

# TRAVEL INSURANCE BUSINESS RECOMMENDATION

---

A presentation by Andromeda



# Tim Andromeda

---



**Elisa Fitri**

Data Science Lead



**Pinandito Wahyu P**

Data Scientist



**Bagus Pamuncak N**

Data Scientist



**Vanny Bowo**

Data Scientist

# Latar Belakang

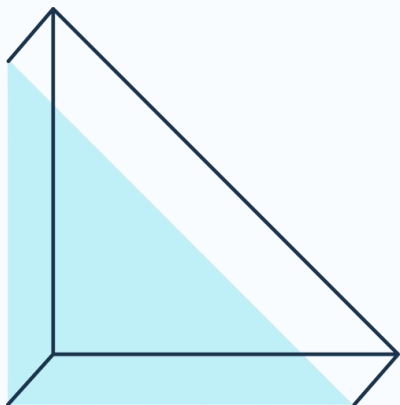


---

Client kami adalah suatu perusahaan asuransi yang memiliki sebuah produk bernama ***Travel Insurance***. Client meminta kami untuk mengambil insight dari database perusahaan dan memberikan rekomendasi bisnis serta model *machine learning* yang dapat mengoptimalkan *customer approach* dan meningkatkan penjualan produk.

## PROBLEM:

Pada tahun 2019, dari 1987 individu yang dihubungi oleh tim marketing, hanya 35% customer yang membeli produk Travel Insurance.



## Goal

Mengefisiensikan dan mengoptimalisasi *costumer acquisition*.

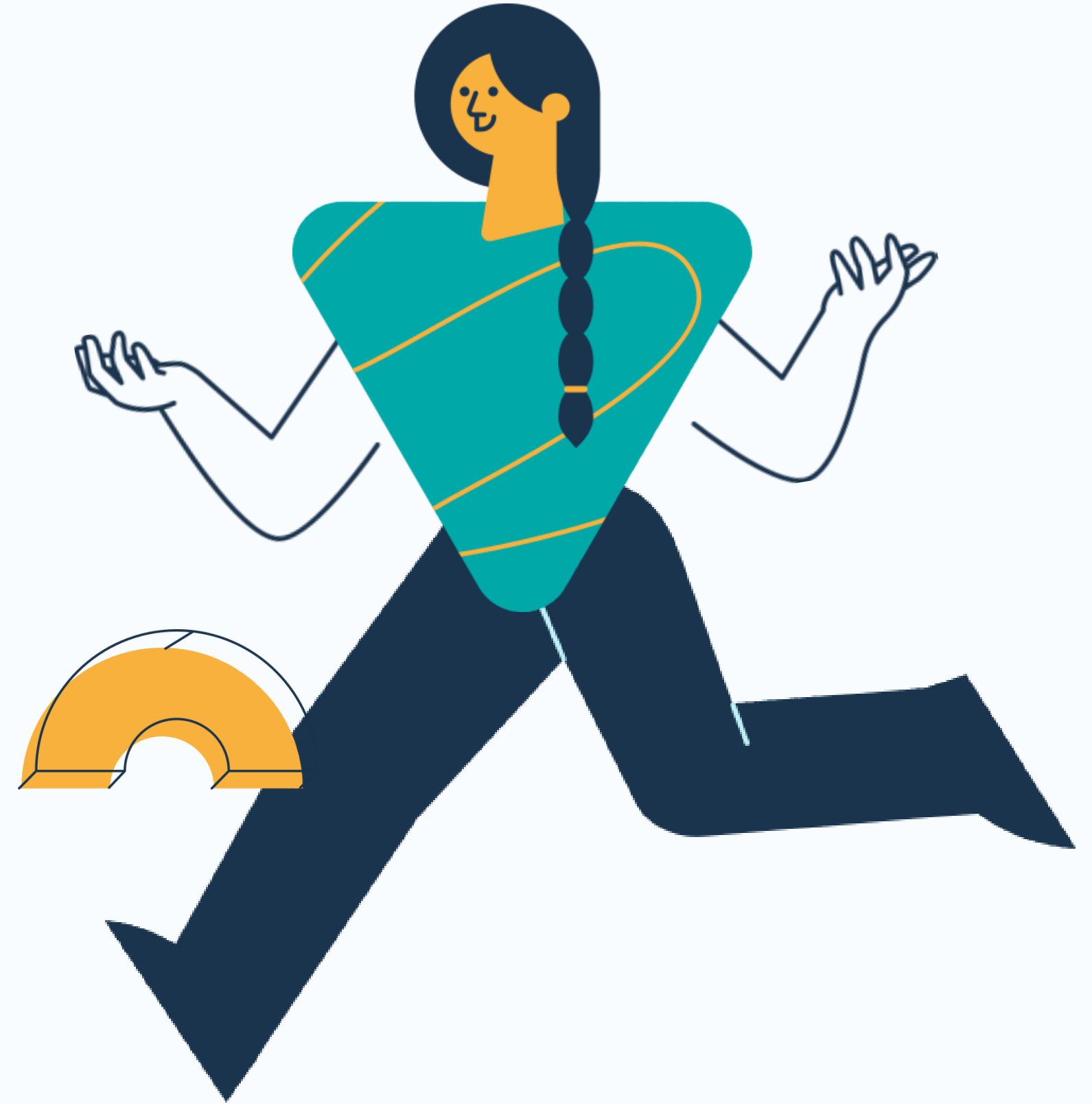
## Objective

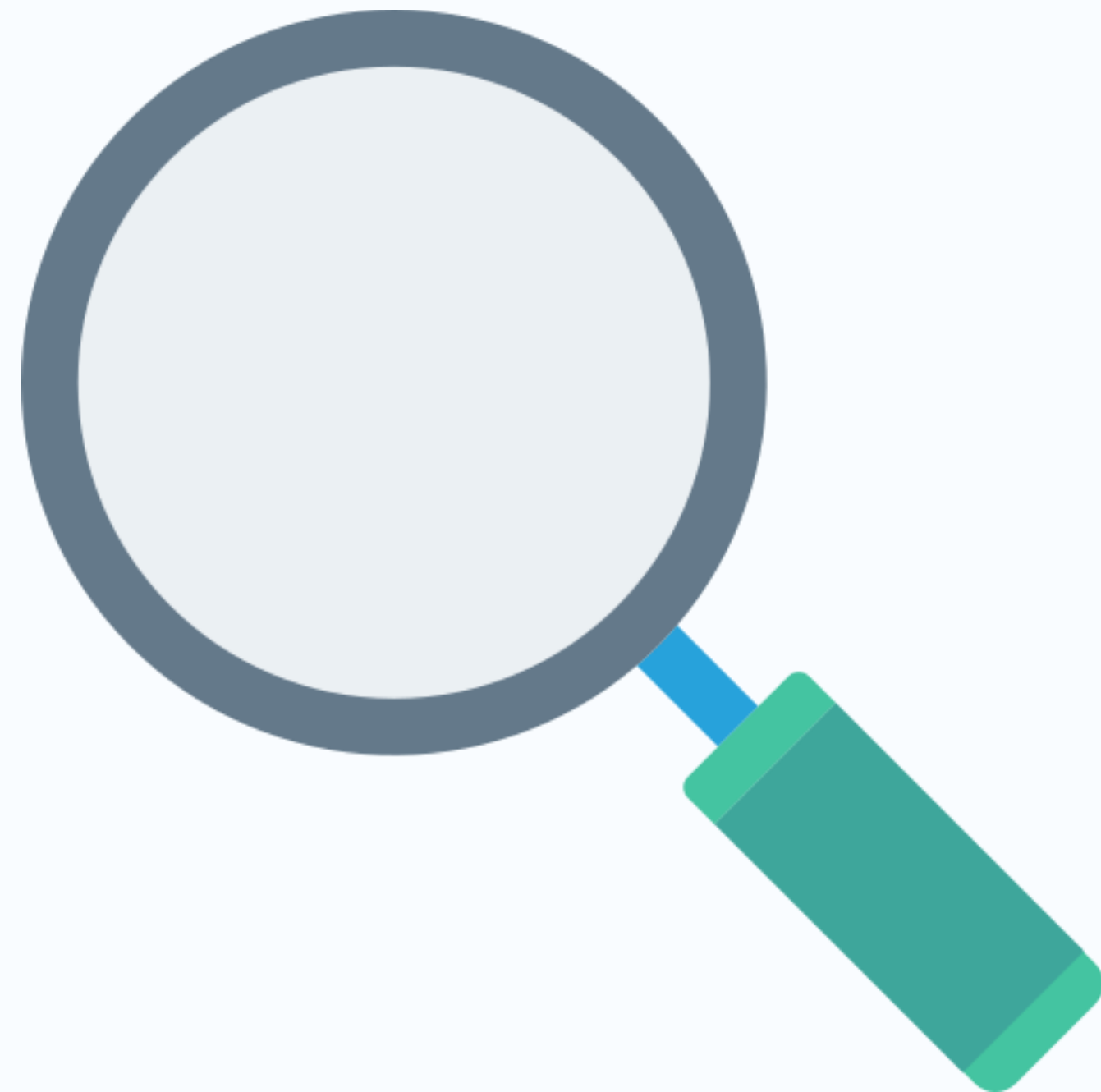
- Menciptakan *model machine learning* yang mampu memprediksi *promising customer*
- Menciptakan produk yang sesuai dengan target market

