customer dari Region Code 28 karena sebagian besar customer yang tertarik dengan asuransi kendaraan berasal dari region code tersebut.

Simulation

Conversion Rate Setelah Model

- Simulasi perhitungan conversion rate setelah menggunakan model dilakukan menggunakan data test hasil split
- Dengan menggunakan model prediksi untuk melakukan penawaran conversion rate naik 47.1% menjadi 59.4%



Conversion rate naik 47.1%

Marketing Cost

No	Keterangan	Jumlah	Biaya	Total
1	Gaji Pegawai	38	2.500.000	95.000.000
2	Customer Service	15	1.250.000	18.750.000
3	Panggilan Telepon	381.109	120	45.733.080
4	Perlengkapan	1	3.000.000	3.000.000
5	Biaya Operasional	1	2.000.000	2.000.000
Total Biaya Marketing				164.483.080

Keterangan

Jumlah target nasabah untuk marketing diasumsikan sesuai dengan banyaknya data train yaitu 381.109 nasabah

Cost Simulation

Sebelum Modeling

Total Marketing Cost	Jumlah User Baru	Customer Acquisition Cost
Rp164.483.080	46.876	Rp3.508



Sesudah Modeling

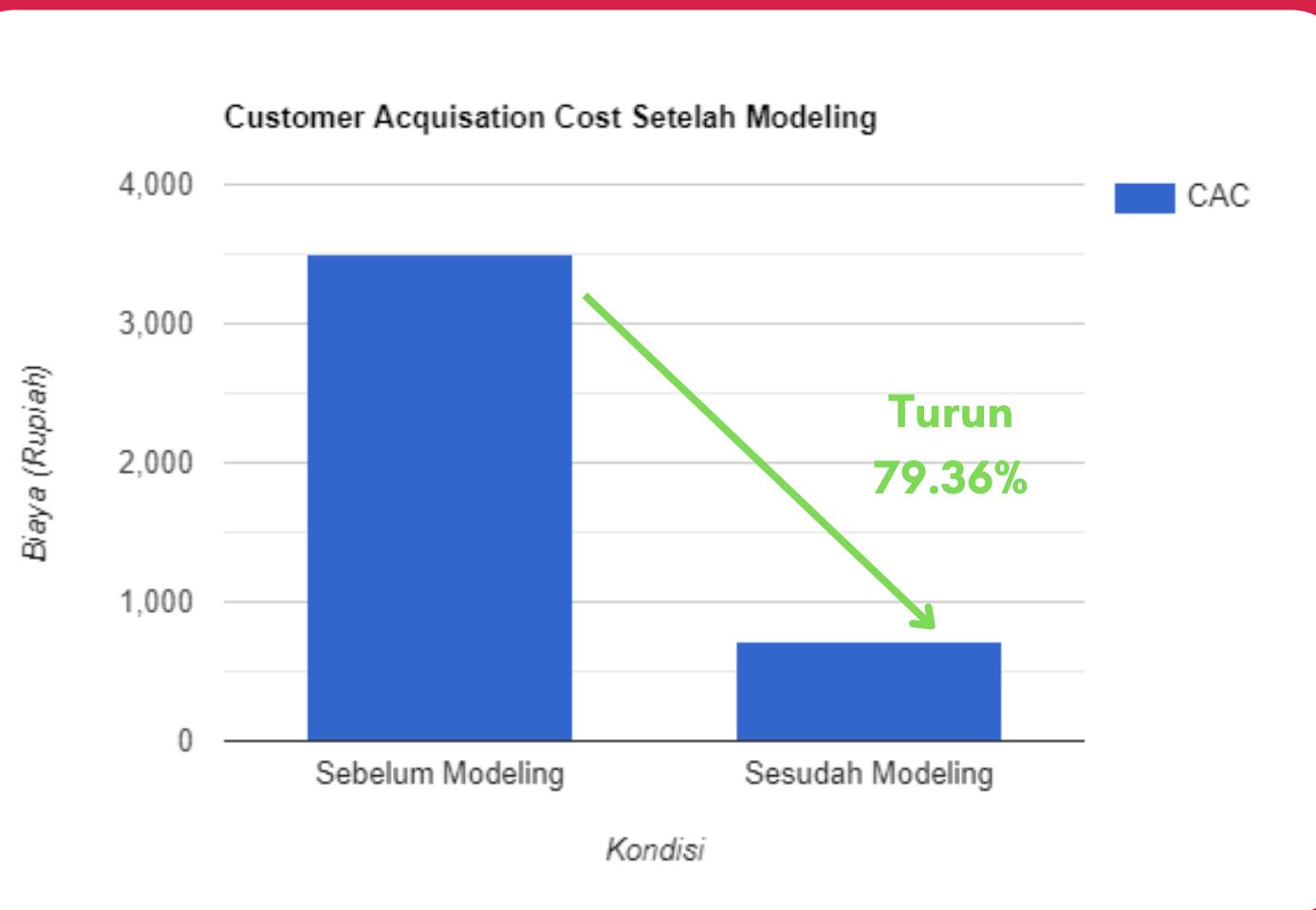
Total Marketing Cost	Jumlah User Baru	Customer Acquisition Cost
Rp164.483.080	227.141	Rp724

