



Rakamin Final Project

Uno Inc.

POTENTIAL CUSTOMER ANALYSIS AND PREDICTION

Analysis Report



MEET THE TEAM

Uno Inc.



**Devriansyah
Sya'ban**

DATA SCIENCE
LEAD



Vira Diana

DATA ANALYST



Adinda Dita

DATA ANALYST



Kun Aggiar

STATISTICIAN



Rahmadani A

STATISTICIAN



Laurenzius Julio

ML ENGINEER



**Farih Afdhalul
Ihsan**

ML ENGINEER



Dignu Akbar

DATA SCIENTIST

Who We Are?

Konsultan yang bergerak di bidang data untuk membantu dalam melakukan analisis permasalahan perusahaan Anda.

OUTLINE

-  **PROBLEM STATEMENT**
-  **DATA OVERVIEW & INSIGHT**
-  **DATA PRE-PROCESSING**
-  **MODEL & EVALUATION**
-  **BUSINESS RECOMMENDATION & IMPACT**

OUTLINE



PROBLEM STATEMENT



DATA OVERVIEW & INSIGHT



DATA PRE-PROCESSING



MODEL & EVALUATION



BUSINESS RECOMMENDATION & IMPACT

Business Understanding

Perusahaan X merupakan perusahaan yang bergerak dibidang **Tour and Travel**

Guna **meningkatkan profit** perusahaan, selain memiliki produk utama berupa layanan/jasa travel, Perusahaan X juga mengambil peluang kesempatan dengan **menambahkan segmen bisnis asuransi (*travel insurance*)** yang berupa penyediaan layanan asuransi (mencakup perlindungan terhadap covid 19) bagi para *customer travel* nya.



Uno Inc.

Dalam melakukan campaign terhadap produk asuransinya,
**Perusahaan X selalu memberikan campaign menyeluru
kehada seluruh customernya**

Diasumsikan bahwa ada biaya yang perlu dikeluarkan oleh perusahaan pada setiap aksi campaign (jenis apapun dan cara apapun) yang diberikan / dilakukan terhadap customer



Problem Statement



Campaign Cost Tidak Efisien dikarenakan campaign produk asuransi diberikan **secara menyeluruh** ke semua customer



Conversion Rate Rendah untuk product travel insurance yang ditawarkan kepada para customer



GOALS



Meningkatkan
conversion rate



Optimasi campaign cost

Objectives

Uno Inc.



Membangun model machine learning yang bisa memprediksi customer potensial untuk membeli paket travel insurance

Menemukan faktor penting dan karakteristik utama customer yang berpotensi membeli paket travel insurance

Memberikan insight dan rekomendasi untuk menurunkan kemungkinan perusahaan merugi dalam campaign berdasarkan hasil yang didapatkan dari proses pengolahan data yang dimiliki oleh perusahaan

