



# **TRAVEL INSURANCE PREDICTION**

**6TH SENSE ANALYST**

# Table Of Content



## **Business Overview**

Problem Statement, Purpose dan Business Metrics



## **Dataset Variables**

Isi dari dataset yang digunakan



## **Exploratory Data Analysis**

Ekstrasi visual dan informasi untuk mencari insights mengenai dataset



## **Data Pre-Processing**

Pembersihan dataset sehingga dataset siap digunakan



## **Modelling & Segmentation**

Pembuatan model Machine Learning & Segmentasi Customer



## **Business Recommendations**

Rekomendasi bisnis berdasarkan uji analisa yang sudah dilakukan



# Business Overview

## Problem Statement

Perusahaan asuransi perjalanan kami ingin mengoptimalkan penjualan paket asuransi perjalanan. Tantangannya adalah memprediksi minat pelanggan secara akurat berdasarkan data historis yang sudah diberikan kepada kami. Kami bertujuan untuk mengatasi hal ini dengan memanfaatkan atribut demografis dan perilaku untuk mengidentifikasi calon pelanggan yang paling mungkin membeli paket asuransi tersebut.

## Purpose

Mengembangkan **model prediksi** yang akurat untuk mengidentifikasi calon pembeli asuransi perjalanan. Melalui pendekatan berbasis data dan analisis yang mendalam, kami bertujuan untuk memberikan perusahaan alat prediktif dan pemahaman yang lebih dalam mengenai perilaku pelanggan, sehingga memungkinkan peningkatan **efisiensi strategi penjualan** dengan menargetkan calon-calon pelanggan yang tepat.



# Business Metrics

- **Accuracy**

Mengukur kemampuan model untuk memprediksi minat pembelian asuransi perjalanan dengan benar.

- **Presisi**

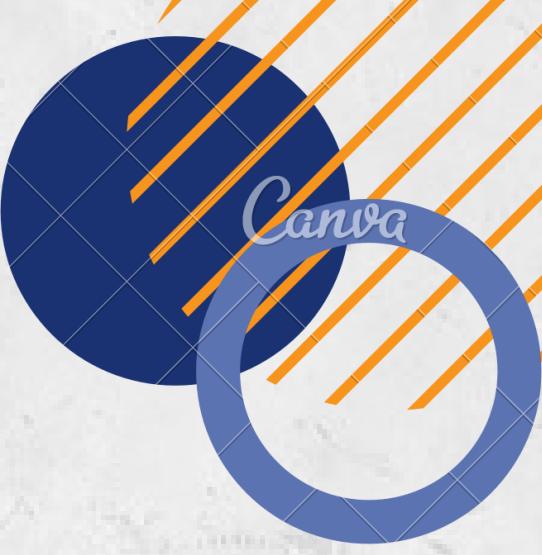
Mengevaluasi kemampuan model untuk mengidentifikasi pelanggan yang benar-benar tertarik.

- **Conversion Rate**

Menilai persentase pelanggan yang membeli asuransi setelah dihubungi berdasarkan prediksi model.



# Dataset



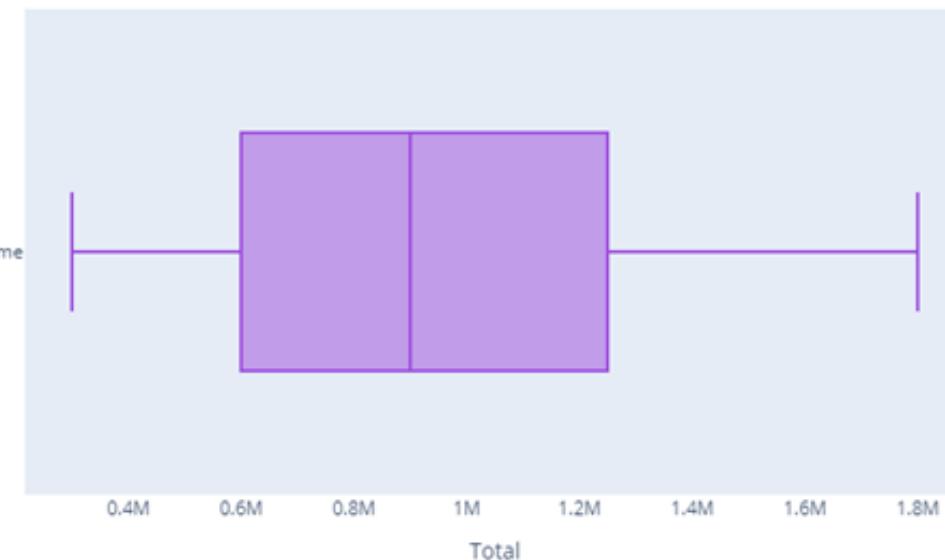
1. **Age**: Numeric, the customer's age
2. **Employment Type**: Categorical, the sector of employment
3. **GraduateOrNot**: Categorical, whether the customer is a college graduate
4. **AnnualIncome**: Numeric, the customer's yearly income
5. **FamilyMembers**: Numeric, the number of family members living with the customer
6. **ChronicDiseases**: Categorical, whether the customer has any chronic conditions
7. **FrequentFlyer**: Categorical, whether a customer books frequent tickets
8. **EverTravelledAbroad**: Categorical, has the customer ever travelled abroad
9. **TravelInsurance**: Target, whether the customer bought travel insurance