# Business Metrics

CUSTOMER ACQUISITION

MARKETING COST





# Data Exploration

---

# Dataset



Age

Employment  
Type

Graduate or  
Not

Annual  
Income

Family  
Members

Chronic  
Diseases

Frequent  
Flyer

Ever Travelled  
Abroad

Travel  
Insurance

# Describe Data

Insight yang dapat diambil

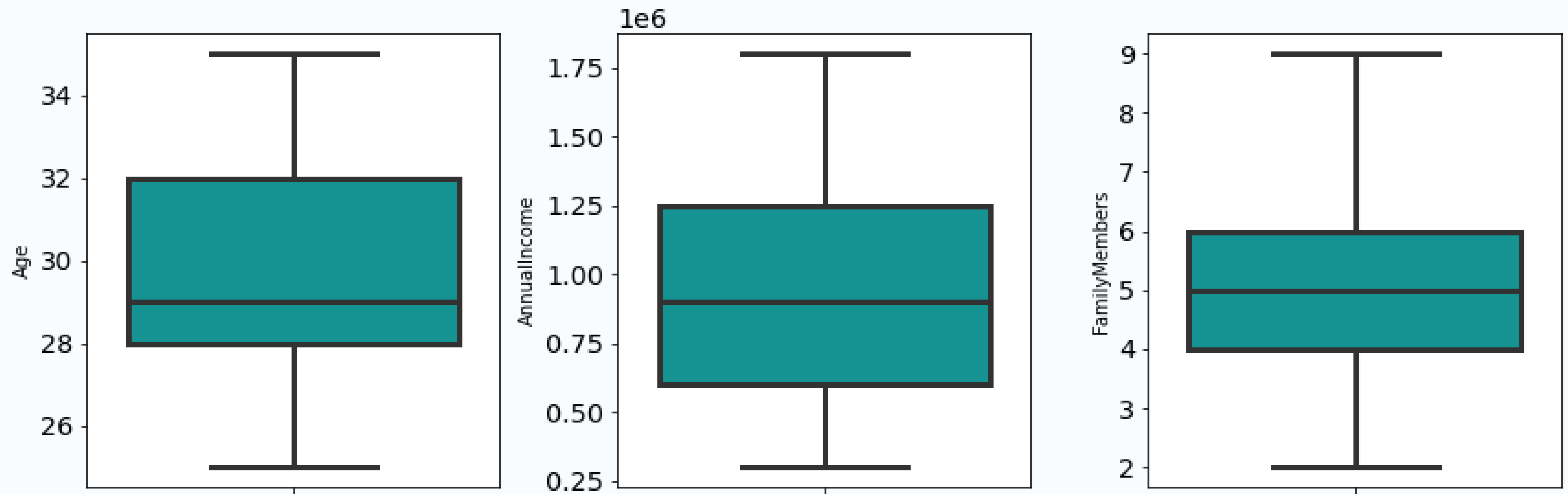
1. **1987 baris**, jumlah customer yang dihubungi oleh tim marketing selama 2019
2. **9 kolom**, jumlah fitur data customer yang dimiliki
3. Tipe Data: **Numeric** (int64/integer) dan **Categorical** (object/string)
4. **Tidak ada Null Values** (data yang kosong)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1987 entries, 0 to 1986
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Age                    1987 non-null   int64
1   Employment Type       1987 non-null   object
2   GraduateOrNot         1987 non-null   object
3   AnnualIncome          1987 non-null   int64
4   FamilyMembers         1987 non-null   int64
5   ChronicDiseases       1987 non-null   int64
6   FrequentFlyer         1987 non-null   object
7   EverTravelledAbroad   1987 non-null   object
8   TravelInsurance       1987 non-null   int64
dtypes: int64(5), object(4)
memory usage: 139.8+ KB
```