Biaya Premi	Jumlah User Baru	Revenue
Rp950.000	46.876	Rp44.532.200.000

Biaya Premi	Jumlah User Baru	Revenue
Rp950.000	227.141	Rp215.783.950.000

Cost Simulation

Customer Acquisition Cost



Revenue

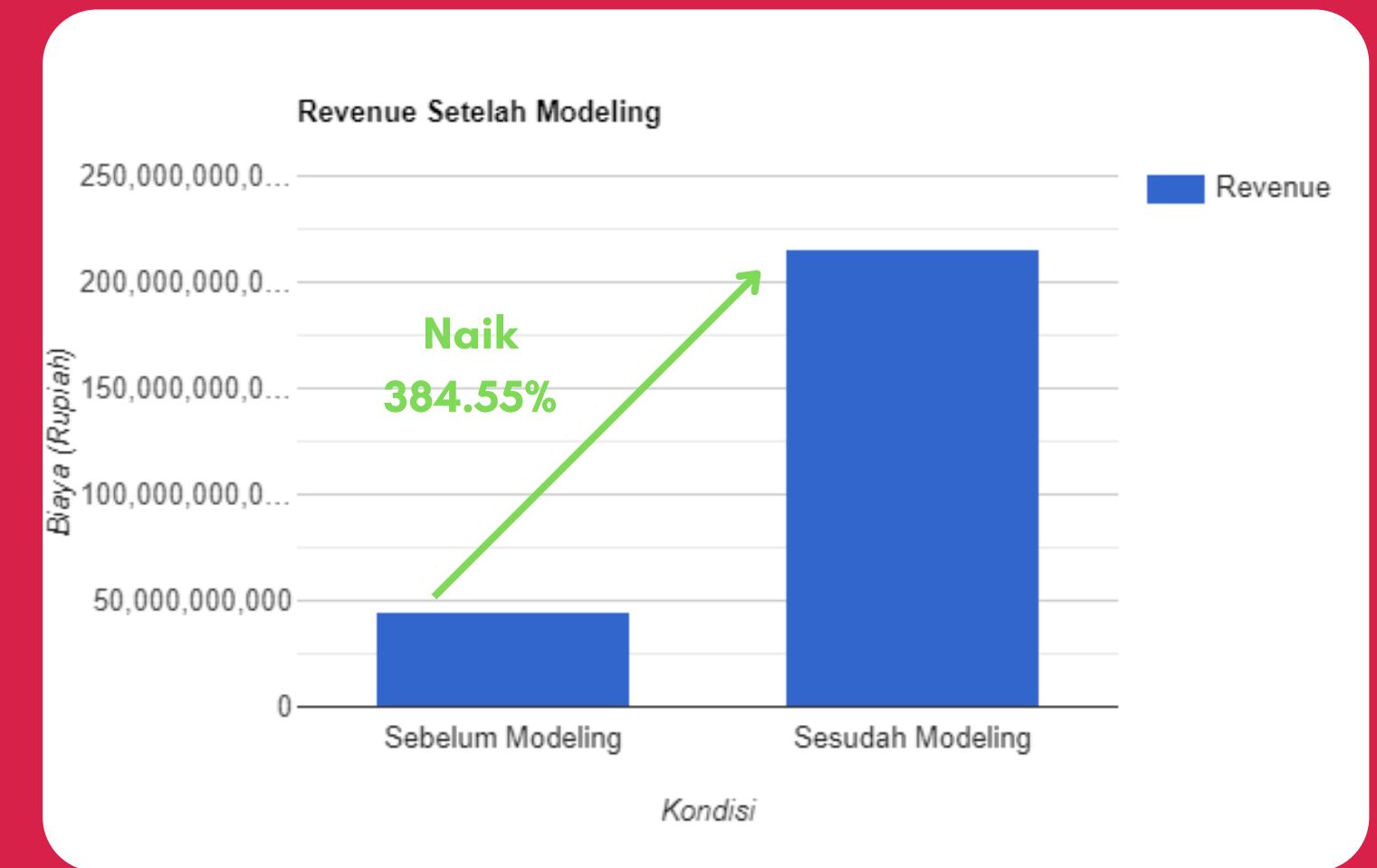


Table Of Content

 Project Background

 EDA & Insight

 Pre-Processing

 Modeling

 Recommendations



Presentation by Team Data 200



Thank you!

Final Project **Data 200**