Business Metrics



Conversion Rate



Revenue

Budget Actual Difference



SALE REPORT



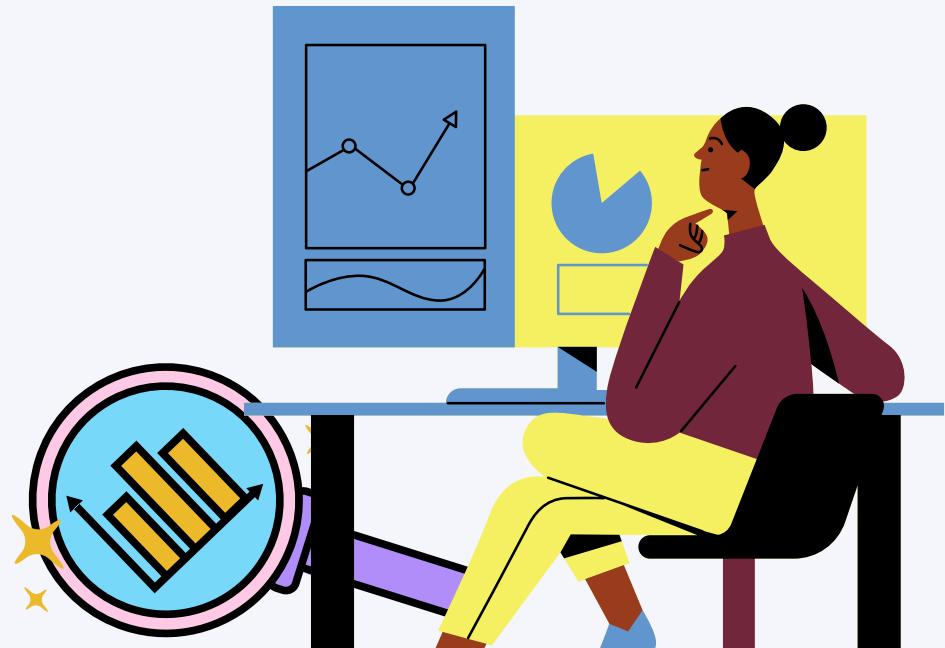
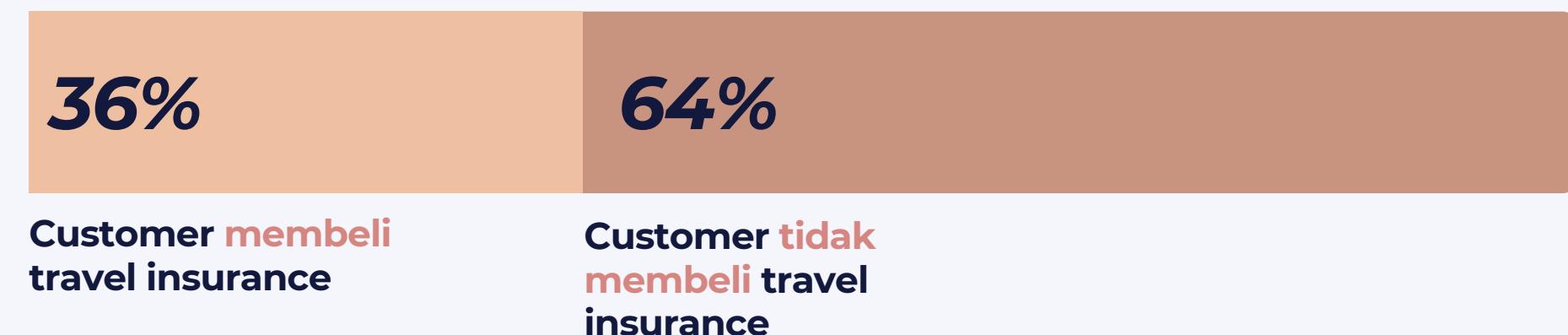
OUTLINE

-  **PROBLEM STATEMENT**
-  **DATA OVERVIEW & INSIGHT**
-  **DATA PRE-PROCESSING**
-  **MODEL & EVALUATION**
-  **BUSINESS RECOMMENDATION & IMPACT**

Data Overview

- 1 Dataset merupakan data penjualan travel insurance sejak 2019
- 2 Dataset terdiri dari **1.986 baris, 9 kolom feature, dan 1 kolom target**
- 3 Terdapat **6 data numerik** dan **4 data kategorikal**
- 4 **Tidak terdapat** data null atau kosong