# **Exploratory Data Analysis**

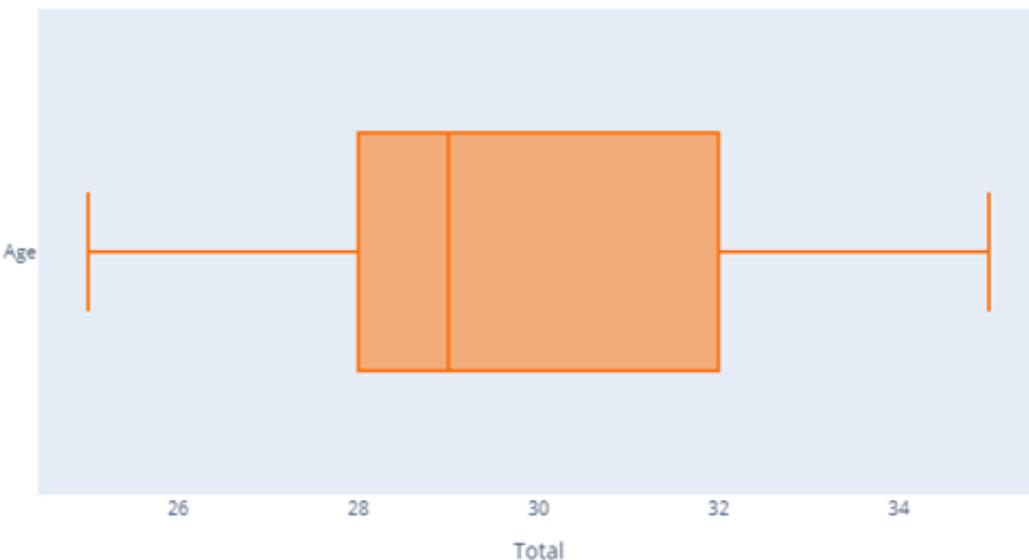
# Univariate Analysis (Numerical)