# Describe [NUM]

Distribusi data pada fitur-fitur yang bersifat Numeric seluruhnya bersifat **Distribusi Normal**



# Describe [CAT]

Data pada fitur-fitur yang bersifat Categorical seluruhnya tergolong *make sense* karena tidak ada data yang anomali

	Employment Type	GraduateOrNot	FrequentFlyer	EverTravelledAbroad
count	1987	1987	1987	1987
unique	2	2	2	2
top	Private Sector/Self Employed	Yes	No	No
freq	1417	1692	1570	1607

# Duplicated Data & Class Imbalance



## Duplicated Data

1. Dengan subset = 885
2. Tanpa subset = 0

Terdapat data yang duplikat namun pada kasus ini dianggap sebagai **customer yang berbeda**



## Class Imbalance

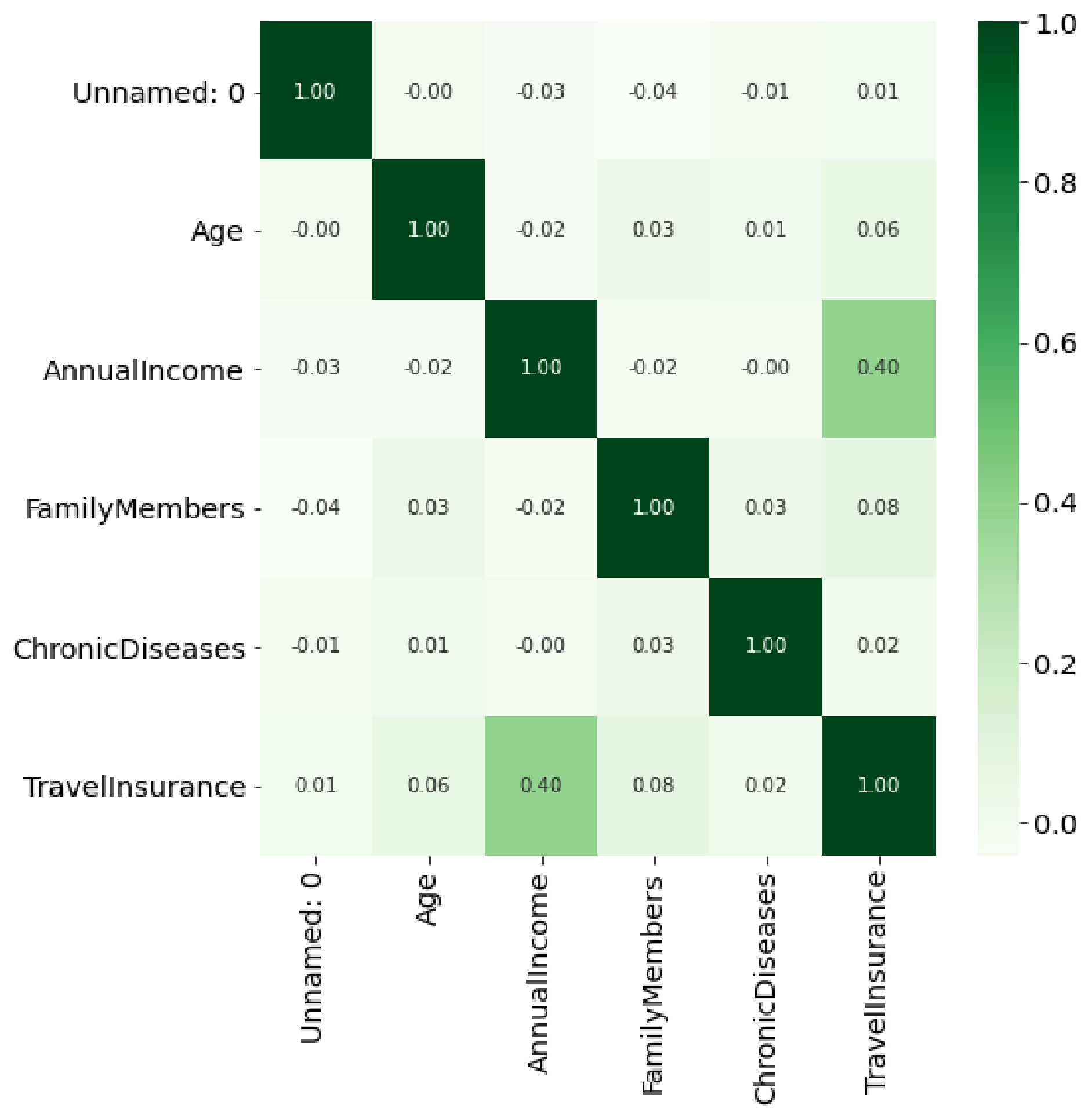
0.1277 = 65%

1.710 = 35%

Tergolong **Balance**

# Corr Heat Map

Tidak ada feature yang *redundant*





# Data Cleansing

Tidak diperlukan

Berdasarkan:

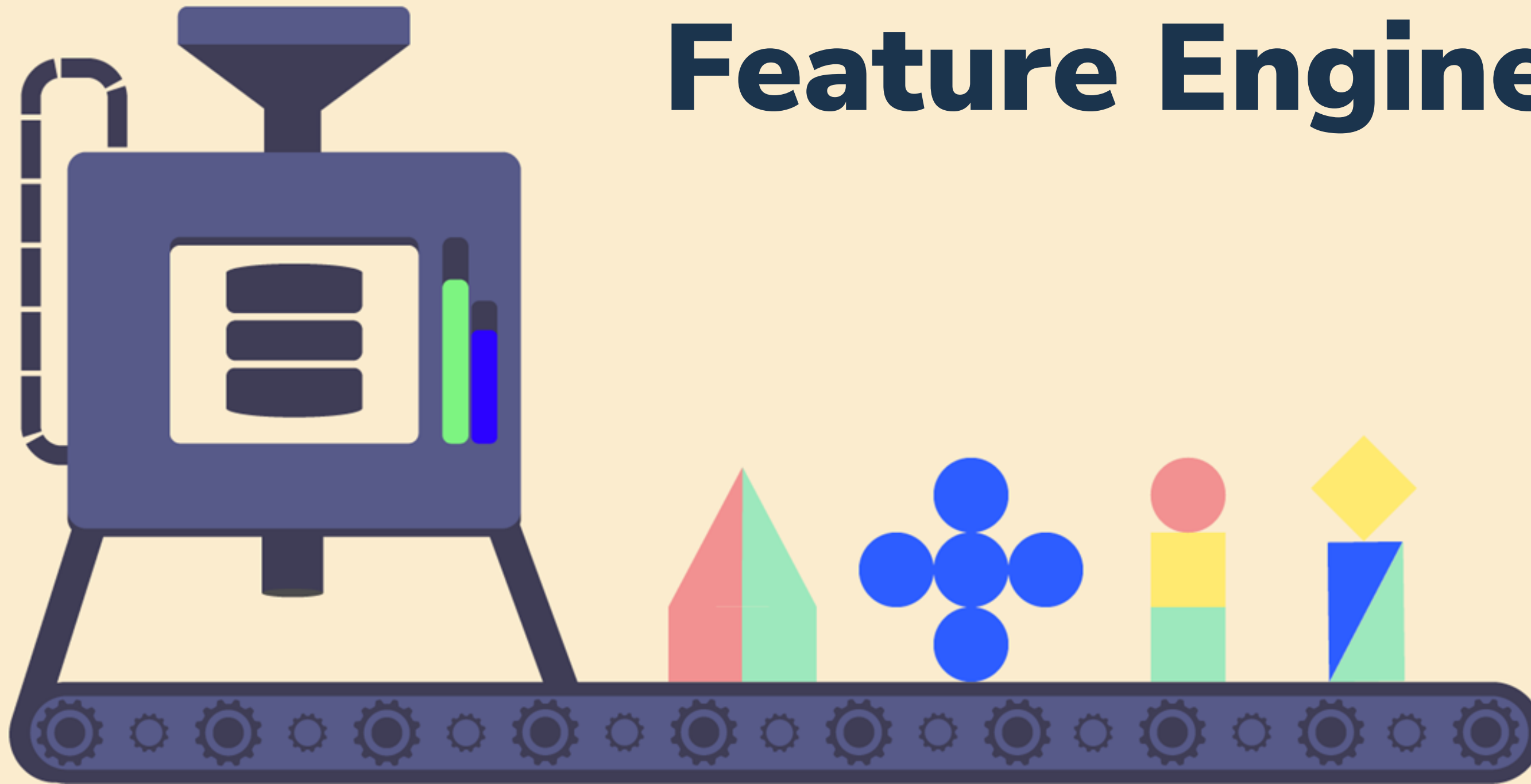
---

1. Null Values
2. Distribution
3. Duplicated Data
4. Class Imbalance
5. Redundant

Data sudah siap masuk ke tahap **feature engineering**



# Feature Engineering



	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravellInsurance
0	31	Government Sector	Yes	400000	6	1	No	No	0
1	31	Private Sector/Self Employed	Yes	1250000	7	0	No	No	0
2	34	Private Sector/Self Employed	Yes	500000	4	1	No	No	1
3	28	Private Sector/Self Employed	Yes	700000	3	1	No	No	0