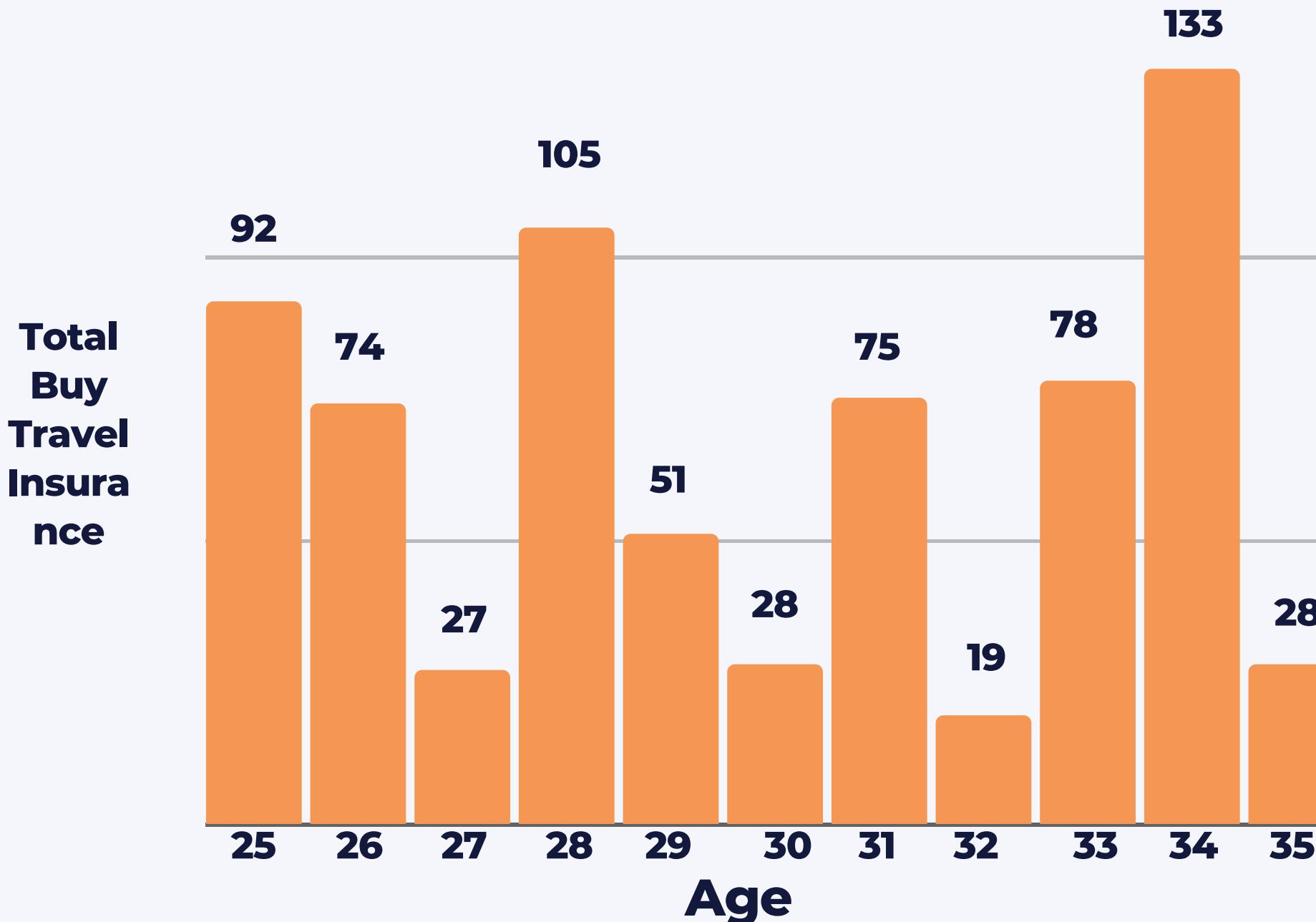
Target Feature



Variabel	Definisi
Age	umur customer
Employment Type	apakah pelanggan bekerja di sektor swasta atau pemerintahan
GraduateOrNot	apakah pelanggan adalah lulusan perguruan tinggi atau tidak
Annual Income	pendapatan tahunan customer
FamilyMembers	Jumlah anggota keluarga yang tinggal bersama customer
ChronicDiseases	apakah customer menderita penyakit kronis atau kondisi seperti diabetes / bp tinggi atau asma, dll.
FrequentFlyer	apakah customer sering menggunakan layanan jasa penerbangan
EverTravelledAbroad	apakah customer pernah berpergian ke luar negeri
TravellInsurance	apakah customer membeli asuransi perjalanan atau tidak