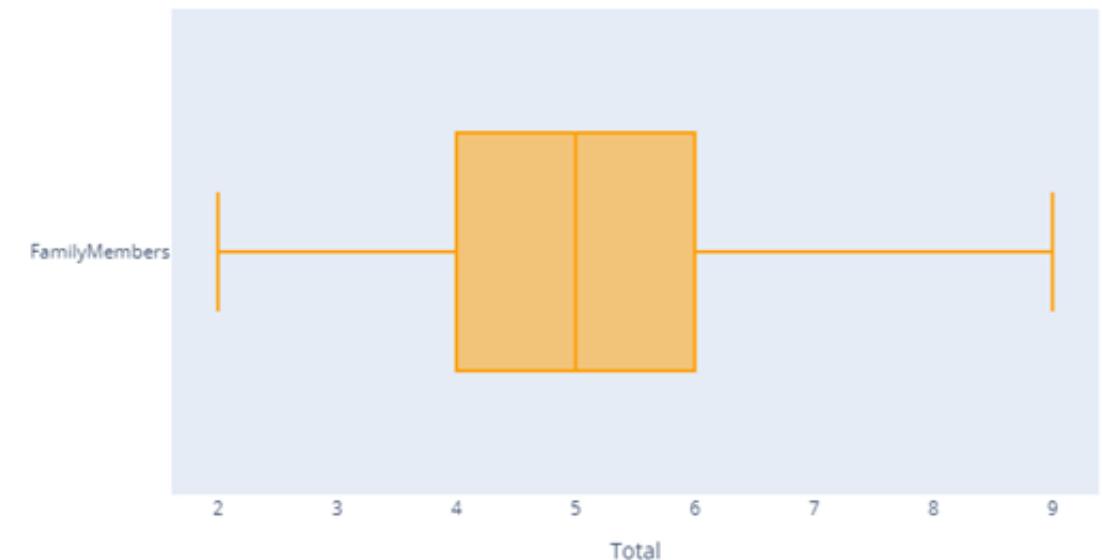
Box Plot AnnualIncome



Box Plot Age



Box Plot FamilyMembers



## No Outliers!

# Univariate Analysis (Categorical)



## Winning Categories

01

Job Sector &  
Chronic Diseases

Mayoritas pembeli bekerja  
di sektor swasta dan tidak  
memiliki penyakit kronis.

02

College  
Graduate

Mayoritas pembeli  
merupakan lulusan  
universitas

03

Flight Status

Mayoritas pembeli jarang  
membeli tiket pesawat  
dan tidak pernah  
berpergian internasional