4	28	Private Sector/Self Employed	Yes	700000	8	1	Yes	No	0
...	...	...	...	...	...	...	...	...	...
1982	33	Private Sector/Self Employed	Yes	1500000	4	0	Yes	Yes	1
1983	28	Private Sector/Self Employed	Yes	1750000	5	1	No	Yes	0
1984	28	Private Sector/Self Employed	Yes	1150000	6	1	No	No	0
1985	34	Private Sector/Self Employed	Yes	1000000	6	0	Yes	Yes	1
1986	34	Private Sector/Self Employed	Yes	500000	4	0	No	No	0

1987 rows × 9 columns

**Employment Type:** 0 = Government Sector; 1 = Private Sector/Self Employed  
**GraduateOrNot, FrequentFlyer, EverTravelledAbroad:** 0 = No; 1 = Yes



## Standardization

Feature *Age* dan *AnnualIncome* dilakukan standarisasi.

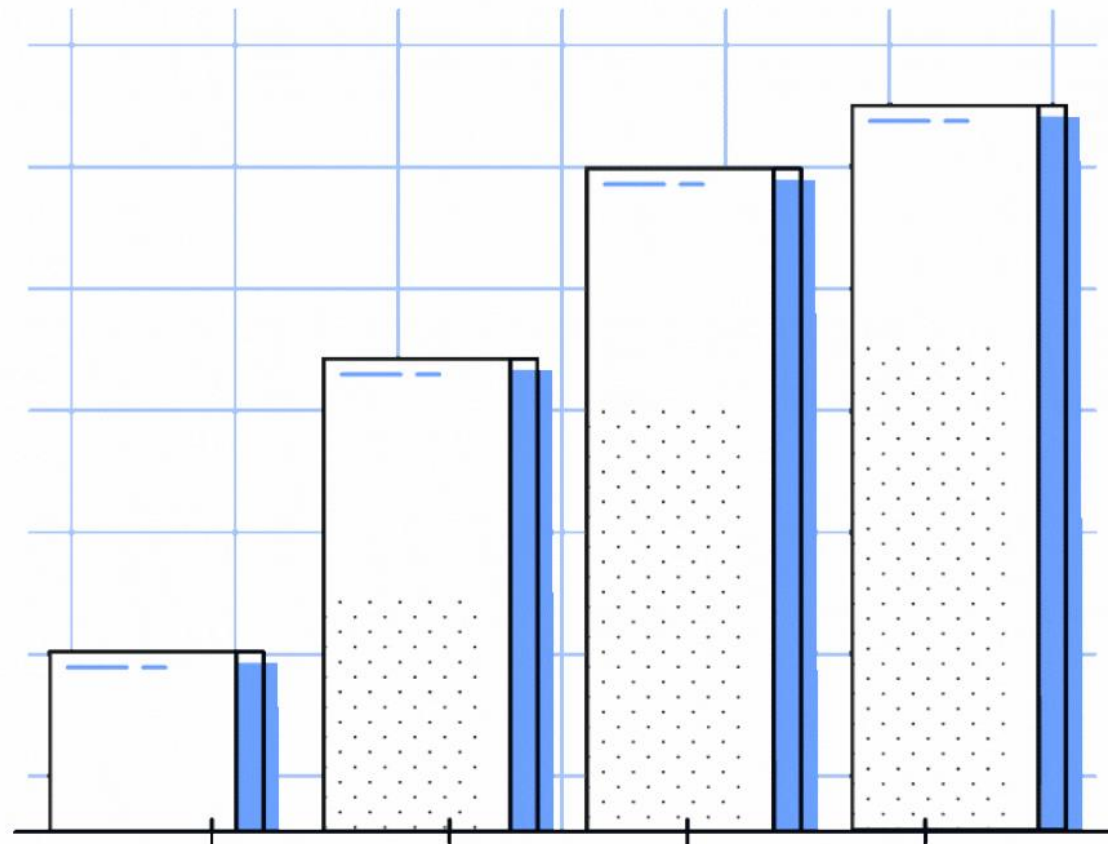
## Normalization

Feature *FamilyMembers* dilakukan normalisasi.



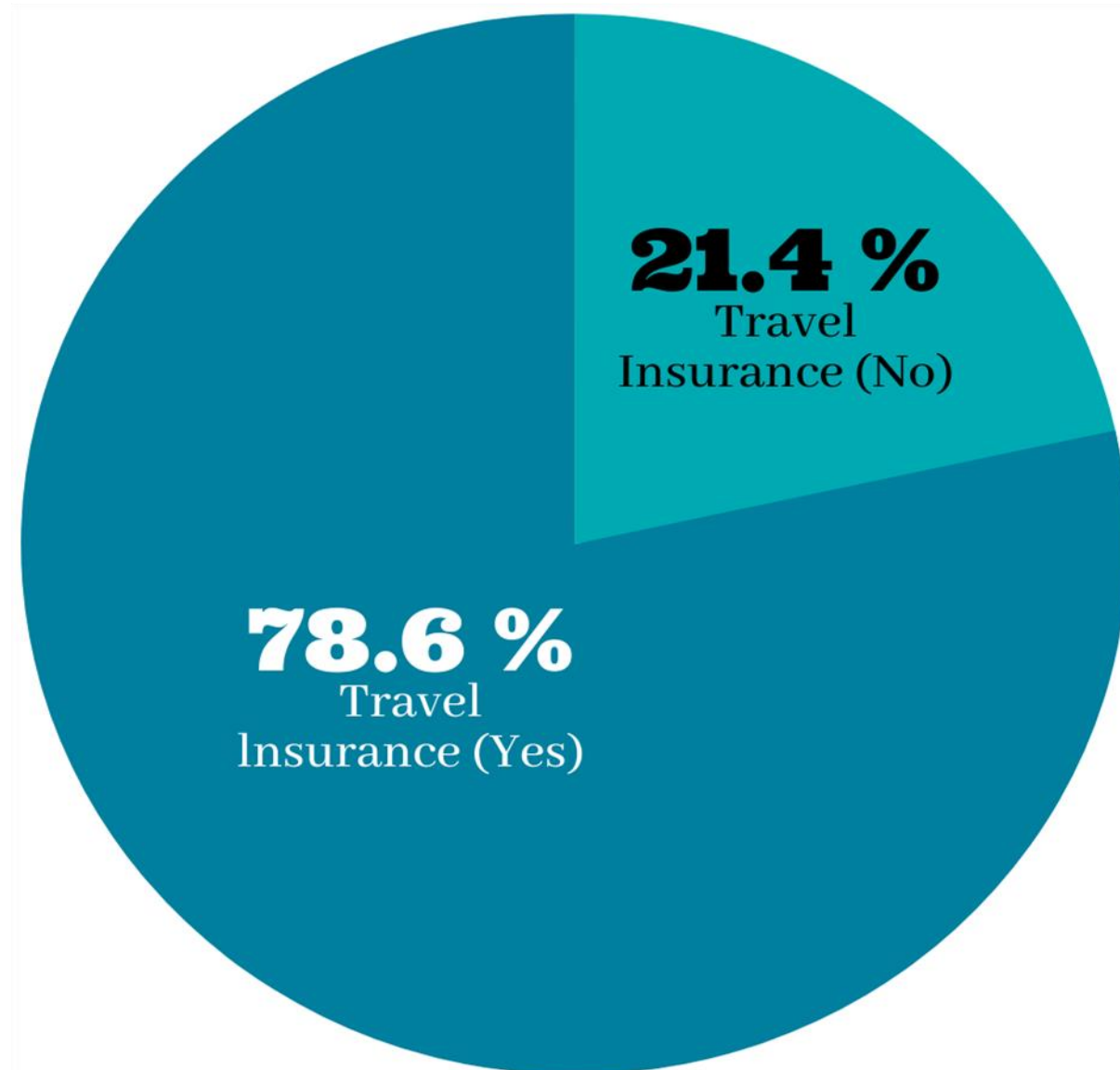


# Business Insight & Visualization

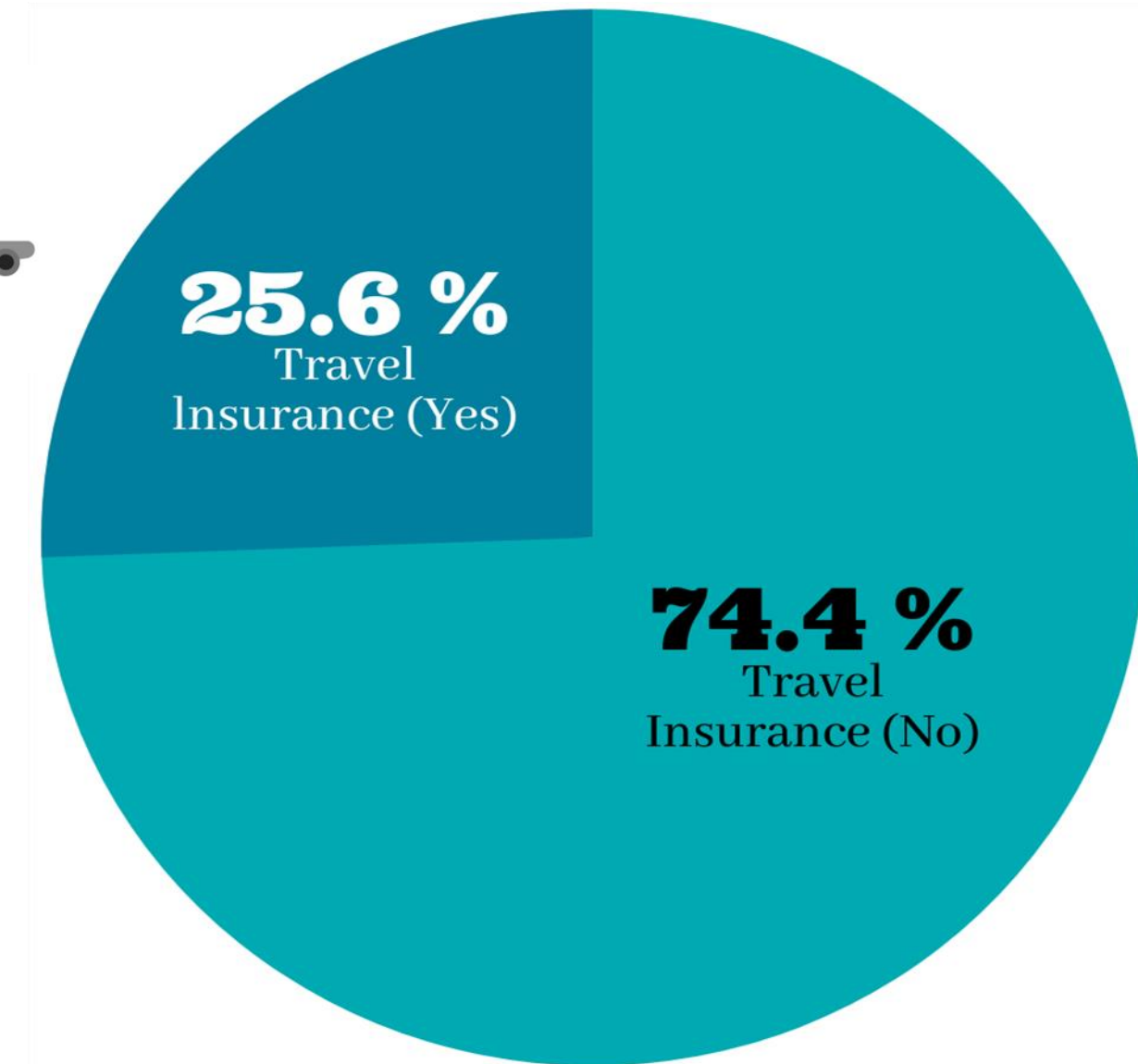


## ***Ever Travelled Abroad to Travel Insurance***

Ever Travelled Abroad

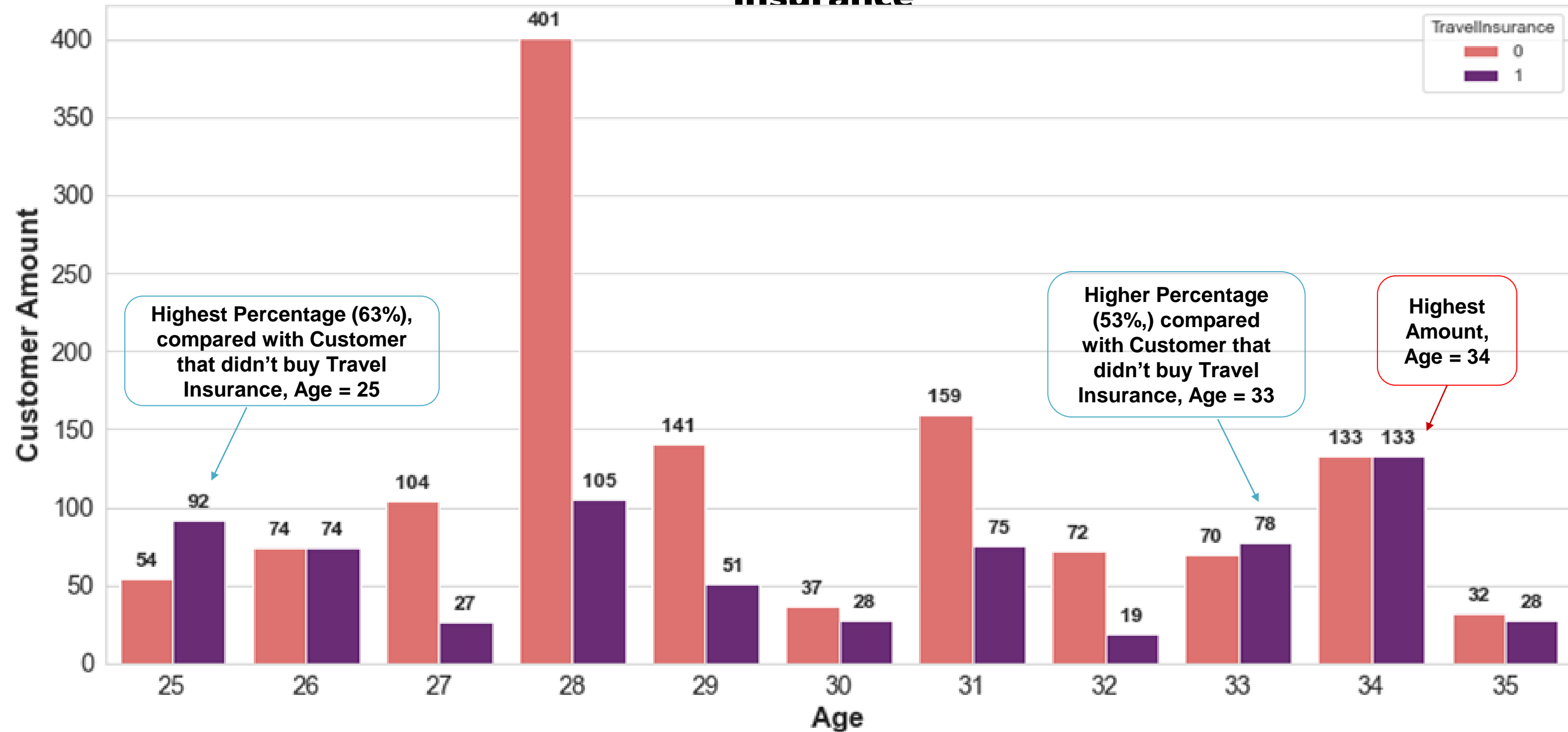


Non-Ever Travelled Abroad



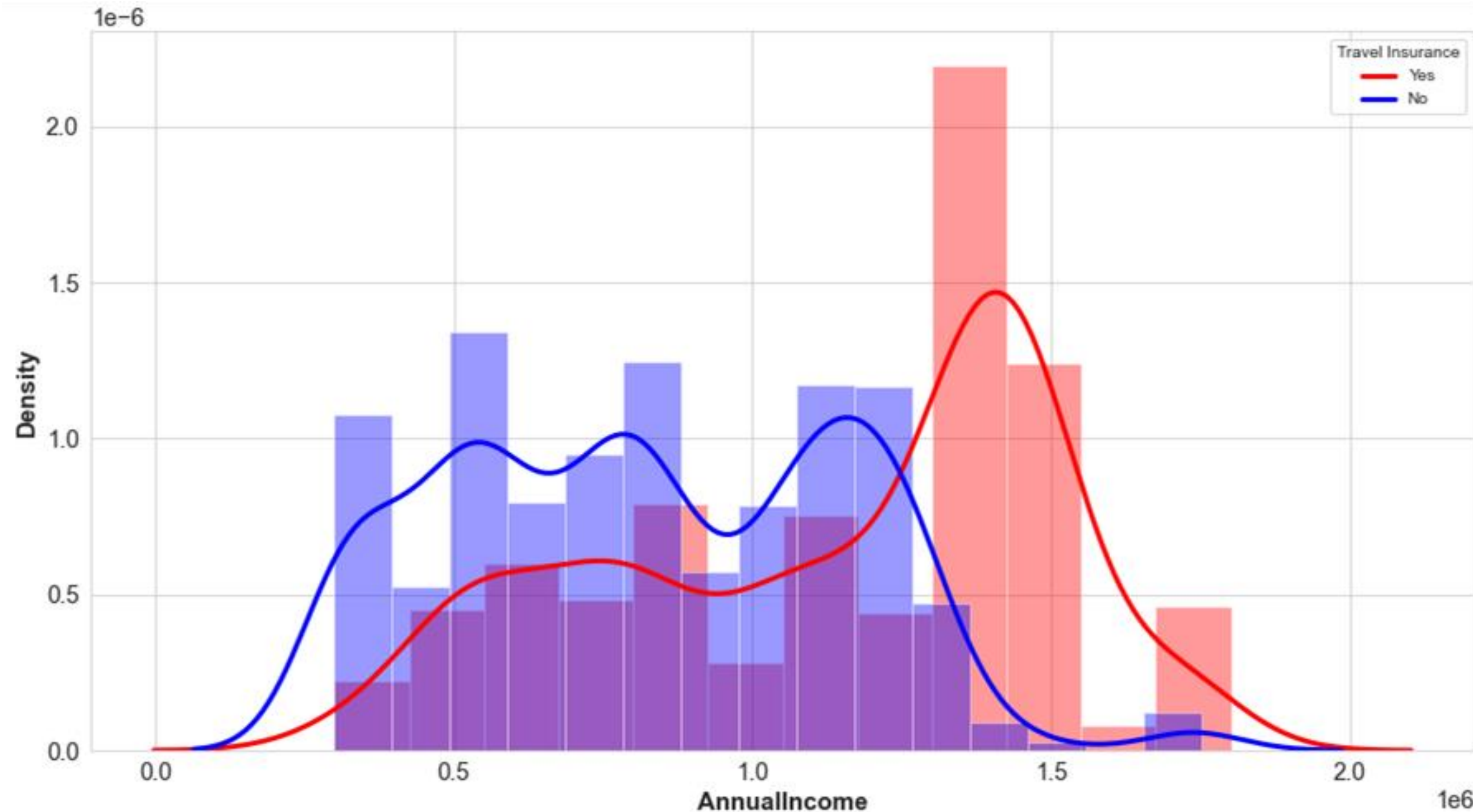
Customer yang pernah melakukan perjalanan ke luar negeri memiliki persentase lebih tinggi sebesar **78.6 %** dalam membeli produk Travel Insurance dibandingkan Customer yang tidak pernah melakukan perjalanan ke luar negeri dengan persentase sebesar **25.6%**.