Data Insight

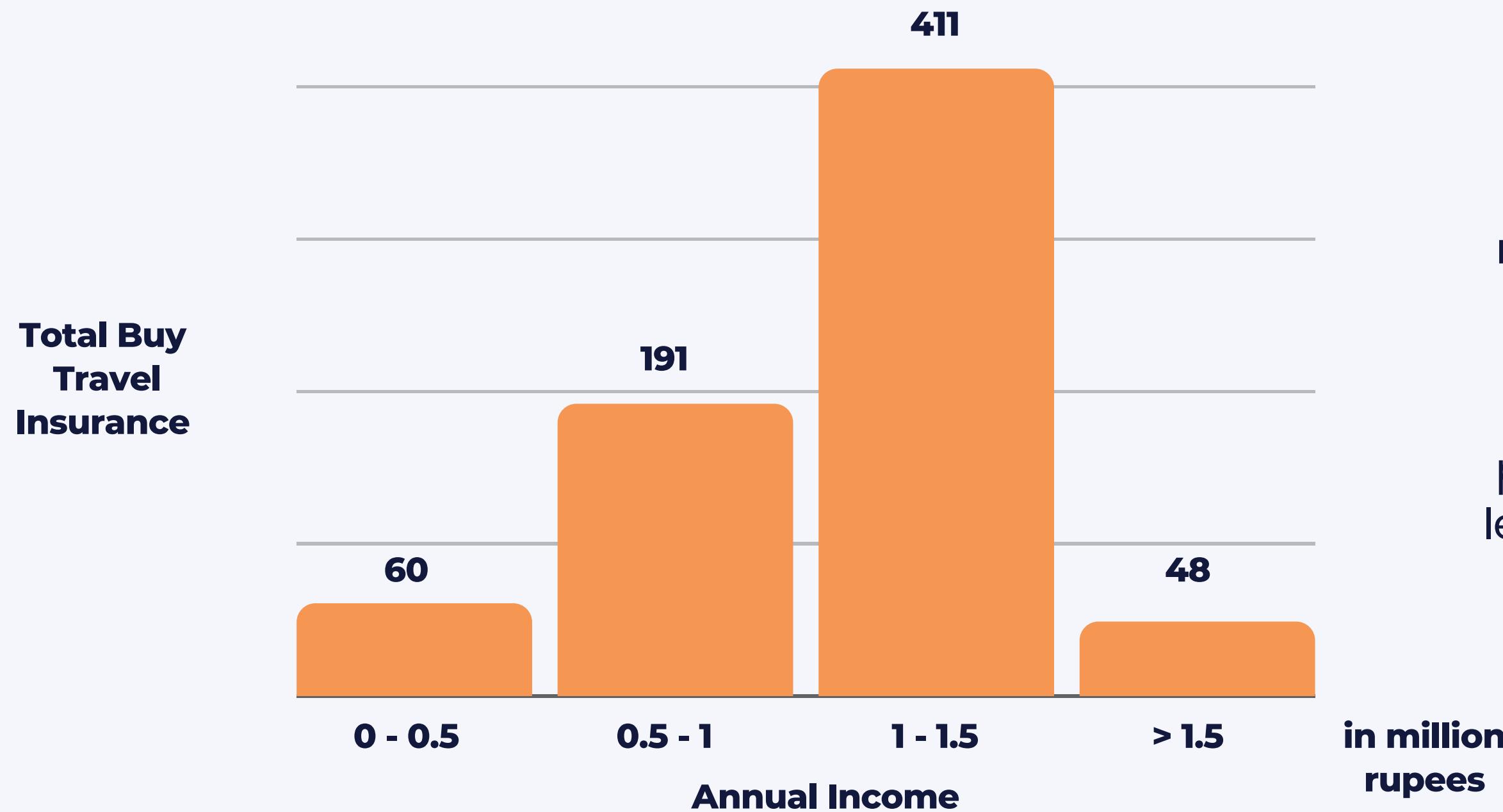
Pengaruh **Age** terhadap Pembelian Travel Insurance



Lima usia tertinggi customer yang memilih untuk membeli paket travel Insurance adalah usia 25, 28, 31, 33, dan 34 dengan pembeli paket Travel Insurance tertinggi adalah customer berusia 34 tahun.