# Multivariate Analysis

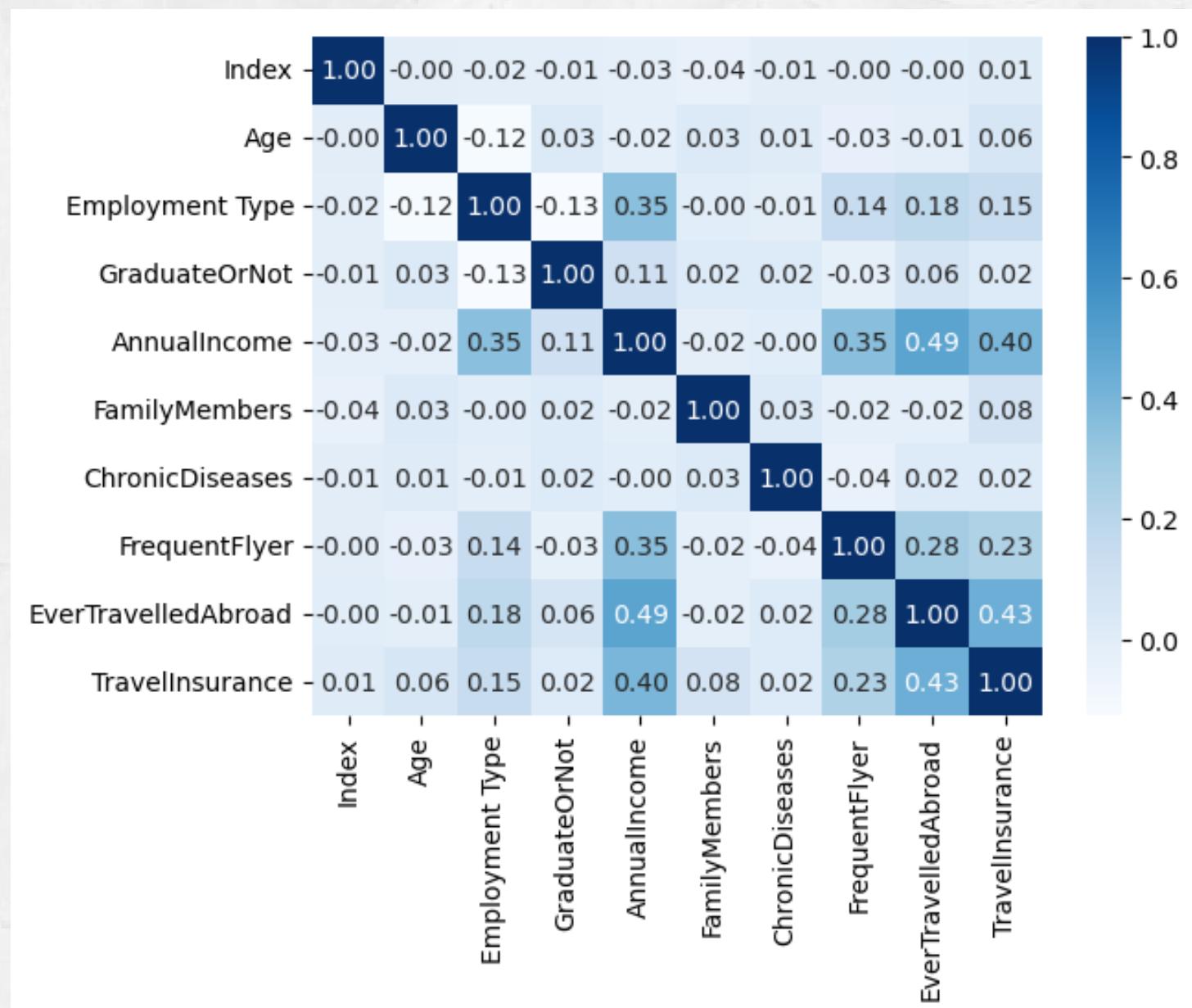
## Feature Importance

	Feature	T-Statistic	P-Value
0	Index	-0.276050	7.825387e-01
1	Age	-2.725506	6.476684e-03
2	Employment Type	-6.660267	3.527704e-11
3	GraduateOrNot	-0.843703	3.989372e-01
4	AnnualIncome	-19.257785	6.635018e-76
5	FamilyMembers	-3.571635	3.632080e-04
6	ChronicDiseases	-0.810542	4.177259e-01
7	FrequentFlyer	-10.631295	1.032551e-25
8	EverTravelledAbroad	-21.413067	1.088979e-91

Berdasarkan hasil T-Test disamping, features "AnnualIncome", "EverTravelledAbroad", "Employment Type", "FrequentFlyer", "Age", dan "FamilyMembers" memiliki significance terhadap target variable yaitu "TravelInsurance"

# Multivariate Analysis

## Multicollinearity Checking



Tidak ada multicollinearity yang terdeteksi dari heatmap di samping. Untuk memastikan lebih lanjut, VIF checking akan dilakukan.

# Multivariate Analysis

## VIF Checking

	feature	VIF
0	Age	24.783328
1	Employment Type	15.485388
2	GraduateOrNot	6.752638
3	AnnualIncome	11.137858
4	FamilyMembers	9.295239
5	ChronicDiseases	1.388029
6	FrequentFlyer	1.482909
7	EverTravelledAbroad	1.640893

Age memiliki VIF yang paling tinggi. Ini mengindikasikan bahwa Age memiliki nilai multicollinearity yang sangat tinggi.

# Multivariate Analysis

## Pemilihan Variabel



	feature	VIF
0	Employment Type	4.333650
1	AnnualIncome	8.676301
2	FamilyMembers	4.908090
3	FrequentFlyer	1.535836
4	EverTravelledAbroad	1.756159

Dengan konsiderasi dari T-Test, VIF Checking, dan pemilihan secara logika, kita akan mengorbankan feature Age, Index, GraduateOrNot, dan ChronicDiseases untuk dibuang. Setelah Features ini dibuang, semua VIF di samping menjadi normal kembali.