# Age to Travel Insurance



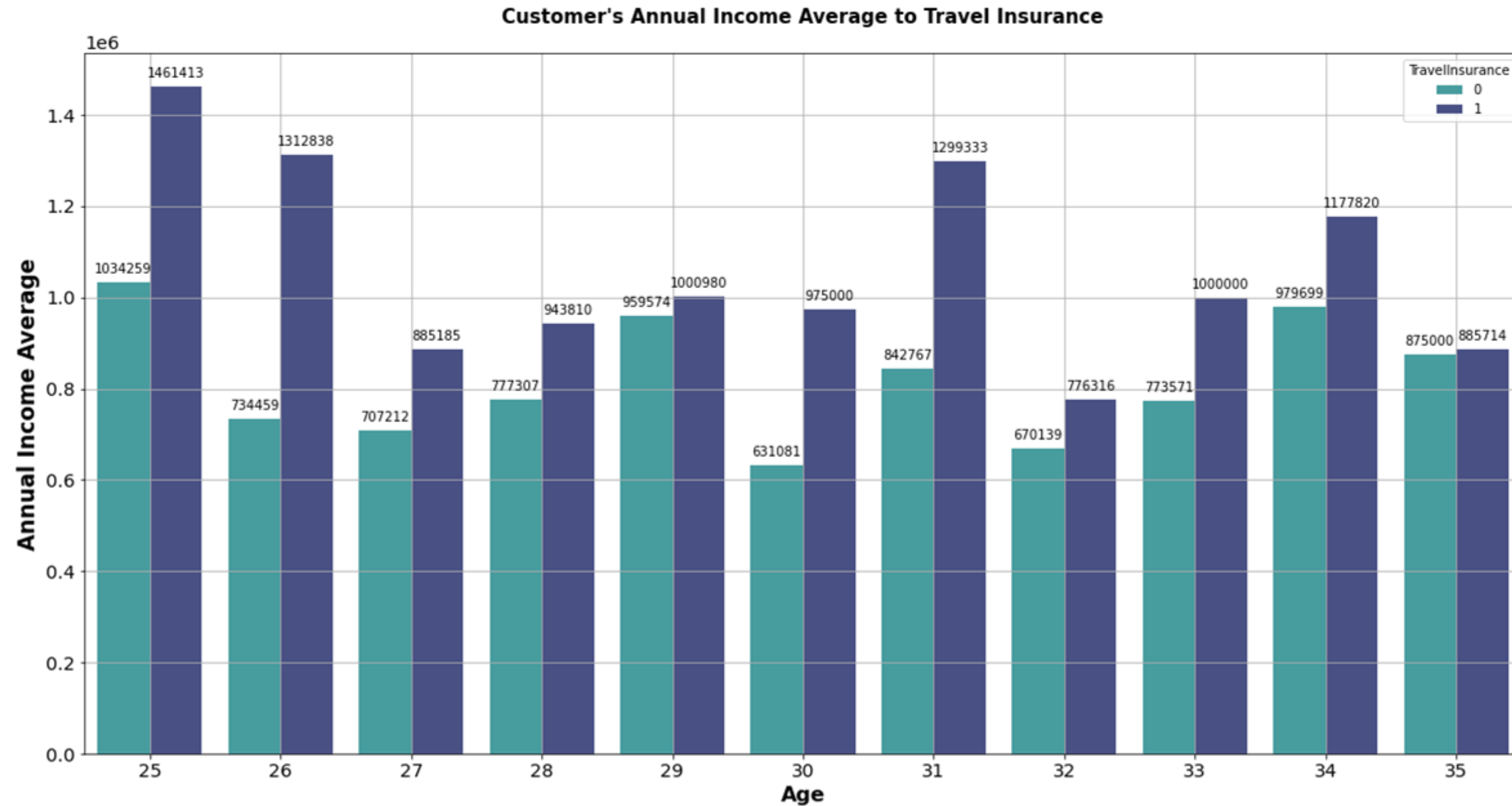
1. Customer yang membeli produk Travel Insurance dengan jumlah tertinggi berada pada umur **34** tahun.
2. Namun perbandingan antara Customer yang membeli produk Travel Insurance dengan yang tidak, pada umur **33** dan **25** tahun memiliki persentase lebih tinggi untuk membeli dengan perbandingan **53%:47%** dan **63%:37%**. Persentase tertinggi pada umur **25** tahun.
3. Umur **27-32** dan **35** memiliki persentase lebih tinggi pada Customer yang **tidak** membeli produk Travel Insurance.

## Annual Income to Travel Insurance



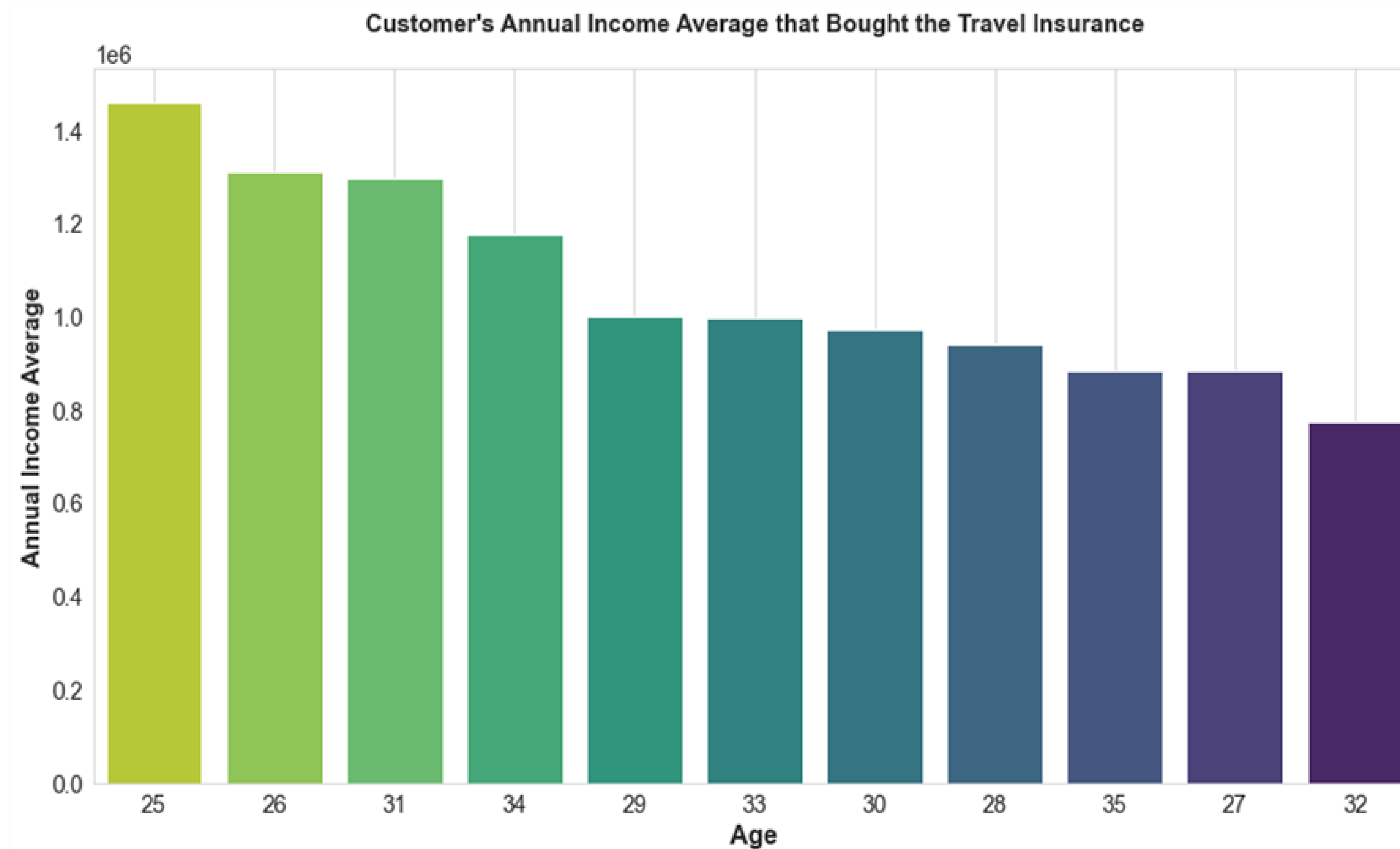
Customer dengan penghasilan tahunan **di atas 1.3 juta** memiliki persentase lebih tinggi dalam membeli produk Travel Insurance dibandingkan dengan yang tidak membeli, dengan jumlah Customer terbanyak pada penghasilan **1.4 juta**. Dari insight ini kita dapat merancang dan merekomendasikan Travel Insurance Package produk dengan kisaran harga **di bawah 1.3 juta**, untuk menjangkau Customer yang sebelumnya tidak membeli paket Travel Insurance.

# Annual Income Average to Travel Insurance



Pada semua umur (25-35 tahun,) nilai Annual Income Average lebih tinggi pada Customer yang membeli produk Travel Insurance