Data Insight

Pengaruh **Annual Income** terhadap Pembelian Travel Insurance

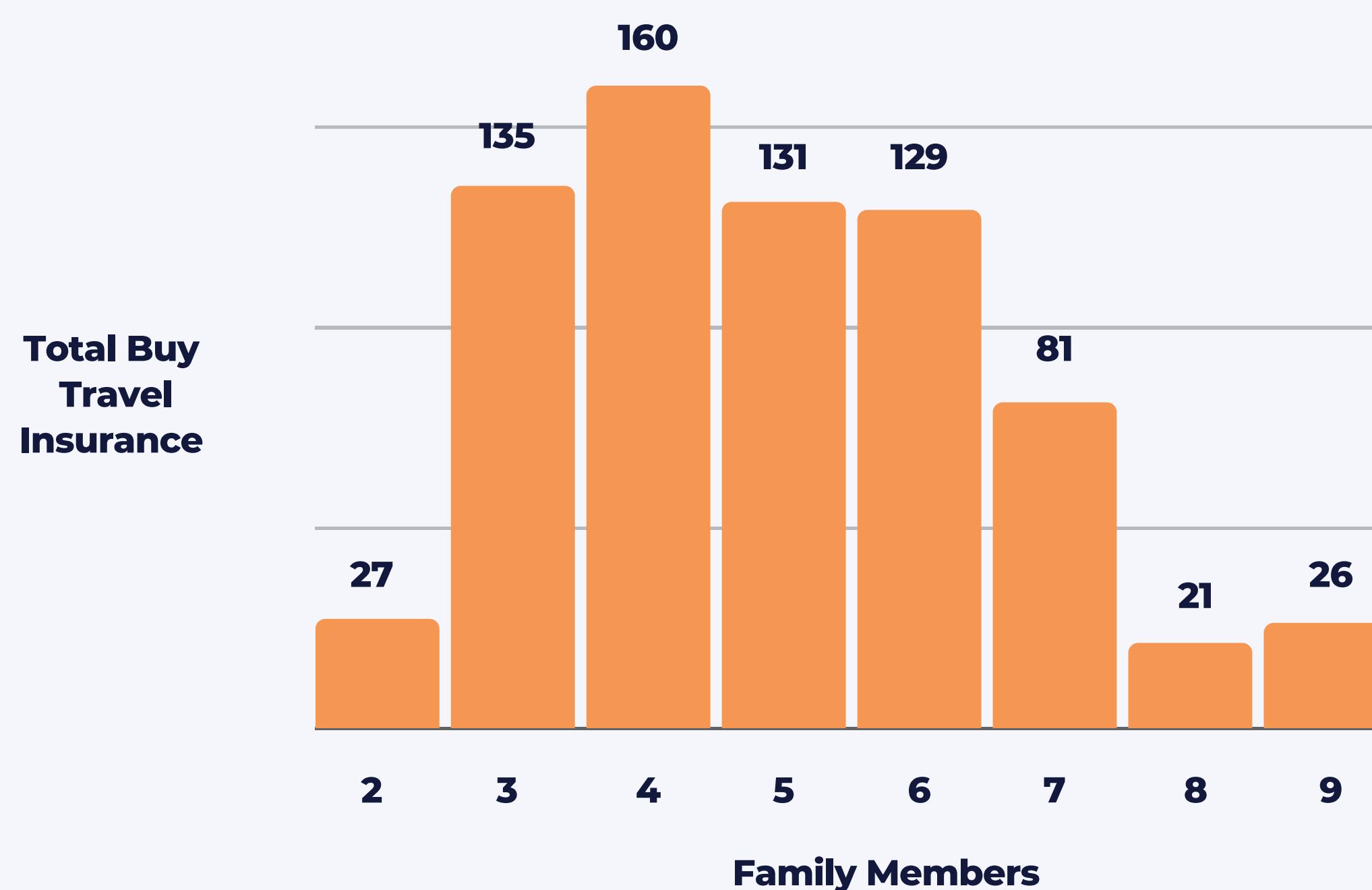


Customer dengan annual income di atas 1.000.000 rupees cenderung membeli paket travel insurance.

Artinya, customer dengan pendapatan tahunan tinggi lebih mampu untuk membeli travel insurance.

Data Insight

Pengaruh **Family Member** terhadap Pembelian Travel Insurance



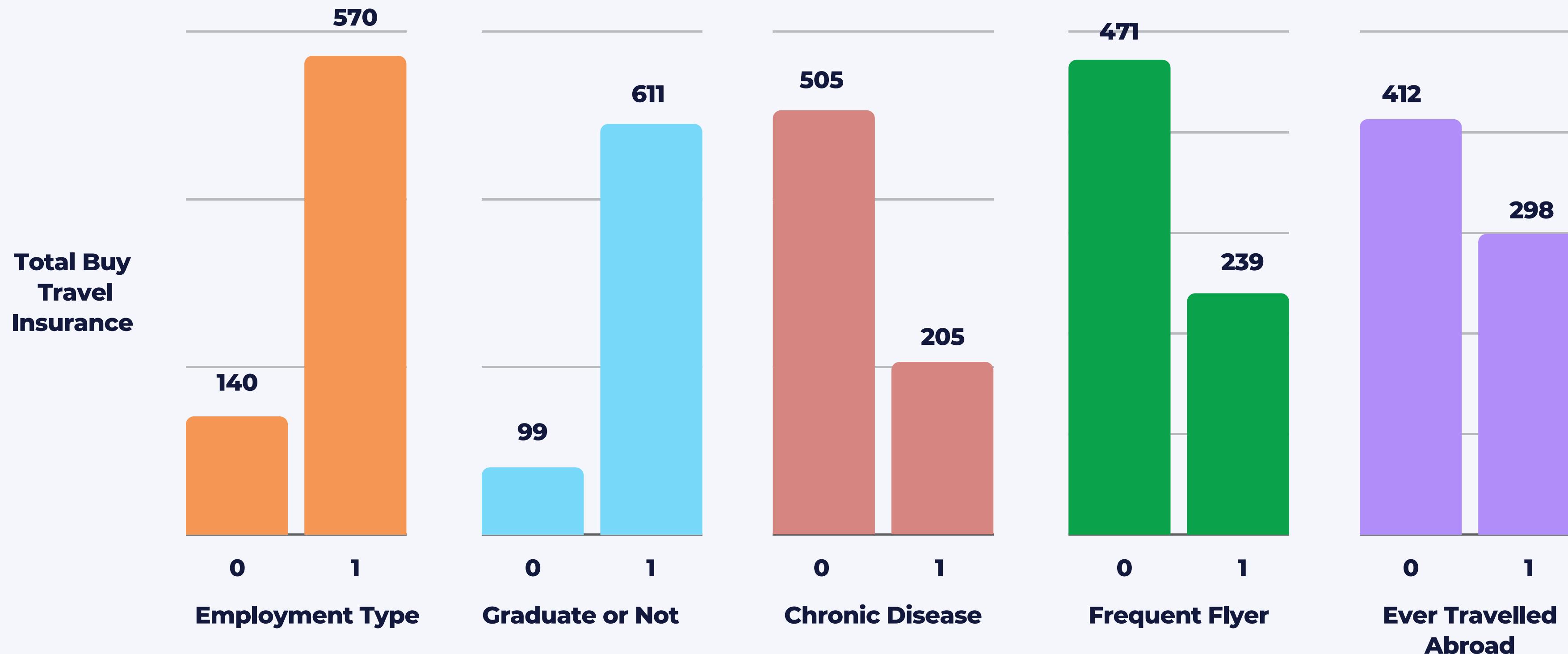
Mayoritas customer yang membeli travel insurance memiliki jumlah anggota keluarga 4 orang.