# Data Pre-Processing

## Handling Missing Values

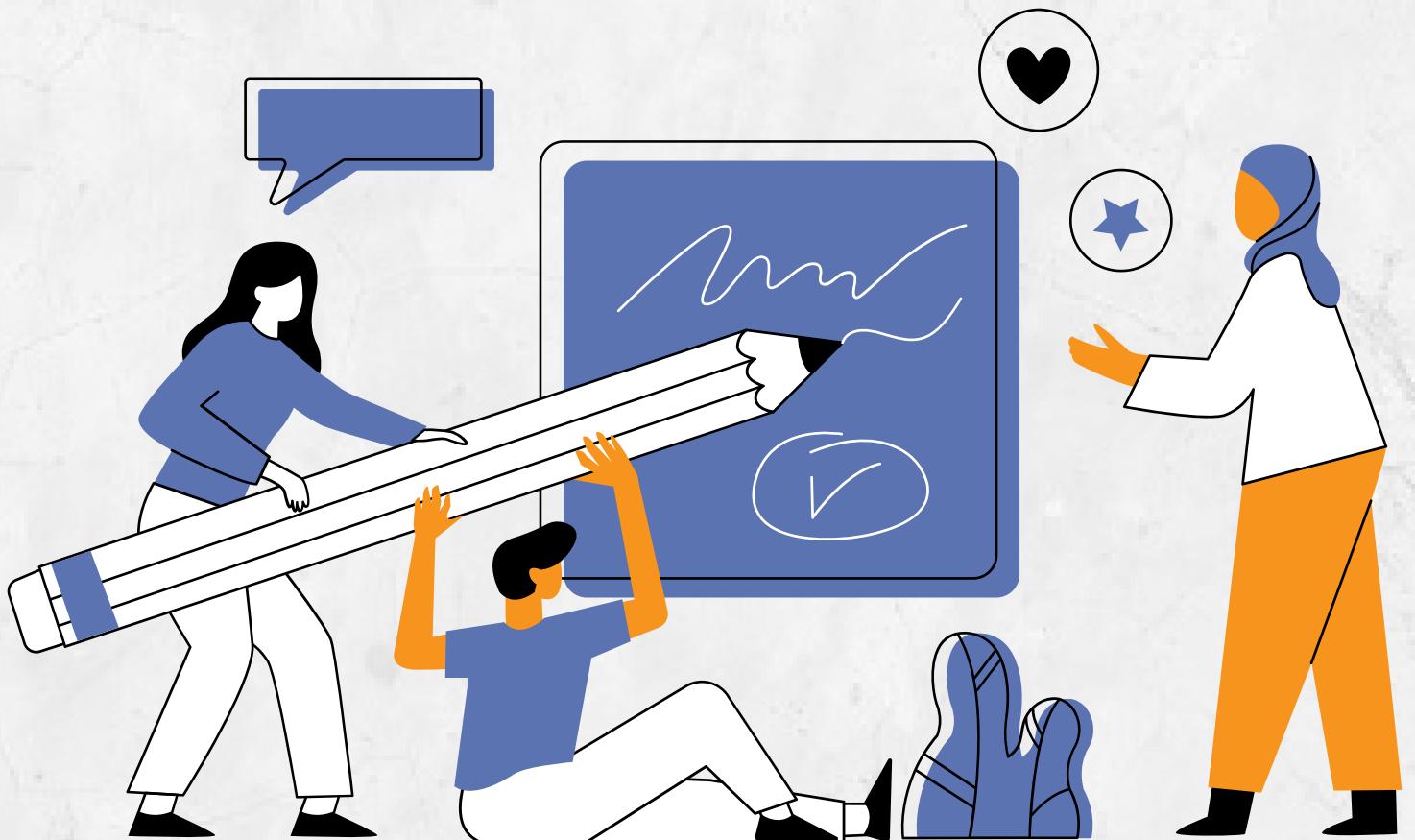
Dilakukan untuk memastikan konsistensi dan akurasi data. Tidak ada missing values di dataset kita.

## Handling Duplicates

Membersihkan duplikat memastikan integritas dan kualitas data yang lebih baik. Tidak ada duplicates di dataset kita

## Handling Outliers

Penanganan outliers membantu menjaga integritas statistik dan validitas interpretasi analisis data. Tidak ada outliers di dataset kita.



# Class Imbalance

**Travel Insurance No = 1277**

**Travel Insurance Yes = 710**

**Not Balanced!**

**Oversampling**

**Travel Insurance No = 1277**

**Travel Insurance Yes = 1277**

**Balanced!**



# **MODELING & SEGMENTATION**

# Data Splitting

Train  
70%

Test  
30%

Cross-Validation = 10 Folds

# Models Evaluations

## Logistic Regression

ROC AUC	Train: 69,7% Test: 70,1%
Precision	Train: 76,1% Test: 75,7%
Accuracy	67,5%
F1 Score	62,7%

## Decision Tree

ROC AUC	Train: 85,9% Test: 80,7%
Precision	Train: 90,3% Test: 87,8%
Accuracy	74,5%
F1 Score	71,6%