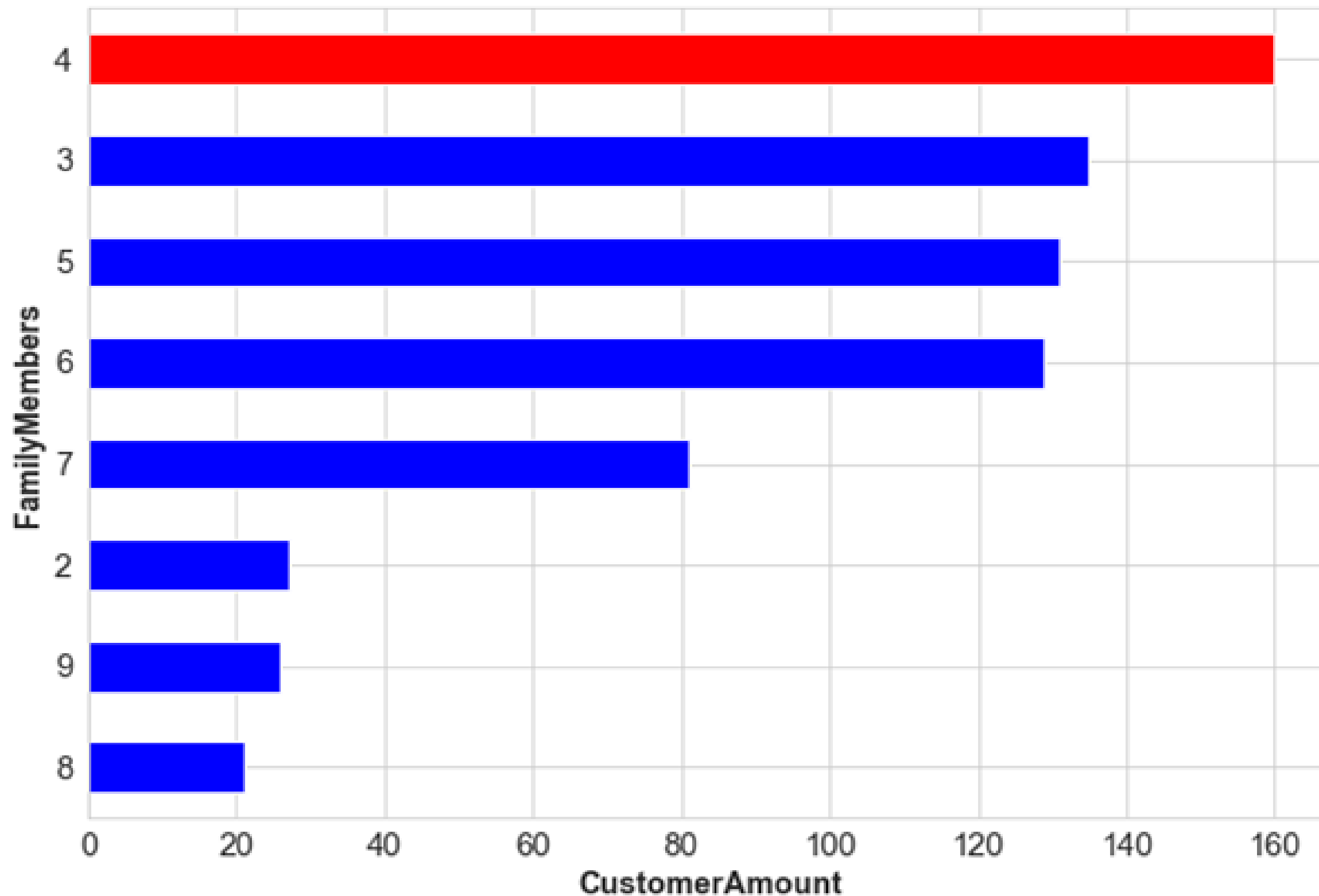
## Annual Income Average to Travel Insurance (2)



- Jika diurutkan berdasarkan data Customer yang hanya membeli produk Travel Insurance, Customer umur 25 tahun memiliki nilai Annual Income Average paling tinggi dengan nilai 1.46 juta
- Hal ini relevan dengan insight sebelumnya, bahwa persentase tertinggi Customer yang membeli produk Travel Insurance pada feature Age berada pada umur 25 tahun yang ternyata dipengaruhi oleh Annual Income.

# FamilyMembers to Travel Insurance

FamilyMembers that Bought the TravelInsurance



Family Members dengan anggota **4 orang**, memiliki jumlah Customer yang membeli produk Travel Insurance terbanyak dengan jumlah **160 Customer**.



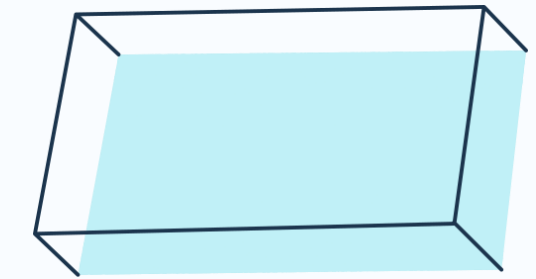


# Modeling





# Evaluation Metrics



AUC

Presisi

Mengurangi *False*

*Positif* →

meminimalkan cost

*marketing*

# Percobaan Model

No.	Model	Data Tes				Data Train			
		AUC	Presisi	Akurasi	Recall	AUC	Presisi	Akurasi	Recall
1	Log Regression	0.74	0.79	0.76	0.48	0.77	0.79	0.79	0.49
2	K-NN	0.79	0.91	0.78	0.45	0.98	0.88	0.79	0.48
3	Decision Tree (1)	0.79	0.81	0.80	0.61	0.87	0.84	0.82	0.62
4	Random Forest	0.80	0.94	0.85	0.59	0.94	0.93	0.85	0.62