Dapat disimpulkan bahwa jumlah anggota keluarga tidak terlalu berpengaruh terhadap travel insurance.

Customer dengan 8 hingga 9 keluarga tidak membeli banyak travel insurance dibandingkan keluarga lainnya

Data Insight

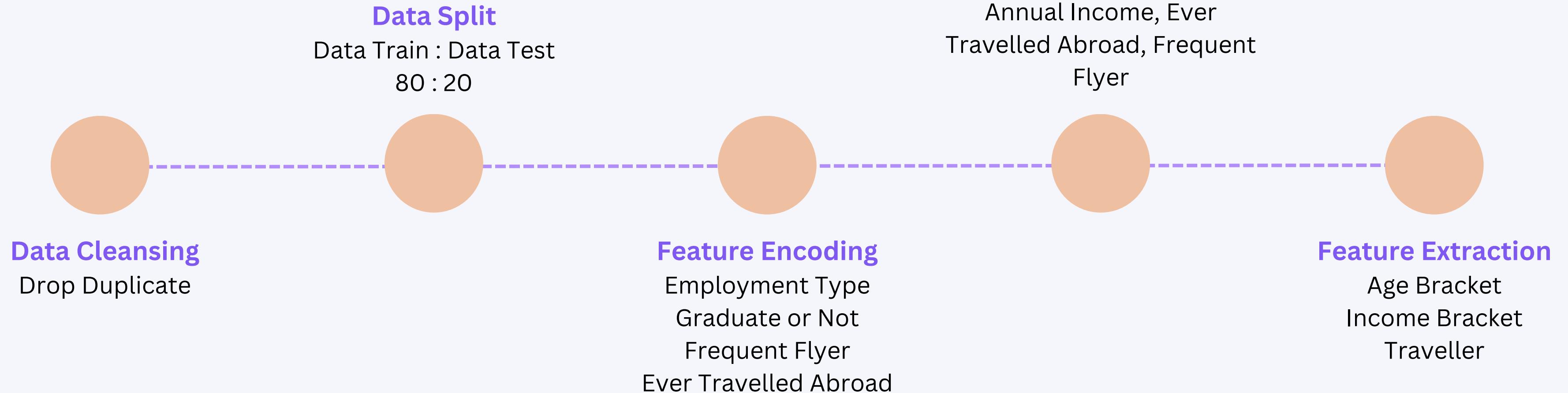
Pengaruh **Feature Lain** terhadap Pembelian Travel Insurance



OUTLINE

-  **PROBLEM STATEMENT**
-  **DATA OVERVIEW & INSIGHT**
-  **DATA PRE-PROCESSING**
-  **MODEL & EVALUATION**
-  **BUSINESS RECOMMENDATION & IMPACT**

Data Pre-Processing.