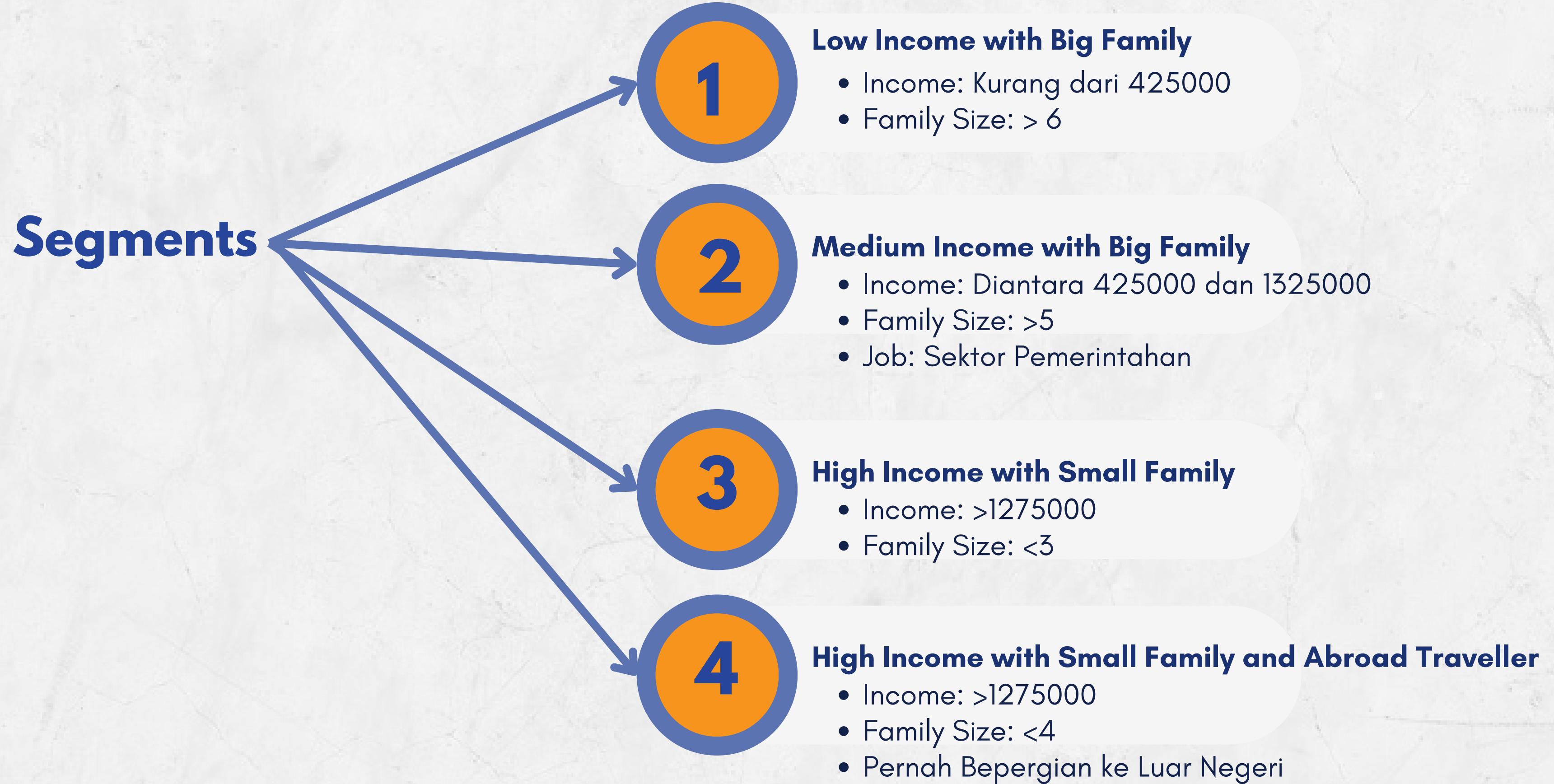
## Random Forest

ROC AUC	Train: 85,8% Test: 80,9%
Precision	Train: 89,8% Test: 87,6%
Accuracy	74,5%
F1 Score	71,7%

Secara keseluruhan, ketiga model memiliki performa yang **baik**.

Tetapi, **Decision Tree** menjadi **model terbaik** dibanding 2 model lainnya sehingga akan menjadi **model final** yang akan dipakai untuk tahap selanjutnya.

# Customer Segmentation



# Business Recommendations

Berdasarkan analisa segmentasi customer, kami merekomendasikan beberapa strategi untuk menambah efisiensi aktivitas penjualan travel insurance:



01

**Memberikan diskon premi kepada customer yang memiliki anggota keluarga dengan jumlah yang besar**

02

**Menjalin kemitraan dengan lembaga atau instansi pemerintah dengan menjadikan travel insurance sebagai benefit tambahan untuk pegawai**

03

**Melakukan promosi travel insurance melalui komunitas atau klub sosial yang biasa diisi segmen high income seperti klub golf dll**

04

**Menyediakan layanan tambahan seperti paket perlindungan perjalanan internasional**

**THANK YOU**