5	Decision Tree (2)	0.80	0.88	0.82	0.60	0.87	0.90	0.83	0.60

Keterangan:

Semua model sudah dilakukan *tuning hyperparameter*

*Decision Tree (1)* menggunakan semua fitur dan *Decision Tree (2)* sudah dilakukan *drop* fitur yang tidak penting

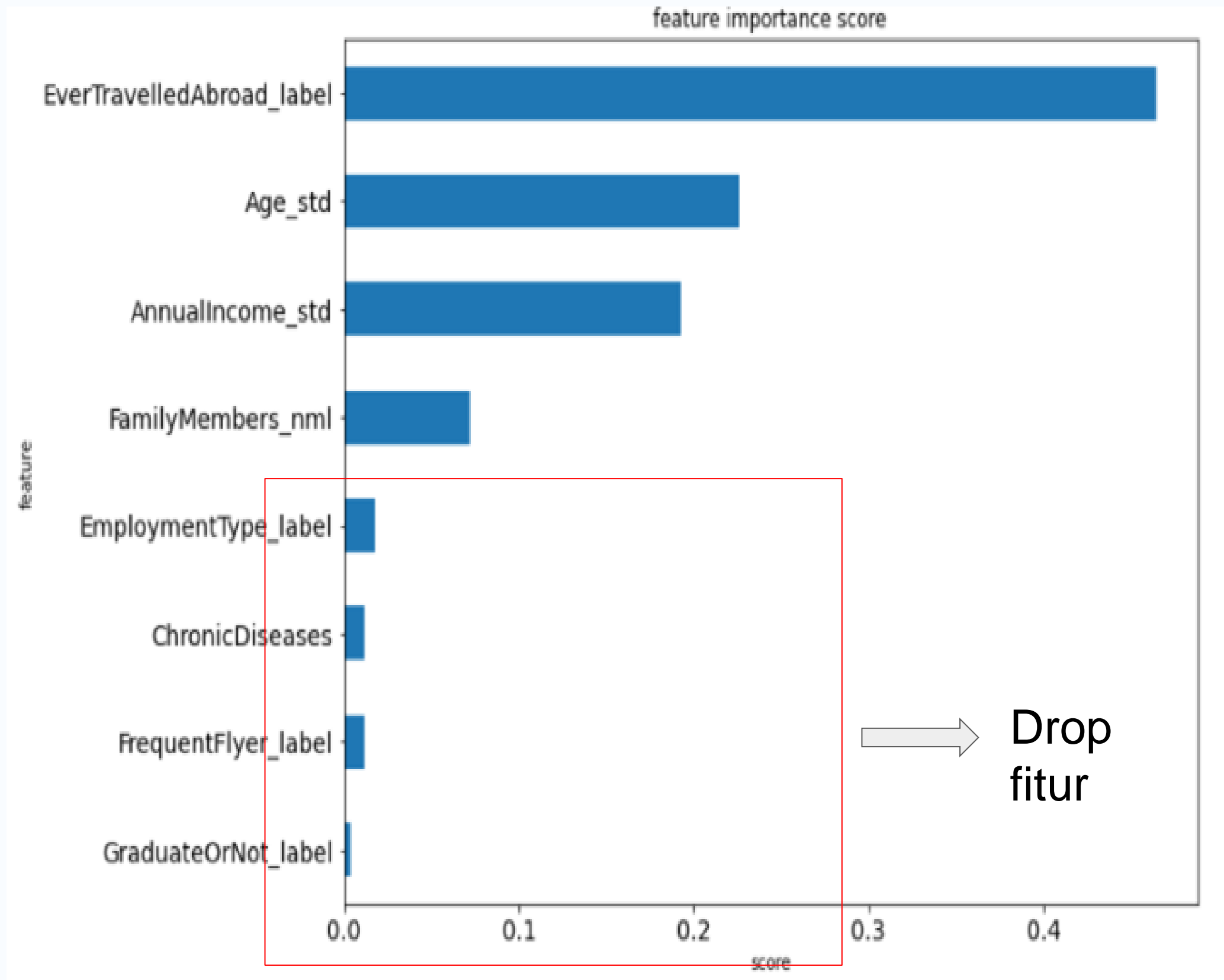
# Percobaan Model

No.	Model	Data Tes				Data Train			
		AUC	Presisi	Akurasi	Recall	AUC	Presisi	Akurasi	Recall
1	Log Regression	0.74	0.79	0.76	0.48	0.77	0.79	0.79	0.49
2	K-NN	0.79	0.91	0.78	0.45	0.98	0.88	0.79	0.48
3	Decision Tree (1)	0.79	0.81	0.80	0.61	0.87	0.84	0.82	0.62
4	Random Forest	0.80	0.94	0.85	0.59	0.94	0.93	0.85	0.62
5	Decision Tree (2)	0.80	0.88	0.82	0.60	0.87	0.90	0.83	0.60

**Keterangan:**

Semua model sudah dilakukan tuning hyperparameter

*Decision Tree (1)* menggunakan semua fitur dan *Decision Tree (2)* sudah dilakukan *drop* fitur yang tidak penting



## Decision Tree (2)

Melakukan pemeriksaan untuk *data importances* guna untuk mengoptimalisasi model.

# Percobaan Model

No.	Model	Data Tes				Data Train			
		AUC	Presisi	Akurasi	Recall	AUC	Presisi	Akurasi	Recall
1	Log Regression	0.74	0.79	0.76	0.48	0.77	0.79	0.79	0.49
2	K-NN	0.79	0.91	0.78	0.45	0.98	0.88	0.79	0.48
3	Decision Tree (1)	0.79	0.81	0.80	0.61	0.87	0.84	0.82	0.62
4	Random Forest	0.80	0.94	0.85	0.59	0.94	0.93	0.85	0.62