OUTLINE

-  **PROBLEM STATEMENT**
-  **DATA OVERVIEW & INSIGHT**
-  **DATA PRE-PROCESSING**
-  **MODEL & EVALUATION**
-  **BUSINESS RECOMMENDATION & IMPACT**

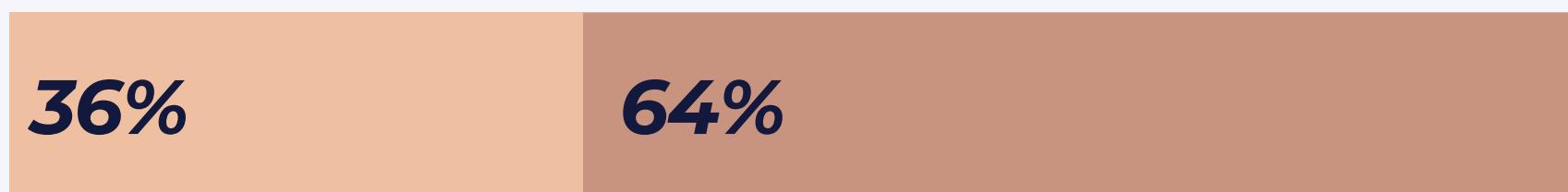
Model Evaluation.

PREDIKSI

AKTUAL

	POSITIVE	NEGATIVE
POSITIVE	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
NEGATIVE	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

Distribusi data target



Customer membeli travel insurance

Customer tidak membeli travel insurance

$$Precision = \frac{TP}{TP + FP}$$

Memperhitungkan nilai **Precision** untuk mengurangi angka **False Positive** (prediksi setuju, tetapi kenyataan menolak) yaitu memperkecil kemungkinan munculnya prediksi customer membeli insurance travel sedangkan data actual menunjukkan customer tidak membeli insurance travel. Hal ini dilakukan untuk memperkecil cost marketing.

Result Modeling.

DATA TRAIN RESULT

MODEL	PRECISION
LOGISTIC REGRESSION	0.84
KNN	0.51
DECISION TREE	0.85
RANDOM FOREST	0.85
ADABOOST	0.84

Berdasarkan hasil modeling, decision tree dan random forest menunjukkan hasil precision terbaik sebesar 85%

Di antara algoritma decision tree dan random forest tersebut, dipilih algoritma **Decision Tree** sebagai algoritma terbaik dengan pertimbangan bahwa algoritma decision tree lebih sederhana, interpretable, dan mudah dipahami dibandingkan dengan algoritma random forest.



Result Modeling.

Kasus	Precision (Train)	Precision (Test)
Df1	0.81	0.55
Df2	0.77	0.59
DF3	0.79	0.72
DF4	0.85	0.89
DF5	0.78	0.78
DF6	0.84	0.78

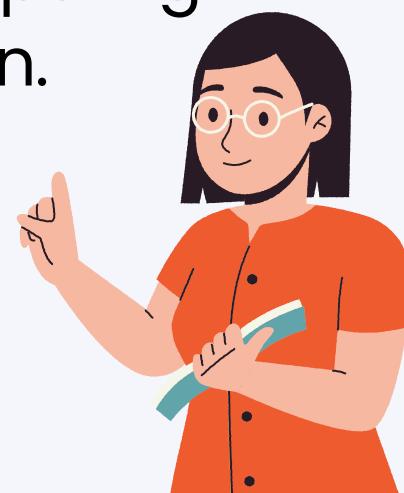
Pengecekan Overfitting dan Underfitting

Underfitting

Terlihat bahwa dari uji coba kasus tiap dataframe tidak memiliki nilai precision dan test yang sama - sama rendah.

Overfitting

Dataframe 1 dan 2 mengalami overfitting. Sedangkan dataframe 3 - 6 tidak overfit. Kami memilih memakai datagramme 4 karena paling *best fit* ketimbang yang lain.



Result Modeling.

DATA TEST RESULT

MODEL	PRECISION
DECISION TREE	0.89

Skor hasil pengujian model menggunakan data test menghasilkan **Precision Score 89%**



Rakamin Final Project

CONFUSION MATRIX DECISION TREE

AKTUAL	PREDICT BUY TRAVEL INSURANCE	PREDICT NOT BUY TRAVEL INSURANCE
	ACTUAL BUY TRAVEL INSURANCE	ACTUAL NOT BUY TRAVEL INSURANCE
ACTUAL BUY TRAVEL INSURANCE	224	7
ACTUAL NOT BUY TRAVEL INSURANCE	89	55

PREDIKSI

OUTLINE

-  **PROBLEM STATEMENT**
-  **DATA OVERVIEW & INSIGHT**
-  **DATA PRE-PROCESSING**
-  **MODEL & EVALUATION**
-  **BUSINESS RECOMMENDATION & IMPACT**

Business Recomendation

Dengan data yang telah didapatkan dari hasil modeling dan disertai dengan data-data *external* yang berkaitan dengan bisnis *travel insurance*. Sebagai data science consultant dapat merekomendasikan serta dapat memberi beberapa insight terkait dengan *travel insurance* yang akan launching.