5	Decision Tree (2)	0.80	0.88	0.82	0.59	0.87	0.90	0.83	0.60

Model Akhir yang dipilih: Decision Tree (2)

# PRODUCT TAKEN SIMULATION (Cost Marketing Optimization)

Biaya untuk menghubungi/calon customer:

Benchmark; sekitar \$2.70-\$5.60 = Rp45.000,-.

Sumber: <https://www.cxtoday.com/contact-centre/how-to-calculate-your-cost-per-inbound-outbound-call-and-why/>

## Sebelum modeling:

Asumsi: Rp 45.000/call

Total keseluruhan calon Customer (1987)

$$\begin{aligned}\text{Cost1} &= 1987 \times \text{Rp } 45.000 \\ &= \text{Rp } 89.415.000\end{aligned}$$

## Setelah modeling:

Asumsi: Rp 45.000/call

Total keseluruhan potensial Customer (473)

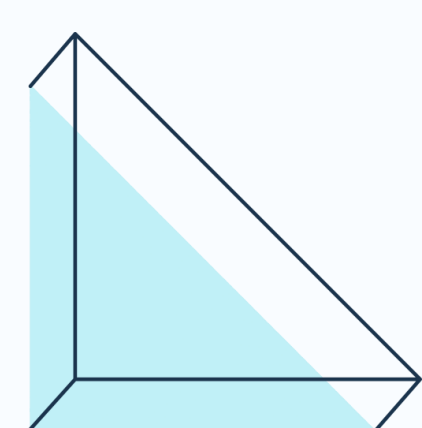
\*berdasarkan perhitungan confusion matrix

$$\begin{aligned}\text{Cost2} &= 473 \times \text{Rp } 45.000 \\ &= \text{Rp } 21.285.000\end{aligned}$$

$$\text{Reduce Cost} = \text{Cost1} - \text{Cost2}$$

$$= \text{Rp } 89.415.000 - \text{Rp } 21.285.000$$

$$= \text{Rp } 68.130.000$$



# Recommendation

Menawarkan harga premi asuransi lebih murah untuk customer yang berpenghasilan dibawah Rp 1,3 Juta

Mengalokasikan lebih banyak marketing cost untuk individu yang sudah berkeluarga dengan 2 anak

Priotitas marketing ditujukan pada individu yang sudah pernah ke Luar Negeri