Rekomendasi & Impact.

1. Tim *marketing* hendaknya dapat menyesuaikan harga *travel insurance* yang akan diluncurkan agar lebih *affordable* karena dari data yang didapatkan menunjukan adanya kecenderungan orang-orang yang memiliki *income* lebih tinggi untuk membeli *travel insurance*.
2. Membuat *campaign* yang bisa mengedukasi dan lebih menarik perhatian bagi mereka yang berasal dari golongan pendapatan *low* agar tertarik untuk membeli *travel insurance*, karena terlihat dari data yang dimiliki bahwa jumlah data yang berasal dari golongan *low income* lebih banyak daripada mereka yang berasal dari golongan *high income*.
3. Diasumsikan bahwa *campaign* diluncurkan pada 2020 saat awal pandemi yang berarti ketakutan dari masyarakat terhadap covid19 masih cukup tinggi, sehingga launching dari *travel insurance* terbaru yang mengcover perlindungan terhadap covid19 bisa dilakukan secepatnya.
4. *Campaign* yang lebih "milenial" serta "gen z" friendly, diharapkan mereka yang berasal dari kategori umur tersebut akan jadi lebih tertarik untuk membeli *travel insurance* dengan *campaign* yang dibuat menyesuaikan umur mereka yang mana hal tersebut berarti mencakup hobby atau hal-hal yang menarik untuk mereka.
5. *Campaign* yang diluncurkan untuk product ini juga bisa dengan menyisipkan "rasa aman" untuk berpergian bahkan ketika pandemi sedang berlangsung karena kelebihan product tersebut yang juga termasuk perlindungan biaya perawatan dan pengobatan apabila customer terkena covid19 sehingga orang-orang yang berpergian akan merasa lebih aman.
6. Offering dapat dilakukan tim *marketing* hanya ke pelanggan potensial saja, yakni konsumen yang tidak membeli *travel insurance* tapi memiliki ciri – ciri seperti customer pembeli (*False Positif*) karena dari sudut pandang model customer ini memiliki kecenderungan bakal membeli *travel insurance* kedepannya.

Business Impact 1

Apabila diasumsikan ada 1000 data baru yang masuk dan diprediksi bahwa ada 500 orang customer potensial. Dengan akurasi model sebesar 89% berarti data yang diprediksi dikurangi 11% atau 55 orang sebagai margin error yang berarti ada 445 orang yang diprediksi pasti membeli travel insurance.

Simulasi:

Travel insurance price = \$5

Campaign cost per person = \$0.1

$445 \times 5 - 50 = \$2175$ (gross profit)

Jika ada tambahan dari false positif yang membeli travel insurance sebanyak:

1. 25% maka = $(445 + 13) \times 5 - 50 = \2240 (gross profit) (meningkat 2.98%)
2. 50% maka = $(445 + 26) \times 5 - 50 = \2305 (gross profit) (meningkat 5.97%)
3. 75% maka = $(445 + 39) \times 5 - 50 = \2370 (gross profit) (meningkat 8.96%)

Business Impact 2

PREDIKSI			
		PREDICT BUY TRAVEL INSURANCE	PREDICT NOT BUY TRAVEL INSURANCE
AKTUAL	ACTUAL BUY TRAVEL INSURANCE	224	7
	ACTUAL NOT BUY TRAVEL INSURANCE	89	55

Before

Campaign Cost = $375 \times \$0.1$ = \$ 37.5
Revenue = $224 \times \$1.5$ = \$ 336
ROAS = $\$ 336 / \$ 37.5$ = 8,96x

After

Campaign Cost = $313 \times \$0.1$ = \$ 31.3
Revenue = $224 \times \$1.5$ = \$ 336
ROAS = $\$ 336 / \$ 31.3$ = 10,7x

*Asumsi revenue yang diperoleh dari setiap customer yang melakukan pembelian = \$5

*Asumsi biaya yang dikeluarkan untuk melakukan campaign terhadap setiap customer = \$0.1

ROAS MENINGKAT 20%!!

Uno Inc.

THANK YOU!!!